The information content of FOMC minutes

Ellyn Boukus **Joshua V. Rosenberg**

Federal Reserve Bank of New York
33 Liberty St.
New York, NY 10045
joshua.rosenberg@ny.frb.org, ellyn.boukus@ny.frb.org

First Draft: January 2005
This Revision: June 2006
This Print: July 6, 2006

Abstract: In this paper, we analyze the information content of Federal Open Market Committee minutes from 1987-2005. We apply an objective, statistical methodology known as Latent Semantic Analysis to decompose each minutes release into its characteristic themes. We show that these themes are correlated with current and future economic conditions. Our evidence suggests that market participants can extract a complex, multifaceted signal from the minutes. In particular, Treasury yield changes around the time of the minutes release depend on the specific themes expressed, the level of monetary policy uncertainty, and the economic outlook.

Keywords: Monetary policy, minutes, transparency, Latent Semantic Analysis, computational linguistics
JEL Codes: C4, D78, E4, E5

---

[*]We would like to thank Tobias Adrian, Michael Fleming, Jiang Wang, and seminar participants at the Federal Reserve Bank of New York for their helpful comments. Any views expressed represent those of the authors only and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

Market reaction to Federal Reserve policy changes has been the focus of many recent studies. For example, the impact of Federal funds target rate adjustments on other interest rates is examined by Cook and Hahn (1989), Kuttner (2001), Cochrane and Piazzesi (2002), Demiralp and Jorda (2004), and Thornton (2004), while Fleming and Piazzesi (2005) and Bernanke and Kuttner (2005) examine the stock market reaction to changes in the Federal funds rate. Fleming and Remolona (1997) cite fourteen studies measuring the influence of money supply announcements on interest rates, while the influence of discount rate announcements on the yield curve is explored in Lombra and Torto (1977), Smirlock and Yawitz (1985), Cook and Hahn (1987, 1988) Dueker (1992), and May (1992).

These previously analyzed quantitative measures provide an important signal of current Fed policy. However, the Federal Reserve also communicates qualitative information to the public and to financial markets through statements, minutes, speeches by officials, and published reports. Increased commitment to transparency, especially since the mid-1990s, has arguably led to a focus as much on the Fed's words as its actions as indicators of future policy.[1] Therefore, an analysis of the financial market impact of Fed communications emerges as a natural extension of previous studies.

Our paper focuses on the market reaction to Fed policy as revealed by the FOMC meeting minutes. We choose to analyze FOMC minutes because they are closely watched by market participants and widely cited by market analysts as a gauge of Fed policy. As discussed by Danker and Luecke (2005), the FOMC minutes include summaries of staff and committee member views on economic conditions as well as a discussion of the rationale for the policy decision. Further evidence of their significance is the recent FOMC decision to expedite the minutes release:

[1] See, for example, Blinder, Goodhart, Hildebrand, Lipton and Wyplosz (2001), Jenkins (2001), Geraats (2002), Thornton (2003), Poole (2003), Carpenter (2004), Bernanke (2004), Posen (2004), and Poole (2005).

> [FOMC] [p]articipants noted that the minutes contained a more complete and nuanced explanation of the reasons for the Committee's decisions and view of the risks to the outlook than was possible in the post-meeting announcement, and their earlier release would help markets interpret economic developments and predict the course of interest rates. [Federal Open Market Committee (2004)]

In this paper, we apply a statistical method known as Latent Semantic Analysis (LSA) to extract meaning from the FOMC minutes. LSA is an algorithm that identifies common themes across a collection of texts and then characterizes the documents by the relative prevalence of each theme. The first step of LSA is to represent a document as term-frequency vector, i.e. a list of the words in the document with their relative frequencies. These term-frequency vectors are then arranged in order to form a term-document matrix, where each row corresponds to a term and each column represents one document. Finally, themes are extracted using singular value decomposition of the term-document matrix. These themes are derived from the correlations between term frequencies across documents.[2]

While LSA is currently used in the area of text interpretation, information retrieval, and search engine technology — e.g., Berry, Dumais, and O'Brien (1995), Landauer, Laham, Rehder and Schreiner (1997), Landauer, Foltz and Laham (1998), Landauer, Laham and Foltz (1998), and Deerwester, Dumais, Furnas, Landauer and Harshman (1990), — this technique has not yet been applied to Federal Reserve communications.[3] With this methodology, we can analyze the

[2] This approach is similar to the extraction of factor loadings using principal components analysis of a return covariance matrix as discussed in Connor and Korajczyk (1986, 1988). Principal components analysis can be applied to the return covariance matrix when there are more observation dates than assets, which ensures the matrix is invertible. In our case, the term-document matrix has many more terms than documents; therefore the covariance matrix of term frequencies is singular. This necessitates the use of singular value decomposition.

[3] There are several papers that implement automatic classifiers to analyze internet stock board messages. Antweiler and Frank (2004), Das and Chen (2004), as well as Das, Martinez-Jerez and Tufano (2005) use these algorithms to measure the sentiment of messages as bullish, bearish, or neutral. They calibrate each algorithm using a large sample of hand-classified messages. Antweiler and Frank (2005) use automatic classifiers to identify the content of corporate news stories and measure the impact of different types of news on stock prices.

Automated text interpretation methods are applied to Central Bank communications in two relatively recent papers. Payne (2001) creates a Beige Book index by scoring the summary sentence of each Beige Book. He assigns values of -

relationship between themes expressed in FOMC minutes and current and future economic conditions. We can also more precisely measure the effect of FOMC communications on market expectations about future policy rates by examining differences in Treasury market reaction to the minutes release based on the particular themes that are expressed.

LSA affords several improvements over previous studies. Existing subjective text interpretation techniques typically represent meaning discretely along a single dimension (e.g. growth versus recession or ease versus tighten), whereas our approach allows extraction of multiple themes that are not specified a priori and that can vary continuously.[4] This allows us to decipher the meaning of documents based on their content as a whole, thereby extracting more information and accounting for variation in theme expression more accurately.[5]

The manual approach is at further disadvantage because subjective scoring is labor intensive, and scores can be affected by reader bias. In contrast, using LSA, it takes only minutes to process hundreds of documents. Because the LSA technique is automated, it necessarily precludes any user bias. Moreover, as opposed to standard event studies, which merely identify the presence of new information on release dates, we are able to detect what features of the news (as

---

1 to 1 to adjectives that describe economic weakness or strength and calculates the sum to measure the outlook. Cecchetti (2003) measures the frequency of 15 keywords in FOMC minutes and transcripts from 1981-1997 to infer how much attention the FOMC pays to the stock market.

[4] Cook and Hahn (1988) use expert readers to classify discount rate announcements into three types, depending on whether the stated goal is to bring the rate in line with market rates, to respond to macroeconomic conditions, or both. Balke and Petersen (2002) as well as Fettig, Rolnick and Runkle (1999) use expert readers to score the economic outlook as described in the Federal Reserve Beige Book. Zavodny and Ginther (2005) use the Balke and Petersen approach to construct a Beige Book index, and they analyze the stock and bond market reaction to the Beige Book release. Ehrmann and Fratzscher (2004) categorize the first Reuters news story pertaining to central bank statements (speeches, interviews or testimony) in terms of economic outlook (strong or weak) and policy inclination (tightening versus easing). Similarly, Guthrie and Wright (2000) sort Reuters news stories about Royal Bank of New Zealand monetary policy statements into two categories: tighten or loosen. Kohn and Sack (2004) classify 16 of the FOMC Chairman's speeches and testimonies as warnings about equity valuations, while Gurkaynak, Sack and Swanson (2005) use Wall Street Journal commentary to identify whether interest rates move in response to the specifics of the FOMC statement.

[5] While not conducive to quantitative analysis, there is a fairly large literature using a narrative approach to measure Fed policy based on Fed communications. See, for example, Friedman and Schwartz (1963), Chandler (1971), Beckner (1996), DeLong (1997), Edison and Marquez (1998) Mayer (1999), Orphanides (2003), Chappell and McGregor (2004), Romer and Romer (2004), and Broaddus (2004).

determined by the themes) are most relevant to market participants.[6] We can also determine whether larger information shocks are associated with larger market reaction.

In our empirical analysis of FOMC minutes over the period 1987 – 2005, we find that the themes extracted from the minutes are indeed correlated with current and future macroeconomic and financial market indicators.[7] As a result, we would expect that the minutes release could affect interest rates by altering expectations about near-term monetary policy expectations and the future economic outlook. To some extent, these effects may be separated by comparing the responses of short- and long-term rates to an information shock as discussed in Kohn and Sack (2003), Gürkaynak, Sack and Swanson (2005), and Fleming and Piazzesi (2005).

We go on to examine the Treasury market reaction to the minutes release. We observe that FOMC minutes release moves Treasury yields, and more importantly, that the particular reaction depends on the specific themes expressed in the minutes. While results in Reinhart and Sack (2005) suggest that market reaction across FOMC communications is fairly homogeneous, we find that the market reaction within a specific communication type (the minutes) is nuanced; the market reaction depends on the specific themes that are communicated and their context. Ultimately, our analysis shows that qualitative information from complex, economically-relevant texts can be usefully quantified using an automated approach that appears to mirror, to some extent, the information extracted by market participants. Therefore, increased central bank transparency is

[6] Asset price changes in response to the release of Federal Open Market Committee (FOMC) statements are examined in Gurkaynak, Sack and Swanson (2005), while Rigobon and Sack (2004) also measure the effect of FOMC Chairman's testimony to Congress. Kohn and Sack (2004) further expand the set of information events to include the FOMC Chairman's speeches, and Reinhart and Sack (2005) add in speeches of individual FOMC members. Guthrie and Wright (2000) examine Royal Bank of New Zealand central bank official statements and speeches, while Ehrmann and Fratzscher (2004) consider Bank of England, European Central Bank, and Federal Reserve speeches, testimony, and interviews.

[7] There is mixed evidence on whether the Fed forecasts are superior to those of private forecasters. Joutz and Stekler (2000) and Faust, Swanson and Wright (2004) find that Fed forecasts are no more accurate than those from standard statistical models or private forecasters, while Romer and Romer (2000) and Gavin and Mandal (2003) find that Fed forecasts are more accurate. Balke and Petersen (2002) show that information from the Beige Book can improve forecasts of macroeconomic aggregates, while Fettig, Rolnick and Runkle (1999) find that the outlook from the Beige Book is correlated with current and future economic conditions but Beige Book information does not improve on private forecasts of macro aggregates.

meaningful in that financial markets respond to many layers of information in qualitative communication. However, our study does not address whether the market reaction is consistent with the objectives of the FOMC.

The rest of the paper is organized as follows. Section 2 describes the Latent Semantic Analysis approach used to quantify the themes expressed in FOMC minutes. Section 3 describes the data, and Section 4 presents the theme estimation results. Section 5 analyzes the Treasury market reaction to FOMC minutes releases. Section 6 concludes.

2. Latent Semantic Analysis

2.1 The Latent Semantic Analysis approach

LSA is a tool that recognizes patterns in large sets of data.[8] It has many useful applications, both in language-related fields and beyond. In practice, LSA has an impressive ability to model human knowledge acquisition: it has been shown to mimic human performance on certain elements of standardized tests, such as the synonym section of the TOEFL, and has proven its adroitness as an automatic essay grader as discussed in Landauer, Laham and Foltz (1998). It is commonly used in the area of information retrieval and search engine technology, e.g. Berry, Dumais, and O'Brien (1995), Deerwester, Dumais, Furnas, Landauer and Harshman (1990). In addition to language tasks, LSA is employed in gene expression analysis and image search and retrieval, e.g. Wall, Rechtsteiner and Rocha (2003) and Heisterkamp (2002).

In our analysis of FOMC minutes, we utilize LSA as an automated method for extracting the conceptual meaning underlying a body of text. Although a variety of automated approaches

[8] An excellent overview of LSA is given in Landauer, Foltz, and Laham (1998).

exist to capture meaning from text documents (see footnote 3), LSA stands out in terms of its robustness and low labor intensity. As opposed to other automatic classifiers, LSA does not require hand coding a test sample nor manual determination of categories of interest. Instead, the technique deciphers similarities and differences in the existing data based solely on word usage. By grouping words and sets of words within and across documents based on frequency of use, LSA reveals the major associative patterns in the data, which we label "themes." It may appear surprising that documents can be meaningfully characterized without respect to grammar or syntax, but LSA has proven effective in a variety of settings. LSA is particularly conducive to the current study as our chief interest is the information content of documents, not style or mechanics, which can vary with different authors over time.

2.1 The LSA algorithm

The first step in performing LSA is to group *d* documents into a collection called a *corpus*. The textual data is then preprocessed in the following manner. First, we eliminate formatting including capitalization, punctuation, symbols, digits, and key codes. Next, we remove extremely common and function-neutral words that do not contribute to the informational value of the document. These words are referred to in the LSA literature as *stop words*. Stop words typically include pronouns, articles, prepositions, conjunctions, auxiliary verbs, numbers, days, weeks, months, and frequently occurring proper names. The remaining terms are then *stemmed*; stemming is the process of removing suffixes so words that share a common etymological root are mapped to a single "stem." We are left with t_i unique terms in each document, $i = 1,\ldots,d$.

Each document is then used to define a t_i x1 *term-frequency vector*. Each component of the vector corresponds to the number of occurrences of that term within the document divided by the

total number of terms. This within-document relative frequency calculation is an example of a *local weighting scheme*.[9] The intuition is that the more often a term appears within a particular document, the more it contributes to the content of that document. The *d* term-frequency vectors are concatenated to form a *txd term-document matrix* which is denoted by X, where t is the total number of unique words that occur in the filtered text corpus. ($t=\max\{t_i\}, 1 \leq i \leq d$). The X matrix is usually sparse, as a given term may not appear in every document. Also, because the number of terms is usually larger than the number of documents, the *txt* matrix XX^T is singular.

If weights are then applied to the entire term-document matrix, they are referred to as *global weights.* Although a variety of *global weighting* schemes are available as noted in Sebastiani (2002), the most common is term-frequency-inverse-document-frequency (TFIDF). For a given term, the *document frequency* is the number of documents in which the term appears divided by the total number of documents. Under TFIDF, each term frequency is multiplied by the log of the inverse document frequency. The underlying principle is that the more documents in which a term occurs, the less discriminating it is.

Once the term-document matrix has been created, the next step is to apply singular value decomposition (SVD). SVD is a generalized form of principal component analysis that can be used on an arbitrary rectangular matrix. SVD factors the term-document matrix (X) into the product of three characteristic matrices such: $X = USV^T$.[10] The columns of U (*txd*) and V (*dxd*) — commonly referred to as left and right singular vectors of X — represent the orthonormal eigenvectors of XX^T and X^TX, respectively. S (*dxd*) is a diagonal matrix of *singular values*. Its elements are equal to the

[9] An alternative example of local weighting for entries in the matrix is to use a simple binary coding where 1 indicates presence and 0 indicates absence of the term in the document.

[10] If SVD is performed on a mean-centered matrix (X) then the columns of U are equal to the eigenvectors of XX^T, the columns of V are the principal component values of X, and the elements of S are equal to the square roots of the eigenvalues of X^TX. The matrix product US yields the principal component scores.

non-negative square roots of the eigenvalues associated with both XX^T and X^TX, ordered by decreasing magnitude.[11]

The columns of the U matrix can be interpreted as *d* pseudo-documents (as described by mutually orthogonal term-frequency vectors) that characterize the themes of the corpus. Note that each original term-frequency vector can be reconstructed as a linear combination of these k themes. The i,j^{th} element of U defines the contribution of term *i* to theme *j*, that is, how characteristic a particular term is of a given theme. In our Figure 1, which is adapted from Berry, Dumais, and O'Brien (1995), the highlighted portion of U is a vector that shows the significance of each term to the first theme. The themes appear in order of decreasing importance, so that first theme accounts for the most variance (i.e. it is the most dominant concept underlying the text), the second theme accounts for the second largest amount of variance, and so forth.

While the U matrix relates terms to themes, the V matrix relates documents to themes. The i,j^{th} element of V^T defines the contribution of theme *i* to document *j*, that is, how strongly a particular theme is expressed within a given document. Therefore, if documents are arranged chronologically, one row of V defines the time series of loadings for a particular theme. In Figure 1, the shaded region of V^T describes significance of Theme 1 over time.

The S matrix contains the singular values arranged in descending order. The magnitude of the i^{th} singular value indicates the importance of the i^{th} theme in explaining differences across documents. The matrix products US and VS rescale the U and V matrices by theme significance.[12]

2.3 Out-of-sample estimation of LSA themes (folding-in)

[11] SVD is unique up to the sign of the singular vectors.
[12] The relation of S and US (V and VS) is analogous to the relation between principal component values and principal component scores. Principal component scores are created by scaling by the square-root of the corresponding eigenvalue while SVD scores are created by scaling by the singular value, which is the square-root of the eigenvalue.

Folding-in is a procedure that characterizes new texts in terms of the themes derived from an existing latent semantic structure. Thus, folding-in can be used to test the explanatory power of themes extracted from a fixed corpus to new documents not used in the estimation procedure. While it is also possible to apply SVD to a term-document matrix created using a rolling or expanding time window for the document corpus, folding-in is more common, and computationally more economical.

To derive the theme loadings for a new document, it is first filtered using the methods previously described. In addition, terms appearing in the new document that are not already in the corpus are deleted. The resulting term-frequency vector (d) undergoes the following transformation: $\hat{d} = d^T U S^{-1}$.

The transpose of the new document vector, d^T, comprised of relative term frequencies, is multiplied by U, which indicates how much each term contributes to each theme, based on the preexisting document collection. This result is right multiplied by the inverse of S, which differently weights the themes. The new term-frequency vector, $\hat{d}$, is therefore a weighted sum of its constituent theme vectors and can now be appended as a column to form an augmented V matrix. New words are deleted because there is no existing row in U corresponding to the new term, so there is no information available to create the weighted sum. This can be a disadvantage for documents that introduce new vocabulary, since the folding-in process will not be able to represent their content to the fullest extent. However, this potential loss of information is greatly outweighed by the time-saving nature of the technique.

3. Data

3.1 The FOMC minutes sample

Our empirical analysis is based on the FOMC meeting minutes over the period from January 1987 to December 2005. The FOMC convenes for eight regularly scheduled meetings per year, held every five to eight weeks. Additional meetings or telephone conferences are called as needed. From the beginning of our sample period through November 2004, the minutes of each meeting are released to the public two or three days following the next scheduled meeting, which corresponds to a six to eight week lag.[13] Since of the meeting held December 14, 2004, the FOMC has released minutes exactly three weeks after each policy decision.

We obtain an electronic version of the FOMC minutes from the Board of Governors website for meetings from 1996 onward, while the pre-1996 minutes are from Factiva. There are a total of 152 meetings over this period, and the minutes usually range from 5 to 10 single-spaced pages of text.

We delete the introductory section of the minutes, which deals primarily with administrative items. This portion generally includes the list of members attending the meeting, details about newly appointed positions, amendments to committee procedures, and approval of minutes from the previous session. As a general rule, our analysis commences with the line, "The Committee then turned to a discussion of the economic outlook and the conduct of monetary policy over the intermeeting period," or, "The information reviewed at this meeting suggested..." or some similar variation. The final section, which describes the voting actions of the committee, is also omitted. For additional details about the structure of the FOMC minutes, see Danker and Luecke (2005).

[13] Until 1997, the minutes release was on the Friday after the next FOMC meeting. Then, through the end of 2004, the release was changed to the Thursday after the next meeting. In December 2004, the committee agreed to expedite the release of the minutes to three weeks following a policy decision. The first such accelerated release occurred on January 4, 2005.

The remaining body of text then undergoes further filtering. First, we strip the formatting from the document. Next, we eliminate stop words. The most common stop words, as shown in Panel A of Table 1, include *the*, *and*, *that*, *for*, and *was*. We then use the Porter Algorithm to stem the terms.[14] Panel B of Table 1 lists common stems and associated terms for the FOMC minutes. For example, our stemmer maps the terms *increase*, *increased*, *increases*, and *increasing* to *increas*. Finally, we reapply the stop word filter to remove any neutral words that reemerge as a result of stemming. This leaves us with a list of all filtered words appearing in the document. We repeat this process for each document in the collection. An example of this procedure is given in Appendix A.

In Panel C, we report the total and unique term count at various stages of filtering. Before any filtering, the average number of words per minutes release is 3,196. On average, 886 are unique terms. The stop filter reduces the average to 1,836 words of which 722 are unique terms. Stemming further decreases the total word count to 1,766 and the unique term count to 559. Across all documents, there are 2,402 unique terms.

3.2 Macroeconomic and financial market indicators

In Section 4, we relate FOMC minutes themes to changes in macroeconomic and financial market indicators. For this purpose, we collect data on nine market indicators and four macroeconomic indicators. Financial market variables are measured on the last day of each FOMC meeting, and macroeconomic variables are measured using the most recent revised data that would have been available at the meeting.

[14] Porter's stemming algorithm is described in Porter (1980). We use the computer code for this stemmer from http://www.tartarus.org/~martin/PorterStemmer/.

As a gauge of short-term expectations about Fed policy, e.g. Krueger and Kuttner (1996), we use daily Federal funds futures settlement prices from the Chicago Board of Trade. We calculate the expected Federal funds rate as 100 minus the contract price, for contracts expiring two months ahead.[15] We collect additional daily interest rate data from the Federal Reserve H.15 release for more distant points on the yield curve. In particular, we choose the three-month and ten-year constant maturity Treasury yields as well as Moody's BAA corporate bond yield.

The credit spread and term spread have been used as business cycle indicators as discussed in Lahiri and Wang (1996) and Estrella and Hardouvelis (1991). We calculate the credit spread based on the difference between the Moody's BAA corporate bond yield and the ten-year constant maturity Treasury yield. We measure the term spread as the difference between ten-year and three-month constant maturity Treasury yields. All data is from the daily Federal Reserve H.15 release. Intermeeting changes for yield and yield spread variables are calculated as first differences.

We also collect daily data on equity market prices as measured by the S&P500 index level from Datastream, the foreign exchange value of the dollar as reported in the Federal Reserve H.10 Foreign Exchange release based on the broad dollar index, and the price of crude oil using the spot price for light, sweet crude from Datastream. Intermeeting price changes for these variables are calculated using logarithmic differences based on daily data.

Data for our four macroeconomic indicators are gathered from the Haver/DLX database. All series represent latest seasonally-adjusted values as reported at the time of FOMC meetings. We calculate growth rates for GDP, the GDP deflator, and industrial production using logarithmic differences and for unemployment as a first difference. Intermeeting changes in these growth rates are then calculated using differences. Summary statistics for the intermeeting changes for these

[15] Data come from Bloomberg's FF3 and FF4 series, i.e. contracts expiring in two and three months, respectively. In calculating daily changes we take the first difference of the FF3 implied rate. For the first trading day occurring in a given month, corresponding to the contract rollover date, we take $FF3_t$-$FF4_{t-1}$. This ensures that each daily change is calculated using the same futures contract.

variables are reported in Table 2, measured in basis points. The GDP and GDP deflator are reported quarterly, which accounts for the lower number of observations corresponding to these variables.

3.3 Treasury market data

In Section 5, we analyze the Treasury market reaction to FOMC minutes releases over the period from 1997-2005. For this purpose, we collect data for on-the-run three-month, two-year, and ten-year Treasuries from the GovPX database.[16] GovPX data includes tick-by-tick bids, offers, and transactions prices from five of six of the major interdealer brokers. Our measure of the new information released that is relevant to market expectations is the absolute value of the yield change; see, for example, Fleming and Piazzesi (2005).

We measure yields over a short interval that brackets the time of the FOMC minutes release. Using a narrow time window helps to minimize influences on yields from other market factors such as macroeconomic news releases. The time of the minutes release is verified using time stamps from Reuters reports, the time-stamp on the Federal Reserve Board website, and Bloomberg news. For Reuters, we find release times range from 2:01pm to 3:22pm, and from 1:56:11pm to 3:08:40pm for Bloomberg. We select 1:55pm to 2:15pm to guarantee that we bracket the actual release time, while allowing sufficient time for the market to absorb the news. To form our non-event sample, we measure yield changes over the same 20-minute interval on non-release dates.

We construct the yield change using the average of on-the-run bid and ask yields closest but prior to 1:55pm and 2:15pm. Since it is not clear that the market reaction to a theme will

[16] We are especially grateful to Michael Fleming for his assistance with the GovPX data.

always be of a particular sign, we take the absolute value of the yield change as a measure of the news content of the event. We limit the maximum length of our event window to minimize the possible effects of other announcements or market events on our results. If there is no reported yield between 1:30 and 1:55 pm or between 2:15 and 3 pm, then we do not include that interval in the sample.

Our summary statistics for the Treasury market data are reported in Table 3. Panel A displays mean absolute yield changes from 1:55 to 2:15 pm on minutes release dates for Treasury instruments of various maturities. We see that volatility tends to increase with maturity, i.e. ten-year yields are more volatile than two-year yields which are also more volatile than three-month yields. The same is true for non-event days, shown in Panel B. Of the 72 minutes releases over the 1997-2005 period, we can construct a valid event window for 64, 65, and 49 observations using three-month, two-year, and ten-year Treasury yields, respectively.[17] For non-event dates, we have 1917, 1821, and 1364 observations.

It is worth noting that each Treasury instrument exhibits greater volatility over the event window on minutes release dates compared to non-release dates. This provides a first, preliminary confirmation that the FOMC minutes contain information not already incorporated in market expectations of future interest rates. On average, three-month yields fluctuate by .48 basis points on minutes release dates, up from .31 bps on all other dates.[18] We find that two-year and ten-year yields nearly double in volatility on minutes release dates.

To gain a deeper understanding of the Treasury market reaction surrounding minutes releases, we examine absolute yield changes in 5-minute intervals from 1:30 to 3 pm on both

[17] The relatively large number of missing observations for the ten-year Treasury is due to the declining share of trades covered by GovPX as described in Fleming (2003). One alternative that could be pursued is replacement of missing observations using yield changes from ten-year Treasury futures traded on the CBOT.

[18] Gurkaynak and Sack (2006) find that yields from Federal Funds futures contracts outperform three-month Treasury yields in forecasting monetary policy for horizons up to six months. Future research could incorporate intraday Federal Funds futures data.

release and non- release days. First, we calculate the absolute yield changes for each Treasury security over 5-minute intervals centered on 1:30, 1:35, 1:40, …, 2:55, and 3 pm. We then create two sets of indicator variables. The first set of dummies specifies whether the midpoint of each interval lies in a particular time range, e.g. the first indicator designates the 5-minute change surrounds 1:30, the second indicator indicates the change takes place around 1:35, and so on. For the second set of indicators, each variable assumes a value of 1 only if it lies within a given time interval *and* the date corresponds to an FOMC minutes release. We now have a one set of time indicators representing all dates, and a second set corresponding to release dates only.

Next, we regress the absolute yield changes against the full set of indicators, omitting an intercept, for each Treasury instrument. The coefficient on each of the indicators from the first set represents the average volatility for each time period across all dates. The coefficients from the second set represent the amount of additional volatility occurring on minutes release dates. The sum of the two coefficients corresponds to the total volatility over each time interval on event days. Figure 2 graphs the coefficients for all dates and for release dates, with a star signifying whether the volatility for each interval on event days is statistically different at the 10% level from its baseline across all dates. A vertical line denotes 2 pm, the approximate time at which minutes are released to the public. Panels A, B, and C show results for the three-month, two-year, and ten-year Treasuries, respectively.

From the graphs, it is clear that there is higher volatility in Treasury yields following the minutes release. On minutes dates, prior to 2 pm, volatility remains near its mean, but then spikes upward within fifteen minutes following the news release. This heightened volatility persists anywhere from 15 minutes for ten-year Treasuries up to an hour after the event for two-year instruments.

4. Extracting themes from FOMC minutes

4.1 Forming the term-document matrix

Using the FOMC minutes data described in Section 3.1, we construct a column vector with term frequencies for each of the 152 minutes releases. There are 2,402 unique terms in the text corpus, so each vector has 2,402 entries. We then align the vectors to form a 2,402 by 152 term-document matrix denoted as X. Next the rows of the matrix are mean-centered so that X^TX can be interpreted as a covariance matrix.

Perhaps the simplest means of summarizing the information in the minutes is to look at the most frequently occurring terms in each document. In Table 4, we list the terms with highest relative frequency (based on the row averages of the term-document matrix, prior to mean-centering) for the full sample and over four subintervals. In Panel A, the most significant terms relate closely to the primary goals and concerns of the Federal Open Market Committee, as well as the instruments available to them for implementing monetary policy. One prong of its core mandate is to promote steady economic growth. Hence it is not surprising that *growth* is the most important term for three of the four subintervals, as well as for the entire collection of documents. Another goal of the Fed is to maintain price stability, i.e. curb excess inflation or deflation. The word *price* consistently appears in the top 3 for all intervals, and *inflat* (inflation and related terms) is present in each of the top 25 lists, appearing as high as number six in the third interval.

A large portion of each meeting centers on the presentation and discussion of the current economic outlook based on staff forecasts. The committee members review data and research that provide a broad perspective of economic conditions. Among the key topics of interest are output growth, industrial production, consumer spending and confidence, business and financial

developments, labor market conditions, and prices and inflation. These issues are clearly represented in the most frequent word lists. General terms include: *growth, price, market, econom, busi, condit, indic, finan, inflat, consum, spend, product*, while directional terms include: *increas, expans, declin, rise*, etc. In addition to reviewing the most recent data releases, the committee also listens to forecasts and makes projections on the future path of the economy (*expect*). In the final stage of the meeting, the committee formulates and votes upon the policy action to be carried out (*rate, polici, monetari*). It is interesting to note that the results remain consistent over time – e.g. *growth, price, rate, market, econ, increas, polici, busi, inflat* are consistently among the top terms.

In Figure 3, we graph the five most frequent terms over time. The top words are based on the highest average relative frequencies across all documents. Panel A plots the top five words over time from the full sample of minute releases. The word *price* spikes up in mid-1990, during a period of recession, and may reflect fluctuating oil prices during that time. The word *growth* tends to diminish in significance during these recession periods, hitting a minimum at the tail end of the 2001 recession. This informal analysis suggests that there may be merit in applying a more structured procedure to extract conceptual patterns from the data. Panel B displays results from folding-in, which are highly similar, though there is less variance in each series over the second half of the sample period.

Panel B of Table 4 shows the most frequent terms after applying the TFIDF global weighting scheme. These lists tend to fluctuate much more over time, thus reflecting the sensitivity of the measure to infrequent terms. TFIDF places greater weight on words that appear in only a few documents as a means of differentiation. As a result, the most important words relate to major news events or economic shocks. For instance, several significant words from the second interval, 1991-1995, appear related to the Mexican financial crisis that sprang from peso devaluation in 1994: *mexico, swap, depositori, credit,* and *monei*, etc. For the years 1996 to 2000, the most

frequent terms involve the Asian and Latin American financial crises and the introduction of the euro in 1999: *asia, turmoil, wealth, equiti, euro, latin, america*. In the most recent portion of our sample, 2001-2005, key words reflect the catastrophic hurricane Katrina, September 11th and its aftermath, disinflation worries and technology growth: *hurrican, softwar, katrina, attack, disinfl, tech, terrorist, geopolit.*

While TFIDF can make it easier to identify unique themes, the uncommon words generally apply to only a small portion of the data. As a result, their associated themes are less applicable to the collection as a whole, which blurs our subsequent analysis of macroeconomic and market reaction to minutes releases. Nonetheless, it provides an interesting point of comparison.

4.2 Extracting the themes using singular value decomposition

Once the term-document matrix has been constructed, the next step in LSA is to extract the characteristic themes expressed in the text collection. To achieve this, we use singular value decomposition (SVD) to factor the term-document matrix into three components: U, S, and V^T. In Panel A of Table 5, we report the amount each theme contributes to the overall understanding of the texts based on the estimated matrix of singular values (S). We divide each squared singular value (eigenvalue) by the sum of all squared singular values to determine the percent variance explained by each theme. According to Table 5, the first ten themes account for nearly half of the total variance, while the first 100 themes explain approximately 94% of total variance. In the interest of parsimony, we focus on the first five themes in further analysis.

Because each document corresponds to a particular date, we can use the rows of the V matrix (document contribution to each theme) to trace theme evolution over time. In Figure 4 Panel A, we graph the first 5 themes over our sample period. Theme 1 exhibits a strong decline

over the duration of the sample. Theme 2 fluctuates in a wave-like pattern, first decreasing until the beginning of 1992, then rising through mid-1997. The series proceeds to dip again, reaching a trough at the end of the 1991 recession before rapidly increasing to the end of the sample. Theme 3 gradually increases until the middle of 2000 and drops off before the 2001 recession. It then remains relatively steady before taking a pronounced downward trajectory beginning in 2004. Theme 4 increases slightly over time, but remains fairly constant in the middle of sample between contraction periods. Lastly, Theme 5 also resembles a wave, first rising until 1994, then decreasing, exhibiting a trough at the end of the 2001 recession, before increasing once more to the end of the sample. Panel B plots Themes 1-5 using folding-in, where the themes are estimated over the first half of the original sample, spanning FOMC meetings from February 1987 to July 1996. The themes estimated using folding-in are similar to the full-sample themes.

We see that the themes are fairly persistent. The first-order autocorrelations (Table 5, Panel B) are significant at the 5% level, ranging from 0.28 to 0.92. In our empirical analysis we are interested in the new information revealed by each minutes release. Therefore, we define innovations as residuals from first-order autoregression on each of the first five themes in the following analysis.

4.3 Interpreting the themes using key terms

Although it is a difficult task, we attempt to attach some specific meaning to each theme based on patterns in their constituent terms. Our analysis hinges on the columns of the U matrix, which define the contribution of each term to a particular theme. We examine the absolute values of the term loadings, since SVD is only unique up to the sign.

Figure 5 graphs the average term loadings for the ten most important terms in Themes 1-5. Key terms for Theme 1 include *spend, consum, domest, expans, pressur,* and *growth*, which suggests a relationship to consumer spending and sentiment, and perhaps economic expansion. Theme 2 highlights the general macroeconomic conditions with the following important terms: *econom, weak, busi, ease, declin, price, growth,* and *increas*. In Figure 4, Panel A, Theme 2 exhibits a cyclical pattern, with a tendency to decline during periods of recession, further supporting a connection to the business cycle. Predominant terms for Theme 3 include *inflat, rise, demand, labor, tight,* and *price*. The presence of *tight, rise, demand,* and *inflat* imply a growing inflationary pressure coupled with contractionary monetary policy. Theme 4 emphasizes concern over the dollar and foreign exchange market, as revealed by the terms *dollar, market, inflat, condit,* and *pressur*. Finally, Theme 5 appears to reflect the monetary policy stance with words such as *polici, slightli, restraint,* and *consider*.

Panel B shows an analogous graph for themes extracted using folding-in. Theme 1 is similar to its full sample counterpart, characterized by *consum, domest, growth, pressur, expans,* and *aggreg,* again suggesting a focus on consumption and economic growth. Major terms for Theme 2 include *growth, debt, pressur, foreign, dollar, oper, credit,* and *market*. It appears to encompass foreign investment and debt, foreign exchange operations, and dollar strength, thus corresponding to Theme 4 from the full-sample results. Theme 3 under the folding-in technique relates strongly to the economic growth cycle, featuring words such as *eas, weak, declin,* but also *progress, increas, and rise*. In this respect the theme is analogous to Theme 2 from the original sample. The fourth theme adds a financial dimension to the data, with *stock, financi*, and *market* figuring prominently. Finally, Theme 5 again highlights inflation *(price, inflat),* with particular emphasis on energy *(oil).*

4.4 Interpreting the themes using financial and macroeconomic indicators

We verify that the LSA methodology is able to extract meaningful information from the minutes by exploring correlations between themes and macroeconomic and financial market indicators. In Table 6, we look at indicator correlations with contemporaneous and lagged theme innovations, as measured by residuals from a first-order autoregression. In this way, we see how the FOMC discussion is shaped by current economic conditions (Panel A) and the extent to which the concerns expressed by the FOMC track future economic outcomes (Panel B).[19] The indicator data is described in Section 3.2.

In Panel A, we see that theme innovations are contemporaneously correlated with financial market indicators. Themes 3 and 5 exhibit especially high degrees of correlation. For Theme 3, of the ten correlations, two are significant at the 1% level, three at the 5% level, and one at the 10% level. Theme 5 produces three correlations significant at the 5% level and two at the 10% level. The second theme has three significant correlations at 1%, while the fourth has one significant correlation at each the 1, 5, and 10% levels. Theme 1 displays two significant correlations, at the 1 and 10% level.

The financial market indicators that are most correlated with theme levels are the ten-year yield, credit spread, and S&P 500 index; each has 3 significant correlations. The expected federal funds rate and the three-month yields demonstrate two significant correlations, along with crude oil price and the corporate bond yield. Overall, we see that the strongest contemporaneous theme correlations involve short- and long-term interest rates, which is to be expected given their role as a policy tool and their association with economic outlook, respectively.

[19] Because the minutes are released considerably later than the actual FOMC meeting, it is not possible to know with certainty that the minutes are based strictly on information available at the time of the meeting.

The contemporaneous relation of macroeconomic indicators with theme changes is not as strong. Perhaps this is because the most recent quarterly change for these macroeconomic indicators is less important for monetary policy discussions than their longer-run trends. In the lower section of Table 4 Panel A, we see that Themes 2 through 4 exhibit one significant correlation apiece, while Theme 1 is not associated with any of the macro variables. The unemployment rate demonstrates a significant negative relationship to FOMC content as expressed by Themes 4 and 5. GDP growth is significant and positively related to Themes 2 and 3.

Moving to Panel B to examine the predictive power of themes, we see less compelling results. Theme 2 remains strongly positively correlated with both short-term interest rate measures, and is also positively related to crude oil price changes. Themes 1 and 5 have one significant correlation apiece, while 3 and 4 exhibit no significant relationships. The themes have the most predictive power for short-term interest rate behavior as captured by the Federal funds yield and the three-month Treasury yield and, to a lesser extent, crude oil prices. We do not find that the themes have predictive power for the macroeconomic indicators.

Overall, Themes 2 and 3 appear most closely tied to current and future financial market and economic conditions. Theme 2 appears to be cyclical; it rises on higher short -term interest and expected policy rates, narrower credit spreads, and faster GDP growth. In terms of predictive content, Theme 2 is positively correlated with future short-term interest rates and higher energy prices. Theme 3 is more closely related to inflation; it is contemporaneously associated with long-term treasury yields, a steepening yield curve, narrower credit spreads, higher BAA yields, and higher GDP growth. These correlations indicate a positive relationship with economic growth. Theme 4 is related to contemporaneous lower long-term interest rates, a narrowing term spread, a lower BAA yield, positive S&P500 returns, and lower unemployment growth. This theme appears to also track macroeconomic conditions. Theme 5 seems to be related to the monetary policy

stance. It is significantly correlated with the two interest rates most affected by Fed policy: the expected Federal funds rate and three-month Treasury rate.

5. Treasury market reaction to themes in FOMC minutes

In this section, we examine the Treasury market reaction to the FOMC minutes release in order to answer three questions. First, is there information content in the minutes release? In other words, are the minutes a valuable communication tool, or is the economic news they contain already incorporated into expectations? Second, is the market response to the minutes homogeneous or does it depend on the specific themes expressed? This question is intended to address whether the market can extract a complex, multifaceted signal from the minutes. Third, is the market reaction to the themes dependent on uncertainty about Fed policy and the economic outlook? This helps to reveal the degree of sophistication in market participants' response to the minutes, based on whether themes are interpreted differently in changing economic environments.

Our analysis is based on absolute changes in three-month, two-year, and ten-year yields occurring daily over the time interval from 1:55pm to 2:15pm. The sample begins in 1997, at which point the FOMC moved the minutes release time from 4:30pm to 2:00pm. The data is described in detail in Section 3.3. We use a regression of absolute yield changes on explanatory variables to measure information content.

Our regression results are reported in Table 7 with separate panels for three-month yields (Panel A), two-year yields (Panel B), and ten-year yields (Panel C). We show six specifications in each panel, beginning with absolute yield changes from 1:55 to 2:15 pm regressed on a constant, an indicator for minutes release dates, and an additional indicator for accelerated minutes release dates (i.e. those occurring on or after January 4, 2005).

We see an increase in Treasury volatility on minutes dates compared to other dates that is significant at the 1% level for three-month, two-year, and ten-year yields as shown in column (1) for each panel. The relative volatility increase is substantial. For example, the average absolute three-month yield change is .30 basis points over this 20 minute interval on non-release dates compared to .43 after the minutes release (Panel A). Given the accelerated schedule, the increase on release dates rises further to .85.[20]

This evidence supports the hypothesis that there is new information in the minutes. It is also worth noting that these explanatory variables only explain a relatively small fraction of yield changes with adjusted R-squareds of 0.4%, 3.5%, and 4.8%. The increase in explanatory power with maturity seems to support the view that minutes provide more information about the FOMC's view of the longer-run outlook and policy path than about the FOMC's likely short-run policy decisions. However, when additional variables are added to the regression, for example in columns (8) and (9), there is no clear pattern of explanatory power across maturities. Overall, the significance of the minutes in explaining longer-term yields indicates that the information in the market's interpretation of minutes goes beyond a simple clarification of the near-term policy aims of the FOMC.

To investigate the influence of monetary policy uncertainty on the Treasury market reaction to the minutes release, we utilize a measure that is closely related to the market's expectation of the standard deviation of future Federal funds rates. For this purpose, we analyze the Eurodollar futures option implied volatility for contracts expiring in two months; this data is from Bloomberg. We use the Eurodollar contract instead of the Federal funds contract, because

[20] Reinhart and Sack (2005) report average absolute yield changes of 1.7 and 2.3 basis points for eurodollar and two-year Treasury rates on minutes releases dates prior to accelerated minutes release. The effects after accelerated minutes release are 3.4 and 3.8 basis points, respectively. They use a longer time window (75 minutes), which would account for their finding of higher total yield variation.

Federal funds futures options data is only available for a portion of our sample period and these two measures are highly correlated.

In our regressions, we rescale the previous day's three-month eurodollar implied volatility to a 20-minute rate in basis points so that it is in the same units as the dependent variable. Looking at the results in column (2) of Table 7, with the exception of three-month Treasury yields, higher eurodollar implied volatility is associated with higher Treasury market volatility, and the relationship is close to one-to-one. This is not surprising since yield volatility is persistent, and there is a strong common factor across yields.

To determine whether the reaction to the minutes is stronger when monetary policy uncertainty is elevated, we interact eurodollar implied volatility with an indicator for the minutes release date. The two-year and ten-year Treasuries demonstrate a positive relationship to this interacted variable, at the 5 and 1% levels, respectively. That is, in a climate of heightened policy uncertainty, market participants appear to react more strongly to the information in the FOMC minutes. With additional control variables, eurodollar implied volatility and the volatility interaction term are positive and significant for three-month yields as well (Panel A, column 3).

We next consider how the market reaction to FOMC minutes varies with the economic outlook. Estrella and Hardouvelis (1991) show that the difference between the ten-year Treasury yield and three-month Treasury yield is a powerful predictor of a future recession. They find that the probability of recession increases as the yield curve flattens (term spread declines). While not a result from their paper, it is also reasonable to expect that a very steep yield curve could be a signal that the economy is overheating and that inflation would accelerate. One might expect that FOMC communications would be particularly important to market participants as the economy moves away from equilibrium towards recession or overheating.

Our analysis uses the term spread as a proxy for expected business conditions. We construct the term spread using the difference between ten-year and three-month constant maturity Treasury yields from the H.15 release. We then augment the column (2) specification with the term spread and the term spread interacted with the minutes release indicator. Results are shown in column (3). We find that a flatter term structure is associated with higher yield volatility for all three Treasury instruments, although the effect is only statistically significant for the three-month maturity. This is consistent with results in Schwert (1989) who finds that stock market volatility is higher during recessions.

The interaction term is negative and significant for three-month yield volatility, showing that the minutes release is most informative for yields over a short horizon when there are concerns about recession. In contrast, the reaction of the two-year yields to the minutes release is larger than average when the term structure is steeper. The minutes apparently provide more information about intermediate-term yields when there are concerns about excessive growth accompanied by inflation.

While the preceding analysis assumes that FOMC minutes releases are homogeneous, we are able to characterize the news in the minutes releases via the theme loadings extracted using LSA. We are interested in whether Treasury market participants react differently to different themes expressed in the minutes. As a measure of the new information introduced by the minutes release, we use the absolute value of residuals from first-order autoregressions for Themes 1 to 5 in the V matrix.

In column (4) of Table 7, we show results for a specification with all of the previous control variables and the first 5 themes. For three-month yields, none of the themes play a significant role. For two-year yields, Themes 1 and 3 are statistically significant at the 1% level with negative coefficients, while Theme 2 is positive and significant at the 5% level. Finally, for

the ten-year instrument, Theme 1 is significant at the 5% level with a negative sign while Theme 3 is significant and positive at the 10% level.

Our results indicate that market expectations of future intermediate- and long-term yields are driven by the specific information conveyed in the minutes. Interestingly, some of the coefficients on the themes are negative, which suggests that an information shock does not always lead to increased yield volatility. It is possible that stronger or weaker expression of certain themes could elucidate the Fed's views and reduce uncertainty.

In our next two sets of regressions, we explore whether the market interpretation of a given theme depends on economic conditions. In the column (5) specification, we interact theme innovations with eurodollar implied volatility to proxy for uncertainty about Fed policy. As shown in Panel A, conditioning on policy uncertainty is clearly important for the three-month yield reaction to the minutes. Now, three themes are significant and all have positive signs, and three theme interaction terms are also significant. Conditioning on the economic outlook (Panel A, column (6)) also increases the explanatory power of the themes. Interactions with the term spread result in one theme and one theme interaction being significant (Panel A, column (6)).

For two-year yields, adding interaction effects with policy uncertainty make Themes 4 significant and positive, and the corresponding interaction effect is significant and negative. Apparently, these themes are interpreted differently according to the level of policy uncertainty. There is also evidence that the implications of the themes for two-year yields are most accurately understood in light of the economic outlook. In column (6), Theme 2, 3, and 4 interactions with the term spread are statistically significant, while Themes 1 and 4 are significant as well. For two-year yields, interactions with economic conditions improve explanatory power more than interactions with policy uncertainty.

Results for ten-year yields are shown in Panel C. Implied volatility interactions are for Themes 1 and 2 are significant and negative while the corresponding themes are significant and positive. Term spread interactions with Themes 1 and 3 are significant, and the interaction effects subsume the importance of the themes alone. In this case, the relevant information from the themes can only be extracted in the context of expected economic conditions.

Overall, our method of characterizing the content of FOMC minutes documents does add value compared to a minutes indicator variable. For long-term yields, the adjusted R-squared for models incorporating theme innovations is higher than that for models with minutes indicators and implied volatility measures alone. For both two-year and ten-year regressions, models with themes outperform those without across the board. Also, individual theme variables exhibit significant coefficients, even in the presence of the minutes indicator, for all three Treasury instruments. We identify increased market reaction to the minutes when policy uncertainty is higher and when there is greater concern about inflation; this effect cannot be detected using an indicator variable alone. Adding theme interactions often causes the minutes indicator to become insignificant, which shows that the themes more accurately measure information content.

In Panel B, we perform the same regressions with themes extracted using the folding-in method. Results based on the folding-in are highly similar to the full sample results. The adjusted R-squared are very close to those in Panel A; the minutes indicators, implied volatility, and term spread variables also are very similar for the full-sample and folding-in results.

5. Conclusions

Understanding how market participants interpret FOMC communications requires an ability to map complex, qualitative information into quantitative measures. Previous studies

examining Federal Reserve communications have generally relied on expert reading and scoring of documents or an indicator variable approach. In this paper, we apply an objective, statistical methodology known as Latent Semantic Analysis to interpret FOMC minutes. With this technology, informational content is captured in the form of themes estimated using singular value decomposition of a matrix containing word frequencies derived from FOMC minutes. We interpret these themes based on their constituent terms and by observing theme correlations with macroeconomic and financial market indicators.

When we measure the Treasury market reaction to these themes at the time minutes are released, we discover a nuanced response. Yield changes depend on the specific themes expressed, the level of monetary policy uncertainty, and the prevailing economic outlook. Our evidence suggests that market participants can extract a complex, multifaceted signal from the minutes.

Appendix A: Example of the document filtering procedure

1. Raw data (excerpt from the FOMC minutes from the meeting of November 17, 1998):

The information reviewed at this meeting suggests some moderation in the expansion of economic activity from a brisk pace during the summer months. growth in nonfarm payroll employment slowed appreciably in September and October; the civilian unemployment rate remained near 4-1/2 percent. Industrial production has declined slightly in recent months. Business inventory accumulation was sizable in the third quarter, and stock- sales ratios rose to uncomfortable levels in some sectors strongly affected by the nation's trade deficit.

2. Formatting is removed and the stop word list is applied (stop words are italicized):

the information reviewed *at this meeting* suggests *some* moderation *in the* expansion *of* economic activity *from a* brisk pace *during the summer months* growth *in* nonfarm payroll employment slowed appreciably *in september and october the* civilian unemployment rate remained near percent industrial production *has* declined slightly *in* recent *months* business inventory accumulation *was* sizable *in the third quarter and* stock sales ratios rose *to* uncomfortable levels *in some* sectors strongly affected *by the* nations trade deficit

2. Stop words are deleted:

information reviewed suggests moderation expansion economic activity brisk pace growth nonfarm payroll employment slowed appreciably civilian unemployment rate remained near industrial production declined slightly business inventory accumulation sizable stock sales ratios rose uncomfortable levels sectors strongly affected nations trade deficit

3. Stemmer is applied (suffixes to be deleted by stemmer are italicized) and stop words are deleted:

inform*ation* review*ed* suggest*s* moder*ation* expans*ion* econom*ic* activ*ity* brisk pace growth nonfarm payrol*l* employ*ment* slow*ed* appreci*ably* civilian unemploy*ment* rate remain*ed* near industri*al* product*ion* declin*ed* slightli busi*ness* inventori accumul*ation* sizable stock sale*s* ratio*s* rose uncomfort*able* level*s* sector*s* strongli affect*ed* nation*s* trade deficit

4. Final alphabetized list of unique terms:

accumul activ affect appreci brisk busi civilian declin deficit econom employ expans growth industri inform inventori level moder nation near nonfarm pace payrol product rate ratio remain review rose sale sector sizabl slightli slow stock strongli suggest trade uncomfort unemploy

Bibliography

Antweiler, Werner, and Murray Z. Frank, 2004, Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *Journal of Finance* 59, 1259-1295.

Antweiler, Werner, and Murray Z. Frank, 2005, The Market Impact of Corporate News Stories, Manuscript.

Balke, Nathan S., and D'Ann Petersen, 2002, How Well Does the Beige Book Reflect Economic Activity: Evaluating Qualitative Information Quantitatively, *Journal of Money Credit and Banking* 34, 114-136.

Beckner, Steven K., 1996. *Back from the Brink: The Greenspan Years* (John Wiley, New York).

Bernanke, Ben S., 2004, Fedspeak, Remarks to the American Economic Association, San Diego.

Bernanke, Ben S. and Kenneth N. Kuttner, 2005, What Explains the Stock Market's Reaction to Federal Reserve Policy? *Journal of Finance* 60, 1221-1257.

Berry, Michael W., Susan T. Dumais, and Gavin O'Brien, 1995. Using Linear Algebra for Intelligent Information Retrieval, *SIAM Review*, 37, 573-595.

Blinder, Alan, Charles Goodhart, Phillip Hildebrand, David Lipton, and Charles Wyplosz, 2001. *How Do Central Banks Talk?* (Centre for Economic Policy Research, London).

Broaddus, J. Alfred Jr., 2004, Macroeconomic Principles and Monetary Policy, *Federal Reserve Bank of Richmond Economic Quarterly* 90, 1-9.

Carpenter, Seth B., 2004, Transparency and Monetary Policy: What Does the Academic Literature Tell Policymakers? Finance and Economics Discussion Series 2004-35, Board of Governors of the Federal Reserve System.

Cecchetti, Stephen G., 2003, What the FOMC Says and Does When the Stock Market Booms, in A. Richards, and T. Robinson, eds.: Asset Prices and Monetary Policy: Proceedings of the Research Conference of the Reserve Bank of Australia.

Chandler, Lester V., 1971. *American Monetary Policy, 1928-1941* (Harper and Row, New York).

Chappell, Henry W. Jr., and Rob Roy McGregor, 2004, Fed Chat: FOMC Transcripts and the Politics of Monetary Policymaking, Manuscript.

Cochrane, John H., and Monika Piazzesi, 2002, The Fed and Interest Rates - A High-Frequency Identification, *American Economic Review* 92, 90-95.

Connor, Gregory, and Robert A. Korajczyk, 1986, Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis, *Journal of Financial Economics* 15, 373-394.

Connor, Gregory, and Robert A. Korajczyk, 1988, Risk and Return in an Equilibrium APT: Application of a New Test Methodology, *Journal of Financial Economics* 21, 255-289.

Cook, Timothy, and Thomas Hahn, 1988, The Information Content of Discount Rate Announcements and Their Effect on Market Interest Rates, *Journal of Money, Credit and Banking* May, 167-180.

Cook, Timothy, and Thomas Hahn, 1987, The Reaction of Interest Rates to Unanticipated Federal Reserve Actions and Statements: Implications for the Money Announcement Controversy, *Economic Inquiry* 25, 511-534.

Cook, Timothy, and Thomas Hahn, 1989, The Effect of Changes in the Federal funds Rate Target on Market Interest Rates in the 1970s, *Journal of Monetary Economics* 24, 331-351.

Danker, Deborah J., and Matthew M. Luecke, 2005, Background on FOMC Meeting Minutes, *Federal Reserve Bulletin* 175-179.

Das, Sanjiv R., and Mike Y. Chen, 2004, Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, Manuscript.

Das, Sanjiv R., Asis Martinez-Jerez, and Peter Tufano, 2005, eInformation: A Clinical Study of Investor Discussion and Sentiment, *Financial Management* 34 103-137.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman, 1990, Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science* 41, 391-407.

DeLong, J. Bradford. 1997, eds. Chicago, pp., 1997, America's Peacetime Inflation: The 1970's, in Christina D. Romer, and David H. Romer, eds.: *Reducing Inflation: Motivation and Strategy* (University of Chicago Press for NBER, Chicago).

Demiralp, Selva, and Oscar Jorda, 2004, The Response of Term Rates to Fed Announcements, *Journal of Money, Credit, and Banking* 36, 387-405.

Dueker, Michael J., 1992, The Response of Market Interest Rates to Discount Rate Changes, *Federal Reserve Bank of St. Louis Review* 74, 78-91.

Edison, Hali J., and Jaime Marquez, 1998, US Monetary Policy and Econometric Modeling: Tales from the FOMC Transcripts 1984-1991, *Economic Modelling* 15, 411-428.

Ehrmann, Michael, and Marcel Fratzscher, 2004, Central Bank Communication: Different Strategies, Same Effectiveness? European Central Bank.

Estrella, Arturo, and Gikas A. Hardouvelis, 1991, The Term Structure as a Predictor of Real Economic Activity, *Journal of Finance* 46, 555-576.

Faust, Jon, Eric T. Swanson, and Jonathan H. Wright, 2004, Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy? *Contributions to Macroeconomics* 4, 1-26.

Federal Open Market Committee, 2004, Minutes of the Federal Open Market Committee (December 14, 2004).

Fettig, David, Arthur J. Rolnick, and David E. Runkle, 1999, The Federal Reserve's Beige Book: A Better Mirror Than Crystal Ball, *Federal Reserve Bank of Minneapolis The Region* 10-13, 28-32.

Fleming, Michael J., Measuring Treasury Market Liquidity, *Federal Reserve Bank of New York Economic Policy Review* 9, 83-108.

Fleming, Michael J., and Monika Piazzesi, 2005, Monetary Policy Tick-by-Tick, Manuscript.

Fleming, Michael J. and Eli M. Remolona, 1997, What Moves the Bond Market? *Federal Reserve Bank of New York Economic Policy Review* 3, 31-50.

Friedman, Milton, and Anna Jacobson Schwartz, 1963. *A Monetary History of the United States, 1867-1960* (Princeton University Press for NBER, Princeton).

Gavin, William T., and R.J. Mandal, 2003, Evaluating FOMC Forecasts, *International Journal of Forecasting* 19, 655-667.

Geraats, Petra, 2002, Central Bank Transparency, *Economic Journal* 112, F532-F565.

Gürkaynak, Refet S. and Brian Sack, 2006, Market-Based Measures of Monetary Policy Expectations, Manuscript.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson, 2005, Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements, *International Journal of Central Banking* 1, 55-93.

Guthrie, G., and J. Wright, 2000, Market Implemented Monetary Policy with Open Mouth Operations, *Journal of Monetary Economics* 489-516.

Heisterkamp, Douglas R., 2002, "Building a Latent Semantic Index of an Image Database from Patterns of Relevance Feedback," *icpr*, p. 40134, 16th International Conference on Pattern Recognition (ICPR'02) - Volume 4..

Jenkins, Paul, 2001, Communicating Canadian Monetary Policy: Towards Greater Transparency, *Bank of Canada Review* 45-50.

Joutz, Fred, and H.O. Stekler, 2000, An Evaluation of the Predictions of the Federal Reserve, *International Journal of Forecasting* 16, 17-38.

Kohn, Donald and Brian Sack, 2004, Central Bank Talk: Does It Matter and Why? *Macroeconomics, Monetary Policy, and Financial Stability* (Ottawa: Bank of Canada).

Krueger, Joel T., Kuttner, Kenneth N., 1996. The Fed Funds Future Rate as a Predictor of Federal Reserve Policy, *Journal of Futures Markets* 16, 865–879.

Kuttner, Kenneth N., 2001, Monetary Policy Surprises and Interest Rates: Evidence from the Federal funds Futures Market, *Journal of Monetary Economics* 47, 523-544.

Lahiri, Kajal, and Jiazhuo G. Wang, 1996, Interest Rate Spreads as Predictors of Business Cycles, in G.S. Maddala, and C.R. Rao, eds.: *Handbook of Statistics* (Elsevier Science, New York).

Landauer, Thomas. K., Peter W. Foltz, and Darrell Laham, 1998, Introduction to Latent Semantic Analysis, *Discourse Processes* 25, 259-284.

Landauer, Thomas K., Darrell Laham, and Peter W. Foltz, 1998, Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge).

Landauer, Thomas K., Darrell Laham, Bob Rehder, and M. E. Schreiner, 1997, How Well Can Passage Meaning be Derived Without using Word Order? A Comparison of Latent Semantic Analysis and Humans, in *Proceedings of the 19th Annual Meeting of the Cognitive Science Society* (Erlbaum, Mahwah).

Lombra, Raymond E., and Raymond G. Torto, 1977, Discount Rate Changes and Announcement Effects, *Quarterly Journal of Economics* 91, 171-177.

May, Don O., 1992, A Reexamination of Market Returns, Discount Rate Changes, and Market Efficiency, *Journal of Macroeconomics* 14, 545-553.

Mayer, Thomas, 1999. *Monetary policy and the great inflation in the United States: The Federal Reserve and the failure of macroeconomic policy, 1965-79* (Elgar, Northampton).

Orphanides, Athanasios, 2003, The Quest for Prosperity without Inflation, *Journal of Monetary Economics* 50, 633-663.

Payne, David R., 2001, Anticipating Monetary Policy with the Federal Reserve's Beige Book: Re-specifying the Taylor Rule, *Business Economics* 21-29.

Poole, William, 2005, FOMC Transparency, *Federal Reserve Bank of St. Louis Review* 87, 1-9.

Poole, William, 2003, Fed Transparency: How, Not Whether, *Federal Reserve Bank of St. Louis Review* 85, 1-8.

Porter, Martin, 1980, An Algorithm for Suffix Stripping, *Program* 14, 130-137.

Posen, Adam, 2004, Six Practical Views of Central Bank Transparency, in Paul Mizen, ed.: *Central Banking, Monetary Theory and Practice: Essays in Honour of Charles Goodhart* (Edward Elgar, London).

Reinhart, Vincent and Brian P. Sack, 2005, Grading the Federal Open Market Committee's Communications, Manuscript.

Rigobon, Roberto, and Brian P. Sack, 2004, The Impact of Monetary Policy on Asset Prices, Manuscript.

Romer, Christina D., and David H Romer, 2000, Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review* 90, 429-457.

Romer, Christina D., and David H Romer, 2004, Choosing the Federal Reserve Chair: Lessons from History, *Journal of Economic Perspectives* 18, 129-162.

Sebastiani, Fabrizio, 2002, Machine Learning in Automated Text Categorization, *ACM Computing Surveys* 34, 1-47.

Schwert, G. W. 1989, Why Does Stock Market Volatility Change Over Time?, *Journal of Finance* 44, 1115-1153.

Smirlock, Michael, and Jess Yawitz, 1985, Asset Returns, Discount Rate Changes, and Market Efficiency, *Journal of Finance* 40, 1141-1158.

Thornton, Daniel L., 2003, Monetary Policy Transparency: Transparent about What? *Manchester School* 71, 478-497.

Thornton, Daniel L., 2004, The Fed and Short-term Rates: Is It Open Market Operations, Open Mouth Operations or Interest Rate Smoothing? *Journal of Banking and Finance* 28, 475-498.

Wall, Michael E., Andreas Rechtsteiner, and Luis M. Rocha, 2003, Singular Value Decomposition and Principal Component Analysis, in D.P. Berrar, W. Dubitzky, and M. Granzow, eds.: *A Practical Approach to Microarray Data Analysis* (Kluwer, Norwell, MA).

Zavodny, Madeline, and Donna K. Ginther, 2005, Does the Beige Book Move Financial Markets? *Southern Economic Journal* 72, 138-151.

Table 1: FOMC minutes data

This table reports summary statistics for FOMC minutes over the period from January 1987 to December 2005. Panel A lists the twenty most frequent stop words. The frequency reported is number of uses of the stop word divided by the total number of stop words in the text corpus. Panel B exhibits some common words in stemmed form along with their unstemmed counterparts. Panel C reports the word count under various stages of preprocessing, as well as the number of unique terms used.

Panel A. Most frequent stop words

Stop word	Frequency
the	22.2%
and	8.0%
that	4.0%
for	3.8%
was	2.5%
committee	1.8%
were	1.7%
with	1.7%
had	1.7%
over	1.4%
some	1.3%
would	1.3%
quarter	1.2%
year	1.2%
period	1.1%
this	1.1%
from	1.1%
further	1.0%
meeting	1.0%
reserve	0.9%

Panel B. Stemmer mappings

Original word	Stem
business	busi
businesses	
consumer	consum
consumers	
continue	continu
continued	
continues	
continuing	
continuity	
decline	declin
declined	
declines	
declining	
economic	econom
economy	economi
economies	
expectation	expect
expectations	
expected	
fund	fund
funds	
funded	
funding	
growth	growth
increase	increas
increased	
increases	
increasing	
market	market
markets	
policy	polici
policies	
price	price
prices	
rate	rate
rates	

Panel C. Word counts during preprocessing

	Words	
	Mean	Std. dev.
Total	3,196	1,093
After stop word filter	1,836	619
After stemming	1,766	595

	Unique terms	
	Mean	Std. dev.
Total	886	196
After stop word filter	722	170
After stemming	559	117

Table 2: Summary statistics for financial market and macroeconomic indicators

This table reports characteristics of financial market and macroeconomic indicators over our sample period from 1987 to 2005. (Federal funds rate, dollar index, and GDP-related series commence in November 1988, February 1990, and 1990Q3, respectively). Mean and standard deviation are presented in basis points. Autocorr is the first-order autocorrelation. Financial market variables are measured on the last day of each FOMC meeting.

The expected Federal Funds rate is estimated using Fed Funds futures prices, Treasury yields are constant maturity Treasury rates from the Federal Reserve H.15 release, the Term spread is the absolute deviation of the 10-year/3-month yield curve slope from its average over the period The BAA corporate yield is from Moody's as reported in the St. Louis Fed FRED database, and the credit spread is the difference between the BAA yield and the ten-year Treasury yield. Intermeeting changes for yields and yield spreads are calculated as first differences.

The S&P500 index level is from Datastream, the Dollar Index is the broad dollar index from the Federal Reserve H.10 release, and the crude oil price is the spot price of light, sweet crude from Datastream. Intermeeting changes for these three asset prices are calculated using log-differences.

Macroeconomic indicators are calculated using the most recent revised data that would have been available at each FOMC meeting. All variables are from the Haver database. The GDP and GDP deflator are measured quarterly, while industrial production and the unemployment rate are measured monthly. Growth rates for the first three variables are calculated using log-differences. Intermeeting changes are calculated using first differences.

Panel A. Financial market indicators (intermeeting change)

Indicator	Num. obs.	Mean	Std. dev.	Autocorr.	
Expected Federal Funds rate	127	-2.961	29.639	0.485	***
Three-month Treasury yield	147	-1.293	32.459	0.237	***
Ten-year Treasury yield	147	-1.197	38.091	-0.109	
Term spread	147	-0.095	35.331	-0.044	
BAA corporate bond yield	149	-1.953	28.360	-0.044	
Credit spread	147	-0.626	19.760	-0.016	
S&P500	151	100.807	469.245	0.054	
Dollar index	123	-7.023	234.892	0.073	
Crude oil	151	79.528	1122.637	0.020	

Panel B. Macroeconomic indicators (intermeeting change)

Indicator	Num. obs.	Mean	Std. dev.	Autocorr.	
GDP growth	60	0.077	56.694	0.219	***
GDP deflator growth	60	-0.856	23.729	0.019	***
Industrial production growth	150	5.281	70.214	-0.515	***
Unemployment rate	151	-1.192	16.690	-0.421	

Table 3: Treasury yield data

This table reports summary statistics for the absolute value of Treasury yield changes from 1:55pm to 2:15pm on minutes release dates and all other dates over the period from 1997-2004. Units are basis points. Yields are bid yields closest to but before the respective cutoff as reported in the GovPX database for on-the-run three-month, two-year, and ten-year Treasuries. Additional detail is given in Section 3.3.

Panel A. Minutes release dates

	Absolute value of yield change (1:55pm - 2:15pm)				
	Num. obs.	Mean	Std. dev.	Skewness	Kurtosis
Three-month yield	65	0.48	0.91	4.32	27.31
Two-year yield	64	1.19	1.35	1.70	5.46
Ten year yield	49	1.27	1.69	2.99	13.20

Panel B. All other dates

	Absolute value of yield change (1:55pm - 2:15pm)				
	Num. obs.	Mean	Std. dev.	Skewness	Kurtosis
Three-month yield	1917	0.31	0.51	3.88	30.12
Two-year yield	1821	0.61	0.73	7.12	128.75
Ten year yield	1364	0.69	0.73	3.00	18.83

Table 4: Most frequent terms used in FOMC minutes

This table shows the 25 most frequently appearing stemmed words in FOMC minutes from 1987 to 2004, divided into four subperiods and for the full sample. In Panel A, the term frequencies (freq.) are the number of occurences of the term divided by the total number of stemmed terms per document, averaged over each interval. In Panel B, the term frequencies are multiplied by the inverse document frequency before they are averaged. The inverse document frequency is the log inverse of the number of documents in which the term appears divided by the total number of documents.

Panel A. No global weighting

1987-1990		1991-1995		1996-2000		2001-2005		Full sample	
Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
growth	2.0%	growth	1.7%	growth	1.8%	price	1.4%	growth	1.7%
rate	1.7%	rate	1.6%	price	1.7%	econom	1.3%	rate	1.5%
price	1.4%	price	1.3%	rate	1.5%	growth	1.3%	price	1.5%
market	1.2%	econom	1.2%	market	1.4%	rate	1.2%	market	1.2%
rang	1.2%	increas	1.1%	increas	1.3%	market	1.2%	econom	1.2%
increas	1.1%	polici	1.1%	inflat	1.2%	increas	1.0%	increas	1.1%
expans	1.0%	busi	0.9%	econom	1.1%	busi	1.0%	polici	1.0%
monetari	0.9%	market	0.9%	continu	0.9%	polici	1.0%	continu	0.9%
pressur	0.9%	continu	0.9%	polici	0.8%	continu	1.0%	busi	0.9%
econom	0.9%	monetari	0.8%	consum	0.8%	spend	1.0%	inflat	0.9%
polici	0.9%	rang	0.8%	rise	0.8%	consum	1.0%	consum	0.8%
continu	0.9%	declin	0.8%	remain	0.7%	inflat	0.9%	expans	0.7%
condit	0.8%	activ	0.7%	moder	0.7%	remain	0.8%	monetari	0.7%
busi	0.8%	expans	0.7%	expans	0.7%	economi	0.7%	activ	0.7%
indic	0.7%	consum	0.7%	level	0.7%	activ	0.7%	remain	0.7%
declin	0.7%	expect	0.7%	product	0.7%	level	0.7%	declin	0.7%
domest	0.7%	economi	0.7%	economi	0.7%	declin	0.7%	economi	0.7%
aggreg	0.6%	direct	0.6%	rang	0.7%	product	0.7%	rang	0.7%
expect	0.6%	inflat	0.6%	busi	0.7%	expect	0.7%	spend	0.7%
develop	0.6%	develop	0.6%	demand	0.7%	pace	0.6%	expect	0.7%
inflat	0.6%	financi	0.6%	labor	0.7%	condit	0.6%	level	0.6%
activ	0.6%	spend	0.6%	activ	0.6%	inventori	0.6%	condit	0.6%
remain	0.6%	remain	0.6%	expect	0.6%	sale	0.6%	product	0.6%
dollar	0.6%	level	0.6%	direct	0.6%	risk	0.6%	financi	0.5%
pace	0.6%	indic	0.6%	financi	0.6%	monetari	0.6%	indic	0.5%

Panel B. Term-frequency-inverse-document-frequency (TFIDF) weights

1987-1990		1991-1995		1996-2000		2001-2005		Full sample	
Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
consult	0.14	recoveri	0.12	asia	0.22	hurrican	0.45	hurrican	0.12
rang	0.14	eas	0.11	tight	0.13	softwar	0.23	eas	0.08
manag	0.13	restraint	0.10	turmoil	0.11	particip	0.22	softwar	0.08
drought	0.12	restructur	0.09	wealth	0.09	katrina	0.20	particip	0.08
depend	0.12	discount	0.09	upper	0.09	upsid	0.18	rang	0.07
left	0.12	rang	0.09	compens	0.09	core	0.18	core	0.07
mainten	0.12	care	0.09	equiti	0.09	directi	0.17	asia	0.07
specifi	0.12	institut	0.09	bound	0.08	attain	0.16	recoveri	0.07
thrift	0.11	mexico	0.09	rang	0.07	releas	0.15	tight	0.06
behavior	0.11	monitor	0.08	core	0.07	roughli	0.15	restraint	0.06
eas	0.11	defens	0.08	share	0.07	downsid	0.14	tax	0.06
specif	0.10	depositori	0.08	strike	0.07	attack	0.13	upsid	0.06
restraint	0.10	credit	0.08	deceler	0.07	disinfl	0.13	releas	0.06
fluctuat	0.10	lesser	0.08	euro	0.07	shortli	0.13	gdp	0.06
sought	0.10	broader	0.07	partner	0.06	perceiv	0.13	discount	0.05
tent	0.09	swap	0.07	tech	0.06	solid	0.13	katrina	0.05
furth	0.09	constraint	0.07	acceler	0.06	target	0.12	target	0.05
procedur	0.09	reduct	0.07	latin	0.06	forese	0.12	credit	0.05
boundari	0.09	prefer	0.07	suppli	0.06	tech	0.12	solid	0.05
sensit	0.09	monei	0.07	care	0.06	terrorist	0.12	announc	0.05
inflationari	0.09	accept	0.06	wane	0.06	press	0.12	downsid	0.05
reserv	0.09	matur	0.06	softwar	0.06	stimulu	0.12	stimulu	0.05
impact	0.09	presumpt	0.06	announc	0.06	geopolit	0.12	behavior	0.05
emphasi	0.09	effort	0.06	subdu	0.06	equal	0.11	tech	0.05
borrow	0.09	depress	0.06	america	0.06	com	0.11	weather	0.05

Table 5: Estimation of FOMC minutes themes

This table presents results from Singular Value Decomposition. Panel A summarizes the amount each theme contributes to the overall understanding of the text corpus based on the matrix of singular values (S). We divide each squared singular value (eigenvalue) by the sum of all squared singular values to determine the percent variance explained by each theme. Panel B displays summary statistics for each of the first five themes as expressed by the V matrix.

Panel A. Variance explained by themes

Theme	Singular Value	Eigenvalue	Variance explained	Cumulative variance explained
1	15.2	231.1	14.6%	14.6%
2	11.6	134.3	8.5%	23.1%
3	9.7	94.7	6.0%	29.1%
4	8.1	64.8	4.1%	33.2%
5	7.4	54.8	3.5%	36.7%
6	6.4	40.7	2.6%	39.3%
7	6.2	38.9	2.5%	41.8%
8	6.0	36.2	2.3%	44.1%
9	5.7	32.2	2.0%	46.1%
10	5.1	26.2	1.7%	47.8%
50	2.6	6.8	0.4%	80.2%
100	1.7	2.9	0.2%	94.3%

Panel B. Summary statistics for theme levels

Theme	Num. Obs.	Mean	Std. dev.	Skewness	Kurtosis	Autocorr.	
1	152	0.00	0.08	0.14	2.17	0.84	***
2	152	0.00	0.08	0.32	2.96	0.92	***
3	152	0.00	0.08	0.05	3.11	0.92	***
4	152	0.00	0.08	-0.23	3.51	0.28	***
5	152	0.00	0.08	0.14	2.31	0.47	***

Table 6: Correlations of minutes themes with macroeconomic and financial market indicators

This table shows the correlations between first-order autoregressive theme residuals and intermeeting changes in macroeconomic and financial market indicators over the period 1987-2005. Further details on the financial market and macroeconomic variables are given in the text in Section 3.2, and details on the estimated themes are given in Section 4.2. Panel A reports contemporaneous correlations to detect the relation between current economic conditions and themes expressed in the FOMC minutes. Panel B reports correlations of themes with the lead of the indicator variable (the realization over the next intermeeting period). These correlations measure predictive content of the minutes themes for indicator variables. Statistical significance is indicated by *, **, *** which correspond to significance levels of 10%, 5%, and 1%.

Panel A. Contemporaneous correlations with AR1 theme residuals

Financial market indicators

Indicator	Theme 1		Theme 2		Theme 3		Theme 4		Theme 5	
Expected Federal Funds rate	-0.02		0.34	***	0.11		0.11		0.17	*
Three-month Treasury yield	0.00		0.23	***	0.07		-0.01		0.20	**
Ten-year Treasury yield	-0.09		0.14		0.22	***	-0.14	*	0.18	**
Term spread	-0.10		-0.07		0.17	**	-0.14	*	0.01	
BAA corporate bond yield	-0.10		0.03		0.15	*	-0.21	***	0.13	
Credit spread	0.04		-0.22	***	-0.21	***	-0.03		-0.17	**
S&P500	0.15	*	0.08		-0.02		0.18	**	0.14	*
Dollar index	-0.02		0.10		0.19	**	0.09		0.09	
Crude oil	-0.21	***	0.03		0.18	**	-0.04		0.08	

Macroeconomic indicators

Indicator	Theme 1		Theme 2		Theme 3		Theme 4		Theme 5	
GDP growth	-0.07		0.31	**	0.22	*	0.05		0.05	
GDP deflator growth	-0.02		0.14		0.00		0.17		-0.17	
Industrial production growth	-0.11		0.00		0.09		-0.04		0.02	
Unemployment rate	0.03		0.10		-0.04		-0.02	**	-0.09	***

Panel B. Lead correlations with AR1 theme residuals

Financial market indicators

Indicator	Theme 1		Theme 2		Theme 3		Theme 4		Theme 5	
Expected Federal Funds rate	-0.07		0.19	**	0.14		0.07		0.16	*
Three-month Treasury yield	0.06		0.23	***	0.13		0.10		0.06	
Ten-year Treasury yield	0.01		0.07		0.01		0.10		0.07	
Term spread	-0.05		-0.13		-0.10		0.02		0.02	
BAA corporate bond yield	-0.06		0.06		0.05		0.04		0.06	
Credit spread	-0.09		-0.07		0.04		-0.13		-0.05	
S&P500	0.04		0.09		0.13		0.00		0.03	
Dollar index	0.13		0.11		0.04		0.11		-0.08	
Crude oil	0.02	*	0.13	*	0.02		0.19		-0.03	

Macroeconomic indicators

Indicator	Theme 1	Theme 2	Theme 3	Theme 4	Theme 5
GDP growth	-0.10	-0.05	0.15	0.08	-0.04
GDP deflator growth	0.00	-0.07	0.00	-0.06	0.03
Industrial production growth	-0.02	0.05	-0.03	0.09	0.05
Unemployment rate	0.08	-0.01	-0.06	-0.02	-0.04

Table 7: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)

This table presents the results of regressions designed to model the Treasury market reaction to the public release of FOMC minutes. The dependent variable is the absolute change in the average of bid and ask yields (measured in basis points) for three-month, two-year, and ten-year treasuries from 1:55pm to 2:15pm over the period 1997 to 2005. Treasury tick data comes from the GovPX database. Further details on the data are given in Section 3.3. The minutes indicator is equal to one on days that FOMC minutes are released to the public and zero otherwise. Eurodollar Implied Volatility estimates the amount of policy uncertainty prevailing in the market and is taken one day prior to a minutes release. Theme variables are based on residuals from first-order autoregression of Themes 1 to 5. Theme estimation is discussed in Section 4.2. Panels A through C display the results from full-sample estimation, while Panel D presents results using the folding-in method. Statistical significance is indicated by *, **, *** which correspond to significance levels of 10%, 5%, and 1%.

Panel A. 3-Month Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.30 ***	0.38 ***	0.20 ***	0.20 ***	0.20 ***	0.20 ***
Minutes release indicator (I_m)	0.13 *	0.31	-0.13	0.03	-3.06 ***	0.26
Accelerated minutes release indicator	0.42 *	0.34	0.55 **	0.47 *	0.01	0.54 *
Eurodollar Implied Volatility		-0.15	0.61 ***	0.61 ***	0.61 ***	0.61 ***
Eurodollar Implied Volatility*I_m		-0.34	1.17 *	1.17 *	7.55 ***	1.12
Term Spread			-0.0014 ***	-0.0014 ***	-0.0014 ***	-0.0014 ***
Term Spread*I_m			-0.0023 ***	-0.0025 ***	-0.0033 ***	-0.0039 **
Theme 1				-5.22	-6.88	-14.59 ***
Theme 2				-1.08	40.70 **	-2.87
Theme 3				-1.35	-0.82	-0.84
Theme 4				0.93	22.37 *	3.80
Theme 5				0.44	26.43 **	-2.90
Eurodollar Implied Volatility*Theme 1					-2.15	
Eurodollar Implied Volatility*Theme 2					-74.63 **	
Eurodollar Implied Volatility*Theme 3					-0.72	
Eurodollar Implied Volatility*Theme 4					-44.10 *	
Eurodollar Implied Volatility*Theme 5					-55.13 **	
Term Spread*Theme 1						0.11 ***
Term Spread*Theme 2						-0.01
Term Spread*Theme 3						0.00
Term Spread*Theme 4						-0.03
Term Spread*Theme 5						0.02
Adjusted R-squared	0.36%	0.44%	8.88%	8.82%	9.10%	9.08%
F-value	4.39 **	3.06 **	31.44 ***	17.49 ***	12.73 ***	12.70 ***
F probability	0.01	0.02	<.0001	<.0001	<.0001	<.0001
Number of observations	1982	1982	1982	1982	1982	1982

Table 7: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)

Panel B. 2-Year Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.62 ***	0.06	0.03	0.03	0.03	0.03
Minutes release indicator (I_m)	0.48 ***	-0.52	-0.05	0.25	-0.26	0.00
Accelerated minutes release indicator	2.07 ***	2.52 ***	2.29 ***	1.80 ***	2.03 ***	1.89 ***
Eurodollar Implied Volatility		1.05 ***	1.18 ***	1.18 ***	1.18 ***	1.18 ***
Eurodollar Implied Volatility*I_m		1.84 **	0.08	-0.34	0.84	0.06
Term Spread			-0.0002	-0.0002	-0.0002	-0.0002
Term Spread*I_m			0.0030 ***	0.0027 ***	0.0028 **	0.0030
Theme 1				-14.48 ***	-13.21	-17.00 **
Theme 2				13.75 ***	9.70	-0.56
Theme 3				-9.55 **	-9.98	1.25
Theme 4				3.13	39.48 **	10.54 *
Theme 5				2.48	-17.40	-0.52
Eurodollar Implied Volatility*Theme 1					2.16	
Eurodollar Implied Volatility*Theme 2					5.02	
Eurodollar Implied Volatility*Theme 3					0.36	
Eurodollar Implied Volatility*Theme 4					-77.73 *	
Eurodollar Implied Volatility*Theme 5					38.88	
Term Spread*Theme 1						0.02
Term Spread*Theme 2						0.09 *
Term Spread*Theme 3						-0.07 *
Term Spread*Theme 4						-0.06 *
Term Spread*Theme 5						0.03
Adjusted R-squared	3.45%	7.09%	7.49%	7.96%	8.24%	8.50%
F-value	32.79 ***	34.97 ***	25.03 ***	15.00 ***	10.98 ***	11.34 ***
F probability	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Number of observations	1885	1885	1885	1885	1885	1885

Table 7: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)

Panel C. 10-Year Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.69 ***	0.20 **	0.16	0.16	0.16	0.16
Minutes release indicator (I_m)	0.37 ***	-1.93 ***	-1.36 *	-1.25	-6.20 ***	-0.38
Accelerated minutes release indicator	2.36 ***	3.01 ***	2.83 ***	2.56 ***	3.08 ***	2.68 ***
Eurodollar Implied Volatility		0.97 ***	1.11 ***	1.11 ***	1.11 ***	1.11 ***
Eurodollar Implied Volatility*I_m		4.57 ***	2.95 *	2.54	13.15 ***	1.06
Term Spread			-0.0003	-0.0003	-0.0003	-0.0003
Term Spread*I_m			0.0022	0.0026 *	0.0013	0.0030
Theme 1				-12.64 **	100.32 **	3.41
Theme 2				1.42	70.04 **	10.49
Theme 3				11.17 *	-8.41	-13.48
Theme 4				0.64	9.75	3.52
Theme 5				0.74	-3.59	-3.86
Eurodollar Implied Volatility*Theme 1					-242.87 **	
Eurodollar Implied Volatility*Theme 2					-126.55 *	
Eurodollar Implied Volatility*Theme 3					48.48	
Eurodollar Implied Volatility*Theme 4					-21.09	
Eurodollar Implied Volatility*Theme 5					0.63	
Term Spread*Theme 1						-0.21 **
Term Spread*Theme 2						0.00
Term Spread*Theme 3						0.25 **
Term Spread*Theme 4						-0.05
Term Spread*Theme 5						0.01
Adjusted R-squared	4.82%	7.92%	8.01%	8.28%	9.92%	8.95%
F-value	34.71 ***	29.65 ***	20.32 ***	11.92 ***	10.16 ***	9.18 ***
F probability	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Number of observations	1413	1413	1413	1413	1413	1413

Table 8: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)
Results from Folding-in

This table presents the results of regressions designed to model the Treasury market reaction to the public release of FOMC minutes. The dependent variable is the absolute change in the average of bid and ask yields (measured in basis points) for three-month, two-year, and ten-year treasuries from 1:55pm to 2:15pm over the period 1997 to 2005. Treasury tick data comes from the GovPX database. Further details on the data are given in Section 3.3. The minutes indicator is equal to one on days that FOMC minutes are released to the public and zero otherwise. Eurodollar Implied Volatility estimates the amount of policy uncertainty prevailing in the market and is taken one day prior to a minutes release. Theme variables are based on residuals from first-order autoregression of Themes 1 to 5 extracted using LSA with folding-in. Theme estimation is discussed in Section 4.2. Panels A through C display the results from full-sample estimation, while Panel D presents results using the folding-in method. Statistical significance is indicated by *, **, *** which correspond to significance levels of 10%, 5%, and 1%.

Panel A. 3-Month Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.30 ***	0.38 ***	0.20 ***	0.20 ***	0.20 ***	0.20 ***
Minutes release indicator (I_m)	0.13 *	0.31	-0.13	0.39	0.79	1.29 ***
Accelerated minutes release indicator	0.42 *	0.34	0.55 **	0.45	0.67 **	0.47 *
Eurodollar Implied Volatility		-0.15	0.61 ***	0.61 ***	0.61 ***	0.61 ***
Eurodollar Implied Volatility*I_m		-0.34	1.17 *	0.89	0.27	0.51
Term Spread			-0.0014 ***	-0.0014 ***	-0.0014 ***	-0.0014 ***
Term Spread*I_m			-0.0023 ***	-0.0024 ***	-0.0025 ***	-0.0068 ***
Theme 1				-3.47	16.09	-10.18 ***
Theme 2				-2.03	-9.40	3.75
Theme 3				1.54	6.76	-7.00
Theme 4				-1.80	-21.54 **	-6.94 *
Theme 5				-2.43	-17.74 *	-7.85
Eurodollar Implied Volatility*Theme 1					-40.64 *	
Eurodollar Implied Volatility*Theme 2					14.10	
Eurodollar Implied Volatility*Theme 3					-13.05	
Eurodollar Implied Volatility*Theme 4					38.75 *	
Eurodollar Implied Volatility*Theme 5					31.75 *	
Term Spread*Theme 1						0.05 **
Term Spread*Theme 2						-0.05 *
Term Spread*Theme 3						0.06 **
Term Spread*Theme 4						0.02
Term Spread*Theme 5						0.03
Adjusted R-squared	0.36%	0.44%	8.88%	8.95%	9.11%	9.74%
F-value	4.39 **	3.06 **	31.44 ***	17.75 ***	12.74 ***	13.64 ***
F probability	0.01	0.02	<.0001	<.0001	<.0001	<.0001
Number of observations	1982	1982	1982	1982	1982	1982

Table 8: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)
Results from Folding-in

Panel B. 2-Year Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.62 ***	0.06	0.03	0.03	0.03	0.03
Minutes release indicator (I_m)	0.48 ***	-0.52	-0.05	0.15	-0.90	0.04
Accelerated minutes release indicator	2.07 ***	2.52 ***	2.29 ***	2.09 ***	2.09 ***	2.32 ***
Eurodollar Implied Volatility		1.05 ***	1.18 ***	1.18 ***	1.18 ***	1.18 ***
Eurodollar Implied Volatility*I_m		1.84 **	0.08	0.30	2.49	0.30
Term Spread			-0.0002	-0.0002	-0.0002	-0.0002
Term Spread*I_m			0.0030 ***	0.0024 **	0.0021 **	0.0046 *
Theme 1				-8.08 **	-8.17	-10.39 *
Theme 2				1.93	47.64 ***	15.52 **
Theme 3				-3.19	-12.65	-5.76
Theme 4				-1.91	14.24	-7.93
Theme 5				6.87 *	-1.91	8.73
Eurodollar Implied Volatility*Theme 1					4.68	
Eurodollar Implied Volatility*Theme 2					-101.00 ***	
Eurodollar Implied Volatility*Theme 3					15.04	
Eurodollar Implied Volatility*Theme 4					-28.38	
Eurodollar Implied Volatility*Theme 5					16.58	
Term Spread*Theme 1						-0.01
Term Spread*Theme 2						-0.11 ***
Term Spread*Theme 3						0.03
Term Spread*Theme 4						0.03
Term Spread*Theme 5						-0.03
Adjusted R-squared	3.45%	7.09%	7.49%	7.88%	8.18%	8.16%
F-value	32.79 ***	34.97 ***	25.03 ***	14.83 ***	10.91 ***	10.88 ***
F probability	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Number of observations	1885	1885	1885	1885	1885	1885

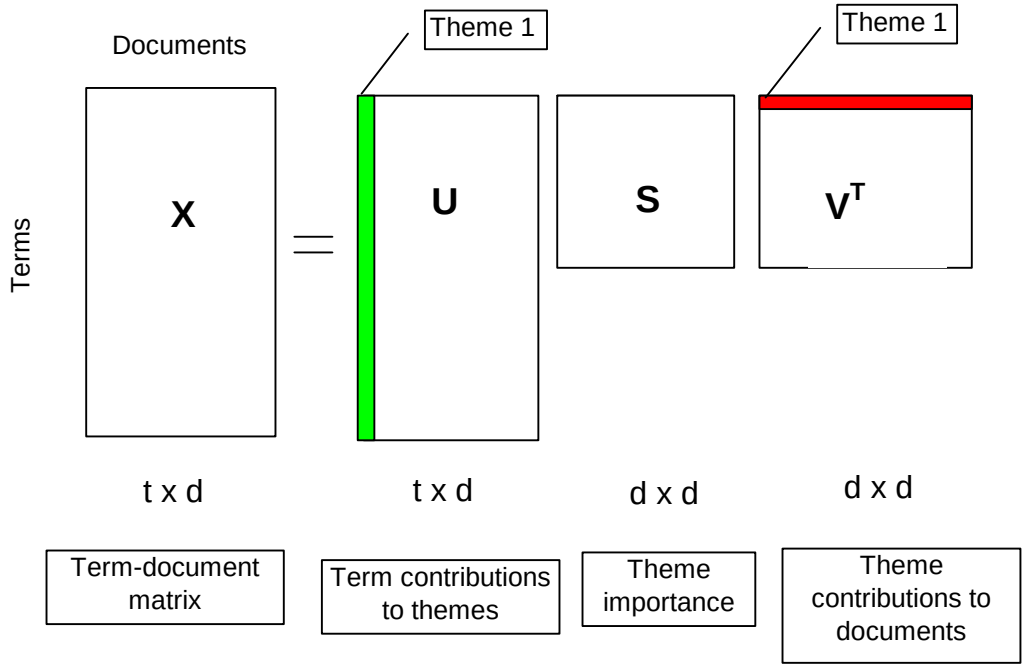
Table 8: Treasury market reaction to FOMC minutes release. Effect on absolute yield changes (1:55pm - 2:15pm)
Results from Folding-in

Panel C. 10-Year Treasury Yields

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.69 ***	0.20 **	0.16	0.16	0.16	0.16
Minutes release indicator (I_m)	0.37 ***	-1.93 ***	-1.36 *	-1.44 *	4.46 **	-0.04
Accelerated minutes release indicator	2.36 ***	3.01 ***	2.83 ***	2.74 ***	2.02 ***	2.34 ***
Eurodollar Implied Volatility		0.97 ***	1.11 ***	1.11 ***	1.11 ***	1.11 ***
Eurodollar Implied Volatility*I_m		4.57 ***	2.95 *	2.57	-9.40 **	2.32
Term Spread			-0.0003	-0.0003	-0.0003	-0.0003
Term Spread*I_m			0.0022	0.0025 *	0.0015	-0.0093 **
Theme 1				-1.26	-52.49 **	-18.90 ***
Theme 2				1.14	53.10 **	13.69 **
Theme 3				-6.87	17.94	4.61
Theme 4				9.93 **	-173.08 ***	-15.77 **
Theme 5				4.69	-9.70	-7.52
Eurodollar Implied Volatility*Theme 1					90.15 **	
Eurodollar Implied Volatility*Theme 2					-109.13 **	
Eurodollar Implied Volatility*Theme 3					-32.87	
Eurodollar Implied Volatility*Theme 4					370.28 ***	
Eurodollar Implied Volatility*Theme 5					36.53	
Term Spread*Theme 1						0.14 ***
Term Spread*Theme 2						-0.08
Term Spread*Theme 3						-0.08
Term Spread*Theme 4						0.22 ***
Term Spread*Theme 5						0.09 *
Adjusted R-squared	4.82%	7.92%	8.01%	8.53%	12.45%	11.24%
F-value	34.71 ***	29.65 ***	20.32 ***	12.29 ***	12.83 ***	11.54 ***
F probability	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Number of observations	1413	1413	1413	1413	1413	1413

Figure 1: Singular value decomposition of the term-document matrix

Figure 1 illustrates the decomposition of the term-document matrix (X) into three matrices: U, S, and V^T. The U matrix defines the contribution of terms to themes, while the V matrix defines the contribution of themes to documents. The S matrix is a diagonal matrix of singular values, which determine the importance of each theme in explaining the documents of the text corpus. The first column of U (green) characterizes the importance of each term in Theme 1. The first row of V^T (red) defines the importance of Theme 1 for each document. When documents are ordered sequentially, this is the time-series of Theme 1 loadings.

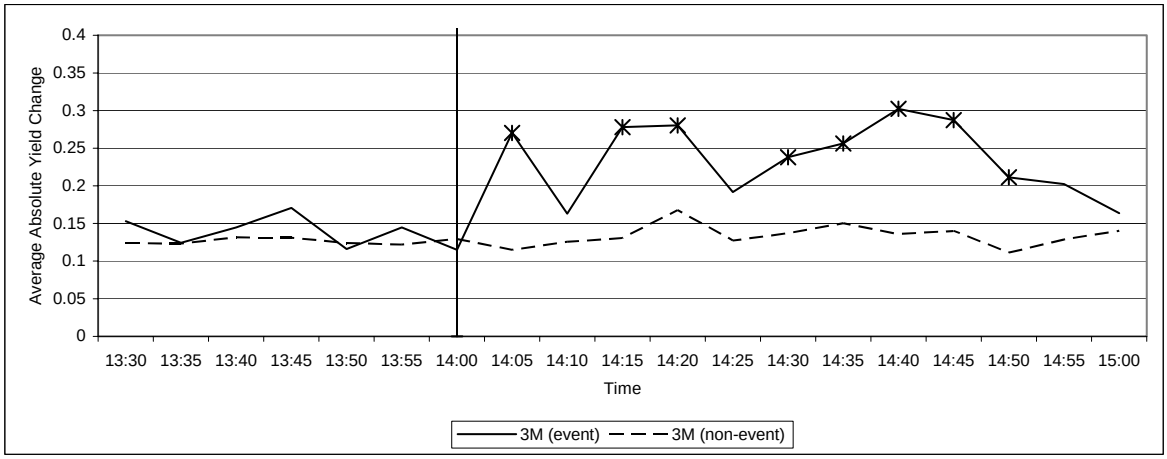


* Adapted from Berry, Dumais and O'Brien (1995)

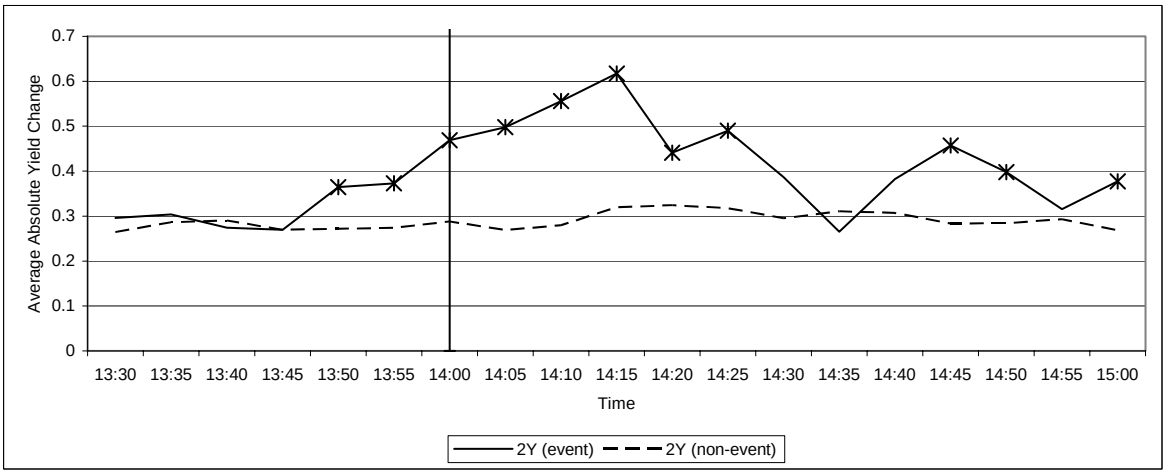
Figure 2: 5-minute Volatility Intervals, 1:30-3:00 pm

This figure presents the absolute yield changes over 5-minute intervals from 1:30 to 3:00 pm. Panels A, B, and C display results for 3-month, 2-year, and 10-year treasuries, respectively. The dotted series represents the average volatility for each interval for all dates over the sample period 1997-2005. The solid line plots the average volatility on event dates. A star denotes whether the volatility for a particular interval on a minutes release date differs significantly from its full sample mean at the 10% level. The vertical line corresponds to 2:00 pm, the approximate time at which the minutes are released to the public.

Panel A: 3-Month Yields



Panel B: 2-Year Yields



Panel C: 10-Year Yields

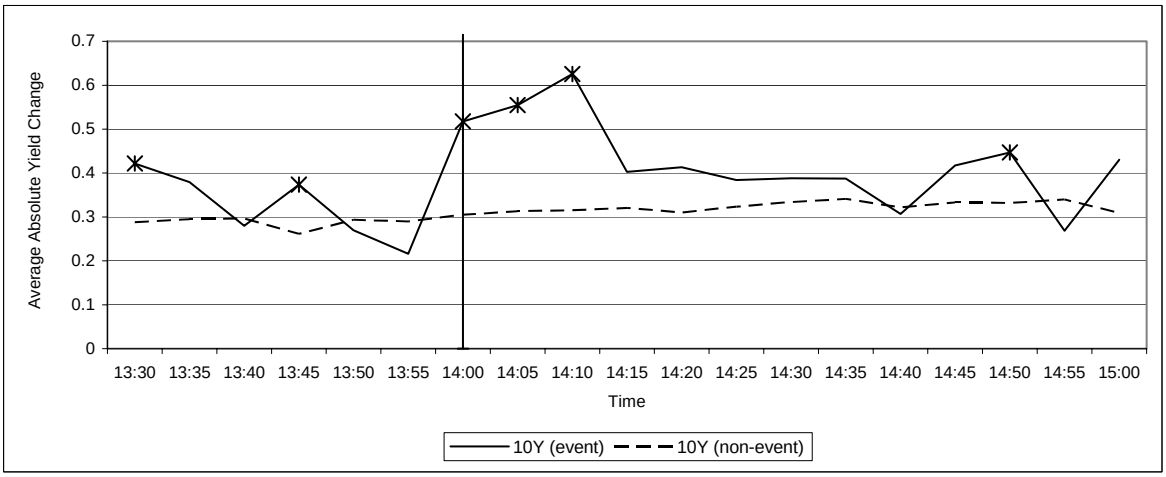
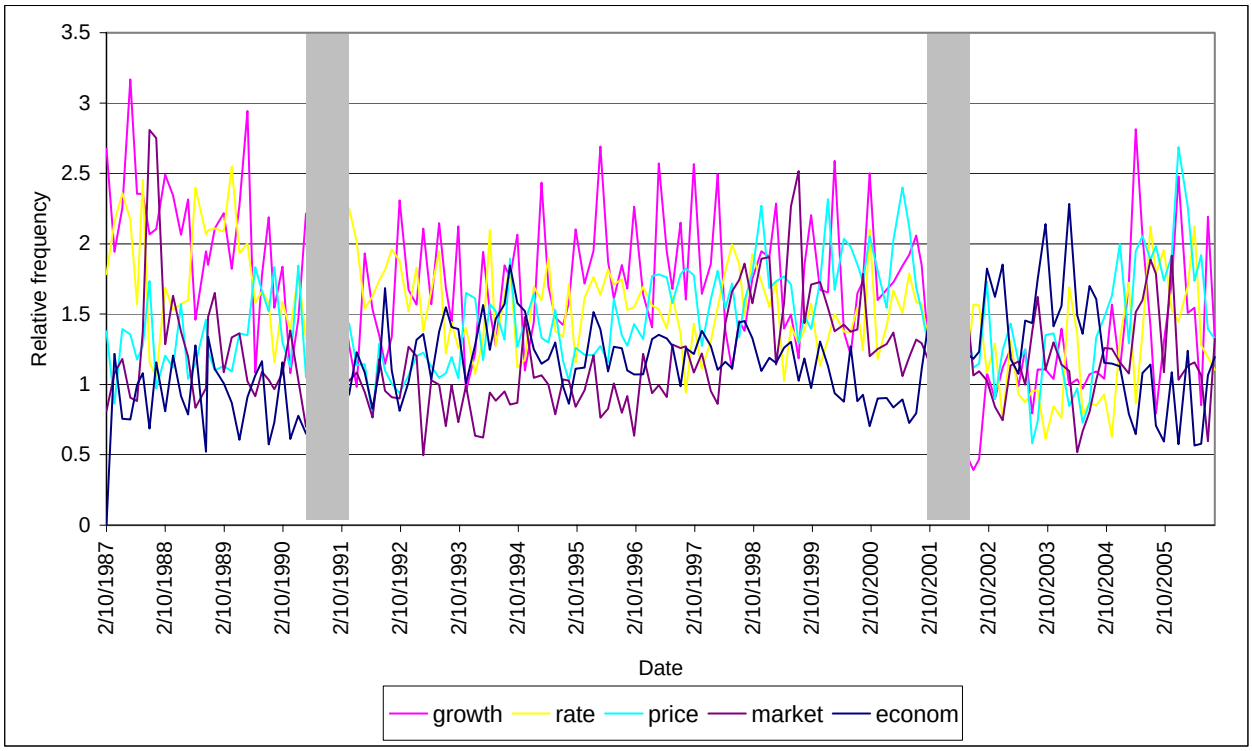


Figure 3: Most frequent terms used in FOMC minutes

This figure graphs the five most frequent terms used in FOMC minutes over the period from 1987-2004. These are extracted from the term-document matrix (X). Panel A presents results from full-sample estimation and Panel B represents results from folding in. Shaded portions correspond to NBER recession periods.

Panel A. Full Sample



B. Results from Folding-in

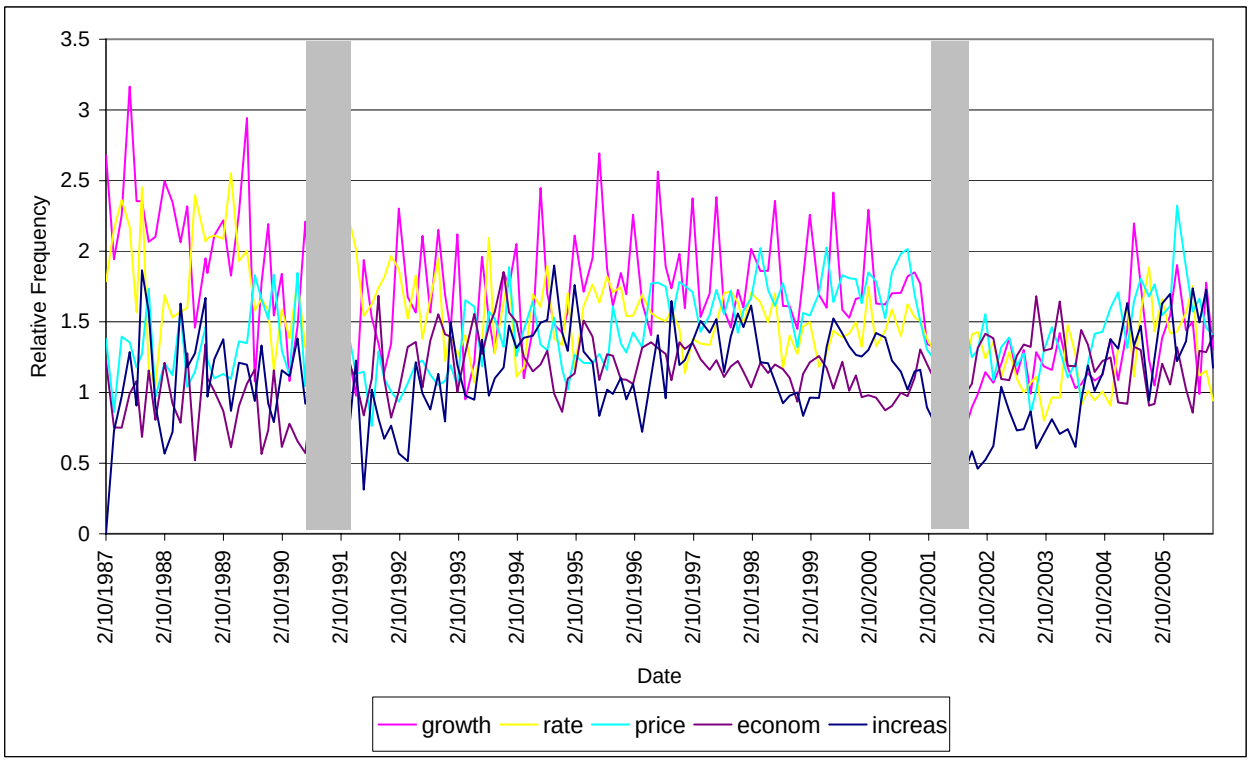
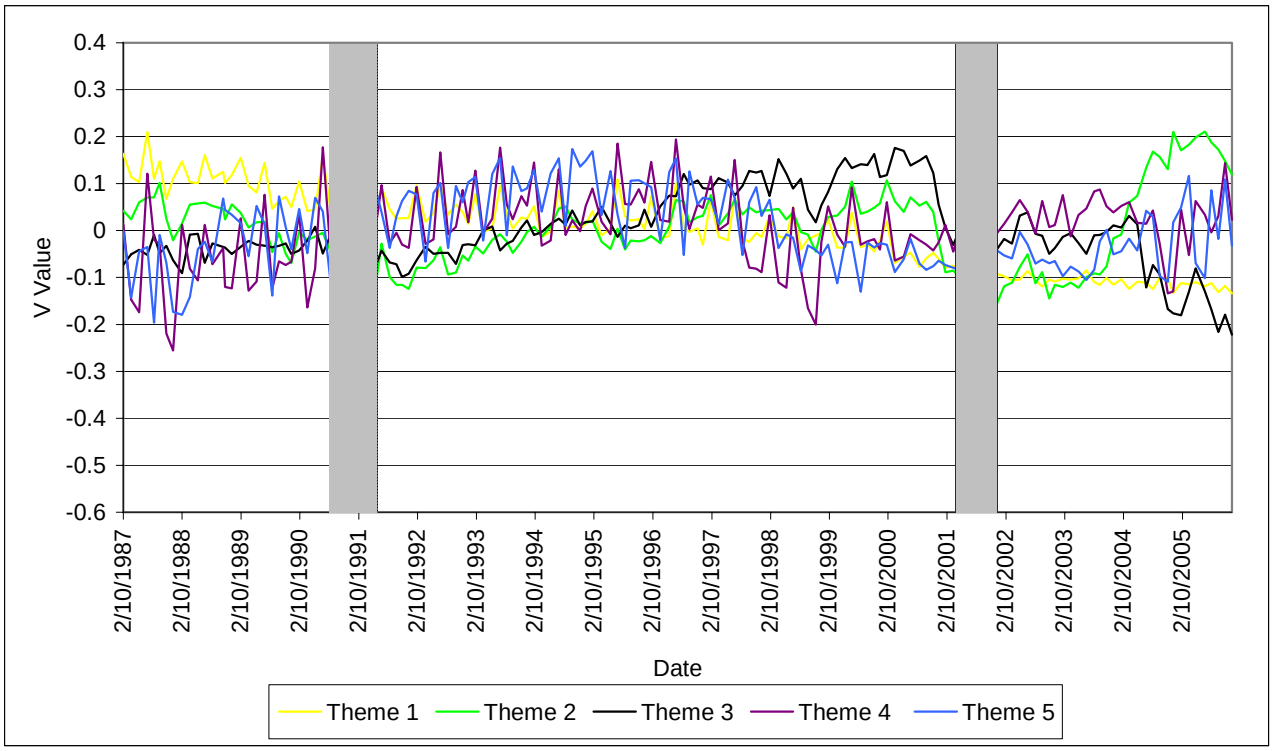


Figure 4: Time-series of theme loadings

This figure graphs the time-series of Themes 1 to 5 based on FOMC minutes over the period from 1987-2005. The theme loadings (rows of the V^T matrix) are estimated using singular value decomposition of the term-document matrix. See Section 4.2 for details. Panel A shows themes estimated using the full sample, while Panel B displays results using the method of folding in. Shaded regions denote NBER recession periods.

Panel A. Full Sample



Panel B. Folding In

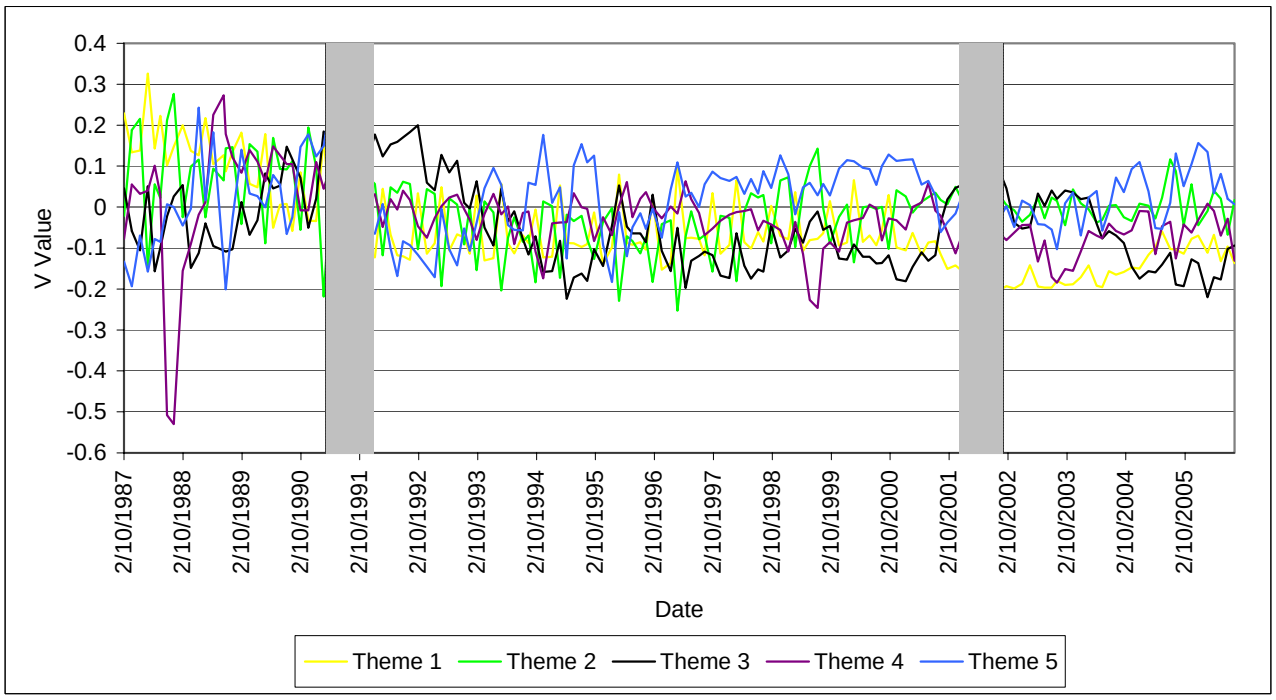
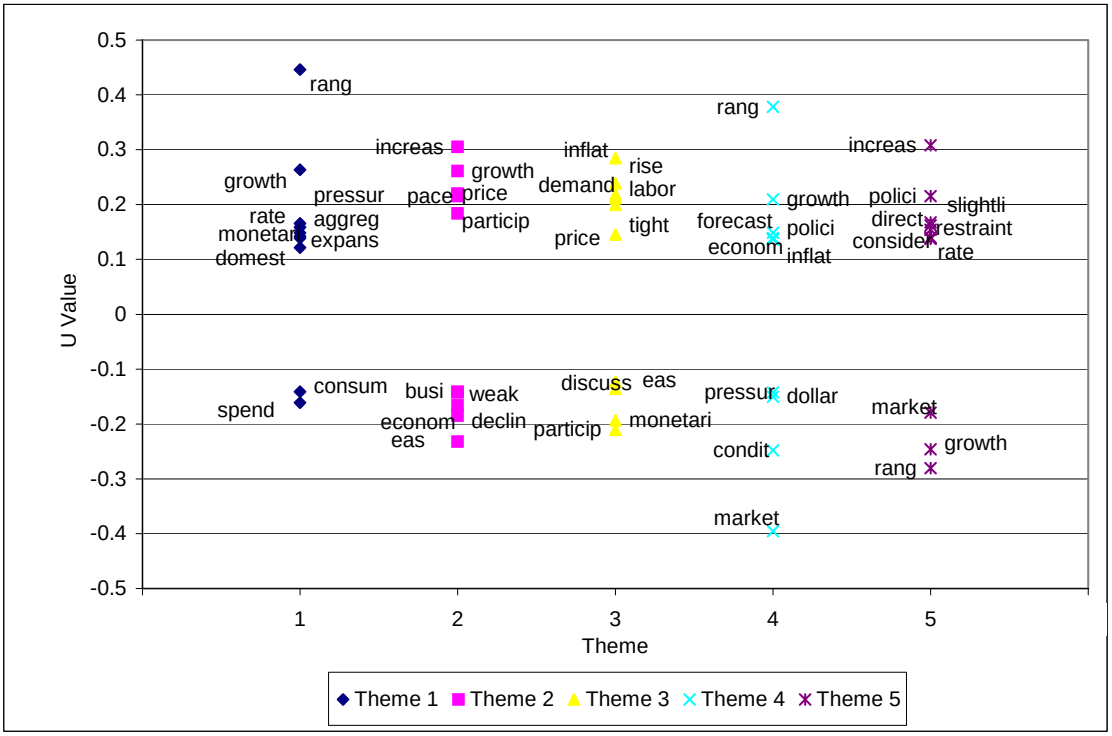


Figure 5: Characteristic terms for Themes 1 through 5

Figure 5 displays the ten most important terms, according to term contributions to themes (U matrix entries), for each of the first 5 themes. Panel A uses the full sample estimate, whereas Panel B applies the folding-in technique.

Panel A. Estimation using full-sample



Panel B. Estimation using folding-in