

FOMC Responses to Calls for Transparency:

Evidence from the Minutes and Transcripts Using Latent Semantic Analysis

J. Miguel Acosta

Department of Economics
Stanford University
Stanford, CA 94305
jmacosta@stanford.edu

Under the direction of
Professor John B. Taylor

May 8, 2014

Abstract

The Federal Reserve has become increasingly transparent in recent years in response to both internal inclinations and external calls from Congress and the private sector, yet it has stopped short of holding open FOMC meetings. Instead, the Fed releases minutes and statements to summarize its policy decisions. This paper applies Latent Semantic Analysis to FOMC transcripts and minutes from 1976—2008 in order to analyze the FOMC's responses to calls for transparency. It reveals that two notable events explain much of the variation in transparency over this 32-year period. First, the 1978 Humphrey-Hawkins Act increased the degree to which the FOMC used meeting minutes to convey the content of its meetings. Historical evidence suggests that this was done because the Act required the Fed to tie its objectives to short-run Congressional and Presidential economic goals. Second, the 1993 decision to publish nearly verbatim transcripts also increased transparency. However, the cost was a decreasing degree of deliberation at each meeting, as evidenced by lower variance in content disagreement at the member level. By applying LSA to FOMC documents, this study presents a unique way of quantifying transparency and suggests LSA's usefulness to other empirical studies in Economics.

Keywords Federal Open Market Committee, Deliberation, NLP, Central Bank

Acknowledgments I would like to first thank my advisor, Professor John Taylor, for all of the support, encouragement and insight offered throughout this process. I also owe much of my understanding of and faith in LSA to Professor Mark Lucianovic. I'm in huge debt to Professor Marcelo Clerici-Arias for the support and assistance from beginning to end. I'm grateful for all of the help from Blake Barton and the other Bing Honors College participants. Professor Allan Meltzer and Dr. Ellen Meade provided me with helpful information, for which I am appreciative. I also owe a special thank you to Lydia Cox for many many hours of discussing cosine-similarities and the FOMC, and a thank you to José Gutierrez for the vector/MATLAB help. I'd also like to thank Donna Hunter, Stephanie Kalfayan and Geoff Cox for their selfless help as I came down the home stretch. Finally, I am hugely indebted to my parents, José Acosta and Mary Grebenc, and my sister, Lourdes, for 22 years of tireless love and support (or 17 in Lourdes's case).

Contents

1	Introduction	1
2	Previous Studies and Theory	3
2.1	Language Analysis of FOMC Communications	4
2.2	Latent Semantic Analysis & Natural Language Processing	5
3	Transcripts, Minutes, ROPAs, MOAs MODs: The History of FOMC Communications	8
4	Measuring FOMC Communications Using Latent Semantic Analysis	11
4.1	Introduction to Document Comparison: A Naïve Approach	11
4.2	Latent Semantic Analysis: Introduction and Application	15
4.3	Additional LSA Parameters	22
5	External Policy Effects on FOMC Transparency	25
5.1	What is Procedural Transparency, and Why Study it?	25
5.2	Measuring Transparency Using Latent Semantic Analysis	28
5.3	Evidence from Minute–Transcript Similarity and Pseudodocuments	33
5.3.1	The Memorandum of Discussion: 1967-1976	34
5.3.2	Closed-Door FOMC Meetings and the Humphrey-Hawkins Act: 1976-1993	34
5.3.3	Opening the Doors: 1993-2008	42
6	The Effect of Public Transcripts: Dampening FOMC Discussion	43
6.1	Measuring Debate in the Board Room	44
6.2	Evidence from the Transcripts	47
6.2.1	Agreement and Debate at the Meeting Level	49
6.2.2	Member-Level Agreement and Debate	52
7	Conclusion	54
A	Appendix	55
A.1	PCA and SVD	55
A.2	Proofs	58
A.3	References	60

1 Introduction

The statements, minutes and transcripts of the Federal Open Market Committee (FOMC) are some of the most closely studied and heavily scrutinized government documents. The subject of many academic studies, they provide one of the few windows into the decision-making process of this influential institution. Perhaps more significant is the attention financial market participants give to these documents—more significant because of the role these words play in managing their expectations and, consequently, those of the entire economy. Given the importance of the topics contained in these documents, it is not surprising that, over the years, both the public and the FOMC itself have favored greater transparency. In practice, this transparency can take two forms: the first is a matter of timing, the second is one of translation. Questions of timing ask which documents should be released when: Should transcripts be thrown away? Released with a five-year lag? Should meetings be broadcast over C-SPAN? Translation refers to the process of distilling detailed information into a concise summary; for example, which words and topics should be included in the minutes? Despite the importance of these documents, topics of timing and, to an even lesser extent, translation, have received little empirical attention. The first goal here is to quantify transparency at the FOMC over time and determine what causes it to change. The second goal is to understand what effect, if any, has increased transparency had on translation procedures and FOMC meeting discussions. To answer these questions, this paper analyzes the minutes and transcripts of FOMC meetings using Latent Semantic Analysis (LSA). LSA has a significant history in the natural language processing (NLP) literature, but has scarcely been used in the Economics literature. This paper is unique in the Economics literature because it takes advantage of the *noise-reduction* properties of LSA in text analysis; LSA was created for this purpose. Reducing noise in natural language allows for better comparisons of the *similarity* of documents—these similarity measures form the basis of much of the analysis presented here.

This research focuses on what Geraats (Geraats (2000); Geraats (2002)) calls "procedural transparency," or the amount of openness that the Fed has exhibited in communicating the content of its meetings over the years.¹ To access this transparency, the minutes and transcripts are considered. The transcripts give the full account of FOMC meetings, yet are not released until 5 years have passed, while the minutes, released with a shorter lag, are a summary of the meetings. Procedural transparency is measured by computing the "similarity" between the minutes and transcripts of each meeting using LSA; the idea is that a higher similarity means that the Fed is presenting more of the content of its meetings. In this sense, translation determines transparency. Next, in order to determine which of the document types was responsible for the increase in similarity, the minutes and transcripts are studied separately so that changes in content can be found. Specifically,

¹See Section 5.1 for a full definition.

all minutes are compared to a single "document"² so that a time series can be created and large changes in content can be seen; the same is done for transcripts. Choosing a single reference point is necessary because all similarity measures are computed using two documents, and creating a time series requires comparing each document pairwise to a single point. These measures reveal two major changes in FOMC procedural transparency since 1976.

First, transparency increased sharply in 1979—this was almost entirely a minute-driven change. Unfortunately, while LSA detects a change in the minutes, it is not able to say *how* or *why* they changed; for this reason, the transcripts are read for historical evidence. Written evidence from the transcripts suggests that this change was in response to the Full Employment and Balanced Growth Act of 1978, known as the Humphrey-Hawkins Act, which amended Section II of the Federal Reserve Act and mandated that the Fed report how its objectives and plans related to the "short-term [economic] goals" of the President and the Congress (Congress, 1978). The smaller change in the content of the transcripts suggests that the new guidelines led the Fed to be more transparent in conveying the conversations that they were already having. In other words, the FOMC was already discussing objectives and plans consistent with short-term Presidential and Congressional goals, so the only change it made was in conveying these to the public. As such, this appears to have been a successful attempt in increasing transparency at the Fed.

The second change occurred in late 1993 and was also marked by an increase in transparency. From 1976, the beginning of the sample period, until 1993, very few people knew that transcripts of FOMC meetings existed. In 1993, Congressional inquiries and Fed discussions revealed that the transcripts had indeed been kept, and at that point the Fed decided to publish the old transcripts with a five-year lag. Within eighteen months they also decided to publish new transcripts going forward, with the same five-year lag. This shift from "closed" FOMC meetings to more public meetings has been the subject of many of the studies discussed below because of the natural experiment that it created. Specifically, this transition allows for a study of FOMC behavior at the committee and member-levels, under different levels of transparency. The analysis used here suggests that this time, the transparency increase was the result of a changing transcript. Because the transcripts give detailed account of meetings, this means that the move to more-public meetings caused the conversation in the Board Room to change. While transparency *did* increase, the worry is that there could be some undesirable impact on the policy-making process. For example, Meade & Stasavage (2008) cites the 1995 words of Kansas City Fed President Hoenig "the tape has had some chilling effects on our discussions. I see a lot more people reading their statements." (Meade & Stasavage, 2008, pp. 704-705). If increasing transparency caused a response like this, then the

²Quotation marks are placed around "document" because all of the minutes are compared to something of a contrived structure that resembles a document. The point is that the minutes from each meeting are compared to a single reference point—metaphorically, this is similar to determining the height of n individuals based on their height relative to one person; the feet and inches of each person are unknown, but there is at least an ordering.

"success" of such an increase should certainly be called into question, and remembered in the future when considering whether to increase procedural transparency at the Fed.

While LSA cannot, on its own, say *why* or *how* transcript changes occurred, it is still helpful in identifying possible answers. One possibility explored here, which has received attention in other studies, is whether the degree of deliberation changed once meetings were made more public. If so, this would support Hoenig's claim about the "chilling effects" of transparency. To study the degree of debate at a single meeting, each member's words at that meeting are compared to the words of all other attendees. Low similarities suggest that people were in disagreement,³ while high similarities suggest that people are in agreement. In the extreme case that there is full similarity, members are simply mimicking the views of every other member. The measure of deliberation is the standard deviation of these member–committee agreement levels for each meeting. By definition, this gives the spread of agreement levels—this is interpreted as a willingness to deviate from the mean level of agreement. The finding, similar to Meade & Stasavage (2008) and Woolley & Gardner (2009), is that the 1993 change resulted in a decrease in the amount of debate within the Board Room—the standard deviation measure fell. In addition, this drop was primarily the result of a decrease in the willingness to offer disagreeing views. Establishing causality is difficult, but visual and econometric tests suggest that, at the very least, there is a correlation.

Perhaps an equally important aspect of this research is the introduction of new uses of LSA into the economics literature. It appears to be a powerful tool that has uses far beyond what can be covered in a single paper. For this reason, a detailed explanation of the methodology is given as a general reference.

The rest of this paper proceeds as follows. Section 2 reviews the literature in economics, political science and natural language processing that are similar to this paper in content and methodology. Next, Section 3 gives an overview of the data used in this paper; in particular, it provides a history of the documents published by the FOMC. Section 4 describes LSA, and Appendix A.1 goes into more detail on the same subject. Sections 5 and 6 contain the primary results. Section 5 starts by defining transparency and its importance to FOMC procedures, then describes the application of LSA used to measure transparency and gives the results. Similarly, Section 6 describes the measurement of deliberation and results from this analysis. Finally, Section 7 concludes.

2 Previous Studies and Theory

This work lies at the intersection of four strands of literature: two in economics, one in political science and the other in the multidisciplinary domain of natural language processing (NLP). The first, in economics, studies the content of FOMC publications to draw conclusions about topics

³If this is unclear, Section 6.1 goes into more detail.

of discussion, and to observe the effects of these publications on external factors such as markets. This strand, and the political science literature, both treated in Section 2.1, are similar in their use of FOMC documents, including minutes, transcripts and statements, as bases for empirical work. Next, the political science topics have been studied by political scientists and economists, and seek to understand the dynamics within the FOMC. Finally, because this paper presents a new application of Latent Semantic Analysis in the field of Economics, some of LSA's other applications are given to show its strength. This is treated in Section 2.2 alongside the recent work in economics, which uses similar text-processing methodology to study the characteristics of firms involved in mergers and acquisitions (M&As) and initial public offerings (IPOs).

2.1 Language Analysis of FOMC Communications

A primary reason for the existence of publicly released FOMC documents is that, over the years, the public has demanded them in order to extract relevant and important information. Because of the value of the content in these communications, they have also been studied by scholars in many contexts. Given the massive amount of content that can be analyzed,⁴ most studies have devised a way to map written text to a single variable.

A number of studies have used text-generated variables to evaluate the impact of FOMC communications on market outcomes. Farka (2011) read news articles released within minutes of FOMC statements in order to classify statements as "informative" or "uninformative," and found that "informative" statements have a larger impact on Treasury yields and equity returns. Rosa & Verga (2007) ranked ECB president statements from hawkish (−2) to dovish (2), and similarly found significant market impacts.

Another group considered these text-variables as objects of study in their own right. Ehrmann & Fratzscher (2007) studied inter-meeting member communications—e.g. speeches, interviews, etc.—and assigned each text an integer value from −1 to 1 for its attitude towards the economic outlook and policy stance.⁵ Their goal was to see which of the world's largest central banks (ECB, BoE, Fed) are more collegial and which are more individualistic. In order to study the extent to which the Philips curve entered into the policy making progress, Meade & Thornton (2012) used extensive word searching and reading to generate an index that identified when the Philips curve was discussed. Chappell *et al.* (2005) used the transcripts to code interest rate preferences of FOMC members under Burns and part of Greenspan—this was a huge improvement over using voting records alone, since dissents are rare. Their book focused more on questions of decision making within the FOMC, but one relevant finding is the disproportionate roll of the chair in guiding Board Room dynamics and decisions. Meade & Stasavage (2008) coded voiced agreement or

⁴Visit <http://federalreserve.gov> to get started.

⁵E.g. −1 for a negative economic outlook and inclination to tighten policy.

disagreement with Greenspan's interest rate proposals in the pre-1993 ('89-'92) and post-1993 ('84-'97) transcripts to study the effect of the FOMC's 1993 decision to publish meeting transcripts after a five year lag. They found that there was less voiced disapproval after 1993; a finding that was predicted by their theoretical "model of deliberation in a committee, where members care both about reaching the correct decision and about convincing an outside audience that they have a high level of expertise." (Meade & Stasavage, 2008, pp. 696). Woolley & Gardner (2009) found similar results by constructing a measure of deliberation from the transcripts. The authors counted the number of speakers per 100 words of transcript; the rationale is that this would capture the amount of "back and forth" among the members. A low number of speakers would likely mean that members were simply reciting their speeches and engaging in little debate. Ultimately, they showed a 9% decrease in deliberation after 1993. It will be shown later (see Figure 13) that the measures of deliberation in Woolley & Gardner (2009) and in this paper look remarkably similar over time ($\rho = 0.79$)—further, both papers estimate a decrease in deliberation on the order of 10%.

The last group took a more holistic approach to studying FOMC documents, taking *all* words into account. Boukus & Rosenberg (2006) used LSA to extract "themes" from the minutes to revisit the issue of the effect of FOMC documents on various financial indicators. However, as the authors note, the "themes" generated by LSA are difficult to interpret—as will be shown later, each theme is essentially a weighted sum of (almost) all of the words in a document. The terms that have the highest "weights" in the theme are typically used to describe the theme, but this ignores the subtlety contained in the weights of other words. With LSA, these themes are represented as vectors, but there is a reason that humans don't read vectors, they read documents. Any specific conclusions drawn from analyzing LSA themes thus rely on a bit of subjectivity—the analysis presented here avoids such interpretations. Finally, the recent work by Schonhardt-Bailey (2013) used software called Alceste to study deliberation within the FOMC and at related congressional hearings. This software is similar to the LSA method employed here in that both methods allow for entire documents to be studied, and both allow for conclusions to be drawn based off of this holistic approach. Of particular interest is her finding that "the change of chairman [is] the most likely cause of changes in the form of deliberation," a conclusion that is corroborated in Section 6.2.

2.2 Latent Semantic Analysis & Natural Language Processing

The primary tool used for analyzing FOMC documents here is Latent Semantic Analysis (LSA), a tool from natural language processing (NLP) first introduced by Deerwester *et al.* (1990) and Dumais *et al.* (1988). A detailed explanation is given in Section 4.1, but in brief, LSA treats documents as vectors of their word counts and uses a linear algebra technique called Singular Value Decomposition (SVD) to reduce the "noise" in the vector that is introduced when people have to

choose words to represent their ideas. The idea is that natural language is a "noisy" way to convey ideas, and LSA seeks to combat this. Before reviewing the literature on LSA, however, there has been some analysis done with the same document-vector approach, without the noise-loss properties of LSA.

The translation of a document into a vector is called a *vector-space model* of that document; this type of document analysis has been employed recently in the financial literature. Hoberg & Phillips (2009) used company 10-K product descriptions in order to create a new industry classification scale based on how firms describe their products in these annual reports. Using the cosine measure of similarity⁶ to compare the similarities between these 10-K forms, the authors were able to classify firms as more or less similar to one another, thus creating a new classification code. This is a more flexible system than the standard SIC or NAICS⁷ codes in that it can vary over time, and it allows each firm to have its own set of competitors. They found that their scale better correlates with accounts from firm management. Hoberg & Phillips (2010) used these classification codes to test the extent to which "product market synergies and asset complementarities" play a role in mergers and acquisitions. They find that firms which are similar to many firms—think of these as general, rather than specialized, firms—tend to merge more often, in order to take advantage of asset complementarities, while firms which have many highly similar rivals tend to merge less because the existence of the similar firms introduces competition for suitable merging firms. They also found that post-merger outcomes are better when the firms were similar and when the merge increased product differentiation between the merged firms and the rest of the industry. Similar tests could have been performed using existing classification codes, but the advantage of the textual approach is the ability to use more detailed and current information. Both Hanley & Hoberg (2010) and Hanley & Hoberg (2012) employ the cosine similarity measure to IPO prospectuses; the first finds that more informative prospectuses lead to more accurate offer prices, and the second ties the issue to litigation risk. Note that none of the work introduced in this paragraph made use of LSA when computing document similarities, but nevertheless found important and significant results—results which suggest the validity of such vector-space model approaches.

Landauer *et al.* (1998) covers, in detail, eight examples of "LSA's ability to model human conceptual knowledge" (Landauer *et al.* , 1998, pp. 270). The real strength of LSA over a simple vector-space model is in dimension reduction, or noise reduction. For example, when computing the similarity between a query and various documents, dimension reduction improved performance by 16% over the non-dimension reduced approach, where "performance" was measured relative to a person familiar with all of the documents. LSA also took the TOEFL⁸ vocabulary test and

⁶This method is used extensively below. Hoberg & Phillips (2009) also consider the TF-IDF weighting scheme used below, but find it of less use.

⁷SIC: Standard Industry Classification. NAICS: North American Industry Classification System.

⁸Test of English as a Foreign Language

scored a 65%, a score "identical to the average score of a large sample of students applying for college entrance in the US from non-English speaking countries" (Landauer *et al.* , 1998). Foltz *et al.* (1999) covers additional examples; first, they cite a study in which "After training on an introductory psychology textbook, [LSA] achieved passing scores on two different multiple-choice exams used in introductory psychology courses" (pp. 940). More closely related to the analysis here is LSA's ability to score student essays—this is more relevant because it demonstrates LSA's ability to compare *documents with documents*, rather than sentences with sentences or queries with documents. Specifically, automated essay grading involves comparing an essay to either (i) a model essay written by the grader or (ii) to a few student essays that have been manually graded. In the first case, higher similarity to the model essay would imply a higher grade; in the second, the pre-graded essays with which the current essay was the most similar would determine the grade of the current essay. Using (ii), they cite numerous cases in which the correlation between the grades assigned by LSA and those assigned by a human grader was nearly identical to the correlation of grades assigned by two human graders. As an example, using GMAT essays, the correlation between human Grader *X*'s grades and LSA's grades (0.86) was exactly equal to the correlation between human Grader *X*'s and human Grader *Y*'s (Foltz *et al.* , 1999, pp. 941). Other examples of LSA's strength are abundant—the key is that LSA presents an *improvement* over simple vector-space models, and even approaches a model of human knowledge and understanding.

Despite the evidence in favor of LSA's effectiveness, it has rarely been explicitly employed in the literature. Aside from Boukus & Rosenberg (2006), it appears that the only other use of LSA in the Economics literature is Hendry & Madeley (2009). This paper takes after Boukus & Rosenberg (2006) in that it extracts themes from the minutes of a central bank; this time, the Bank of Canada. The authors also found significant correlations between certain "themes" and the returns and volatility of interest rate markets—they also showed that textual effects are larger than policy rate surprises. In addition, the authors offered their opinion on the "true power" of the LSA technique:

LSA typically is used to create numerical representations of documents (vectors) that are used to find similarities between documents (e.g. to find all documents on a similar topic, to find all existing documents closest to the document represented by a few key-words typed into an internet search engine) although this is a rather simplistic view of what a very powerful tool actually achieves. Our study investigates whether extracted themes move markets and then attempts to offer some interpretation of the ideas those themes may represent by examining the most important words in each theme. (Hendry & Madeley, 2009, pp. 9–10)

This approach to using LSA contrasts with the earlier discussion about the power of LSA and the subjectivity that is introduced when attempting to attach meaning the mathematical structures that

it produces. The analysis presented in this paper relies heavily on this proven strength of LSA: the ability to more accurately measure document similarities. This approach also removes the less precise aspect of the approach used by Boukus & Rosenberg (2006) and Hendry & Madeley (2009); namely, assigning interpretations to themes. In choosing to use this alternative strength of LSA, this paper offers a perspective and methodology not yet explored by these studies.

3 Transcripts, Minutes, ROPAs, MOAs MODs: The History of FOMC Communications

This section describes the primary source of data that underlies the empirical analysis of this paper: FOMC documents. In order to better understand the various documents which have been published by the FOMC, a history of its publications is given. This will also lay the foundations for comprehending the various circumstances under which the FOMC has made its publication decisions.

The modern day Federal Open Market Committee took shape in the Banking Act of 1935. Since its inception, the FOMC has always communicated in some way the content of its meetings. Figure 1, based off of the timeline in Danker & Luecke (2005), gives a visual timeline of the various FOMC publications since 1935, including the various document names and information about the timing of their release. While the nomenclature of the various documents has undergone several changes over the past eighty years, there have, in general, been two types of documents: detailed accounts of FOMC meetings, and summaries. In general, the latter were more readily available to the public. Most of the changes in FOMC documentation and timing have arisen from public pressure for openness and transparency at the Fed. This trend has been felt by governments in general, particularly by agencies which appear to be cloaked in secrecy. These pushes for transparency have formally come from Congressional pressure, legislation and litigation, and are what ultimately allow for the study of the FOMC's responses to calls for transparency; that is, the FOMC's behavior under different degrees of pressure. Over the years, there are four notably distinct periods of FOMC publications (more detail is given for documents of interest to the present study):

- 1. 1935-1967** During the FOMC's early period, the Committee maintained a transcript-like document, then called "minutes," which was kept confidential. Annually, it would publish the Record of Policy Actions (ROPA) for each meeting; this brief summary was the committee's method of communicating its policies and rationales for deciding upon said policies.
- 2. 1967-1976** In light of the 1967 Freedom of Information Act (FOIA), the FOMC overhauled its publication schedule. Beginning with the April 1967 meeting, the ROPA, which had grown to be about a seven page summary of the FOMC's policy actions and rationales, would be

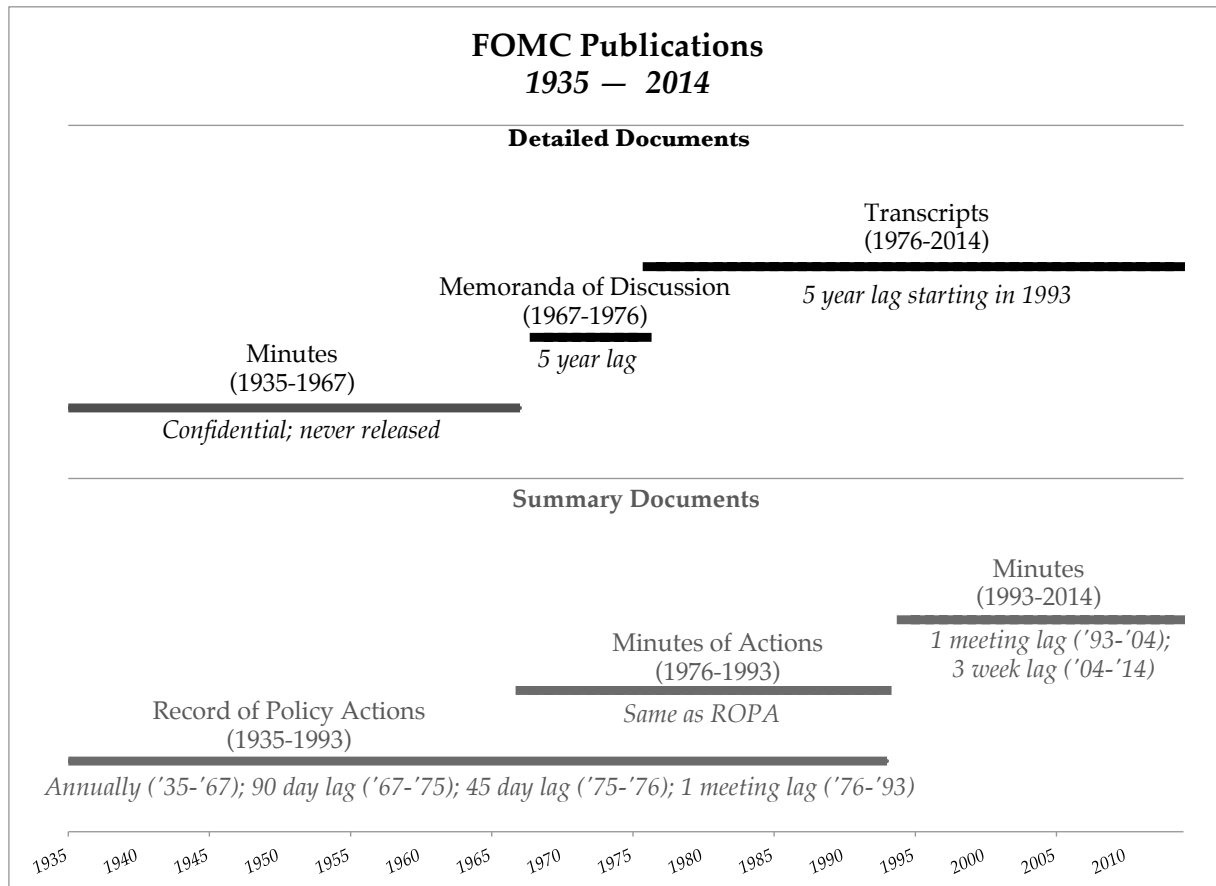


Figure 1: FOMC Publications: Banking Act of 1935 to Present (2014). Adopted from Danker & Luecke (2005). *Release lags in italics.*

published after a 90 day lag (Danker & Luecke, 2005). A new document, the Minutes of Actions (MOA) would be published alongside the ROPA, and announced any policies (monetary, procedural, or otherwise) taken by the FOMC, and attendance. Finally, the Memorandum of Discussion (MOD), another transcript like document, was to be released with a five year lag. The combination of the MOD and MOA was roughly equivalent to the former transcript-like "minutes," while the ROPA served as the public's most immediate window into FOMC decision making. However, it is important to note that the MOD was a heavily edited account of the meeting discussions; therefore, this time period is excluded for the majority of the analysis in this paper for reasons elaborated in Section 5.3.1.

3. **1976-1993** In 1976, the FOMC again updated its communication policies, largely in response to a lawsuit. David Merrill, a law student at Georgetown University, filed suit against the FOMC in March of 1976 for Freedom of Information violations—namely for failing to provide access to the ROPA or MOD of the January and February meetings. After 5 years of litigation, including a trip to the Supreme Court, the US Court of Appeals found in favor of

the FOMC in 1981. Despite winning, the FOMC discontinued the MOD largely at the request of Chairman Burns (Lindsey, 2003). By this time, the ROPA's lag had been shortened to 45 days. At this point, the Committee decided to release an expanded ROPA shortly after each subsequent meeting; at the time, this effectively meant a 30 day lag. At the time, the reason cited for the discontinuation of the MOD was that "the benefits derived from them did not justify their relatively high costs, particularly in light of the changes made in the policy record" (ROPA, '76-'93, 05/18/76). However, the real reason seems to be "'fear that Congress would request access' [to the MOD] promptly" (Lindsey, 2003, pp. 8) and, as an FOMC subcommittee indicated, "concern about the ability to conduct monetary policy, if the court required prompt release of the memoranda of discussion" (Meltzer, 2010, pp. 976).

The discontinuation of the MOD started a nearly 20 year period in which the FOMC published no detailed account of its meetings. Most FOMC members were aware that meetings were recorded, but they also believed that these tapes, used only for the production of minutes by Board staff, were recorded over after each meeting.

- 4. 1993-2014** Contrary to what most members believed, Congressional inquiries (primarily headed by Congressman Henry González) and internal Fed investigations revealed that, in fact, these tapes had been maintained since 1976. In November of 1993, the Committee agreed to publish all of the transcripts since 1976; by 1995 the decision was made to reinstate the publication of meeting transcripts after a five year lag. In addition, the ROPA and MOA were now combined to form the "minutes." In 2004, these minutes began to be released with a 3 week lag.

In order to remain consistent in the document formats over time, the principal analyses in this paper rely on the post-1976 documents—the transcripts, ROPA, and minutes. This protects against the possibility of attributing any results to a change in document type. For example, if the decision was made to treat the (heavily edited) MOD and transcripts both as "detailed accounts of meeting discussions," then it is possible⁹ that conclusions could be made, or differences found, which are purely the result of the fact that the MOD and transcripts are fundamentally different documents. For this reason, the procedural information (voting records, attendance) is removed from the 1993-2008 minutes, which are a combination of the ROPA and MOA, so that these documents are essentially the same as the ROPA.

It was stated earlier that these documents form the "primary source of data" for this analysis. As such, it is important to know their origins. First, all post-1967 ROPAs, MOAs, MODs, minutes and transcripts were downloaded from <http://federalreserve.gov>. Those which were not available in PDF format were "scraped" from their respective webpages using Beautiful Soup,

⁹In fact, it *is* the case; see the introduction to Section 5.3.1.

a Python library. Documents in PDF format were converted to plain text using OCR¹⁰ software.¹¹ The documents were then quickly processed using more Python code: procedural information was removed, and the names of the voting members were extracted. At this point, the documents were ready to be converted to numerical data using the MATLAB[®] toolbox Text Matrix Generator (Zeimpekis & Gallopoulos, 2006).

4 Measuring FOMC Communications Using Latent Semantic Analysis

4.1 Introduction to Document Comparison: A Naïve Approach

The goal of this section is to define a way to compare documents; in particular, the focus will be on comparing the documents released by the FOMC. The minutes and transcripts of the FOMC are rich in content; they are also long. Thus, any attempt to study how their content has changed over time must be holistic, systematic and scalable. Therefore, a second goal of this section is to show that the defined method satisfies these three criteria.

The methodology employed in this research is based on using word counts to deduce meaning from documents. Whereas previous textual analyses of FOMC documents have classified texts along a uni-dimensional scale (e.g. from "hawkish" to "dovish"), using word counts permits more of a document's complexities to be considered in analysis. As a baseline and introduction, this section gives a description of a simple word-count, or vector-space, method for document comparisons, and accompanying examples.

The Setup Suppose that there are three documents which have been summarized by the documents x , y and z (i.e. x represents the gist of what was said in the first document):

$$\begin{aligned}
 x &= \boxed{\text{Inflation inflation inflation unemployment.}} \\
 y &= \boxed{\text{Inflation unemployment unemployment unemployment.}} \\
 z &= \boxed{\text{Unemployment unemployment unemployment.}}
 \end{aligned}$$

There are a few things to notice here. First, the size of the vocabulary set is two; see Equation (1). The use of two words is helpful for visualizing these documents in two-dimensional space; in practice, documents have many more words. Next, notice that documents y and z are relatively

¹⁰Optical Character Recognition

¹¹OCR software is not perfect, but LSA helps to combat some of its imperfections by reducing the weight of words that don't seem to fit.

similar in that they are both heavy in their use of the word *unemployment*; conversely, x and z are the most dissimilar because x focuses heavily on inflation. This imprecise notion of similarity will be more formally defined later, but for now it provides the necessary intuition underlying the methodology. Finally, capitalization and punctuation will be ignored; in fact, using word counts means that the order of the words is also irrelevant, i.e., the following documents are equivalent to x , y and z respectively:

$$x' = \boxed{\text{inflation unemployment inflation inflation}}$$

$$y' = \boxed{\text{Unemployment! unemployment! unemployment! inflation!!}}$$

$$z' = \boxed{\text{unemployment uNeMpLoYmEnT unemployment?}}$$

The Method The first step in the process of comparison is to create a *term x document* matrix. This requires first constructing the vocabulary set of all of the documents. For these documents, the vocabulary set \mathcal{V} is

$$\mathcal{V} = \{\text{unemployment, inflation}\} \quad (1)$$

The next step is to build a matrix with two rows, one for each word in \mathcal{V} , and three columns, one for each document. The entries in the matrix are the number of times that a given word appears in a given document. For this example, the matrix is as follows:

$$\begin{array}{l} \text{Unemployment} \\ \text{Inflation} \end{array} \begin{array}{ccc} x & y & z \\ \left[\begin{array}{ccc} 1 & 3 & 3 \\ 3 & 1 & 0 \end{array} \right] \end{array}$$

In a similar fashion, each document can be represented as a vector with as many elements as there are elements in the vocabulary set, and whose entries are the number of times that each word appears. Each row in the vectors corresponds to the same vocabulary word. Here, the document vectors are given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

This approach satisfies the aforementioned criteria. It is clearly systematic; it is holistic in that it uses every part of every document; and it is scalable in that, given sufficient paper or computer memory, any number of documents can be similarly converted into such a matrix.

The next step is to calculate a measure of similarity between the documents. For these documents, consider Figure 2, where each document has been plotted as a vector in *inflation-unemployment* word count space. With this visualization, it is clear that y and z look "closer

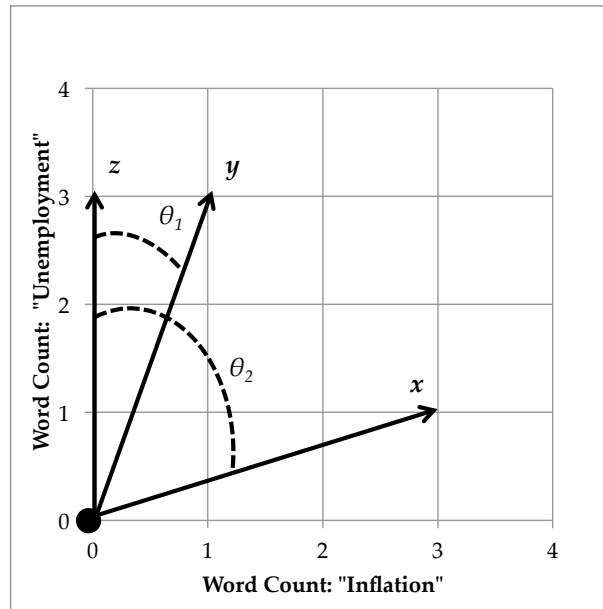


Figure 2: Vector representations of the documents x , y and z .

together" than do x and z . This corresponds to the earlier claim that y and z are the most similar, while x and z are the least similar. In order to quantify this similarity, while controlling for the relative magnitude (length) of each vector, the angles, θ_i , that lie between each of the documents can be measured. Controlling for the length of each document allows for a comparison of documents of any length; thus, it is possible to compare one word or phrase to an entire document to see how well that word captures the meaning of the document, as will be shown in Figure 4 below. Now because θ is some measure between 0 and 90° , in practice, the cosine of the angles is used so that the similarity measure lies between 0 and 1 , where documents with cosine similarity 1 are very similar and documents of cosine similarity 0 are not similar. Note that while it is true that for any angle θ , $\cos(\theta) \in [-1, 1]$, because word counts are always non-negative, the cosine will, in practice, always lie between 0 and 1 . Equivalently, documents with a cosine similarity of 1 use the same words in the same proportions; they are scalar multiples of one another. On the other hand, documents with a cosine similarity of 0 use none of the same words; they are orthogonal. To see this, consider the definition of *cosine similarity*:

Definition 1. For any two vectors \mathbf{v}_1 and \mathbf{v}_2 , separated by angle θ , the **cosine similarity** of \mathbf{v}_1 and \mathbf{v}_2 is given by

$$\text{sim}(\mathbf{v}_1, \mathbf{v}_2) = \cos(\theta) = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}$$

where $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ is the inner product of \mathbf{v}_1 and \mathbf{v}_2 , and $\|\mathbf{v}_i\| = \sqrt{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}$ is the norm of \mathbf{v}_i .

(A proof of why this measure captures the cosine for two vectors in \mathbb{R}^2 is given in Appendix A.2.)

Thus, for this example, $\text{sim}(\mathbf{z}, \mathbf{y}) = \cos(\theta_1) = 0.95$ and $\text{sim}(\mathbf{x}, \mathbf{z}) = \cos(\theta_2) = 0.32$, which confirms that z and y are more similar than are x and z . Again, by definition, the cosine similarity of documents which share no terms is zero, and if documents use exactly the same terms in the same proportions, then the cosine similarity is one. Later, when LSA and the other weighting schemes are applied, cosine similarities *can* range between -1 and 1 . However, since FOMC documents generally discuss the same topics, negative similarities are not seen in this paper.

Why "Naïve"? While this technique, creating a *word x document* word-count matrix and then comparing document similarities using the cosine measure, provides an effective means of comparing documents, it is too simple. Suppose that this technique was being used to grade essays, a technique validated by Foltz *et al.* (1999); a grader could create a "model" essay, then compare each student's essay to the model essay. Student essays with a higher cosine similarity would be given a higher score and vice versa. Unfortunately, this method ignores many crucial aspects of written language. First, it does not take writing style or usage into account—the student essay could be written as cryptically as this paper, but if it contained the right number of the correct words (i.e. those that matched the model essay), the student would receive a high grade; no vector-space model of documents, including LSA, will be helpful in correcting this problem. Next, it's possible that an essay that uses the correct words, but whose meaning is unrelated, will receive a high grade. For example, the essays

Model Essay: I was freed from jail, so I booked a hotel reservation.

Student Essay 1: I was caught in a hotel, so I was booked and sent to jail.

would receive a cosine similarity of 0.6. While this is not particularly close to a perfect score of 1, a score of 0.6 seems high given that the essays are unrelated in all but vocabulary. These essays suffer from a problem called polysemy, which means that one word can describe many concepts; in this case, the word *booked* causes the problem. The final, and most subtle problem, is that any concept can be expressed by any number of words; this problem is called synonymy. For example, the following essay would not receive a high score:

Student Essay 2: Once released from prison, I elected to stay in a hotel.

Clearly this essay is more similar to the model essay in meaning, yet its cosine similarity would be around 0.35. Notice that this essay would receive a lower grade than Student Essay 1, the unrelated essay. In order to combat some of these problems, specifically polysemy and synonymy, Latent Semantic Analysis (LSA) is used. As will be seen, once LSA has been applied to documents, cosine similarity measures will better capture the human notion of *meaning*; for this example, it

would ideally decrease the similarity of the Model Essay and Essay 1, and would increase the similarity between Essay 2 and the Model Essay.

4.2 Latent Semantic Analysis: Introduction and Application

Motivating Example As a guide to understanding LSA, consider the following fictional group of FOMC documents, creatively titled d_1 , d_2 and d_3 . The words spoken in these meetings are given in the following term-document matrix:

$$\mathbf{M} = \begin{matrix} & d_1 & d_2 & d_3 \\ \text{Inflation} & \left[\begin{array}{ccc} 10 & 2 & 8 \\ 2 & 10 & 3 \\ 4 & 8 & 10 \\ 0 & 9 & 2 \\ 1 & 5 & 2 \end{array} \right. \\ \text{Growth} \\ \text{Price} \\ \text{Housing} \\ \text{Laughter} \end{matrix}$$

Given this matrix, it is possible to gather some idea of the circumstances of each meeting. While inflation was of primary concern in Meetings 1 and 3, Meeting 3 opted to use both *inflation* and *price* to discuss this topic. However, *price* was also used heavily in Meeting 2, and given the heavy usage of *housing*, one might infer that this *price* is related to housing prices instead. If this is considered in conjunction with the heavy usage of *growth* and *laughter*, it seems plausible that Meeting 2 took place amidst a relatively positive macroeconomic environment (no worries about inflation), while Meetings 1 and 3 were perhaps troubled by an inflation problem. Therefore, any text analysis performed should find the similarity between d_1 and d_3 to be greater than the similarity of either d_1 or d_3 with d_2 . However, as is apparent from the discussion of the word *price*, a simple word count might mistake the true meaning of a document; in this case, the polysemy of *price* and the synonymy of describing inflation is at fault. Before proceeding, notice that cosine similarity between the documents in \mathbf{M} can be visualized in the following symmetric **similarity matrix** \mathbf{C} :

$$\text{sim}(\mathbf{M}) \equiv \mathbf{C} = \begin{matrix} & d_1 & d_2 & d_3 \\ d_1 & \left[\begin{array}{ccc} 1 & & \\ 0.42 & 1 & \\ 0.86 & 0.69 & 1 \end{array} \right. \\ d_2 \\ d_3 \end{matrix}$$

where each element $c_{ij} \in \mathbf{C}$ represents the cosine similarity between documents i and j . The goal from this point forward is to find a way to manipulate the document-vectors such that these similarities more closely reflect the "true similarities"—those which an all-reading human being would compute.

Introduction Latent Semantic Analysis, first introduced by Deerwester *et al.* (1990) and Dumais *et al.* (1988), is a method that addresses the question of meaning. The method, first introduced as a way of matching user queries to text results, is particularly concerned with overcoming the problems of polysemy and synonymy. To do this, LSA considers how words are used together, with the claim that the words which co-occur with a particular term help one to capture the meaning of that term; knowing that the word *book* co-occurs with the word *hotel* gives an idea of the intended use of *book*—it provides context. The idea, as Deerwester *et al.* (1990) explain, is that word choice is, within certain bounds, random. One could imagine, for example, this paper being written in a completely different manner than its present writing, while still conveying the same ideas. In this sense there exists a great deal of noise when using words, or word-counts, alone to capture meaning. Theoretically, however, there is an underlying "latent," or hidden, semantic structure to which any text can be mapped.¹² Thus, to the extent that language is used to convey an idea in some underlying semantic space, randomness is introduced. The correctness of this argument, that the introduction of language inherently adds "noise" to the conveyance of an idea, is not the primary concern of this paper. However, overcoming the randomness of word choice, which can be introduced through typographical errors, polysemy, synonymy or something else, will certainly aid in the examination of FOMC documents, and it is for this reason that LSA is used.

Dimension Reduction and Connection to PCA So, if the goal is to reduce noise, and since, it has been shown, a collection of texts can be transformed into a data matrix of word frequencies, then familiar linear-algebraic approaches present natural solutions. The reader familiar with such approaches might first consider principal component analysis (PCA). In PCA, the eigenvectors and eigenvalues of the covariance matrix of a dataset with n variables are computed. The eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ are called factors, and each is a linear combination of all variables in the dataset; thus, the same data can be described using the factors in place of the original variables. The eigenvalues $\lambda_1, \dots, \lambda_n$, which are ordered so that $\lambda_1 > \dots > \lambda_n$, rank each corresponding factor based on how much variance in the original data that factor captures. In other words, the first factor is a straight line fitted through the data such that that factor explains as much of the variance in the data as possible.¹³ There typically is a point in $1, \dots, n$ where relatively little of the variance is explained by additional factors. Thus, the dataset can be represented using fewer dimensions by eliminating the components (factors) with the smallest corresponding eigenvalues *without losing*

¹²To continue in the language of functions, the functions from language to ideas f_1 and from ideas to language f_2 are both non-injective. The domain of the first function f_1 , from language to ideas, is an infinite set of possible expressions and the range is the infinite set of "ideas." The other function f_2 has these flipped. The randomness arises from the non-injectivity of the functions; a speaker can express a single idea any multitude of ways (f_1 : synonymy), and many ideas can be expressed by a single expression (f_2 : polysemy).

¹³The first factor is very similar to an OLS estimate; however, the OLS line minimizes the error perpendicular to the independent variables, and PCA minimizes the error perpendicular to the fitted line.

much information. In other words, if t components were used to describe the data, then the most important $s < t$ factors could be retained in order to remove the relatively "noisy" ($t - s$) factors. It is in this aspect of dimension reduction that PCA is similar to singular value decomposition (SVD), which provides the defining feature of LSA. In the case of word counts, the components would be constructed of terms, and could be thought of as "themes," although the interpretation of each theme would be difficult as the number of terms increased. The difficulty arises because each theme, by definition of PCA, is a linear combination of all words. Therefore, interpretations of themes involve studying the factor loadings of all words in the vocabulary of the text collection, which is beyond the scope of this paper. Instead, dimension reduction is the tool that will be used. See Appendix A.1 for a further development of the connection of PCA to SVD.

Singular Value Decomposition Manning *et al.* (2008) deftly define the problem which is to be solved. The purpose of LSA, via SVD, is to solve a constrained optimization problem; namely, the goal is to approximate the original data matrix, which has rank r , with a new matrix with some rank $k < r$, subject to the constraint that not too much information is lost about the original data. In forcing the data (documents and terms) to lie in fewer dimensions, "the SVD should bring together terms with similar co-occurrences." In other words, if terms tend to occur together at a certain rate, say the ratio is $\frac{x}{y}$, then any time that the ratio is something other than $\frac{x}{y}$, the SVD will re-estimate these word counts to be closer to $\frac{x}{y}$. Formally, the constraint is to minimize the Frobenius norm.

Definition 2. For any $m \times n$ matrix \mathbf{X} with elements x_{ij} , the Frobenius Norm of \mathbf{X} is given by

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$$

Therefore, the constrained optimization problem is the following:

Optimization Problem. The goal of LSA, through SVD, is to approximate \mathbf{X} , a matrix with rank r , by \mathbf{X}_k , a matrix with rank $k < r$, such that $\|\mathbf{X} - \mathbf{X}_k\|$ is minimized.

So, for example, if the approximation of \mathbf{X} is \mathbf{X} , then $\|\mathbf{X} - \mathbf{X}\|_F = 0$. Rank reduction corresponds to the previous discussion of the randomness of word choice. Typically, a *term x document* matrix will have rank equal to the number of documents. In the original state of this matrix "what you see is what you get" when computing cosine similarities between the columns of this matrix. If two documents discuss motorized vehicles and one uses the term *car* while the other uses *automobile*, then too bad—these documents will not be considered similar. However, the power of dimension (rank) reduction is that in the new approximation of \mathbf{X} by \mathbf{X}_k , the "word count" assigned to *car* and *automobile* is not a true word count, but what that word count *should be*, given the other terms in that document. This is the strength of LSA; a term doesn't have to occur explicitly in a document,

but once a low-rank approximation is made, that term will have an estimated occurrence in that document—what the occurrence *should be* in the latent semantic space.¹⁴ More concretely, if a word tends to co-occur with certain words in a certain ratio, then the "estimated word counts" in the documents where this ratio is respected will be the same as the actual word counts. If there is an anomalously high use of the term in a certain document (i.e. one that doesn't respect the ratio), then the term will be down-weighted by the SVD in that document (this anomalous ratio will also up-weight the term in other documents). The solution to the optimization problem can be found using singular value decomposition of the *term x document* matrix.

Definition 3. Singular Value Decomposition: An $m \times n$ matrix \mathbf{X} can be decomposed into the product of three matrices:

$$\mathbf{X}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n} \mathbf{V}_{n \times n}^T$$

Where \mathbf{U} and \mathbf{V} are orthonormal matrices and $\mathbf{\Sigma}$ is a diagonal matrix. The columns of \mathbf{U} are the eigenvectors of $\mathbf{X} \mathbf{X}^T$, the columns of \mathbf{V} are the eigenvectors of $\mathbf{X}^T \mathbf{X}$, and the non-zero elements of $\mathbf{\Sigma}$ are the square-roots of the eigenvalues of $\mathbf{X} \mathbf{X}^T$ (and of $\mathbf{X}^T \mathbf{X}$).

A formal explanation of why SVD provides the solution to the optimization problem is beyond the scope of this paper. Instead, an intuitive explanation of how SVD does this will be given by relating SVD to PCA. For a more detailed treatment of this connection, see Appendix A.1.

The appearance of the matrices $\mathbf{X} \mathbf{X}^T$ and $\mathbf{X}^T \mathbf{X}$ hints that SVD is related to PCA; there is a great deal of significance in these matrices. The following discussion will concentrate on the importance of $\mathbf{X}^T \mathbf{X}$, but the discussion would be analogous for $\mathbf{X} \mathbf{X}^T$. Recall the matrix \mathbf{M} from the example above, for which the matrix $\mathbf{M}^T \mathbf{M}$, denoted by \mathbf{M}^C , is

$$\mathbf{M}^C \equiv \mathbf{M}^T \mathbf{M} = \begin{matrix} & \begin{matrix} d_1 & d_2 & d_3 \end{matrix} \\ \begin{matrix} d_1 \\ d_2 \\ d_3 \end{matrix} & \begin{bmatrix} 121 & & \\ 77 & 274 & \\ 128 & 154 & 181 \end{bmatrix} \end{matrix}$$

Each component, \mathbf{m}_{ij}^C , of this matrix gives a measure of overlap between the words in document i and document j . The more terms that two documents share, the higher the corresponding entry in $\mathbf{M}^T \mathbf{M}$ will be; if a term isn't shared, then that term does not contribute at all. In contrast, the highest overlaps, generally, are seen in the ii -th positions, which is expected since each document shares a large number of terms with itself. The significance of this matrix is that it is a few steps away from being the variance-covariance matrix of the three documents; a key aspect of PCA (see Appendix A.1) and a good place to form some intuition. To see this, let \mathbf{M}^D be the column-demeaned representation of \mathbf{M} ; that is, for column i , let each of the entries be original entry, minus

¹⁴For a further discussion of the "should be" language and LSA, see Landauer *et al.* (1998).

the mean of that column. For the current example,

$$\mathbf{M}^D = \mathbf{M} - \begin{bmatrix} 3.4 & 6.8 & 3 \\ 3.4 & 6.8 & 3 \\ 3.4 & 6.8 & 3 \\ 3.4 & 6.8 & 3 \\ 3.4 & 6.8 & 3 \end{bmatrix} = \begin{bmatrix} 6.6 & -4.8 & 3.0 \\ -1.4 & 3.2 & -2.0 \\ 0.6 & 1.2 & 5.0 \\ -3.4 & 2.2 & -3.0 \\ -2.4 & -1.8 & -3.0 \end{bmatrix}$$

Then, if $\frac{1}{n-1}$ is multiplied by $(\mathbf{M}^D)^T \mathbf{M}$, where n is the number of terms in the vocabulary of the text collection, the result is the covariance matrix of the documents .

$$(\mathbf{M}^D)^T \mathbf{M}^D = \begin{bmatrix} 63.2 & -38.6 & 43.0 \\ -38.6 & 42.8 & -16.0 \\ 43.0 & -16.0 & 56 \end{bmatrix} \Rightarrow \frac{1}{4} (\mathbf{M}^D)^T \mathbf{M}^D = \begin{bmatrix} 15.80 & -9.65 & 10.75 \\ -9.65 & 10.70 & -4.00 \\ 10.75 & -4.00 & 14.00 \end{bmatrix} = \text{cov}(\mathbf{M})$$

Thus, while in practice this transformation of $\mathbf{X}^T \mathbf{X}$ is not used in SVD, it is important to understand the significance of $\mathbf{X}^T \mathbf{X}$; namely, that it gives a measure of overlapping terms between documents. That it is a few steps from being the covariance matrix is useful in seeing the link between SVD and PCA.

To see the connection, let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be the eigenvectors of $\mathbf{X}^T \mathbf{X}$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, where the eigenvalues have been arranged such that $\lambda_1 > \dots > \lambda_n$, and let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be the eigenvectors of $\mathbf{X} \mathbf{X}^T$. Also, let the $\sqrt{\lambda_i} \equiv \sigma_i$ be the elements in the diagonal of the diagonal matrix Σ . Thus, from the definition of SVD, $\mathbf{X} = [\mathbf{u}_1 \cdots \mathbf{u}_n] \Sigma [\mathbf{v}_1 \cdots \mathbf{v}_n]^T$. The connection to PCA should be clear; as in PCA, each σ_i gives a relative ranking of the importance of each \mathbf{u}_i and \mathbf{v}_i in explaining the variance in the columns of \mathbf{X}^T and \mathbf{X} respectively. That is, the smallest σ_i are analogous to the smallest λ_i in PCA; they correspond to the factors which explain the least about the variation in the data—they are little more than noise. Using the same language above, these noisy factors correspond to the randomness introduced by using language in the conveyance of ideas in the latent semantic space. For this example, the SVD is given by

$$\mathbf{U} = \begin{bmatrix} -0.48 & -0.74 & 0.45 & 0.16 & 0.00 \\ -0.45 & 0.41 & 0.41 & -0.60 & -0.31 \\ -0.62 & -0.08 & -0.77 & -0.08 & -0.07 \\ -0.35 & 0.50 & 0.16 & 0.77 & -0.12 \\ -0.24 & 0.19 & 0.10 & -0.11 & 0.94 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 21.17 & 0 & 0 \\ 0 & 10.78 & 0 \\ 0 & 0 & 3.41 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} -0.40 & -0.62 & 0.67 \\ -0.70 & 0.68 & 0.22 \\ -0.59 & -0.39 & -0.71 \end{bmatrix}$$

Dimension Reduction It is clear now that the σ_i play a role similar to the eigenvalues in PCA in determining the extent to which factors explain variance. In this regard, the following fact proven by Eckart & Young (1936) should be relatively clear. This fact, which solves the constrained optimization problem described above, says that

Theorem 1 (Eckart & Young (1936)). *For any matrix \mathbf{X} with rank r and singular value decomposition as described above,*

- (1) *the best approximation (i.e. the approximation \mathbf{A} such Frobenius norm of $\mathbf{X} - \mathbf{A}$ is minimized) of \mathbf{X} by a matrix \mathbf{X}_k with rank $k < r$ is found by “zeroing out” the $r - k$ smallest singular values in Σ , now called Σ_k , and letting $\mathbf{X}_k = \mathbf{U} \Sigma_k \mathbf{V}^T$, and*
- (2) *the size of the Frobenius “error” $\|\mathbf{X} - \mathbf{X}_k\|_F$ is given by the $k + 1^{\text{st}}$ singular value.*

While Eckart & Young (1936)’s will not be given, a proof that \mathbf{X}_k in (1) is a low-rank approximation of \mathbf{X} is instructive; see Appendix A.2.

Finally, it is easy to see that \mathbf{X}_k has the same number of rows and columns as \mathbf{X} . For this example, the reduced Σ_k is given by

$$\Sigma_k = \begin{bmatrix} 21.17 & 0 & 0 \\ 0 & 10.78 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so it follows that

$$\mathbf{M}_k = \mathbf{U} \Sigma_k \mathbf{V}^T = \begin{array}{ccc} & d_1 & d_2 & d_3 \\ \text{Inflation} & \left[\begin{array}{ccc} 8.97 & 1.67 & 9.07 \\ 1.06 & 9.70 & 3.99 \\ 5.77 & 8.57 & 8.14 \\ -0.36 & 8.88 & 2.38 \\ 0.77 & 4.93 & 2.24 \end{array} \right] \\ \text{Growth} & & & \\ \text{Price} & & & \\ \text{Housing} & & & \\ \text{Laughter} & & & \end{array}$$

After computing the multiplication to achieve \mathbf{X}_k (or \mathbf{M}_k in this example), the strengths of LSA can be seen. Because this is a low-rank approximation of the original matrix which uses the most “important” term co-occurrences in its composition (recall the discussion of $\mathbf{X}^T \mathbf{X}$), the rows in the matrix (which still correspond to terms) are no longer word-counts, but are instead the representations of those terms in the “latent semantic space”; i.e. the lower dimension space—that is, they are approximations. For example, it was mentioned earlier that d_1 appears to have taken place in less than favorable economic times, and LSA has determined the entry for *laughter* in d_1 should be lower than it actually was (it was originally 1). Many other interesting relationships can

be seen by comparing \mathbf{M} to \mathbf{M}_k ; a last example is that *inflation* has been down-weighted by LSA in d_2 and up-weighted in d_3 , which is in line with the original discussion of these documents.

Finally, the cosine similarity matrix of the low-rank approximation is given by

$$\mathbf{C}_k = \begin{array}{c} \begin{array}{ccc} & d_1 & d_2 & d_3 \\ d_1 & \left[\begin{array}{ccc} 1 & & \\ 0.42 & 1 & \\ 0.94 & 0.71 & 1 \end{array} \right] \\ d_2 \\ d_3 \end{array} \end{array}$$

This results in exactly the goal set out in the beginning of this section: to find a way of *better* computing document similarities. The cosine similarity between d_2 and d_1 is unchanged, while the d_2, d_3 similarity increased by 0.02. More notable, however, is the 0.11 increase in the cosine similarity between d_1 and d_3 ; recall that these were two documents which used different words to describe the same problem (inflation). This aligns with the initial belief that documents 1 and 3 are relatively similar in meaning, and overcomes the problem that documents 1 and 3 were not as similar in language. It is this power of gaining precision, in a relatively easy to compute way, that motivates LSA's use in this paper.

Summary The choice to apply LSA to a *term x document* matrix implies an assumption that word counts alone are insufficient for determining document similarities. This follows because words are a noisy way of conveying ideas, and ideas live in the hidden (latent) semantic space. Thus, describing documents in the latent semantic space requires approximating the original matrix by one with lower rank; that is, in a lower dimensional space which more closely resembles the latent semantic space. LSA reaches this space by reducing the noisy occurrences of terms in hopes of deducing the "true meaning" of words and documents. The retained dimensions are chosen carefully in LSA—they are those which explain the most about the variations in term usage, while the dropped dimensions are claimed to be noisy and add little to the comprehension of the data. Thus, the effect of noisy, or random, term co-occurrences is lessened. For instance, if term 1 almost always co-occurs with term 2, but fails to in one anomalous document, then the factor (from PCA) or basis vector (from SVD) that is dropped will be one that does well in explaining this deviation, but is almost always irrelevant in describing the relationship between terms 1 and 2. Once the original matrix is approximated by one of lower rank, because noisy co-occurrences have been reduced, the lower-rank matrix better reflects the true meaning or intention of the documents and terms. Thus, document (or term) similarity measures taken at this stage will better approximate the true similarities of the documents in the latent semantic space, or, perhaps more importantly, in the human understanding.

4.3 Additional LSA Parameters

While LSA alone goes a long way in improving the accuracy with which document similarities are measured, there is a bit more that can be done. The following adjustments, consisting primarily of term weightings and elimination of irrelevant terms, are standard fare in the LSA and retrieval literature. The reasoning behind each of the adjustments will likely be more clear than was the reasoning behind LSA. Most importantly, these noise-reducing techniques, which are applied to a *term x document* matrix before LSA is applied, are designed to complement LSA.¹⁵ They are, in a way, a "first cut" through the noisy data at a level which is simple to understand and requires no use of abstract concepts like the latent semantic space. There are always options in specifying the following adjustments, but the ones used in this paper's analysis reflect those most commonly used in LSA literature. All of these adjustments, and the LSA itself, are all performed using the MATLAB[®] toolbox Text Matrix Generator (Zeimpekis & Gallopoulos, 2006). This toolbox takes `.txt` files as input and outputs a *term x document* matrix with the appropriate user-selected weightings.¹⁶

- **Front and End Matter:** For minutes and transcripts, the front and end matter of the document is removed. This primarily includes the titles of the documents, attendance and voting records. These procedural segments of the documents are not of interest when studying the substantive content of the meetings and their minutes.
- **Stopwords:** When parsing documents into a *term x document* matrix, terms that are included in a "stoplist" are excluded. As is customary, this list contains "common" words that contribute little meaning to the document since they are used so often. The excluded words are predominantly prepositions, conjunctions and pronouns. Also excluded are FOMC member names, months, and Federal Reserve District numbers (*first* through *twelfth*). The complete list is found in Table 1. Numbers and alphanumerics (e.g. *1st*) are also removed.
- **Minimum and Maximum Term Length:** Words must contain at least three characters, and fewer than 20. The lower bound should catch any relatively common words not caught by the stop list, or any small typographical errors. Conversely, the upper bound should catch some typographical errors or errors in the OCR¹⁷ processing of the original files, which were originally in PDF format. These types of errors would typically be a conjoining of words, e.g. *federalreservesystem*.

¹⁵They can also stand alone; much of Hoberg's work (Hoberg & Phillips (2009), Hoberg & Phillips (2010), Hanley & Hoberg (2010), Hanley & Hoberg (2012)) choose a few of these adjustments and then calculate cosine similarities off of these adjustment document vectors.

¹⁶It can also be used to perform the SVD dimension reduction, and to query a corpus of documents.

¹⁷Optical Character Recognition

Table 1: The stopwords used in the document processing.

Stopwords (Alphabetically L to R)						
a	about	again	all	am	an	and
angell	any	april	are	aren't	as	at
august	balles	baughman	be	because	been	being
bernanke	between	bies	black	blinder	boehne	both
boykin	broaddus	burns	but	by	chairman	coldwell
corrigan	december	did	didn't	do	does	doesn't
doing	don't	down	during	each	eastburn	eighth
eleventh	evans	february	ferguson	few	fifth	first
fisher	for	ford	forrestal	fourth	from	further
gardner	geithner	governor	gramley	gramlich	greenspan	guffey
guynn	had	hadn't	has	hasn't	have	haven't
having	he	he'd	he'll	he's	heller	hendricks
her	here	here's	hers	herself	him	himself
his	hoeing	hoenig	holland	horn	hoskins	how
how's	i	i'd	i'll	i'm	i've	if
in	into	is	isn't	it	it's	its
itself	jackson	january	johnson	jordan	july	june
keehn	kelley	kimbrel	kohn	kroszner	lacker	laware
let's	lilly	lindsey	march	martin	may	mayo
mcdonough	mcteer	me	melzer	meyer	milller	minehan
mishkin	more	morris	moskow	most	mr	mr.
mrs	mrs.	ms	ms.	mullins	mustn't	my
myself	ninth	no	nor	not	npvember	october
of	off	olson	on	once	only	or
other	ought	our	ours	ourselves	out	over
own	parry	partee	phillips	pianalto	poole	rice
rimbrel	rivlin	roberts	roos	rosengren	same	santomero
schultz	second	secretary	seger	september	seventh	shan't
she	she'd	she'll	she's	should	shouldn't	sixth
so	solomon	some	stern	such	syron	teeters
tenth	than	that	that's	the	their	theirs
them	themselves	then	there	there's	these	they
they'd	they'll	they're	they've	third	this	those
through	timlen	to	too	twelfth	under	until
up	very	vice	volcker	walich	wallich	warsh
was	wasn't	we	we'd	we'll	we're	we've
were	weren't	what	what's	when	when's	where
where's	which	while	who	who's	whom	why
why's	willes	winn	with	won't	would	wouldn't
yellen	you	you'd	you'll	you're	you've	your
yours	yourself	yourselves				

Total: 283

- Minimum and Maximum, Local and Global Term Frequency: Terms need only show occur once over the entire corpus to be included. There is no upper bound on the number of times a term can appear.
- Local Term Weighting: Strict word-counts, also known as *term-frequencies* TF_t , are used for weighting a term t 's relevance in each document. Other local weighting schemes typically involve weighing each term's frequency by some function such as the log function.
- Global Term Weighting: "Inverse Document Frequency" is used to weigh terms globally. This is another tool to minimize, but not eliminate, the weight given to terms that are used too frequently to add much insight into a document's meaning, primarily because they are used frequently in all documents (e.g. *monetary, policy*, etc.). The first step is to calculate the number of documents in which a term t occurs, or its document frequency DF_t . The inverse document frequency for a corpus of n documents is given by $IDF_t = \log\left(\frac{n}{DF_t}\right)$, where the log-base used by TMG is 2—an immaterial fact. This yields a lower weight for terms that appear in many documents, and vice versa. Multiplying the IDF_t by the term's frequency TF_t in each document yields the commonly used $TF - IDF$ weighting scheme that is employed here. In other words, a document vector \mathbf{d} which contains word counts (t_1, \dots, t_m) in a collection of n documents becomes

$$\mathbf{d}^* = \left(t_1 \log\left(\frac{n}{DF_{t_1}}\right), \dots, t_m \log\left(\frac{n}{DF_{t_m}}\right) \right).$$

- Stemming: A technique called "stemming" is used to reduce terms with the same base down to that base. For example, the terms *different, differ, differing* and *differential* would all be reduced to *differ*. The tradeoffs here are obvious, but its widespread use in the NLP and LSA literature gives hope that the benefits outweigh the costs. The unconvinced reader might consider the fact that without stemming, the two-word documents $[I \text{ play}]$ and $[she \text{ played}]$ would have a cosine similarity of 0.

From this point forward, the computation of any cosine similarities will be between documents which have been transformed by LSA and the points just listed; in other words, no longer is the vector representation of a document d simply a word-count vector, but a vector with word-counts which have been weighted by the points above and estimated by LSA.

5 External Policy Effects on FOMC Transparency

This section is concerned with measuring transparency in FOMC communications; that is, the extent to which the FOMC reveals the content of its meetings to the public via summary documents like the minutes. To make this precise, Section 5.1 gives an explicit definition of transparency as it is used here, and gives evidence that it is a worthwhile topic of study. The main finding of the analysis, given in Section 5.3, is that the Humphrey-Hawkins Act of 1978 increased the transparency of information transmission. Once the FOMC became subject to the Humphrey-Hawkins guidelines, the minutes and transcripts of each meeting became much more similar. Because this similarity was principally the result of changing minutes, it can be inferred that at this point in time, the FOMC decided to provide more transparent access to the content of its meetings through the minutes. Before reaching this result, however, the methodology behind the result is given in Section 5.2.

5.1 What is Procedural Transparency, and Why Study it?

Defining Transparency When used in the context of monetary policy, the word "transparency" can carry different interpretations. Thus, to understand how the term is used here, Table 2 presents the five forms of transparency relevant to central banks, as defined by Geraats (2002).¹⁸ Procedural transparency is what allows for the main analysis of this paper—the study of FOMC minutes and transcripts. It encompasses the procedure by which the accounts of FOMC decisions are released to the public via documents. It was a change in this type of transparency in the 1993 transition from private to more-public meetings that led to the natural experiment studied here and elsewhere (e.g. Woolley & Gardner (2009), Meade & Stasavage (2008)). Specifically, this change allows for a study of how the Committee responded to increased procedural transparency. What makes procedural transparency interesting is that increased procedural transparency, presumably, leads to increases in the other four types of transparency; it is the mechanism through which the other four are manifested. Specifically, as one of the broadest categories, it demands that the other forms be augmented; openly discussing how decisions are made requires an explanation—this would come in the form of, at the very least, greater political and economic transparency (i.e. explicitly stating the data and objectives used). Of course, the Fed is known for its carefully chosen words and strategic lack of specificity at times, but it would seem to be the case that procedural transparency is the channel through which the other, perhaps more economically important forms of transparency are augmented. For example, mandating that the Fed release the rule it uses to determine policy hinges on effective procedural transparency, though it is a form of economic, policy and political transparency. Therefore, understanding changes in the Fed's behavior in response to increased

¹⁸Geraats has written much about central bank transparency. See Geraats (2000), where these terms were first defined, or Geraats (2007) for other examples.

<i>Type of Transparency & Description</i>	
1.	<i>Political Transparency</i> refers to openness about policy objectives and institutional arrangements that clarify the motives of monetary policy makers. This could include explicit inflation targets, central bank independence and contracts.
2.	<i>Economic Transparency</i> focuses on the economic information that is used for monetary policy, including economic data, policy models and central bank forecasts.
3.	<i>Procedural Transparency</i> describes the way monetary policy decisions are taken. This includes the monetary policy strategy and an account of policy deliberations, typically through minutes and voting records.
4.	<i>Policy Transparency</i> means a prompt announcement and explanation of policy decisions, and an indication of likely future policy actions in the form of a policy inclination.
5.	<i>Operational Transparency</i> concerns the implementation of monetary policy actions, including a discussion of control errors for the operating instrument and macroeconomic transmission disturbances.

Table 2: The types of transparency relevant to central banks (Geraats, 2002, pp. F540).

procedural transparency, is a crucial aspect to consider when crafting policy aimed at increasing transparency.

At a more general level, there are a few reasons to believe that procedural transparency is worth studying. First, the public pays attention to the documents released by the Fed. Some anecdotal evidence here is a graph from Google Trends—a service from Google which plots the "interest over time" of any term. The graph, shown in Figure 3, shows how popular "fomc statement" and "fomc minutes" were over a 12 month period between 2013 and 2014. As expected, peaks in interest in the terms is in one-to-one correspondence with FOMC meetings; the 3-week lagged release of minutes is also clearly noticeable. So, at the very least, there appears to be public interest in the content of FOMC documents.

Next, while not typically on the FOMC agenda, discussions about its document releases have occurred numerous times in the Board Room. These are meaningful discussions, going beyond the procedural and suggesting that, at a minimum, the FOMC believes that the public cares about its discussions. The following quote from former Vice Chairman Blinder supports this claim:

MR. BLINDER. I believe... that we now have a situation where the people that speak the least about the Fed's decisions are those at the Fed, and we are interpreted voluminously. There is nothing wrong with that. We will still be interpreted voluminously even if we say things. But our statement is a chance for us to say what we are up to and why. (Transcripts, '76-'08, 01/31/1995)

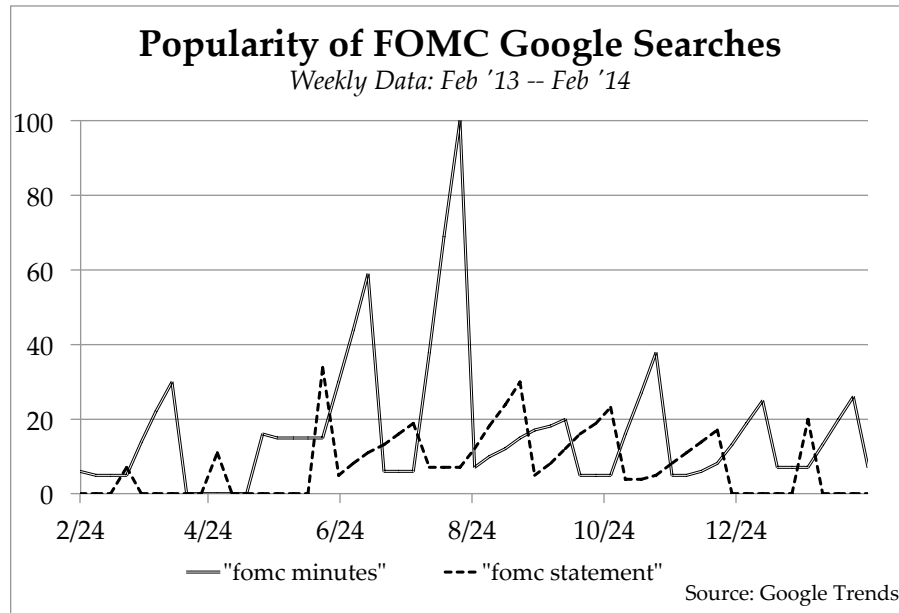


Figure 3: Results from the Google Trend queries "fomc statement" and "fomc minutes." The y-axis represents the frequency with which a given term is searched on Google, and is normalized so that the highest frequency is 100. Thus, this graph does not say how these search terms rank among all other terms, but it does give information on when the terms are searched.

Further, more comprehensive evidence shows that the FOMC was indeed discussing issues of procedural transparency in its meetings. Figure 4 uses the text-analysis techniques described below to infer the extent to which topics of procedural transparency were discussed at each FOMC meeting. It does this by comparing a query of procedural-transparency related words (*transcripts, meetings, statement* etc.) to the transcripts of each meeting. In order to determine the relevance of a query to each transcript, the cosine similarity of that query is taken against each of the transcripts, in the low dimension space—this was the original motivation behind the development LSA. If the *term x document* matrix $\mathbf{M} \in \mathbb{R}^{t \times d}$, then the query vector $\mathbf{q} \in \mathbb{R}^t$ has components which are mostly 0's except for the terms used in the query. Mapping the query q into its lower-dimensional space representation \mathbf{q}_k is done by letting $\mathbf{q}_k = \Sigma_k^{-1} \mathbf{U}^T \mathbf{q}$; at this point, it now has the "estimated" term values discussed above—this process is called "folding in." The cosine similarity is then computed against the other documents in the matrix $\Sigma_k \mathbf{V}^T$; see Appendix A.1 to see why this equivalent to computing the similarities in the reconstructed \mathbf{M}_k matrix. It is clear that topics of procedural transparency have, at times, accounted for significant portions of the discussion at FOMC meetings; without smoothing, some of the similarity measures are over 20% up the scale. Furthermore, the discussion coincides with procedural changes in policy—in 1976 marked the temporary end of transcript publication, and topics of procedural transparency persisted for a few years after that change. Since the early nineties, changes in publication policy have been relatively frequent—in

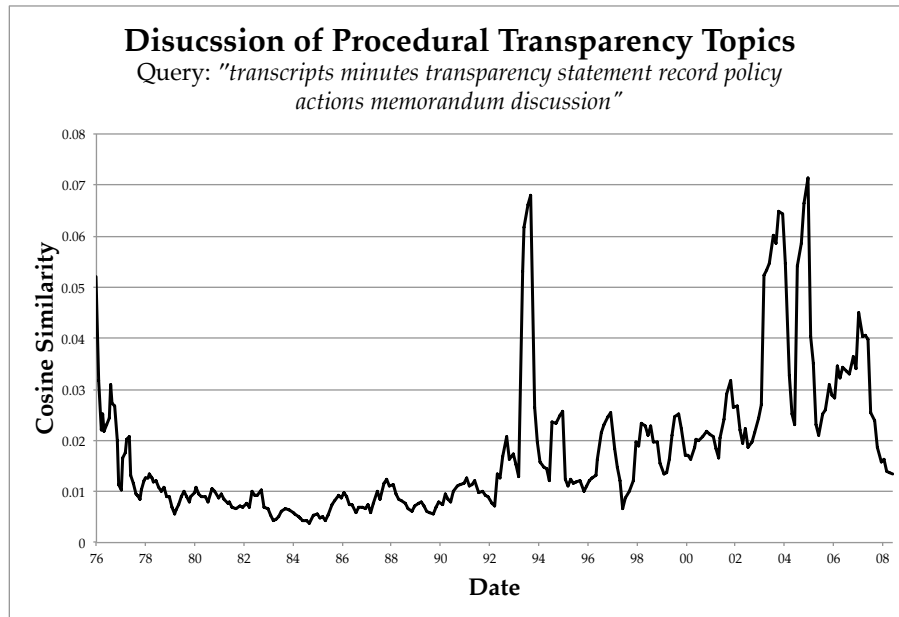


Figure 4: Similarity of the query with the content of each FOMC meeting. The "cosine similarity" gives a measure between 0 and 1 of how similar documents are; in this case, it says how similar the query is to each document, which in turn describes how much the ideas in the query were discussed in the meetings. Four meeting averages plotted.

1993 old transcripts were released, in 1995 new transcripts began to be released, and in 2005 minutes were released with shorter lags. Most of these changes are visible on the graph, indicating that a significant amount of discussion was behind each decision, and lends support to the validity of the methodology of this paper.

In summary, this evidence supports the claim that studying procedural transparency is worthwhile. Both the public and the FOMC pay careful attention to these documents, so understanding how and why this type of transparency changes is certainly meaningful.

5.2 Measuring Transparency Using Latent Semantic Analysis

Minute-Transcript Similarity The first relationship to consider is that of the minutes and transcripts. Specifically, for each meeting, the cosine similarity of that meeting's minutes and transcripts are computed. Because the minutes and transcripts for each meeting are contained in two separate documents, the cosine similarity is well defined. Thus, a variable called $MTsim$ is created which has one observation for each FOMC meeting that was physically held in Washington; phone meetings are not included because they do not have explicit minutes accompanying their transcripts. Because every document, it has been shown, can be represented as a vector, call the vector representing the minutes for meeting i m_i , and call the transcript vector t_i . Then each observation in the

variable $MTsim$ is given by

$$MTsim_i = \text{sim}(\mathbf{m}_i, \mathbf{t}_i).$$

This variable is interesting because it gives a measure for the extent to which the meeting minutes reflect what was said at the meeting. However, minute-transcript similarities can be difficult to interpret without also knowing something about the minutes and transcripts on their own. Minute-transcript similarities alone can hide the underlying changes in content of either document. A useful analogy is a comparison of economic growth, measured by GDP per capita ($\frac{Y}{n}$), between two countries, A and B . To know *only* that $(\frac{Y}{n})_A$ and $(\frac{Y}{n})_B$ are converging is not informative of $(\frac{Y}{n})$ in either country. The same problem is true for document comparisons; for example, consider the extreme case that the content of the minutes never changes—say they are constantly filled with uninformative and undifferentiated "Fed Speak." The content of transcripts, however, fluctuates over time, perhaps in response to macroeconomic conditions. A similarity measure which measured the distance between the content of each document, shown in Figure 5, would measure only the fluctuations in FOMC discussion. The similarity, however, could also be (mistakenly) interpreted

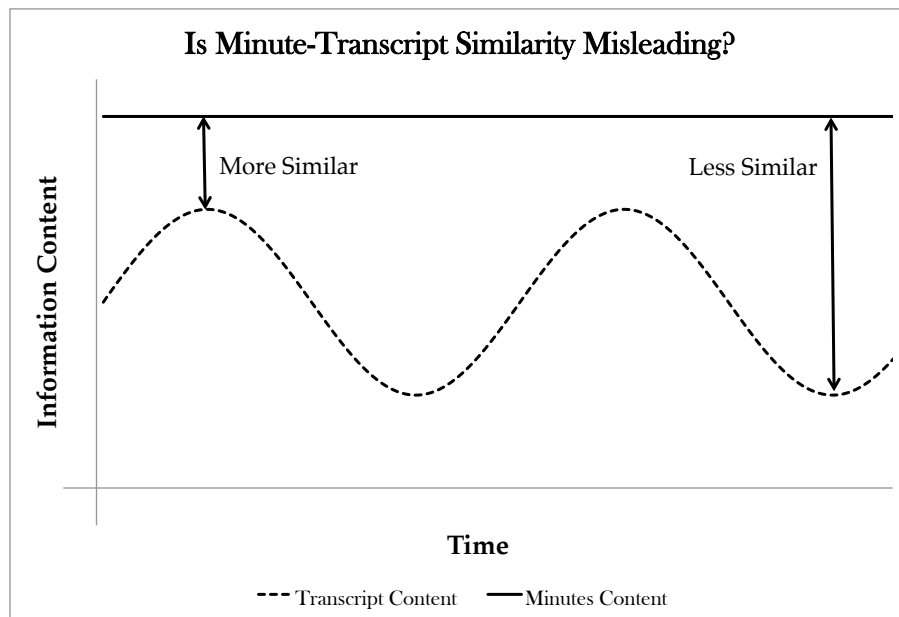


Figure 5: Minute-Transcript similarity can be misleading.

as the Fed choosing to tailor the *minutes* depending on macroeconomic circumstances. Thus, making causal inferences based on minute-transcript similarity alone can be misleading. Because the purpose of this analysis is to understand the translation process from transcripts to minutes, more must be done.

Minute-Minute and Transcript-Transcript Similarity In order to understand how documents have changed over time, this section introduces a way of comparing document content over time through the use of a *summary document*. Focusing on the exact content of meetings, or the topics of discussion of each meeting, would certainly be an interesting endeavor and would involve looking more closely at the particular elements in \mathbf{U} and \mathbf{V} from the SVD. The work of Boukus & Rosenberg (2006) did precisely this; they examined the prevalence of particular "themes" over time in FOMC minutes.¹⁹ However, the scope of this research calls only for observing changes in minutes and transcripts over time at a very high level. More concretely, studying how the content of a particular document type has changed over time can be done by comparing all documents of that type to a single document.

Thus, the next similarities computed are those between each meeting's transcript (minutes) with one *summary transcript* (*summary minutes*); this allows for an understanding of the changing nature of transcript (minutes) content across time, and will help determine the driver behind the minute-transcript similarity.²⁰ The particular choice for this single document will have an effect on the relative magnitude of the cosine similarity measures, although any document should allow for large level shifts to be seen. A first choice might be to randomly chose a transcript in the set of all transcripts and compare every other transcript pairwise to the chosen transcript. While this approach would allow for comparisons to be made, interpretation would be difficult because it is likely that most relatively contemporaneous transcripts would be highly correlated in meaning to the chosen transcript. Transcripts that were further away in time would, presumably, have a low similarity with the chosen transcript, making level-shifts difficult to see. To support this claim, comparisons were made between all transcripts, and the first transcript in the sample. The same was done with the last transcript. In both cases, there was relatively high, but falling, similarity between transcripts within two years; two years away, the similarities became relatively stable at low levels, while continuing to decline as time passed. Panels [2] and [4] in Figure 9 show these results.

Correcting for the low similarity present when choosing a single document motivates the use of the word *summary* in this discussion. One benefit of representing documents as vectors is that they are easy to manipulate, and it is easy to create "documents" without the need to actually write anything; one need only construct a document *vector*. Even better, the ability to easily create such a *pseudodocument* allows one to create a "document," in the form of a vector, with any desired properties. In this case, the creation of a summary document will be useful because such a document is sure to have the highest pairwise similarities with all other documents in the

¹⁹The "themes" are evaluated by looking at the columns of the \mathbf{U} matrix from the SVD; each element in the column tells how much each term contributes to that theme. The columns of \mathbf{V} give the theme contributions to each document.

²⁰The rest of this section describes summary *transcripts*, although the discussion for minutes is completely analogous.

set. For an example, consider again Figure 2; here, the vector representations \mathbf{x} and \mathbf{z} are relatively dissimilar (there is a large angle between them), but the document \mathbf{y} between the two is relatively similar to both documents; the angle between \mathbf{y} and \mathbf{z} and that between \mathbf{y} and \mathbf{x} are close in size. Because the magnitude (length) of a vector does not change its angle relative to other vectors, the simplest way to fit a vector \mathbf{v} between other vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ is to let $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2 + \dots + \mathbf{w}_n$. This new vector \mathbf{v} would be a kind summary document of all documents, a very bad summary in terms of length, but a very good summary in terms of portraying all of the information of all of the documents. Thus, for the present study, a summary transcript and summary minutes are created as sums of all transcripts and minutes, respectively. All documents are then compared pairwise to these summary documents. An expectation for the shape of a graph of these comparisons over time is an

upside-down parabola with the maximum near the middle of the time period. This follows from the idea that, for example, if one were to summarize the history of a country over a 100 year period, the most representative few years would likely be in the middle of the sample. For the FOMC, this argument assumes that the members discuss topics relevant to the time. If conversations are random and disconnected, then this argument would be irrelevant. The evidence, however, suggests that this assumption is reasonable. As will be shown below, the use

of such summary documents dramatically increases the general similarity levels across time. Table 3 gives a summary of the average similarity measures over time for each of the document types considered in the last two paragraphs (first, last and summary); as predicted, the highest similarities are found when comparing minutes/ transcripts to the summary documents.

Finally, to be clear, the variable of interest when studying the content of transcripts across time is given by

$$\text{TSTsim}_i = \text{sim}(\mathbf{t}_i, \mathbf{s}_t)$$

where \mathbf{t}_i is the vector representation of the transcript from the i^{th} meeting, and \mathbf{s}_t is the summary transcript vector. Similarly, for the summary minute vector \mathbf{s}_m and vector representation of the i^{th} meeting \mathbf{m}_i , the variable of interest for studying minutes is

$$\text{MSMsim}_i = \text{sim}(\mathbf{m}_i, \mathbf{s}_m)$$

Document Type	Average Similarity
First Minutes	.26
Summary Minutes	.70
Last Minutes	.44
First Transcripts	.58
Summary Transcripts	.80
Last Transcripts	.55

Table 3: Average of similarities between the relevant document (i.e. the first transcripts) and all other documents of that type (i.e. all transcripts).

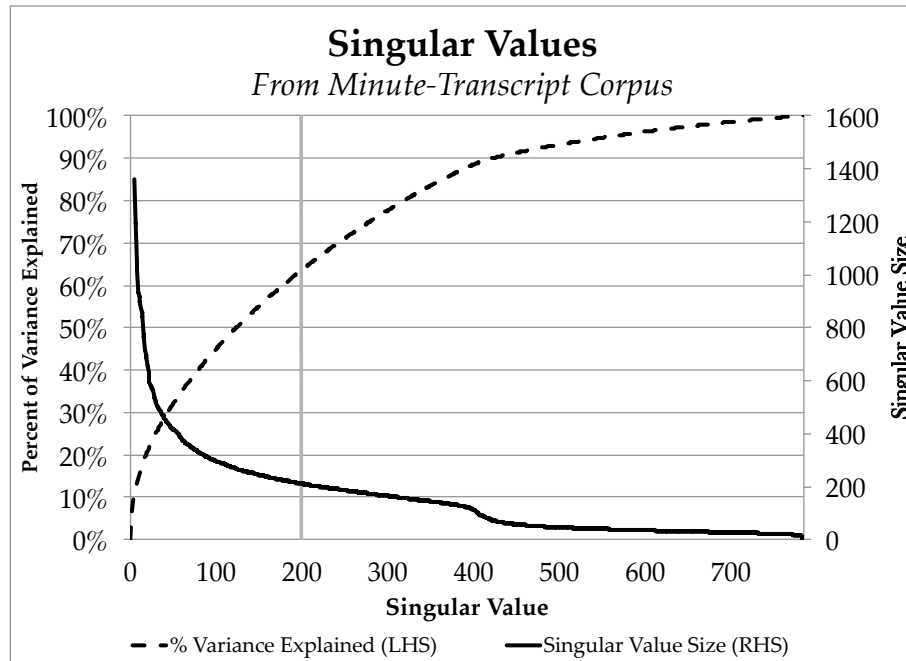


Figure 6: This graph shows the singular values (the elements $\sigma_i \in \Sigma$, $i \in [1, 780]$ from the SVD) and the percentage of the variance explained by these singular values. The plot of the singular values (solid line) excludes the first five SVs, so that more detail can be seen in the other SVs. For reference, they are 7274, 2515, 2330, 1483 and 1358. They account for 10% of the variance.

Latent Semantic Analysis Applied to Minute-Transcript Corpus The first application of LSA in this paper is to the corpus of minutes and transcripts since the beginning of the sample; 1967. The "transcripts" include the Memoranda of Discussion from 1967-1976 and the transcripts until the end of the sample in 2008. The "minutes" include the Record of Policy Actions from 1967-1993, and the minutes until 2008; see Figure 1 for a review of these documents.

The critical part of performing LSA is picking the number of dimensions to retain in the data. The ability to drop noisy dimensions is the reason that LSA is used in the first place. Figure 6 plots the singular values, the elements $\sigma_i \in \Sigma$, from the singular value decomposition of the minute-transcript *term* \times *document* matrix. First, notice that the data initially lies in 780 dimensions, one for each document in the 390 meeting sample, and the inclusion of all 780 dimensions accounts for all of the variation in the data. The LSA literature has not conclusively determined the "optimal number" of dimensions to retain, although generally a few hundred dimensions are chosen.²¹ For this sample, "a few hundred" seems to be a good choice, especially given the relatively sharp drop in singular value significance around the 400th dimension and the stabilization of the singular values between 100 and 300. At this point, just under 90% of the variance in the original data has been captured. This drop by itself is interesting and indicative of the structure of the documents, and is

²¹According to (Landauer *et al.*, 1998, pp. 269), "the number of dimensions retained in LSA is an empirical issue."

another validation of the LSA method; because there are 390 meetings in the sample, and because minutes and transcripts are, ideally, accounts of the same meetings, then this corpus should lie in 390-dimensional space. This space is not the "latent semantic space," it is simply a reflection of the fact that there were 390 meetings. Entering the "latent semantic space," then, requires retaining fewer than 390 dimensions.²² Therefore, for these reasons, and further evidence shown in Section 6.1, the first 200 dimensions are retained; to phrase it in the language of Section 4.2, the noisiest 580 dimensions have been discarded when recomposing the *term x document* matrix. Despite the elimination of so many dimensions, the remaining dimensions account for 64% of the variance in the data. Further, each additional dimension accounts for less than one-tenth of one percent of the variance.

5.3 Evidence from Minute–Transcript Similarity and Pseudodocuments

A look at the time series of the similarity between FOMC minutes and transcripts in Figure 7 reveals two distinct phases of this similarity over the past 40 years:

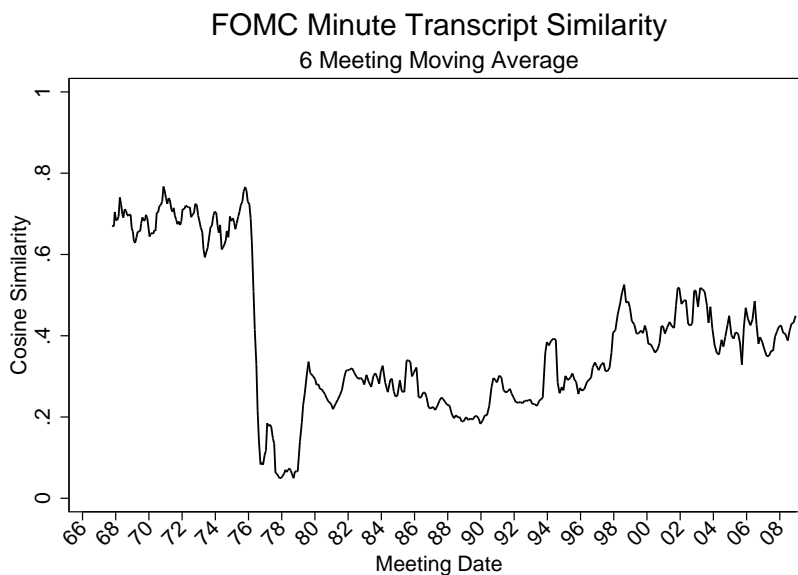


Figure 7: This figure gives the time-series representation of the MT_{sim} variable described in Section 5.2.

²²The use of "should" and "required," and the discussion in general, is not so much a scientific result, but more of a philosophical or linguistic approach that assumes the existence of such semantic spaces and noise in natural language.

5.3.1 The Memorandum of Discussion: 1967-1976

The first time period, from 1967-1976, reflects the time when the FOMC published the heavily edited "Memoranda of Discussion" (MOD), and a minute-like document called the Record of Policy Actions (ROPA). These documents correspond roughly to the transcripts and minutes of the present day, although the ROPA and present-day minutes are much closer in form than are the MOD and transcripts. Despite these differences, it is convenient to call the ROPA "minutes" and the MOD "transcripts" in order to characterize the change in these *types* of documents over time.

The noticeable drop in minute-transcript similarity at the end of this period reflects the change from the MOD to a true transcript; it also lends credence to the LSA methodology. The preparation of the MOD and ROPA was the responsibility of the FOMC secretary; Arthur Broida was the sitting secretary at the time of the transition. This negates the fear that this similarity drop occurred because of a change in the author of the documents. Woolley (1986) aptly describes the editing process:

Editorial revisions were made in preparing the [MODs]- which took place in the days immediately following each meeting —and they may not always be regarded as minor. One former governor described this as "toning down" words and making debates sound less emotional. A further and equally important part of the editorial process consisted of giving more coherence to the statements of some [members] than had actually been true in debate. [A] former governor reported going to [Broida] to congratulate him on his editorial skill: "I didn't say that, but that's what I meant!" (Woolley, 1986, pp. 196)

It is not surprising, then, that this period exhibited exceptionally high similarity between the two documents; they were both highly edited documents, prepared by the staff in the same office in the Fed. Unfortunately, however, because the MOD was not truly a transcript, it is of less use in the present study since it, like the minutes, is a "prepared" document, rather than "raw" like the transcripts. From this point forward, all documents produced before the April 20, 1976 meeting, the first meeting accompanied by a true transcript, are eliminated.

5.3.2 Closed-Door FOMC Meetings and the Humphrey-Hawkins Act: 1976-1993

The second period consists of the post-1976 section of Figure 7, which is shown in Figure 8: Recall that the MOD was discontinued in 1976 in response to the attention brought upon the FOMC by the Merrill case (see Section 3), a change that was largely pushed by Chairman Burns (Lindsey, 2003). The post-MOD period also has three distinct periods, with one obvious change and one less-obvious change. The first change occurred in early 1979. Chairman Miller explained the recent events in the transcripts of the February 6th meeting:

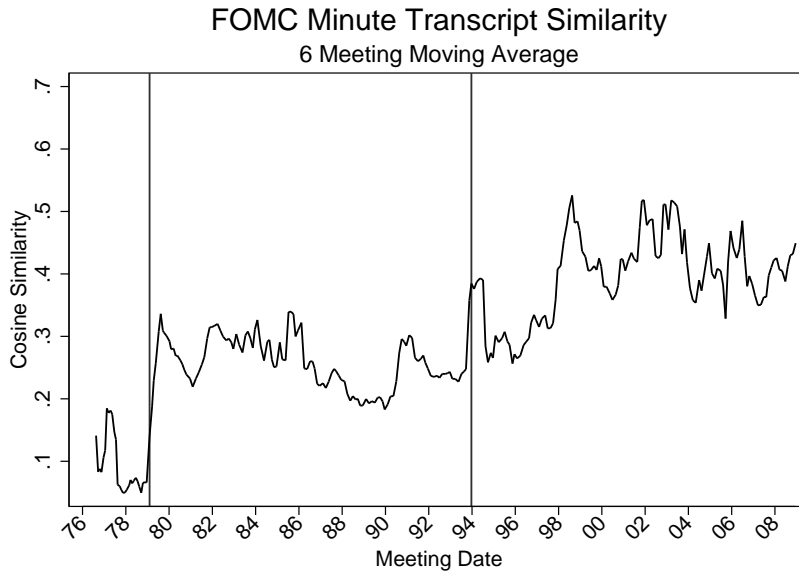


Figure 8: Zooming in on Figure 7.

CHAIRMAN MILLER. Good morning, ladies and gentlemen. Welcome to our historic meeting. It’s not only earlier [in the month than usual] but also involves for the first time the new Humphrey- Hawkins process. I assume that doesn’t mean much [will be different], but it does mean, [since we canceled our regularly scheduled meeting in January], that we haven’t met as frequently. (Transcripts, ’76–’08, 02/06/1979)
 [Bracketed text was inserted by the FOMC secretary at time of publication.]

The 1978 Humphrey-Hawkins Act, created in response to the economic uncertainties and underperformance of the 1970s, called on the Federal Reserve to report biennially to Congress its "objectives for the monetary and credit aggregates" for the year ahead (Meltzer, 2010, pp. 990) and "aim for lower inflation" and unemployment (Meltzer, 2010, pp. 846). Thus, the FOMC changed the schedule of its meetings, electing to schedule meetings that would enable them to prepare for the new oversight hearings (Meltzer, 2010, pp. 990). Figure 8 indicates that at this time, the transcripts and minutes became more similar in meaning; to make this conclusion more robust, Table 5 estimates the following model:

$$\begin{aligned}
 \text{MTsim}_t = & \alpha + \theta_1 \text{Post93}_t + \theta_2 \text{PreHH}_t + \theta_3 t + \beta_1 \text{GB Inflation}_t + \beta_2 \text{GB Inflation}_{t+1} \\
 & + \beta_3 \text{GB GDP}_t + \beta_4 \text{GB GDP}_{t+1} + \beta_5 \text{GB Unemp}_t + \beta_6 \text{GB Unemp}_{t+1} \\
 & + \sum_{C=\text{Miller}}^{\text{Bernanke}} \gamma_C C_t + \sum_{S=\text{Broida}}^{\text{Madigan}} \varphi_S S_t + u_t
 \end{aligned} \tag{2}$$

θ°	0	5	10	15	20	25	30	35	40	45
$100(\cos(\theta))$	100	99.6	98.5	97	93	91	87	82	77	71
θ°	50	55	60	65	70	75	80	85	90	
$100(\cos(\theta))$	64	57	50	42	34	26	17	9	0	

Table 4: Geometric Interpretation of Cosine Similarity: each column has the angle θ (in degrees) and its cosine—this gives the cosine similarity measure ($100(\cos(\theta))$) a geometric interpretation (θ). For example, a cosine similarity of 64 means that there is a 50° angle separating the documents.

where Post93 and PreHH are indicator variables for post-1993 (discussed later) and pre 1979 respectively, t is a trend variable, macro variables at time t are the Greenbook forecasts for the current quarter, macro variables at time $t + 1$ are the Greenbook forecasts for one year ahead of the present meeting, the variables C are indicator variables for the current chairman (Burns omitted) and the S variables are dummy variables for the FOMC secretary (Bernard omitted). Because the FOMC secretary is less well-known, Table 6 lists the secretaries over time. The similarity measures have been scaled by 100 to make the results more easily understood; for example, in (4), a 1% increase in inflation in $t + 1$ would decrease the cosine similarity by 3.8 (or 0.038 before scaling). Regressions (3) and (4) omit the trend variable, and equations (1) and (3) omit the chairmen and secretary dummies. FOMC forecasts for macroeconomic variables from the Fed's Greenbook are used in lieu of realized values because these best reflect the knowledge available to the FOMC concerning economic indicators. Because the primary focus of this paper is understanding why the members and FOMC behave a certain way, it is best to use the economic information that they expect (or, at a minimum, that the Board staff has suggested to them). As is mentioned by Romer & Romer (2008), a significant portion of each meeting, the "economic go-around," is devoted to discussing forecasts and current economic conditions.²³ Furthermore, each member has access to the Fed staff's forecasts before the meetings; thus, within a certain margin of error, the forecasts represent what the members believe to be the economic outlook. The most important coefficients here are θ_1 and θ_2 , the dummies on post-1993 and pre-1979 respectively. In particular, under each specification, the effect of the Humphrey-Hawkins act was at least 16 cosine-similarity points, and in the main regression (2), the effect was a highly significant 27. Geometrically, using Table 4 as a reference, these correspond to a 10° and 25° decrease in the angles between the minute and transcript document vectors after 1979. The other variables are not discussed here because of the aforementioned difficulty in interpreting minute-transcript similarity measures (the effect could be minute *or* transcript driven). However, the evidence confirms that the minutes and transcripts became more similar after 1979—transparency increased.

²³Romer & Romer (2008) discuss the member-level forecast that proceed the Fed's biannual testimony before Congress, not the Greenbook forecasts, but the cited information still applies.

Table 5: FOMC Minute and Transcript Changes in Content after 1993

	(1)	(2)	(3)	(4)
	MTsim	MTsim	MTsim	MTsim
Post93	9.131**	-1.994	13.12***	5.831*
Pre Hum-Haw.	-16.29***	-27.33***	-20.93***	-30.59***
Inflation _t	0.413	-0.393	0.569	-0.0674
Inflation _{t+1}	1.233	-0.950	-0.281	-3.849**
GDP _t	0.0547	-0.281	-0.0965	-0.645
GDP _{t+1}	1.556	-0.447	1.799	0.110
Unemp _t	-0.484	-1.630	-1.677	-4.047
Unemp _{t+1}	1.577	1.417	1.735	1.943
Mtg. # (t)	-0.0862*	-0.227***	[0.0655]	
MILLER		-5.605	[4.770]	-2.734
VOLCKER		-18.87**	[7.032]	-12.43
GREENSPAN		-26.76**	[9.565]	-15.49
BERNANKE		-35.05**	[11.90]	-19.33
BROIDA		21.58*	[9.019]	19.15*
ALTMANN		18.82**	[6.792]	18.72**
AXILROD		12.64**	[4.711]	14.18**
KOHN		2.045	[6.748]	3.327
REINHART		-2.125	[8.073]	10.49
MADIGAN		1.673	[12.08]	16.22
Constant	22.92***	89.67***	20.49**	63.68***
Observations	235	235	235	235
R ²	0.512	0.583	0.498	0.559

Standard errors in brackets

The macro variables are Greenbook forecasts for the current meeting's quarter (t) and one year ahead (t+1).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As was discussed in Section 5.2, it is necessary to also study the underlying structure of the minutes and transcripts to determine in which of these documents (minutes or transcripts) the change occurred. Figure 9 shows the cosine similarities of the minutes and transcripts with the summary minutes and transcripts, respectively. Before determining of the impact of the Humphrey-

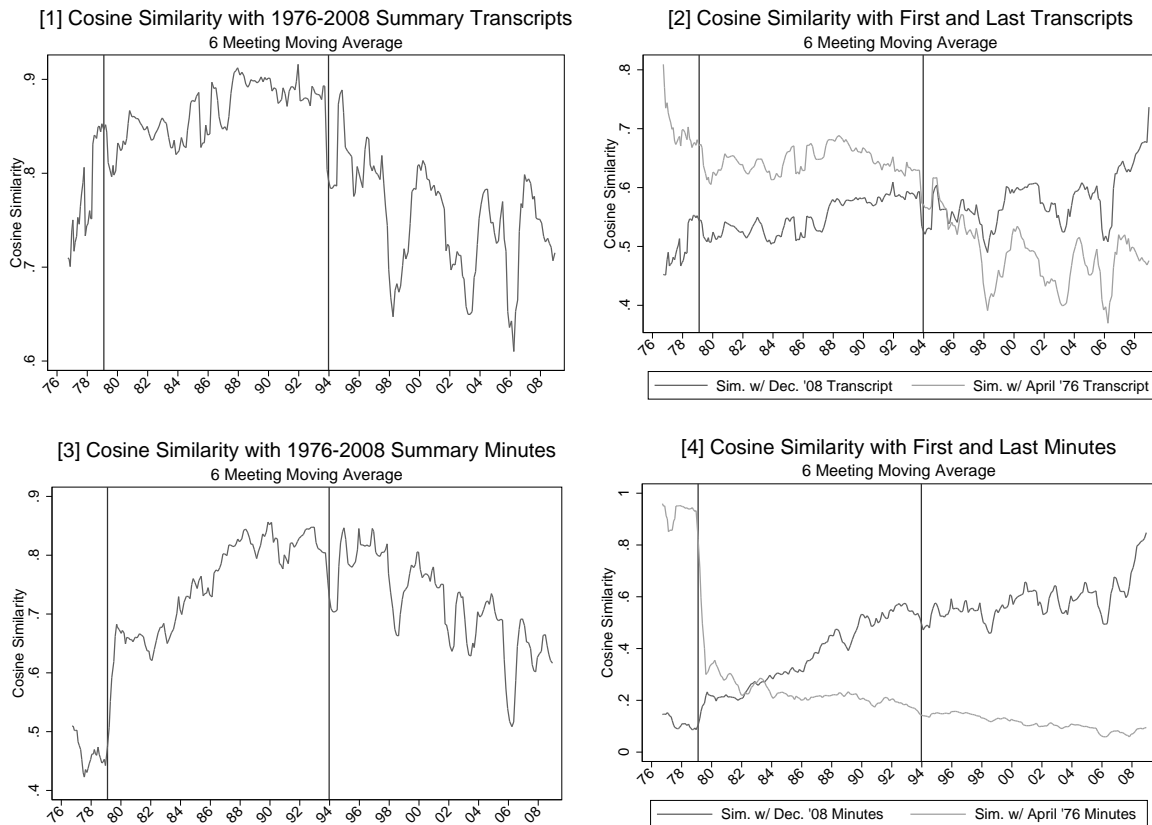


Figure 9: Cosine similarities of transcripts ([1] & [2]) and minutes ([3] & [4]) with summary documents ([1] & [3]) and first/ last documents ([2] & [4]).

Hawkins act, notice that these comparisons behave almost exactly as was predicted in Section 5.2. In both panels [2] and [4], which plot the similarity of the documents with the first and last documents of the sample period, the lines are downward sloping with respect to the reference document; this was the prediction made about comparing documents pairwise to a single document, namely, that the similarities would decrease as documents became more distant in time from the reference document (first or last). Comparisons against the summary documents, shown in panels [1] and [3], also exhibit the predicted upside-down parabola shape.

Despite the general fit of the graphs to the predictions, there are a few notable deviations, the first of which is related to the Humphrey-Hawkins change in minute-transcript similarity in 1979. Specifically, there is evidence of a drastic change in the content of the minutes, while the transcripts stay relatively stable for at least the next decade, and generally follow the predicted hump

shape. Again, there was no change in leadership (Chairman Miller) or secretary (Altmann) at this time, negating the fear that this large change in minute content was caused by a change in either leadership or in who was preparing the minutes. There was also no official policy made concerning the minutes, or at least not one that was made public. In the absence of any other information, it seems reasonable to conclude that this difference in minute content is the result of the "Humphrey-Hawkins process," as it was called by Chairman Miller. Taken in conjunction with the lack of any noticeable difference in the content of the transcripts, it seems that the effect of the law was not to cause the discussion to change within the Board Room, but, rather, to change the way that the FOMC reported its proceedings to the public. Recall that, at the time, the minutes (ROPA) were thought to be the only records of FOMC proceedings, which helps ensure that the transcript content was not purposefully manipulated in response to Act.

That the minutes changed is interesting, but what about the Act would cause such a response? The portion of the Humphrey-Hawkins Act that was relevant to the Federal Reserve was essentially an amendment of Section 2A of the 1913 Federal Reserve Act, which had recently been amended by the Federal Reserve Reform Act of 1977 (FRRA). The FRRA was the act that legally instituted the well-known dual mandate²⁴ and semiannual reports to congressional committees (Congress, 1977). The Humphrey-Hawkins Act emphasized certain points that were to be reported to Congress semiannually (2) and added a new "objective" to be met by the Fed; it states that the Fed should report:

(2) the objectives and plans of the [FOMC] with respect to the ranges of growth or diminution of the monetary and credit aggregates for the calendar year during which the report is transmitted, *taking account of past and prospective developments in employment, unemployment, production, investment, real income, productivity, international trade and payments, and prices*; and (3) the relationship of the aforesaid objectives and plans to the *short-term goals set forth in the most recent Economic Report of the President . . . and to any short-term goals approved by the Congress*.
(Congress, 1978) (*emphasis added*)

While these points are rarely cited when discussing late-70's Federal Reserve legislation, they are indicative of the atmosphere in which the FOMC was operating; essentially, there was pressure from Congress for the Fed to increase its transparency and more fully explain its actions. This, paired with Congress' desire to emphasize and be informed of *short-term* goals are likely explanations of the Fed's decision to increase transparency of its meetings. That no change in the transcripts was seen indicates that the Fed did not change what it said in its meetings, only what it reported, in this case through the minutes—exactly what Congress wanted.

²⁴Which was initially a tri-mandate—unemployment, inflation and moderate interest rates.

Secretary	Dates in Office	Number of Meetings	Chair
Broida	07/73 – 07/78	28 [†]	Burns, Miller
Altmann	08/78 – 02/83	41	Miller, Volcker
Axilrod	03/83 – 04/86	25	Volcker
Bernard*	05/86 – 02/87	7*	Volcker
Kohn	03/87 – 06/02	123	Volcker, Greenspan
Reinhart	08/02 – 06/07	40	Greenspan, Bernanke
Madigan	08/07 – 06/10	12 [†]	Bernanke

*Bernard was the assistant secretary acting the interim.

[†]Number of in-sample meetings.

Table 6: FOMC Secretaries and Acting Secretaries Since 1973

In order to make better conclusions about the nature of the effect of the Humphrey-Hawkins Act, regression (2) in Table 7 estimates the coefficients of the following model

$$\begin{aligned}
 \text{MSMsim}_t = & \alpha + \theta_1 \text{Post93}_t + \theta_2 \text{PreHH}_t + \theta_3 t + \theta_4 t^2 + \beta_1 \text{GB Inflation}_t + \beta_2 \text{GB Inflation}_{t+1} \\
 & + \beta_3 \text{GB GDP}_t + \beta_4 \text{GB GDP}_{t+1} + \beta_5 \text{GB Unemp}_t + \beta_6 \text{GB Unemp}_{t+1} \\
 & + \sum_{C=\text{Miller}}^{\text{Bernanke}} \gamma_C C_t + \sum_{S=\text{Broida}}^{\text{Madigan}} \varphi_S S_t + u_t \quad (3)
 \end{aligned}$$

where the variables are the same as in Equation (2), with the addition of t^2 to control for the "hump-shape" present in the minute-summary minute comparisons. Again, the similarity measures are scaled by 100; for example, in (3), a 1% increase in inflation in $t + 1$ would increase the cosine similarity by 2.9 (or 0.029 before scaling). Regressions (1) and (2) study minute similarity over time with MSMsim as the dependent variable, while (3) and (4) use TSTsim for studying transcript changes. Before discussing the Humphrey-Hawkins effects, however, there are some interesting results to be seen.

First, the only significant chairman dummy variable (Bernanke) is in the transcript regression, while the only significant secretary dummy variable (Altmann) is in the minutes regressions. This is important because it suggests that, as expected, the Chairman has a role in the discussion in the Board Room, and the secretary does in fact have some impact on the content of the minutes. This is not surprising because the preparation of the minutes is the job of the secretary and the Monetary Affairs division of the Fed.²⁵ Nevertheless, it is reassuring to see that this type of relationship was picked up using LSA and this model specification. Finally, the coefficients on the chairmen are large, suggesting a significant role of the chairman in determining Board Room discussion.

²⁵I'd like to thank Dr. Ellen Meade for this information.

Table 7: FOMC Minute and Transcript Changes in Content after 1993

	(1)	(2)	(3)	(4)
	MSMsim	MSMsim	TSTsim	TSTsim
Post93	-0.0709	5.342	-5.963*	-1.510
Pre Hum-Haw.	-17.80***	-23.48***	2.991	5.318
Inflation _t	-1.203	-0.687	-0.377	-0.844
Inflation _{t+1}	0.749	0.770	2.913*	1.914
GDP _t	-0.490	-0.563	0.0861	0.133
GDP _{t+1}	0.679	0.777	1.217	2.468**
Unemp _t	1.152	2.278	0.488	-0.899
Unemp _{t+1}	-3.213	-3.397	-0.229	-0.168
Mtg. # (t)	0.385***	0.330**	0.307***	0.509***
t ²	-0.00127***	-0.000653	-0.00128***	-0.00150***
MILLER		0.593		6.655
VOLCKER		1.560		12.10
GREENSPAN		5.360		13.29
BERNANKE		13.12		25.41*
BROIDA		-11.48		1.611
ALTMANN		-12.39*		5.988
AXILROD		-5.596		0.369
KOHN		3.019		4.301
REINHART		3.570		11.81
MADIGAN		-4.457		16.31
Constant	65.09***	45.11***	57.60***	25.42*
Observations	235	235	235	235
R ²	0.691	0.718	0.363	0.421

Standard errors in brackets

The macro variables are Greenbook forecasts for the current meeting's quarter (t) and one year ahead (t+1).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, regressions (1) and (2) show that the minutes changed as a result of Humphrey-Hawkins somewhere on the order of $\theta_2 \approx 20$ cosine similarity points; this is a large jump—it represents a 15° decrease in the angle between the minutes and summary minutes (see Table 4). On the other hand, there is no statistically significant change in the transcripts. Taken together, the regression results give more evidence that the effect of Humphrey-Hawkins was to increase transparency through better reporting via the minutes.

5.3.3 Opening the Doors: 1993-2008

The second discontinuity in the post-MOD era occurs in 1993. The policy change in 1993 has been taken advantage of as a natural experiment in many studies of both the FOMC, and deliberative policy-making in general. In this year, the FOMC made public, in response to congressional pressure, that tape recording had been kept of its meetings since 1976. Whether or not the Fed *was made aware* of this fact in 1993 or earlier is still a topic of debate,²⁶ but by the end of 1993, the FOMC began releasing the 1976-1988 transcripts, and by 1995 they had agreed to start releasing transcripts from that point forward, with a five year lag. The congressional inquiries that started in 1993 spurred much debate in the FOMC over the coming years regarding minutes and transcripts, and the entire episode led to another period in document similarity—this time in the upwards direction. This again suggests that a call for transparency had the desired effect; namely, minutes more closely reflected the transcripts. The coefficient θ_1 from Equation (2) (estimated in Table 5) confirms this increase.

Again, drawing inferences from minute-transcript similarity alone does not give the full story. Panels [1] and [3] in Figure 9 suggest that the change this time was in the transcripts, not the minutes. While both the content of the minutes and transcripts began to fluctuate more significantly after 1993, the minutes essentially continued along the predicted hump shape, while the transcripts experienced a relatively sharp drop in similarity from the summary transcript. The simple regressions, again in Table 7, helps to verify this difference. The coefficient of interest θ_1 in both regressions is the one associated with the dummy variable `Post93` which indicates if an observation is before or after the November 1993 meeting, when it was decided that historical transcripts would start to be made public. This date is used not only because old transcripts were released, but because it was a time when significant attention was paid to the potential publication of the transcripts—presumably, this was a time when changes in behavior would start to be seen, if they were to be seen at all. While it took until 1995 for the decision to be made about current transcripts, the November 1993 decision suggested that a new path was being paved towards greater meeting transparency—a path which was decided in 1995. The coefficient θ_1 on the `Post93` dummy in the transcript regression

²⁶Auerbach (2009) suggests that some of the members had known for years before admitting to it.

A2 THE WALL STREET JOURNAL THURSDAY, NOVEMBER 18, 1993

<p>ANNUAL REPORTS</p> <p>WHAT SOME PEOPLE DO ANNUALLY... WE DO EVERYDAY... TO FIND OUT WHY WE'RE CHOSEN TO PRODUCE MORE ANNUAL REPORTS THAN ANY DESIGN FIRM IN AMERICA.</p> <p>CALL SCOTT GREENBERG AT: (516) 435-0400</p> <p>CURRAN & CONNORS</p>	<p>ECONOMY</p> <p>Housing Starts Climbed 2.7% During October</p> <p>Post-Flooding Construction In Midwest Puts Activity</p>	<p>Federal Reserve to Release Transcripts Of Past Sessions After Five-Year Delay</p> <p>By DAVID WESSEL Staff Reporter of THE WALL STREET JOURNAL</p> <p>WASHINGTON—The Federal Reserve, under pressure from Congress, said it will release edited transcripts of past monetary-policy discussions after a five-year delay, but hasn't decided whether to do the same for future meetings.</p> <p>of the transcripts as well as the public release of edited transcripts older than three years.</p> <p>In a letter sent yesterday to the House Banking Committee, Mr. Greenspan said Fed board members and regional-bank presidents who constitute the Open Market Committee decided at their meeting on</p>
---	---	--

Figure 10: The news of the November 1993 decision.

(3) is both higher (in absolute terms) than the θ_1 in the minute regression and is more statistically significant; in fact, the 1993 change appears to have had no statistically significant effect on the content of the minutes by this measure (θ_1 in (1) and (2)). Thus, the effect of the 1993 call for transparency was to decrease the similarity of the transcripts with the summary document; in other words, there was a change in content. This change of about 6 cosine similarity points is smaller than the 1979 minutes-change, but still notable. While this shift increased the level of transparency used in this analysis (MT_{sim}), a serious worry is that instead of the members of the FOMC being more open about the content of their discussions, they opted to alter what they said in meetings to be closer to the carefully worded minutes; this would be consistent with a changing transcript and unchanged minutes. For example, the theory in Meade & Stasavage (2008) predicts that under increased transparency, members would have the incentive to offer fewer dissenting views. Alternatively, they may feel compelled to prepare statements in advance in order to avoid misspeaking, because any error or less-than-intelligent remark would be reported to the public. Thus, the question becomes whether the fact that board members knew that their individual words and opinions would be published would dampen what they were willing to say in meetings. It is said that, before 1976 and after 1993, members were far more likely to essentially read off their prepared remarks. Section 6 attempts to answer this question.

6 The Effect of Public Transcripts: Dampening FOMC Discussion

The dynamics of FOMC meetings, a topic which has received much attention in the literature, have typically been studied by examining voting patterns and policy preferences of FOMC members; both typically use voting data itself or code preferences into an interest rate figure (e.g. Meade & Sheets (2002); Chappell *et al.* (2005); Meade (2005); Meade & Stasavage (2008)). Another approach to understanding the discussion dynamics of FOMC meetings is to use the text of the transcripts themselves, not as single entities, but, rather, as accounts of discussion that can be bro-

ken apart by member. Consider, for example, the August 19, 1986 meeting which had 12 voting members present. Breaking the transcript up into 12 *member documents*, each of which holds the content of one member's utterances at that meeting, then allows for a comparison of this *member document* to the words of all other members at the meeting. In other words, each comparison is between what member i said and what everyone else said. Section 6.1 formally defines the similarity measure, but to get an idea of the output, see Table 8 for the results of the August 1986 meeting. Of

Member	Similarity to Other Members	Member	Similarity to Other Members
Boehne	0.87	Rice	0.53
Boykin	0.58	Schultz	0.88
Corrigan	0.92	Solomon	0.92
Gramley	0.94	Teeters	0.83
Keehn	0.46	Volcker	0.97
Partee	0.89	Wallich	0.79

Table 8: Measuring the extent to which FOMC members "agreed" with other FOMC members at the August 19, 1986 meeting. Figures closer to 1 represent more agreement. The measure is formally defined in Section 6.1.

course, these numbers are relatively high, given that a cosine similarity of 1 signifies identical documents. However, these figures have a sufficient range and variance in their distribution—this allows for a study of their behavior over time. In this context, the figures (similarities) show the extent to which FOMC members "agreed" with other FOMC members at each meeting. With this, a measure of the amount of debate/deliberation, or differences in opinion and discussion topics present in the Board Room, can be achieved. Specifically, to measure deliberation, the standard deviation of these measures is taken for each meeting; this gives a measure of the average *willingness to deviate* from the average level of agreement. To foreshadow the results, given in Section 6.2, this measure fell after 1993, which implies that deliberation decreased in the Board Room. Perhaps more significant is that this drop came primarily from a decrease in willingness to offer *disagreeing* views, where disagreement is defined as below-mean similarity.

6.1 Measuring Debate in the Board Room

This section focuses solely on the transcripts of FOMC meetings after 1976; that is, after the full account of the FOMC meeting released to the public was a true transcript, not the more heavily edited MOD. For each meeting, the transcripts are then separated by voting member. For example, if the transcript, represented by a vector \mathbf{t} , were

CHAIRMAN X: Good morning everybody. The market is up.

GOVERNOR Y: I beg to differ. Things aren't going well.

then it would be separated into two vectors, \mathbf{x} and \mathbf{y}

CHAIRMAN X: Good morning everybody. The market is up.

GOVERNOR Y: I beg to differ. Things aren't going well.

respectively. So $\mathbf{t} = \mathbf{x} + \mathbf{y}$. In order to measure the degree of debate or disagreement in this meeting, the cosine similarity between \mathbf{x} and \mathbf{y} could be taken; this would say how similar the two documents were, and an inference could be made about the extent to which the two members agreed. More generally, if a transcript \mathbf{T} is given by

$$\mathbf{T} = \mathbf{m}_1 + \mathbf{m}_2 + \cdots + \mathbf{m}_n$$

where each \mathbf{m}_i are the words of FOMC voting member i , then for each member i , a new document $\mathbf{m}_{(-i)}$ can be created such that

$$\mathbf{m}_{(-i)} = \mathbf{T} - \mathbf{m}_i$$

so that the similarity $\text{sim}(\mathbf{m}_i, \mathbf{m}_{(-i)})$ can be computed in order to measure the extent to which that member was in agreement with the others present at the meeting—this is a comparison of what member i said (\mathbf{m}_i) with what everybody else said ($\mathbf{m}_{(-i)}$). The only assumption that underlies this methodology is that FOMC meetings consist of dialog and discussion; for if each person came to the meeting solely to "say their piece," and that "piece" was unrelated to every other member's "piece," then these similarity measures would also be low. If, however, each member is essentially discussing the same topics as the rest of the members with varying degrees of agreement, then this measure of debate or disagreement is satisfactory, as any differences would come from the "varying degrees of agreement."

The document set, then, consists of all the member documents \mathbf{m}_i and all of the excluded member documents $\mathbf{m}_{(-i)}$ for all meetings held in Washington D.C. and all voting members. For the 278 meetings, this results in 6152 documents total; half (3076) of which are member documents \mathbf{m}_i . The average number of voting members per meeting is 11.1. This average is less than twelve for two reasons: absenteeism and silence. While there are twelve voting positions on the FOMC, there were often meetings where either a spot had not been filled yet, or the person was simply absent. There are also a couple of cases in which a member did not speak for the entire meeting:

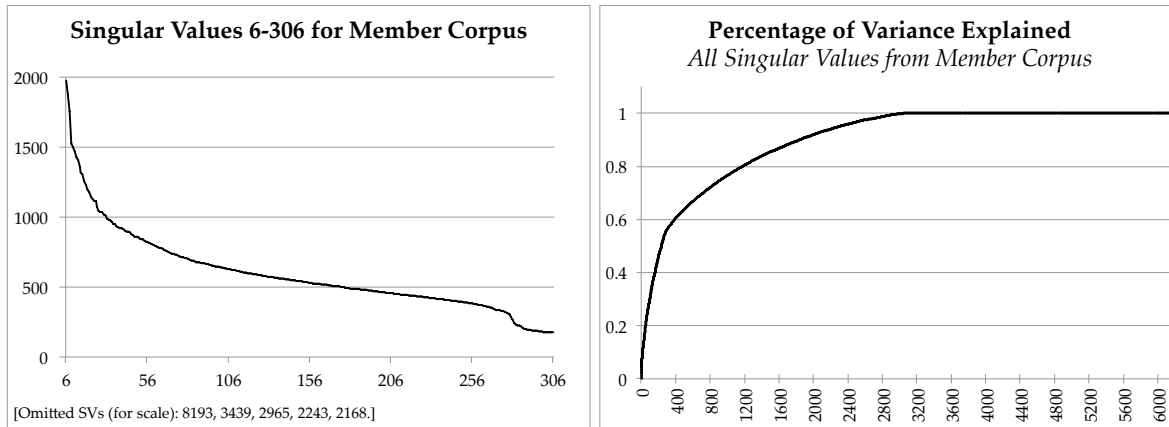


Figure 11: The left hand side shows the 6th through 306th singular values (the elements $\sigma_i \in \Sigma$ from the SVD) from the member corpus. The right hand side graph show percentage of the variance explained by all 6152 singular values for the member corpus.

during the August and September meetings in 1978, the record indicates that Governor David Lilly was silent.²⁷

Choosing LSA/ SVD Dimensionality A brief examination of the singular values (SV) $\sigma_i \in \Sigma$ from the SVD of the member corpus again lends credence to the LSA methodology, and is useful in determining the number of dimensions to retain. Figure 11 visually portrays the singular values; the left hand side graph focuses on the first 300. Similarly to the findings in Section 5.2, there is noticeable drop in SV sizes at around 278; this again confirms that the documents are all accounts of the same 278 meetings. Thus, in order to eliminate the noise created by separating transcripts into \mathbf{m}_i and $\mathbf{m}_{(-i)}$, it is necessary to keep no more than 278 factors; at this point, 55% of variance in the data has explained. In order to enter the latent semantic space, then, 200 factors are retained. The right hand side graph confirms the decreasing explanatory power of the factors after the "kink" at 278. The choice of 200 is in line with the "few hundred" factors that are typically retained; further, the singular values have largely stabilized by this point, and 46% of the variance has been explained. Notice that after 3076, the singular values have reached zero, as evidenced by the right graph in Figure 11; this happens because the rank of the original *term x document* matrix is 3076, which is half of the total documents. To see why this is true, notice that any member document \mathbf{m}_i can be written as $\mathbf{m}_i = \mathbf{m}_j + \mathbf{m}_{(-j)} - \mathbf{m}_{(-i)}$ for some other member document \mathbf{m}_j in the same meeting; but this violates the definition of linear independence and thus lowers the rank of the matrix. Eliminating all member documents \mathbf{m}_i while leaving $\mathbf{m}_{(-i)}$ would leave a full rank matrix which had half as many documents as the original; therefore, the rank of the *term x document* is half the number of documents, or 3076. To be precise, calling the SVs past 3076 "singular values"

²⁷This is the only case of this occurring in the data.

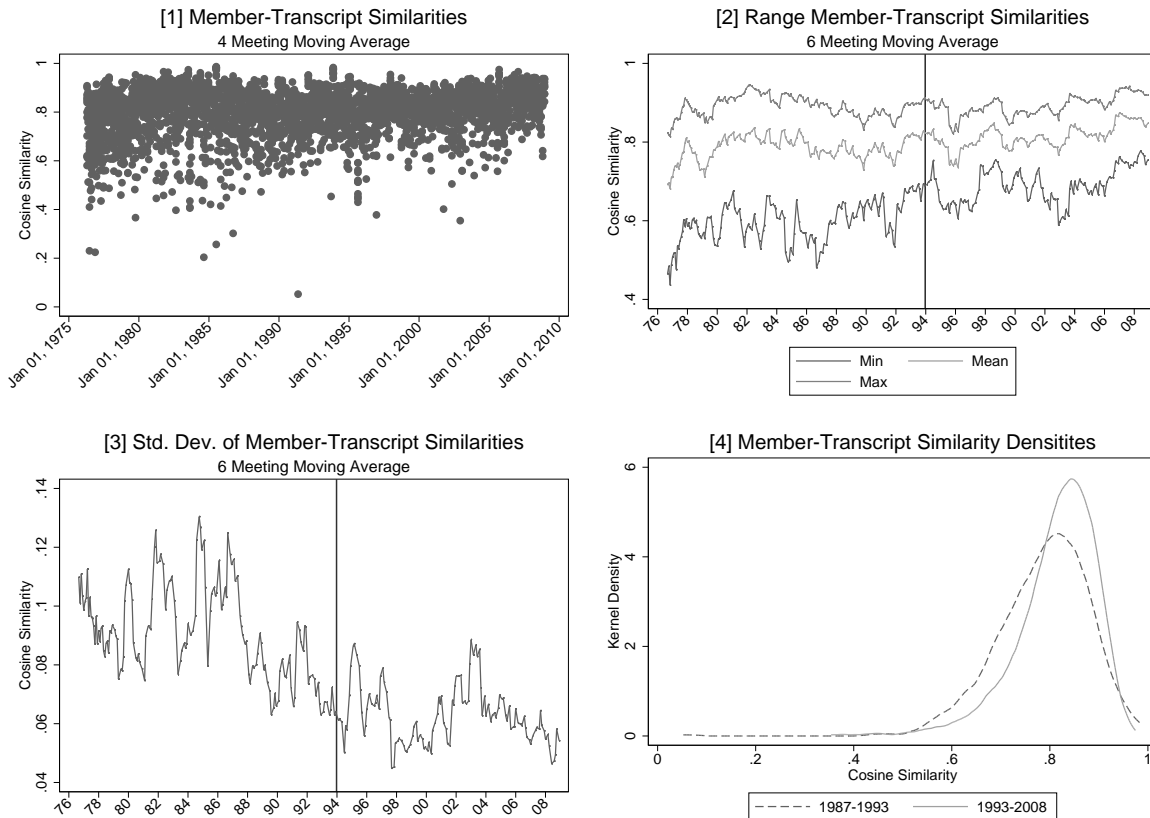


Figure 12: [1] Shows the values for MembSim , [2] plots its range and mean, [3] its standard deviation, and [4] shows its kernel density estimate during Greenspan’s term before and after November 1993.

is incorrect, because there are only as many SVs as the rank of the matrix. The use of this term serves to ingrain the concept of a singular value and rank. Interestingly, the only reason that the second graph Figure 11 has *any* values past 3076 is because of an apparent rounding error in the MATLAB[®] SVD algorithm.

6.2 Evidence from the Transcripts

The results from Section 5.3 showed that the transition from "transcriptless" (closed) meetings to recorded and published (open) meetings resulted in some change in the content of FOMC transcripts. The primary finding here is that at least some of that change is the result of a decrease in member disagreement within meetings. Figure 12 gives four visual representations of the variable MembSim_{it} , which is defined using the same notation from the previous section:

$$\text{MembSim}_{it} = \text{sim}(\mathbf{m}_{it}, \mathbf{m}_{(-i)t})$$

for each member i and time t . Before any statistical analysis of this variable is made, however, there are a few observations that will certainly help guide the thinking behind the statistical/econometric tests. Panel [1] shows all the member-level observations over time, and the second panel ([2]) plots a six meeting moving average of the range and mean of MembSim_{it} of each meeting t . Looking at these in conjunction, an eyeball test suggests that there was a narrowing of the distribution as time progressed; in particular, *it appears that there were fewer members who expressed views which conflicted with the rest of the committee*. Statements that were in disagreement would presumably be those below the mean level of agreement, and the increasing minimum shown in Panel [2] confirms that there have been fewer disagreeing views expressed over time.

Panel [3] plots the standard deviation of member-disagreement levels for each meeting across time (again, a moving average); this is perhaps the best summary measurement of agreement or disagreement at each meeting, as it, tautologically, gives a measurement for the variance in how much members agreed at each meeting. In other words, it gives a measure of how far members were willing to stray from the mean level of agreement. The standard deviation is also helpful to consider at this stage given that changes in the mean level of agreement/disagreement are not obvious in the graphs. The standard deviation changes, however, are notable and easy to pinpoint using the graphs. First there is a large drop after Chairman Volcker left the Fed, consistent with the commonly held view that "Alan Greenspan's quiet authority was rarely challenged during his 18-year rule. His predecessor, Paul Volcker, clashed with governors appointed by President Ronald Reagan (Hilsenrath, 2013)". A similar, but smaller, drop in standard deviation and increase in mean member-disagreement is also seen when Chairman Bernanke came into office, although this is not given much attention because of a lack of data (only 2 in-sample years). Another interesting, albeit not as striking, change in the data, which is more relevant to the changes which took place in the early to mid-90's, is the drop in standard deviation after 1993. This suggests that once members were aware that their words might—and, eventually, would—be made public, the discussion within the Board Room became significantly less varied. This decrease in variance is indicative of the change predicted by theory when meetings are made essentially open to the public (see Meade & Stasavage (2008) and Woolley & Gardner (2009) for theory and other results); for example, if someone's words are consistently out of line or contradictory with the FOMC's general discussion, that member might worry that their reputation with the public will suffer—they would be the odd-member out. Finally, panel [4] presents two kernel density estimates of this data under Chairman Greenspan's leadership—one before 1993 and one after. While perhaps not 100% convincing, it appears that the pre-1993 group had a lower mean, slightly wider distribution, and much longer left-tail. In summary, the hypothesis that the making-public of transcripts caused the degree of disagreement (debate) to decrease is at least worth studying further, certainly with more statistical power than an eyeball test.

6.2.1 Agreement and Debate at the Meeting Level

The first set of tests are some simple t-tests to see whether the observations noted in the previous paragraph concerning panels [2] and [3] in Figure 12 hold; namely that within Greenspan's time as Chairman,²⁸ the effect of the 1993 reform was a lower dispersion of views expressed (lower standard deviation) and higher levels of overall agreement (higher mean). To test the first hypothesis, let SD_{sim_t} be the standard deviation of $MembSim_{it}$ for all members i at meeting t . Then the null hypothesis is that the mean of SD_{sim_t} before 1993 (51 meetings) equals the post-1993 mean (98 meetings). A t-test reveals a significant²⁹ difference of 1 cosine similarity point—the mean of the first period is 6.5 and the second 7.5; the null hypothesis is rejected. Thus, while 1 is not a huge mean difference, it represents a 13%³⁰ decrease in the average SD_{sim_t} between the two periods. Alternatively, an F-test could be used to test the hypothesis that the variance in $MembSim_{it}$ is the same in the pre and post-1993 periods of Greenspan. Again, the null hypothesis is rejected and finds that the variance was higher in the pre-1993 period.³¹ Testing the second hypothesis would require a similar t-test on the variable $MeanSim_t$, which is the mean of $MembSim_{it}$ for all members i at meeting t . However, as the eyeball test predicts, there is no statistically significant difference between the means, although the mean is slightly higher in the second period (80.3) than the first (78.5).

In order to better interpret the changes in the variables $MeanSim$ and SD_{sim} over time, the following sections study the results of OLS estimated models of the three variables, the results of which are given in Table 9. To summarize the results first, each model shows a significant ($p < .05$) increase in the mean member-transcript similarity for each meeting and decrease in the standard deviation after the November 1993 meeting, the meeting which signified the beginning of a short road to open-meetings.

MeanSim Recall that the mean-similarity at each meeting is a summary statistic for the general level of agreement at the meeting. In order to find the "drivers" of agreement at each meeting t , the

²⁸Using only observations in Chairman Greenspan's term helps to control for leadership effects, which seem important given the large difference seen between Volcker and Greenspan's Fed.

²⁹Significant at the 1% level— $t = 2.25$.

³⁰ $.13 = \frac{1}{7.5}$

³¹The F-Stat is 1.29 with degrees of freedom 576 and 1064 for the two periods. This is significant at the 0.1% level. Ehrmann & Fratzscher (2007) use a similar variable (standard deviations) to measure dispersion—theirs is a standard deviation of the content of statements, where the content was coded to be either -1 , 0 or 1 .

following model is estimated:

$$\begin{aligned} \text{MeanSim}_t = & \alpha + \theta_1 \text{Post93}_t + \theta_2 \text{PreHH}_t + \theta_3 t + \beta_1 \text{GB Inflation}_t \\ & + \beta_2 \text{GB GDP}_t + \beta_3 \text{GB Unemp}_t + \sum_{C=\text{Miller}}^{\text{Bernanke}} \gamma_C C_t + u_t \end{aligned} \quad (4)$$

where PreHH and Post93 are dummy variables for pre-1979 and post-1993, the macroeconomic variables GBx are Greenbook forecasts for the current quarter, t is included as a trend variable and each C is an indicator variable for the sitting chairman. The results of the OLS regression are given in column (4) of Table 9, and column (3) estimates the same model without the trend variable. The trend variable removes all statistical significance in the estimates, and adds little in the way of explanatory power (very small contribution to R^2), so the rest of the discussion focuses on regression (3). The relatively high constant $\alpha = 70$ suggests that, in general, FOMC members are discussing the same topics and have a relatively high level of agreement of the issues. Of course this is not a perfect measure, but the belief is that applying LSA to the documents would help distinguish, in a meaningful way, documents that discuss the *same* topics but disagree about them. So to the extent that LSA captures meaning and can make such distinctions, this variable is informative. It is important to see, however, that the magnitude of the coefficient on the post-1993 variable, 2.6, is small relative to the mean of MeanSim which has a mean close to 80. Unfortunately, it's difficult to say how a difference of 2.6 in cosine similarity translates into meaning, especially compared to the relative high cosine similarity of 80. Nevertheless, it is possible that because FOMC members generally discuss the same topics—broadly speaking, of course, but the expectation is that they discuss issues related to the economy, which makes up a relatively small subspace in the space of all discussable objects—it is difficult for the LSA algorithm to distinguish more than this small difference.

The rest of the coefficients lack statistical power, including the macro variables. The significant coefficient of 11.4 on the Bernanke indicator may be indicative of the more collegial style that he was said to have brought to the Board Room.

SDsim Regressions (5) and (6) study the SDsim variable using the same independent variables as the model given in equation (4), and also find significantly ($p < .05$) negative results from the move towards open meetings in 1993, again only when considering the regression (5) with no trend-effects. The coefficient on the post-93 variable in (5) is 1.4 cosine similarity points, which, given the small mean of that variable (always less than 15, often less than 10) suggests a bigger change than was apparent in the the MeanSim regressions. Whereas the mean-similarity measured the average level of agreement/disagreement, the standard deviation measures, by definition, how far

Table 9: Measuring Debate at FOMC Meetings

	(1)	(2)	(3)	(4)	(5)	(6)
	MembSim	MembSim	MeanSim	MeanSim	SDsim	SDsim
Post 1993	2.556*** [0.625]	-0.273 [0.907]	2.594* [1.159]	-0.0472 [1.672]	-1.305* [0.615]	-1.100 [0.897]
Pre Hum-Haw.	-0.216 [1.699]	0.692 [1.707]	-0.0141 [3.037]	0.825 [3.037]	1.451 [1.613]	1.386 [1.629]
Inflation _t	0.162 [0.148]	0.497** [0.167]	0.147 [0.273]	0.456 [0.306]	-0.0724 [0.145]	-0.0965 [0.164]
GDP _t	-0.108 [0.109]	-0.0848 [0.109]	-0.0980 [0.203]	-0.0778 [0.202]	-0.0359 [0.108]	-0.0375 [0.108]
Unemp _t	0.576** [0.216]	0.655** [0.216]	0.564 [0.403]	0.646 [0.402]	-0.152 [0.214]	-0.158 [0.215]
MILLER	0.881 [1.165]	-0.301 [1.193]	1.088 [2.148]	-0.00168 [2.188]	-1.166 [1.140]	-1.081 [1.174]
VOLCKER	4.354* [1.893]	2.553 [1.933]	4.752 [3.421]	3.062 [3.481]	1.671 [1.816]	1.803 [1.867]
GREENSPAN	4.545* [2.004]	0.809 [2.179]	4.613 [3.629]	1.148 [3.937]	-1.635 [1.927]	-1.365 [2.112]
BERNANKE	11.24*** [2.470]	4.602 [2.907]	11.38* [4.462]	5.243 [5.249]	-2.433 [2.369]	-1.955 [2.816]
IsChairman	2.336*** [0.688]	2.315*** [0.686]				
IsBankPres	-1.391*** [0.417]	-1.436*** [0.416]				
Mtg. # (t)		0.0477*** [0.0111]		0.0443* [0.0204]		-0.00345 [0.0109]
Constant	70.67*** [2.827]	12.96 [13.74]	70.09*** [5.154]	16.50 [25.17]	10.34*** [2.737]	14.52 [13.50]
Observations	2612	2612	235	235	235	235
R ²	0.061	0.068	0.149	0.167	0.265	0.266

Standard errors in brackets

MembSim, MeamSim and SDsim have been scaled to be between 0 and 100, as are the macro-percentages
CHAIRDUM and BANKPRES indicate whether the member is chairman or bank president (governors omitted).

The macro variables are Greenbook forecasts for the current meeting's quarter (t).

The chairmen variables indicate the sitting chairman

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

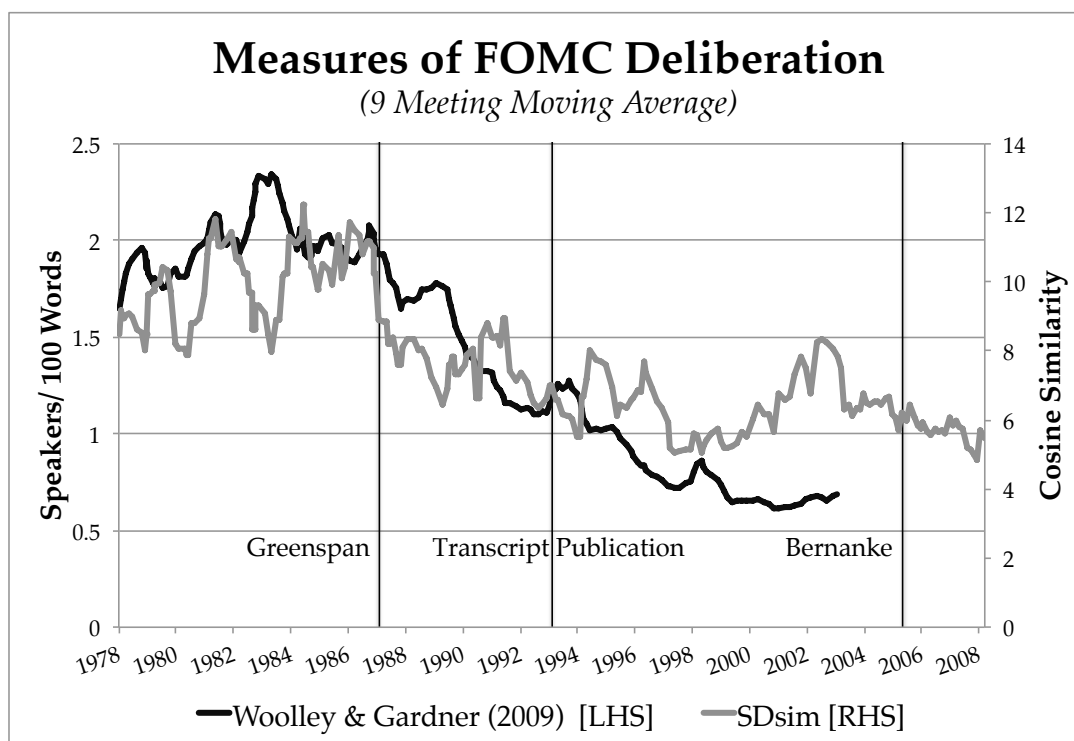


Figure 13: Comparing the SD_{sim} measure of FOMC deliberation (cf. Figure 12) with that of Woolley & Gardner (2009). The correlation coefficient $\rho = 0.79$.

away members were willing to depart from this mean level of agreement. Furthermore, that the *difference between the maximum and mean similarity stayed relatively constant over the period* while the *minimum-mean difference noticeably shrunk* suggests that the lower standard deviation is due primarily to decreased disagreement (where *disagreement* is similarity below the mean, and *agreement* above)—see panel [2] in Figure 12. Thus, the results are consistent with the theory mentioned above; namely that openness leads to a lower willingness to offer disharmonious opinions. In the case of the 1993 event, the discussion, by this measure, "dampened" by roughly 10%.³² This corroborates the figure found by Woolley & Gardner (2009); Figure 13 plots their measure of transparency with SD_{sim} —the measures have a correlation coefficient $\rho = 0.79$.

6.2.2 Member-Level Agreement and Debate

Aside from the meeting level statistics studied in regressions (3)-(6), the first two regressions in Table 9 study the member-level responses to the variables mentioned above, in addition to a few others. This is a less crucial aspect of the analysis of this section, but serves to tie the data to results

³²This is a rough ratio of $\frac{\theta_1}{\alpha}$ in regressions (5) and (6) in Table 9.

found in other studies. Specifically, for member i at meeting t ,

$$\begin{aligned} \text{MembSim}_{it} = & \alpha + \theta_1 \text{Post93}_t + \theta_2 \text{PreHH}_t + \theta_3 t + \beta_1 \text{GB Inflation}_t + \beta_2 \text{GB GDP}_t \\ & + \beta_3 \text{GB Unemp}_t + \delta_1 \text{IsChairman}_{it} \delta_2 + \delta_3 \text{IsBankPres}_{it} + \sum_{C=\text{Bernanke}}^{\text{Bernanke}} \gamma_C C_t + u_{it} \end{aligned} \quad (5)$$

where IsChairman_{it} and BankPres_{it} are dummy variables which indicate whether the member i has the title of Chairman or Reserve Bank President during the meeting t . Regression (1) estimates (5) without the trend variable t and (2) includes it. The estimates in (1) show that after 1993, members were more likely to express opinions which were similar to those of the other members. Again the relevant coefficient $\theta_1 = 2.6$ is small relative to the mean, but suggests that there was at least some change after members could feel the impending move towards transparency. Again, without having a feeling for what a "2.6 cosine similarity change" means for a discussion, especially in a room devoted to discussing a narrow set of topics, it is difficult to know how tangible this difference was. Anecdotal evidence, primarily member accounts, suggest that the change was in fact noticeable; members tended to read their remarks more often in meetings.

Next, the coefficients δ_i on the member characteristic variables (IsChairman and IsBankPres) are both in line with the literature mentioned above, which suggests that Reserve Bank presidents tend to dissent more by voting measures. Here, presidents are more likely to present differing opinions from the rest of the group; this is consistent with the general observation that Reserve Bank presidents tend to be the more outspoken members of the Board (Chappell *et al.*, 2005). Chairmen, on the other hand, are less likely to offer dissenting opinions, and Governors (omitted) are in between. That these coefficients are highly significant and of the same magnitude in both equations (with and without the time effect) suggests that these are important factors in understanding discussion dynamics.

Finally, interesting insights are offered by the chairman indicator variables, i.e. those that describe who the chairman is at the time of the meeting (notice that Burns is the omitted former chairman for this sample). From these results, it appears that the Chairman, more than any other factor, is the main change agent in the dynamics of FOMC discussion over time—at least in (1), the regression with no trend. Most notable is Bernanke's effect; at 11 cosine-similarity points, it is the largest effect seen in this section. It is again indicative of the collegial style for which he was well known. Of course such an interpretation is something of a best guess given the general descriptions of his leadership style; it could be that he was so dictatorial that everyone chose to agree with him. However, equally likely is that the discussion was dampened, narrowed or more tightly focused. Regardless, chairman effects are significant and should certainly be kept in mind when considering ways to affect FOMC transparency and discussion.

7 Conclusion

The analysis presented in this paper has made only a small dent in showing what techniques in Natural Language Processing have to offer to the study of empirical questions in Economics. Text offers an alternative perspective from which to approach problems; because text is generally prepared by a human being, the choices made in composing the text, both in content and style, reflect the preferences and mindset of the author(s). For FOMC documents, this quality offers new insights into the decision-making process of the institution that has always been surrounded by a bit of mystique.

After a thorough description of FOMC document history and Latent Semantic Analysis, the main results of the paper were given in Sections 5 and 6. First, the passing of the Humphrey-Hawkins Act (1978) coincided exactly with a large increase in procedural transparency, as measured by the cosine similarity between the minutes and transcripts of each meeting. The textual evidence also shows that this change was primarily driven by the minutes. Historical evidence suggests that the requirement that the Fed report *how* its objectives are related to the short-term goals of the President and Congress caused the FOMC to more accurately (transparently) report the content of its meetings, while making few adjustments to the content of those meetings. In this sense, the transparency increase was relatively uncontaminated by some of the problems that accompanied the 1993 change.

The next major finding is that the 1993-1995 move from completely private meetings to semi-public (with a lag) meetings also increased transparency; a shift which was driven primarily by changing transcripts. Of course, LSA cannot say *why* or exactly *how* this change occurred. Given the complexity of these documents, it is difficult to pinpoint which changes occurred post-1993 that explain the change in transcripts, so historical accounts again served as a guide. FOMC members expressed concern about potential "chilling" effects—as one bank president put it—of open-meetings, suggesting that *deliberation* may have been affected. Using the similarity of each member's words with those of everyone else at each meeting to get a sense of the level of debate, the evidence suggests that there was a substantial degree of "dampening" in deliberations after 1993. The deliberation case is particularly relevant today because it suggests how the FOMC responds to increased procedural transparency.

Most papers that study FOMC transcripts lament that more data is not available due to the five-year lag; the sentiment is shared here. The majority of the similarity measures computed here are relatively volatile, which makes it difficult to see trends until a substantial period of time has elapsed. While the 2008 transcripts, for instance, were included in the sample, it will likely take a few years until conclusions can be drawn using these techniques about the behavior of the Fed during the financial crisis. Nevertheless, the richness of these documents lend themselves to further analysis in the interim. □

A Appendix

A.1 PCA and SVD

The appendix shows more explicitly the connection to PCA. Before doing that, it's important first to understand what PCA does and why it's so good when much, potentially noisy, data is present. While the link between PCA and SVD is shown more rigorously, the explanation of PCA is not, for this typically takes an entire chapter/ section in a textbook. For nice derivations, see Lay's textbook (Lay, 2000) or the tutorial by Shlens (2005).

PCA takes an $m \times n$ data matrix X (i.e. X has m observations and n variables) and returns the eigenvectors and eigenvalues λ_i of the covariance matrix S_x . The eigenvectors are called the *principal components*. Each principal component is a linear combination of all the n variables. For each principal component, the size of the associated eigenvalue gives a "ranking" of the amount of variance explained by that component. To say what percentage each of the PCs i explains, then, is just a matter of taking

$$\frac{\lambda_i}{\sum_{j=1}^n \lambda_j}.$$

PCA is a method of explaining data using the fewest dimensions possible; that is, each dimension explains *as much of the variance in the data* as possible. To do this, the first PC is essentially a least-squares fit through the data. Subsequent PCs are orthogonal to each other, so that the next possible maximum amount of variance is explained by each. Just consider the two variable case, shown in Figure 14; i.e. consider \mathbb{R}^2 . Obviously only two dimensions describe the data, so the goal is to re-orient these dimensions such that the most variation in the data is explained by the first dimension. In this case, the data seem to lie along the line $y = x$, but there is some noise in the data that makes the line $y = x$ not a perfect fit. Thus, the first PC is $y = x$, and the second is orthogonal to this (in order to explain the noise), or $y = -x$. In general, the further along you move down your set of PCs, the more each PC is explaining noise in the data.

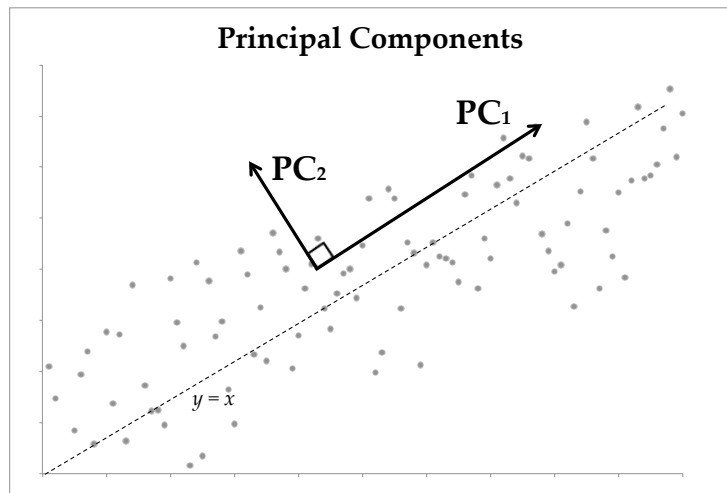


Figure 14: Some data noisily distributed around $y = x$, and its principal components. PC_1 is shifted a bit away from the line $y = x$ so that it can be seen, but in principal it should lie exactly on that line.

The power of PCA should now be clear; it allows for the creation of new variables that better

describe, with less noise, why given data behave as it does.³³ Now note that for any column de-meaned data matrix X , using the term *demeaned* as defined in section 4.2, the covariance matrix S_x is given by

$$S_x = \frac{1}{n-1} X^T X$$

Now let $Y_{(m \times n)} = \frac{1}{\sqrt{n-1}} X$ so that

$$\begin{aligned} Y^T Y &= \left(\frac{1}{\sqrt{n-1}} X \right)^T \left(\frac{1}{\sqrt{n-1}} X \right) \\ &= \frac{1}{n-1} X^T X \\ &= S_x. \end{aligned}$$

Now let the SVD of Y be given by $Y = U \Sigma V^T$. Then because the columns of V are the eigenvectors of $Y^T Y = S_x$, they are the principal components of X . *This is the connection.* With only a bit of scaling, this shows that PCA and SVD are very similar, while SVD is just a more general form.

It should now be clear why it is OK to "throw out" the eigenvectors associated with the smallest eigenvalues—it loses the least amount of data, and most efficiently reduces noise. This applies to eigenvalues in PCA and the singular-values in SVD alike. As was mentioned in Section 4.2, throwing out the smallest singular values is equivalent to setting the associated columns in V from the SVD equal to 0. So once the matrix multiplication ΣV^T , where the smallest singular-values have been set to zero, the noise reduction has been complete. At this point, however, there might be the question of how the re-multiplication of the SVD matrices (i.e. multiplying ΣV^T by U) effects the "throwing out" of dimensions, especially as related to cosine similarities. The answer is: it doesn't. To prove this, first consider an intermediary claim:

Claim For any orthogonal matrix $C_{(m \times n)}$ and vectors $v_{(n \times 1)}$ and $w_{(n \times 1)}$, $\langle Cv, Cw \rangle = \langle v, w \rangle$.

Proof. Thanks to Professor Lucianovic. Note that Cv is $m \times 1$ as is Cw . So the inner product $\langle Cv, Cw \rangle$ is the same as the matrix multiplication $(Cv)^T Cw$. Also, since C is orthogonal, $C^T = C^{-1}$. So,

$$\begin{aligned} \langle Cv, Cw \rangle &= (Cv)^T Cw \\ &= v^T C^T Cw \\ &= v^T w \\ &= \langle v, w \rangle \end{aligned}$$

³³The following explanation follows closely the one given in Shlens (2005).

□

Thus, the the claim to prove is

Claim: In the singular value decomposition of X given by $X = U\Sigma V^T$, the cosine similarity of any two vectors x_i and x_j in X equals the cosine similarity of the corresponding vectors b_i and b_j in ΣV^T .

Proof. Also thanks to Professor Lucianovic. The SVD is $X = U\Sigma V^T$, which, multiplying both sides on the left by U^{-1} , implies $\Sigma V^T = U^{-1}X = U^T X$ by the orthogonality of U . So the question now is to show that for any two columns b_i and b_j in $U^T X$ and corresponding x_i and x_j in X , $\text{sim}(b_i, b_j) = \text{sim}(x_i, x_j)$. Since each column $b_i = U^T x_i$, then by the first claim it follows that

$$\langle x_i, x_j \rangle = \langle U^T x_i, U^T x_j \rangle = \langle b_i, b_j \rangle$$

and

$$\langle x_i, x_i \rangle = \langle U^T x_i, U^T x_i \rangle = \langle b_i, b_i \rangle.$$

Thus,

$$\text{sim}(x_i, x_j) = \frac{\langle x_i, x_j \rangle}{\sqrt{\langle x_i, x_i \rangle} \sqrt{\langle x_j, x_j \rangle}} = \frac{\langle b_i, b_j \rangle}{\sqrt{\langle b_i, b_i \rangle} \sqrt{\langle b_j, b_j \rangle}} = \text{sim}(b_i, b_j)$$

as desired. □

So the final multiplication of the ΣV^T matrix by U does not violate the noise-loss properties ensured by the PCA. In summary, PCA is closely tied to the matrices Σ and V^T , so the same characteristics of PCA can be discussed when discussing the dimension reduction of matrix by SVD. Note that no dimension reduction was shown in the proofs nor in the discussion, but the claims still hold. Finally, to reiterate the point made in section 4.2, the vectors with the smallest associated singular values do the best job at explaining anomalous (noisy) term uses. For example, if in 50 documents *car* and *automobile* are used in a 1:1 ratio, but this ratio in the 51st document is 100:1, then this entry of 100 in the term-document matrix is anomalous, or "noisy." PCA guarantees that this noise will be best explained along one of the axes with a small associated eigenvalue (or singular-value in SVD). If the smallest singular values have been set to zero, then this noise is reduced and, thus, the value of the anomalous term in the term-document matrix will be adjusted so that the entry is closer to how that word is normally used in other documents. In the *car* example, the 100 entry would be set much lower.

A.2 Proofs

Cosine Similarity

Claim: If x and y are two non-zero vectors in \mathbb{R}^2 with angle θ between them, then $\langle x, y \rangle = \|x\| \|y\| \cos \theta$; i.e. $\cos(\theta) = \frac{\langle x, y \rangle}{\|x\| \|y\|}$.

Proof. There are two cases to consider:

1. First, consider the case where x and y form a triangle with third side z . The Law of Cosines says that $\|z\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\| \|y\| \cos \theta$.

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then the distance between their endpoints, i.e. z , is given by the distance formula

$$\begin{aligned} \|z\| &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\ &= \sqrt{(x_1^2 - 2x_1y_1 + y_1^2) + (x_2^2 - 2x_2y_2 + y_2^2)} \\ &= \sqrt{\langle x - y, x - y \rangle} \\ &= \sqrt{(x - y)^t(x - y)} \\ &= \sqrt{(x^t - y^t)(x - y)} \\ &= \sqrt{x^t x + y^t y - x^t y - y^t x} \\ &= \sqrt{\langle x, x \rangle + \langle y, y \rangle - 2\langle x, y \rangle} \\ &= \sqrt{\|x\|^2 + \|y\|^2 - 2\langle x, y \rangle} \end{aligned}$$

So equating these two equations for $\|z\|^2$ gives

$$\begin{aligned} \|x\|^2 + \|y\|^2 - 2\|x\| \|y\| \cos \theta &= \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle \\ \|x\| \|y\| \cos \theta &= \langle x, y \rangle \end{aligned}$$

As desired.

2. Next, consider the case where x and y are scalar multiples of one another; consider WLOG, that $y = cx$. Then $\langle x, y \rangle = cx_1^2 + cx_2^2$. Also, $\|x\| = \sqrt{x_1^2 + x_2^2}$ and $\|y\| = \|cx\| = \sqrt{c^2 x_1^2 + c^2 x_2^2} = |c| \sqrt{x_1^2 + x_2^2}$.

If $c > 0$, then $\cos \theta = 1$, so $\langle x, y \rangle = cx_1^2 + cx_2^2 = |c|x_1^2 + |c|x_2^2 = \sqrt{x_1^2 + x_2^2} \sqrt{c^2 x_1^2 + c^2 x_2^2} = \|x\| \|y\| 1 = \|x\| \|y\| \cos \theta$.

If $c < 0$, then $\cos \theta = -1$, so $\langle x, y \rangle = cx_1^2 + cx_2^2 = -|c|x_1^2 - |c|x_2^2 = -\sqrt{x_1^2 + x_2^2} \sqrt{c^2 x_1^2 + c^2 x_2^2} = -\|x\| \|y\| = \|x\| \|y\| \cos \theta$.

□

Low-Rank Approximation

Claim: For the matrix \mathbf{X}_k as defined by the SVD, $\text{rank}(\mathbf{X}_k) = k < r \text{rank}(\mathbf{X})$.

Proof. Keep in mind that for any *term x document* matrix N , the rank of N is given by the number of documents since there are typically fewer documents than terms. For X as defined, $\text{rank}(X) = n$. Because $X^T X$ is a symmetric matrix, its eigenvectors form an orthonormal basis for some subspace \mathcal{S} of \mathbb{R}^n . Thus, because V is the matrix with these eigenvectors as its columns, and because these vectors are orthogonal, they are linearly independent. Thus, by definition of rank, V has rank n . When performing the multiplication ΣV^T , the resulting matrix is just V^T with each v_i scaled by σ_i . If all but k of the σ_i 's have been set to zero in Σ (call this vector Σ_k), then $\Sigma_k V^T$ has rank k . This is because with some of the $\sigma_i = 0$, the vectors $\sigma_i v_i = 0 v_i = 0$ in V , which are no longer linearly independent from the other vectors $\sigma_j v_j$, where $\sigma_j \neq 0$. Because U has rank m (use the same argument as above), the multiplication $U \Sigma_k V^T$ results in a matrix with rank k still, for it is a fact that for any matrices A and B , $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$. Thus, $\text{rank}(X_k) = k < n = \text{rank}(X)$. \square

A.3 References

- Auerbach, Robert D. 2009. *Deception and Abuse at the Fed: Henry B. Gonzalez Battles Alan Greenspan's Bank*. University of Texas Press.
- Boukus, Ellyn, & Rosenberg, Joshua. 2006. "The information content of FOMC minutes. *Federal Reserve Bank of New York*.
- Chappell, Jr. Henry W., McGregor, Rob Roy, & Vermilyea, Todd. 2005. *Committee Decisions on Monetary Policy*. Cambridge: The MIT Press.
- Congress, US. 1977. *H.R. 9710–95th Congress: Federal Reserve Reform Act*. United States.
- Congress, US. 1978. *H.R. 50–95th Congress: Full Employment and Balanced Growth Act*. United States.
- Danker, Deborah J, & Luecke, Matthew M. 2005. Background on FOMC meeting minutes. *Fed. Res. Bull.*, **91**, 175.
- Deerwester, Scott C., Dumais, Susan T, Landauer, Thomas K., Furnas, George W., & Harshman, Richard A. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, **41**(6), 391–407.
- Dumais, Susan T, Furnas, George W, Landauer, Thomas K, Deerwester, Scott, & Harshman, Richard. 1988. Using latent semantic analysis to improve access to textual information. *Pages 281–285 of: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM.
- Eckart, Carl, & Young, Gale. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**(3), 211–218.
- Ehrmann, Michael, & Fratzscher, Marcel. 2007. Communication by central bank committee members: different strategies, same effectiveness? *Journal of Money, Credit and Banking*, **39**(2-3), 509–541.
- Farka, Mira. 2011. The asymmetric impact of "informative" and "uninformative" federal open market committee statements on asset prices. *Contemporary Economic Policy*, **29**(4), 469–493.
- Foltz, Peter W., Laham, Darrell, & Landauer, Thomas K. 1999. Automated Essay Scoring: Applications to Educational Technology. *Pages 939–944 of: World Conference on Educational Multimedia, Hypermedia & Telecommunications*, vol. 1999. EdMedia.

- Geraats, Petra M. 2000. Why Adopt Transparency? The Publication of Central Bank Forecasts. *Center for International and Development Economics Research, UC Berkeley*.
- Geraats, Petra M. 2002. Central Bank Transparency. *The economic journal*, **112**(483), F532–F565.
- Geraats, Petra M. 2007. The Mystique of Central Bank Speak. *International Journal of Central Banking*.
- Hanley, Kathleen Weiss, & Hoberg, Gerard. 2010. The information content of IPO prospectuses. *Review of Financial Studies*, **23**(7), 2821–2864.
- Hanley, Kathleen Weiss, & Hoberg, Gerard. 2012. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, **103**(2), 235–254.
- Hendry, Scott, & Madeley, Alison. 2009. "Text Mining and the Information Content of Bank of Canada Communications". *Bank of Canada Working Paper*.
- Hilsenrath, Jon. 2013. Yellen Would Bring Tougher Tone to Fed. *The Wall Street Journal*, Sept 22.
- Hoberg, Gerard, & Phillips, Gordon. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, **23**(10), 3773–3811.
- Hoberg, Gerard, & Phillips, Gordon M. 2009. *Text-Based Network Industries and Endogenous Product Differentiation*. Tech. rept. National Bureau of Economic Research.
- Landauer, Thomas K, Foltz, Peter W, & Laham, Darrell. 1998. An introduction to latent semantic analysis. *Discourse processes*, **25**(2-3), 259–284.
- Lay, David C. 2000. *Linear Algebra and its Applications*.
- Lindsey, David E. 2003. A Modern History of FOMC Communication: 1975-2002. *Board of Governors of the Federal Reserve System, Washington, DC*.
- Manning, Christopher D, Raghavan, Prabhakar, & Schütze, Hinrich. 2008. *Introduction to information retrieval*. Vol. 1. Cambridge university press Cambridge.
- Meade, Ellen. 2005. "The FOMC: Preferences, Voting, and Consensus. *Federal Reserve Bank of St. Louis Review*, **87**, 93–101.
- Meade, Ellen E, & Sheets, D Nathan. 2002. *Regional influences on US monetary policy: some implications for Europe*. Centre for Economic Performance, London School of Economics and Political Science.

- Meade, Ellen E, & Stasavage, David. 2008. Publicity of Debate and the Incentive to Dissent: Evidence from the US Federal Reserve. *The Economic Journal*, **118**(528), 695–717.
- Meade, Ellen E, & Thornton, Daniel L. 2012. The Phillips curve and US monetary policy: what the FOMC transcripts tell us. *Oxford Economic Papers*, **64**(2), 197–216.
- Meltzer, Allan H. 2010. *A History of the Federal Reserve, Volume 2*. University of Chicago Press.
- Romer, Christina D., & Romer, David H. 2008. The FOMC versus the Staff: Where Can Monetary Policymakers Add Value? *American Economic Review*, **98**(2), 230–35.
- ROPA. '76–'93. *FOMC ROPA 1976-1993*. Federal Reserve.
- Rosa, Carlo, & Verga, Giovanni. 2007. On the consistency and effectiveness of central bank communication: Evidence from the ECB. *European Journal of Political Economy*, **23**(1), 146–175.
- Schonhardt-Bailey, Cheryl. 2013. *Deliberating American Monetary Policy: A Textual Analysis*. The MIT Press.
- Shlens, Jonathon. 2005. A tutorial on principal component analysis. *Systems Neurobiology Laboratory, University of California at San Diego*, **82**.
- Transcripts. '76–'08. *FOMC Transcripts 1976-2008*. Federal Reserve.
- Woolley, John T. 1986. *Monetary politics: The Federal Reserve and the politics of monetary policy*. Cambridge University Press.
- Woolley, John T, & Gardner, Joseph C. 2009. Does Sunshine Reduce the Quality of Deliberation? The Case of the Federal Open Market Committee. *The Case of the Federal Open Market Committee*.
- Zeimpekis, D., & Gallopoulos, E. 2006. TMG: A MATLAB Toolbox for Generating Term-Document Matrices from Text Collections. *Pages 187–210 of: Kogan, Jacob, Nicholas, Charles, & Teboulle, Marc (eds), Grouping Multidimensional Data*. Springer Berlin Heidelberg.