Pontifícia Universidade Católica
do Rio de Janeiro

Lucas Zaniboni

Connecting the Dots: Assigning FOMC Members to Fed Dots Through Speech Quantification

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia .

Advisor : Prof. Carlos Viana de Carvalho
Co-advisor: Prof. Marcelo Cunha Medeiros

Rio de Janeiro
April 2019

Pontifícia Universidade Católica
do Rio de Janeiro

Lucas Zaniboni

Connecting the Dots: Assigning FOMC Members to Fed Dots Through Speech Quantification

Dissertation presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia . Approved by the undersigned Examination Committee.

Prof. Carlos Viana de Carvalho
Advisor
Departamento de Economia – PUC-Rio

Prof. Marcelo Cunha Medeiros
Co-advisor
Departamento de Economia – PUC-Rio

Prof. Leonardo Rezende
Departamento de Economia – PUC-Rio

Prof. Diogo Abry Guillen
Departamento de Economia – PUC-Rio

Rio de Janeiro, April 5th, 2019

Lucas Zaniboni

B.A. in Economics, University of Sao Paulo (USP-SP), 2016.

Bibliographic data

Zaniboni, Lucas

Connecting the Dots: Assigning FOMC Members to Fed Dots Through Speech Quantification / Lucas Zaniboni; advisor: Carlos Viana de Carvalho; co-advisor: Marcelo Cunha Medeiros. – Rio de janeiro: PUC-Rio , Departamento de Economia , 2019.

v., 74 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia .

Inclui bibliografia

1. Economia – Teses. 2. Macroeconomia – Teses. 3. Comunicação de Bancos Centrais;. 4. Política Monetária;. 5. Alocação Dirichlet Latente (LDA);. 6. Rastreamento de Alvos;. 7. *Dots* do Fed.. I. Carvalho, Carlos. II. Medeiros, Marcelo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia . IV. Título.

CDD: 620.11

To my family, for their support
and encouragement.

Acknowledgments

I thank my advisor, Prof. Carlos Viana de Carvalho, and my co-advisor, Prof. Marcelo Cunha Medeiros, for the effort allocated to this project and the extremely helpful insights.

I also thank my parents Angelo and Isabel, and my brother Gabriel, for the constant support and incentive, without which I would definitely never have gone so far.

This study was financed in part by the Coordenação de Aperfeiçoamento Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. I also thank the institutions PUC-Rio and *Vinci Partners* for the financial support which enabled me to keep continuing my studies as an academic.

Finally, I would like to thank all my friends and colleagues that took part in my life during these two years, who helped me to grow up not just as a researcher, but also as a person.

Abstract

Zaniboni, Lucas; Carvalho, Carlos (Advisor); Medeiros, Marcelo (Co-Advisor). **Connecting the Dots: Assigning FOMC Members to Fed Dots Through Speech Quantification**. Rio de Janeiro, 2019. 74p. Dissertação de Mestrado – Departamento de Economia , Pontifícia Universidade Católica do Rio de Janeiro.

As (1) points out, monetary policy predictability can enhance a Central Bank stabilization policy efficacy. In this paper we aim to reduce uncertainty about one Federal Reserve forward guidance instrument by estimating full association probabilities distributions between members and the interest rate dot plot for each FOMC meeting. Our contribution to the literature is twofold: first, we propose a general Bayesian algorithm which estimates these association hypotheses between agents and actions whenever they are not observed. Second, we elaborate a novel and less subjective technique for quantifying text into data, using *Latent Dirichlet Allocation* (LDA) and shrinkage econometric tools. This method shows some desirable features such as positive correlation between the FOMC chair and the rest of the committee, and a policy stance ordering which partially reflects analysts and market participants views on this hawk-dove spectrum. Our tracking algorithm performs successfully in a simulated environment, in a sense that it on average considers the correct member-to-dot association as the most likely one. Using real data on speeches and Fed dots, it is also able to attribute the highest probability to the correct assignment hypothesis in the only meeting it is known for sure.

Keywords

Central Bank Communication; Monetary Policy; Latent Dirichlet Allocation (LDA); Target Tracking; Fed Dots.

Resumo

Zaniboni, Lucas; Carvalho, Carlos; Medeiros, Marcelo. **Ligando os Pontos: Associando Membros do FOMC aos *Dots* do Fed Através da Quantificação de Discursos**. Rio de Janeiro, 2019. 74p. Dissertação de Mestrado – Departamento de Economia , Pontifícia Universidade Católica do Rio de Janeiro.

Como (1) aponta, a previsibilidade acerca da política monetária pode melhorar a eficácia da política de estabilização de um Banco Central. Nesse artigo, procuramos reduzir a incerteza a respeito de um instrumento de *Forward Guidance* do Banco Central norte-americano (o *Federal Reserve*) estimando distribuições de probabilidade completas sobre todas as associações possíveis entre seus membros e o *dot plot* de taxa de juros para cada reunião. Nossa contribuição para a literatura ocorre em duas frentes: primeiro, propomos um algoritmo Bayesiano geral que estima essas hipóteses de associação entre agentes e ações sempre que elas não são observadas. Além disso, elaboramos uma maneira nova e menos subjetiva para quantificar textos em dados numéricos, usando *Alocação Latente Dirichlet (LDA)* e modelos econométricos de seleção. Esse método apresenta algumas características desejáveis como uma correlação positiva entre o presidente do FOMC e o resto do comitê, e um ordenamento na postura de política monetária que reflete, ainda que parcialmente, visões de analistas de mercado a respeito desse espectro entre membros mais duros e mais lenientes com a taxa de juros. Nosso algoritmo de rastreamento de alvos também tem bom desempenho num ambiente simulado, no sentido em que, em média, considera como mais provável a verdadeira associação entre membros e *dots*. Usando dados reais de discursos individuais e *dots*, ele também consegue atribuir a maior probabilidade para a associação correta na única reunião em que ela é conhecida de fato.

Palavras-chave

Comunicação de Bancos Centrais; Política Monetária; Alocação Dirichlet Latente (LDA); Rastreamento de Alvos; *Dots* do Fed.

Table of contents

List of figures

List of tables

List of Abreviations

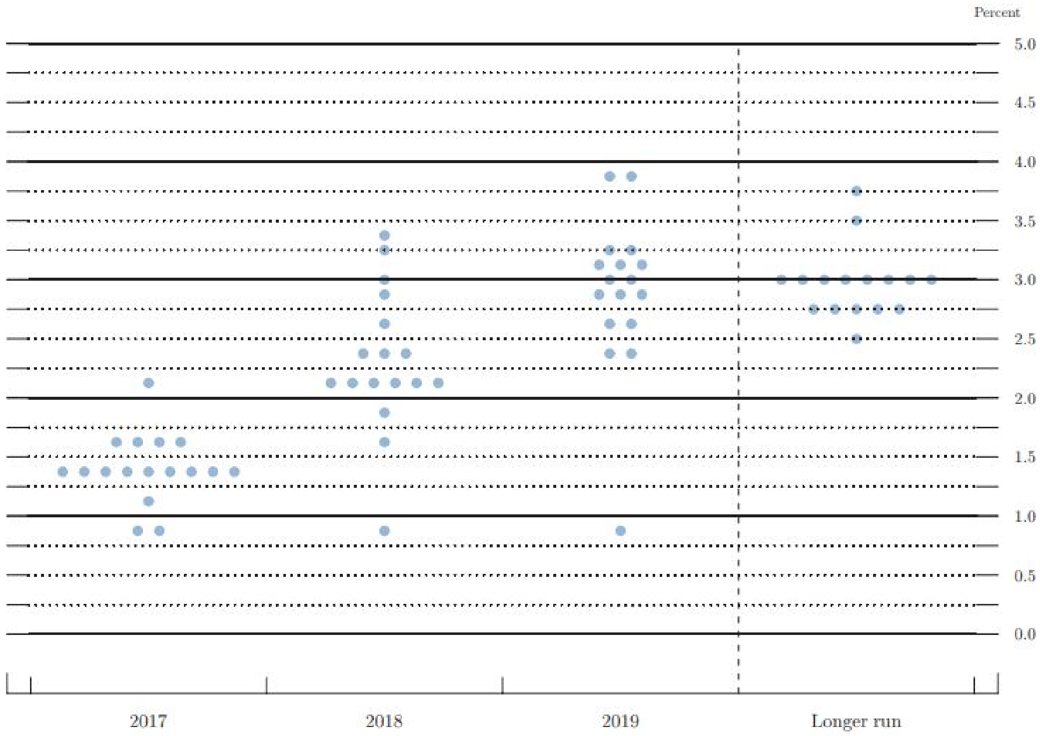
Fed – *Federal Reserve Bank*
FOMC – *Federal Open Market Committee*
SEP – *Summary of Economic Projections*
JPDAF – *Joint Probabilistic Data Association Filter*
LDA – *Latent Dirichlet Allocation*

1
Introduction

After the 2008 Great Financial Crisis, the proximity to the *Zero Lower Bound* on nominal interest rates called for a new set of monetary policy instruments, and found in *Forward Guidance* - the creation of the right kind of expectations about how monetary policy will be used *after* the economy returns to normality - one interesting alternative. It is under this context that the Fed implemented in early 2012 the *interest rate dot plot* (henceforth *dot plot*) in its *Summary of Economic Projections* (SEP). Basically, it is a set of individual and anonymous assessments of each FOMC member for the most appropriate nominal rate to be set at the current year-end, the next 2 year-ends and the long-run, according to his/her own forecasts and economic conditions evaluation. An illustration for March 2017 *dot plot* is provided in Figure 1.1.

The *dot plot* is our main object of study. This paper elaborates a new *target tracking* algorithm that estimates full probability distributions over all possible combinations between present members and current-year-end *dots* at each meeting. It also allows us to extract the probability of a given member having emitted each *dot*, as well as the probability that a certain *dot* corresponds to a determined member. The former will be called *member-to-dot* probabilities, and the later, *dot-to-member* probabilities. To achieve that, we quantify individual official speeches into a *hawk-dove* scale, also proposing a new method for turning text into data.

Figure 1.1: March 2017 meeting interest rate *dot plot*. According to the Fed, *"This chart is based on policymakers' assessments of appropriate monetary policy, which, by definition, is the future path of policy that each participant deems most likely to foster outcomes for economic activity and inflation that best satisfy his or her interpretation of the Federal Reserve's dual objectives of maximum employment and stable prices. Each shaded circle indicates the value (rounded to the nearest $\frac{1}{8}$ percentage point) of an individual participant's judgment of the midpoint of the appropriate target range for the federal funds rate or the appropriate target level for the federal funds rate at the end of the specified calendar year or over the longer run".*



Our research question is important for a number of reasons. First, the FOMC heterogeneity between committee members plays an important role in monetary policy decision, as shown in (2). The *dot plot* brings individual but *anonymous* information about one important kind of heterogeneity - the preference towards inflation fighting ("*hawks*") or growth promoting ("*doves*")[1]. Moreover, it holds this information for the whole FOMC, not only the 12 voters. If one gets to know exactly which *dots* are linked to the committee voting part, it is some extra useful information to infer about future monetary policy. This argument is also reflected when considering the turnover in regional Feds president chairs, and in the *hawk-dove* composition change it may cause on the committee.

[1]More information about the *hawk-dove* composition within the Federal reserve and its determinants can be found in (3) and (4).

Second, although voting members have equal weight on their votes towards monetary policy implementation, market agents give more importance to what some of them say. For instance, one of (5) main findings is that speeches emitted by the Central Bank president or its vice-president have significant impact on market rates moves, while in general speeches from other members fail to do so. This may happen for several reasons, from the theoretical solid background which puts that person in a more prominent position, to the institutional role defined by the Central Bank - while the Fed follows a *collegial* approach (6), with relative freedom of speech for regular members to express their own views about the conducting of monetary policy, the FOMC chair is more stuck to an *institutional* role, taken to represent the consensual committee view more than any other member. If different members' opinions have discrepant impacts on agents' views, linking *dots* to members may add some relevant information about this issue.

Third, if a solid relation between the *dots* scattering and individual speeches is discovered, since all speeches are publicly available prior to any meeting, our tracking algorithm should be capable of generating good forecasts about the full cross-sectional *dots* distribution before they are published. Monetary policy predictability and its benefit to welfare have been widely studied and documented (1). (7) argue that, since the economy is affected by expectations as much as it is by any decision on short-term rates, an effort in turning policy decisions more predictable could enhance the effectiveness of monetary policy. (6) follows the same logic, deliberating on a simple theoretical framework in which the total effect of any central bank action for different maturity rates operates not only by the direct *overnight rate change* channel on aggregate demand but also through the direct or indirect effect of Central Bank *signals* on expected future short rates.

In fact, one can extend the case for monetary policy predictability to the first and second arguments above. Central banking is also about managing expectations[2]. In essence, this paper seeks and produces information that may reduce uncertainty around an important monetary policy instrument. (1) says: "*Better information on the part of market participants about central-bank actions and intentions should increase the degree to which central-bank policy decisions can actually affect these expectations and so increase the effectiveness of monetary stabilization policy.*"

Our results are satisfactory in several dimensions. The set of quantified speeches obey a number of pre-established criteria, such as the positive correlation between their average and the *dots* median, the positive correlation

[2](8)

between the FOMC president and the rest of the committee, and even some anecdotal evidence collected from media vehicles. We are also successful in capturing the *true* association hypothesis as the most likely one in the only meeting in our sample for which we know for sure which member corresponds to each *dot*.

This work is mainly inserted in central bank communication literature, a topic which began to attract more attention since the early 2000s. "*When I was at the Federal Reserve, I occasionally observed that monetary policy is 98 percent talk and only two percent action*"[3], former FOMC chair Ben Bernanke would point out after he left the Fed presidency, in January 2014. While less theoretical work has been developed (e.g., (9) and (10)), the empirical part of this field covers a wide range of subjects. These can range from differences in *how* to communicate to the impact of communication in market variables, for instance. In the former, we can fit (5), in which the difference between collegial and individualistic approaches in policy making and communication is investigated; and (11), which deals with the effects from press conferences that explain policy decisions. On the other hand, in the later subject, (12) uses a high-frequency event-study analysis to identify separate market impacts from changes in the overnight rate and those originated from future policy guidance contained in FOMC statements; (13) studies term structure impacts originated from different sources of signals; and (14) focus on verbal interventions impacts to support the euro. A great survey on central bank communication can be found in (6).

More specifically, we are closer to articles that apply text mining techniques in order to quantify textual data into numerical data. (15) creates a glossary that maps expressions from the European Central Bank statements into numbers, where each document can take five values: -2 (very dovish), -1, 0, +1, +2 (very hawkish), to later test the consistency between wording and posterior policy decisions. (16) follow a similar dictionary method approach to create an index for FOMC minutes that ranges from -1 to +1. In a distinct manner, (17) recur to Google search engines and the *Dow Jones FACTIVA* news media database to quantify central bank language in terms of intensity and direction in a more automated (less subjective) way. For more information on text mining and central bank text, see also (18).

This work applies *Latent Dirichlet Allocation* (LDA hereafter) in its text quantifying procedure. Introduced in (19), this unsupervised machine learning tool extracts fewer hidden variables than the feature space dimension and has been widely used in natural language studies. To the best of our knowledge, it

[3]https://www.brookings.edu/blog/ben-bernanke/2015/03/30/inaugurating-a-new-blog/

was first introduced in central bank text analysis by (20), to extract meaning *diminishers* or *intensifiers*. (21) also use LDA to identify sentences from specific topics, applying word-counting methods in these phrases to create an index that translates FOMC views on the economic situation. In a similar manner, (22) use LDA to study ambiguity in Bank of Japan's monthly report texts, finding evidence that it deliberately introduces dubiety when the economy is in a bad state. (23) use the same technique but to study a different (but far from unimportant) question: deliberation patterns between FOMC members before and after transcripts became public. Our contribution to this branch of literature consists of providing a novel way for quantifying text into data - in this case, central bank communication into a *hawk-dove* scale -, in a manner less sensitive to researcher's arbitrariness and subjectivity. Our method is similar to (24) in reference of combining LDA and *shrinkage* econometric tools, with the difference that their interest relies more on market rates impact from news (volatility) than the implied interest rate level *per se*.[4]

This article also fits in the *zero lower bound* and *forward guidance* literature, subjects which have been well documented specially after the 2008 Crisis. In the seminal work by (26), agents expectations influence over the optimal monetary policy path is explored in an intertemporal equilibrium framework. Due to its *dynamic expectations* nature, the model also develops an optimal way of handling the ZLB which relates to the *forward guidance* concept. Since agents bring future expectations to present decisions, a *commitment* to keeping the main interest rate lower for a longer period even after the economy recovers may have direct positive results on current aggregate demand and, consequently, social welfare[5].

Although the *dot plot* is one important forward guidance tool, we did not find much developed work so far, probably due to its short data spam. In (28) the authors try to estimate the most likely timing for the *lift off* (the rates were still close to the ZLB by that time) by quantifying the uncertainty degree around these individual forecasts. They also use the *dots* as an input in deciphering the degree of *commitment* with low rates held by the committee so far. In a similar fashion, (29) also investigate the *dots* influence over the *lift off* occurrence, in comparison with the verbal influence contained in the post-meeting statements, finding statistical relevance in both. Finally, (30) uses the *dots* cross-sectional dispersion over different meetings as a measure for disagreement *within* the FOMC, postulating that the higher this degree, stronger is the sensitivity of market-based interest rates to macroeconomic

[4]For more information about text mining and economics in general, we suggest checking on (25).

[5]For a broader discussion about this instrument, see also (27).

news. All these studies point out to the relevance regarding the *dot plot*, in such a way that our contribution towards shedding some light on it should not be discarded.

There is a growing literature on how monetary policy decisions are formulated, specially regarding different types of Committees (as (31) defines, individualistic, autocratic, collegial or genuinely collegial), together with their consequences. (32) elaborate an empirical model which allows the classification of several Central Banks under (31) defined classes, providing micro foundations behind different types of decision-making processes, such as individual Taylor-rule-like preferences and the chairman influence over the rest of the Committee. (33) digs deeper in those preference determinants, proposing an empirical model to test which idiosyncratic FOMC members characteristics holds more influence over monetary policy decisions. Features like the time spent within the Federal Reserve System and the FOMC tenure seems to have a greater weight than ones like education, age and work experience. We believe the techniques proposed in this paper along with the new data we generate can be of great use in this stretch of literature, allowing to test more hypotheses or assess individual behaviour in a different way.

Finally, our work is also related to the multiple object tracking literature, a branch from the field of *computational science*. According to (34), *"The task of Multiple Object Tracking is largely partitioned to locating multiple objects, maintaining their identities, and yielding their individual trajectories given an input video"* - tasks which are similar to ours in many dimensions, specially if we think of each meeting as a frame from this video. Our tracking algorithm is inspired in works such as (35), (36) and, more specifically, (37), where a new filter is developed to keep track of unidentified targets in a cluttered environment along time, taking in its formulation joint posterior association probabilities. In a controlled environment, this algorithm shows dramatic improvement in estimating states and associations.

We contribute to this literature by creating a filter that may be applied in a situation where the researcher has a variable which behaves in a hybrid manner: not a fully *latent* variable, but also not a fully *observed* one. More specifically, in our scenario we are able to observe the *dots* set at each meeting, but we cannot individually link them neither *over* time nor to any observed individual (FOMC member). Our routine aims to solve that problem by estimating full distribution probabilities for these association hypotheses. Moreover, we believe our framework is general enough so it can be applied to any environment in which one can observe actions along time but not the agent behind it, however, there is still interest in estimating this information

- i.e. any secret voting setting.

This paper is structured as it follows. Chapter 2 presents our adapted *target tracking* algorithm developed to estimate association hypothesis between *dots* and members for each meeting, together with the theory and assumptions behind it and some simulations to show its effectiveness. Chapter 3 holds a brief analysis of all data used in this paper: FOMC statements, nominal rate decisions, individual speeches and cross-Chapter year-end *dots* distribution for each meeting. Chapter 4 introduces our new text quantifying method, with a brief description of all dimension reduction tools we use in our steps, such as the *Elastic Net* operator and the *Latent Dirichlet Allocation* (LDA) soft clustering device. In Chapter 5 we present the results both for our text quantifying technique, comparing it with some desired outcomes; as well as for our tracking algorithm, analyzing its answer for the last meeting of our sample, in which we know the true association between members and *dots*. Finally, Chapter 6 concludes and leaves suggestions for further research.

PUC-Rio - Certificação Digital Nº 1712591/CA

2
Target Tracking Algorithm

2.1
Theory and Recursive Steps

This study primary goal is to assign an occurrence probability for each feasible combination between interest rate *dots* and present members for each FOMC meeting, allowing us as well to extract both *member-to-dot* and *dot-to-member* individual probabilities. As an example, we may be interested in knowing how likely is the highest *dot* to represent member's A, B or C opinion; or if it is more likely for member D to be under, over or with consensus.

To achieve this objective, we first assume a structure between these interest rate individual assessments and the set of *quantified speeches* for each member - which, from now on, will be called *Policy Stance Index* (PSI). For now, suppose there is a way to convert these individual speeches into personal measures of appropriate interest rates. Chapter 4 explains in detail how we aim to do that.

This algorithm associates members with anonymous *dots* by observing the PSI over time, and is highly based on the *Target Tracking* literature. Works such as (35), (36) and specially the *Joint Probabilistic Data Association Filter* (JPDAF) derived in (37) were of great inspiration.

Consider the following space-state model:

$$\begin{aligned} x^t_{k+1} &= \Gamma x^t_k + \epsilon^s_k \\ y^t_k &= \Upsilon x^t_k + \epsilon^r_k \end{aligned} \tag{2-1}$$

Where k denotes the meeting (a time-series dimension); the *state* x^t_k represents member's t *dot* at meeting k, the *measurement* y^t his/her PSI; Γ and Υ are known parameters (which estimation will be dealt later in this Chapter), and both ϵ^s_k and ϵ^r_k are *white noise* processes with variance σ^s and σ^i, respectively. We also assume the initial state to be $\hat{x}_{0|0}$ and co-variance $\Sigma_{0|0}$.

Note that in this setting there is an important distinction between a state and a measurement - the later is a common variable, about which we know both its value and to whom it belongs; however, although we know the states values, we can't *a priori* link them to any *target* - the FOMC members -, and neither link them across time, obtaining at least their time-series. These are the uncertainty sources of our problem.

Define $Y_k \equiv \{y_k^1, ..., y_k^T\}$ as the set of PSI collected at meeting k, and $X_k \equiv \{x_k^1, ..., x_k^{m_k}\}$ as its set of *dots*. Note that, since we don't know which member corresponds to each *dot*, we index them by generic $j \in \{1, ..., m_k\}$ measurements instead of the known targets set, $\{1, ..., T\}$. Also, let $Y^k \equiv \{Y_n\}_{n=1}^k$ denote the story of measurements received up to time k and $X^k \equiv \{X_n\}_{n=1}^k$ the story of all *dots* (remember they are not linked through time), while m_k denotes the number of *dots* observed in meeting k. We shall also define what is an *individual event*:

$$\chi_j^{t_j} = \{\text{measurement } j \text{ originated from target } t_j\} \tag{2-2}$$

We depart from the following approach: suppose we are interested in the *expected dot* for each member *conditioned on all speeches received up to time k and all dots received so far, including k*:

$$\begin{aligned}\mathbb{E}\{x|Y^k, X^k\} \equiv \tilde{x}_{k|k} &= \int xp(x|Y^k, X^k) \\ &= \sum_{j=0}^{m} \mathbb{E}\{x|\chi_j^{t_j}, Y^k, X^k\} P\{\chi_j^{t_j}|Y^k, X^k\}\end{aligned} \tag{2-3}$$

Where the last step holds because all *feasible events* - those in which no more than one measurement originates from each target - are exhaustive and mutual exclusive. We also drop member t and time k indexing in x_k^t for notation simplicity.

The time line of what happens between meetings $k-1$ and k helps us keep track of what will be handled in each step from our algorithm:

i In the end of period $k-1$, we have all calculated conditional expectations $\tilde{x}_{k-1|k-1}$, which are functions of all received *dots* and PSI so far[1];

ii Between meetings $k-1$ and k, we collect the PSI values for each member[2]. We shall call this "inter meeting" period in which we have the values for all y_k^i but not yet for any x_k^j period $k - \frac{1}{2}$.

iii Finally, the set of *dots* for meeting k, x_k, is received, and the event probabilities are collected.

[1]Except for $k = 0$, in which an arbitrary prior needs to be assigned by the researcher
[2]the methodology for that will be described in chapter 4

The left hand-side term of equation (2-3) last step, $\mathbb{E}\{x|\chi_j^{t_j}, X^k, Y^k\}$, is simply the value of *dot* x^j associated to target t_j in hypothesis $\chi_j^{t_j}$. The only term left to be found then is the one on the right, the conditional probability of event $\chi_j^{t_j}$, $P\{\chi_j^{t_j}|X^k, Y^k\}$, for which we will use the notation β_j^t - this is the term we are most interested in.

The key factor for this approach is the evaluation of the following joint events conditional probabilities:

$$\chi = \cap_{j=1}^{m} \chi_j^{t_j} \tag{2-4}$$

Given that they obey the *feasibility condition*:

$$j \neq l \cap t_j > 0 \Rightarrow t_j \neq t_l \tag{2-5}$$

In the simplest example, if we had 2 *dots* (labeled $d1$ and $d2$) and 2 members (A and B), the two feasible joint events would be $\{d1$ is A and $d2$ is $B\}$ and $\{d2$ is A and $d1$ is $B\}$. Using Bayes' Rule we can rewrite the conditional probability of a joint event as it follows:

$$\begin{aligned} P\{\chi|Y^k, X^k\} &= P\{\chi|x_k^1, ..., x_k^m; X^{k-1}; Y^k\} \\ &= p(x_k^1, ..., x_k^m|\chi, X^{k-1}, Y^k) P\{\chi|X^{k-1}, Y^k\}/c \end{aligned} \tag{2-6}$$

Where k from m_k notation was dropped for simplicity and the normalization constant c is the joint prior density of all measurements, and is easily obtained by summing the numerators over all χ.

Appendix A contains all algebra and assumptions necessary to derive our algorithm. The most important hypothesis regards a *Gaussian Approximation* for the *dots*' likelihood functions:

$$p(x_k^j|\chi_j^t, Y^k, X^{k-1}) = \mathcal{N}(x_k^j; \hat{x}_{k|k}^{t_j}, Q_k) \tag{2-7}$$

In words, it is equivalent to state that the measurement-updated expectancy $\hat{x}_{k|k}$ is distributed in a Gaussian manner around the true received *dot*, x_k^j[3], with co-variance Q_k detailed below.

We can then finally compute the conditional probability of a *joint hypothesis:*

$$P\{\chi|Y^k, X^k\} = \frac{1}{c} \prod_j \mathcal{N}(x_k^j; \hat{x}_{k|k}^{t_j}, S_k) \tag{2-8}$$

Which also allows us to extract individual *member-to-dot* probabilities - we just need to sum the probabilities of all events in which *dot* j is assigned to member t:

[3]This can be assured by the *symmetry* property of a Normal distribution, since assumption (2-7) actually states that the true dot is distributed around our previous expectancy.

$$\beta_j^t = \sum_{\chi:\chi_j^t \in \chi} P\{\chi | Y^k, X^k\}; \quad j = 1, ..., m; \; t = 0, ..., T. \tag{2-9}$$

The following Bayesian algorithm is then proposed to keep track of all assignments probabilities and conditional expected dots for each meeting k. Parameters written under hats means they are estimated (more details later in this chapter). Start with $\tilde{x}_{0|0}$ and $\Sigma_{0|0}$ (initial priors), set $k = 1$. For each k and each t:

I Run the *prediction update* (period $k-1$):

$$\begin{aligned} \hat{x}^t_{k|k-1} &= \hat{\Gamma}\tilde{x}^t_{k-1|k-1} \\ \Sigma^t_{k|k-1} &= \hat{\Gamma}\Sigma^t_{k-1|k-1}\hat{\Gamma}' + \hat{\sigma}^s \end{aligned} \tag{2-10}$$

II Run the *measurement update* (period $k - \frac{1}{2}$):

$$\begin{aligned} \hat{x}^t_{k|k} &= \hat{x}^t_{k|k-1} + K^t_k(y^t_k - \hat{y}^t_{k|k-1}) \\ \Sigma^t_{k|k} &= \Sigma^t_{k|k-1} - K^t_k Q^t_{k|k-1} K^{t'}_k \end{aligned} \tag{2-11}$$

where:

$$\begin{aligned} \hat{y}^t_{k|k-1} &= \hat{\Upsilon}\hat{x}^t_{k|k-1}; \\ Q^t_{k|k-1} &= \hat{\Upsilon}\Sigma^t_{k|k-1}\hat{\Upsilon}' + \sigma^r; \\ K^t_k &= \Sigma^t_{k|k-1}C'(Q^t_{k|k-1})^{-1} \end{aligned} \tag{2-12}$$

III Calculate the likelihood for each received dot j (period k):

$$p^t_{k|k}(x_j) \equiv \mathcal{N}(x^j_k; \hat{x}^t_{k|k}, \Sigma_{k|k})$$

IV Generate $\tilde{x}^j_{k|k}$ - the expectation of a member's *dot* conditioned on all PSI and *dots* received so far, as described in (2-3):

$$\tilde{x}^t_{k|k} = \sum_j x^j_k \beta^t_j$$

Although the variable $\tilde{x}^t_{k|k}$ is not our primary focus here, it takes an important role in calculating any association probability we might be interest in. Note, by the algorithm's recursions above, that it will have an indirect impact in $\hat{x}^t_{k|k}$, the center of the Normal *pdf* which will serve as distance norm between realized *dots* and what they were expected to be. Their existence is important since no previous feasible hypothesis is totally discarded as the algorithm is run, such as other algorithms that, for instance, selects only the most likely *association hypothesis* at each k and assign full probability to it, throwing out most likely useful information each time a new set of data is received[4].

A last feature from our tracking algorithm is worth noticing. Although this paper is focused on anonymous *dots* and members' quantified speeches as input, the developed framework is general enough so it can be applied in any situation in which there is interest to link actions with agents when these assignments are not directly observed, but there are co-related variables that can help - such as an anonymous voting setting.

2.2 Parameters Estimation

Estimating parameters Γ, Υ, σ^s and σ^r would be an extremely hard task in a regular environment in which we knew the links between *dots* and members, due to the short time-series and the big and alternating number of members per meeting. If we can't even properly assign the time series to individuals, it turns the direct estimation impossible, forcing us to take some extra assumptions.

For the first equation in (2-1), we estimate a simple AR(1) process over the FOMC nominal target interest rate, using its coefficient of approximately 0.99 for $\hat{\Gamma}$ value and its estimated standard deviation for $\sqrt{\hat{\sigma^s}}$, 0.048. The assumption here is that the committee process for setting the nominal rate and the one for individually assess the appropriate end-of-the-year nominal rate behaves similarly. Although they refer to slightly different terms, it seems a fairly reasonable assumption.

As for the measurement equation, due to the methodology employed in chapter 4 to quantify speeches and generate the PSI, it is sufficient to assume $\hat{\Upsilon} = 1$. Notice that, by doing that, we model it as a noisy measure *around* the true value of member's t individual preferred interest rate, with a signal variance σ^r, which will also be dealt with in chapter 4.

Table 2.1 below summarizes all variables and parameters described above, acting as a glossary to help the reader navigate through the algorithm.

[4]For an example, check (38)

Table 2.1: Target tracking algorithm variables summary table.

Variable	Type	Description	Status
x_k^t	Scalar	Member's t *state* (*dot*) at period k	Not linked to t
X_k	Vector	*Dots* collected at meeting k	Not linked to t
X^k	Matrix	Story of *dots* collected up to meeting k	Not linked across time
y_k^t	Scalar	Member's t *measurement* (PSI) at period k	Fully observable
Y_k	Vector	PSI collected at meeting k	Fully observable
Y^k	Matrix	Story of PSI collected up to meeting k	Fully observable
$\tilde{x}_{k\|k}$	Scalar	$\mathbb{E}\{x\|Y^k, X^k\}$	Estimated along the algorithm
m_k	Scalar	Number of members and *dots* in meeting k	Fully observable
Γ,Υ,σ^r,σ^s	Scalars	Model parameters	Estimated in 2.2
$\chi_j^{t_j}$	Definition	Event *"dot j represents member t_j"*	Defined
χ	Definition	Joint event ($\chi \equiv \cap_{j=1}^{m} \chi_j^{t_j}$)	Defined
β_j^t	Scalar	Probability that dot j represents member t	Estimated along the algorithm

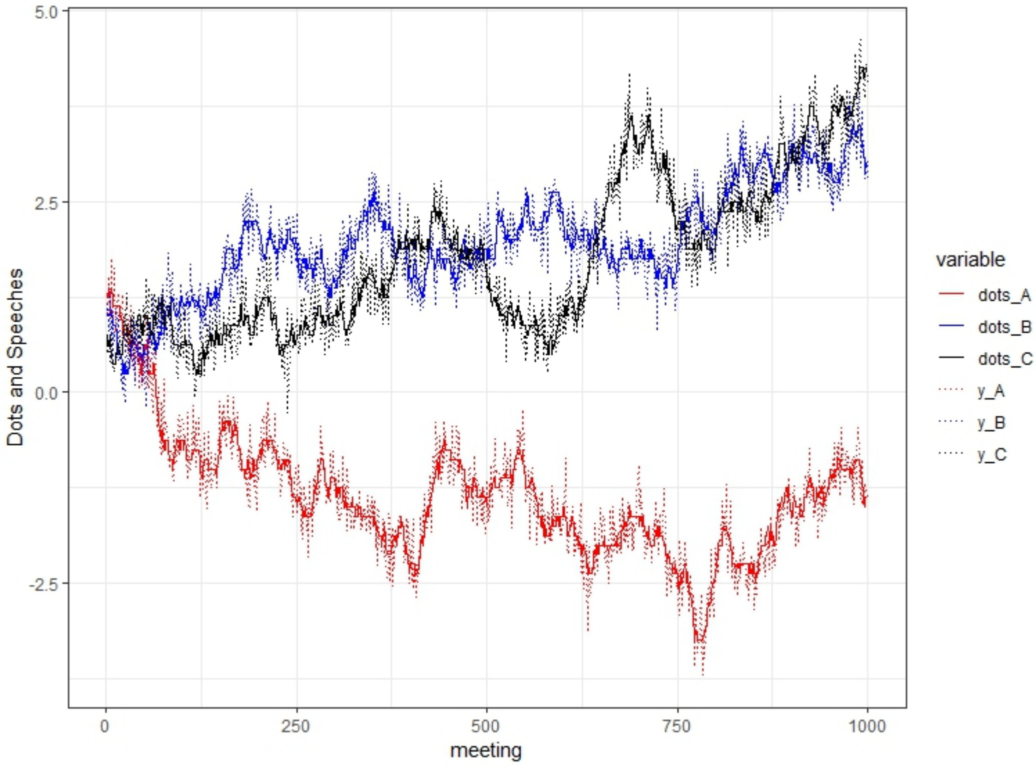
2.3 Simulations

In order to test our algorithm efficiency, some controlled experiments are run. We build a setting in which we know the *true* association between *dots* and members, but feed the algorithm only with the exact information that would be available when running it with real data (without the correct answer).

As a simplification, consider a setting where there are 3 FOMC members - labeled A, B and C - and 3 *dots* - 1, 2 and 3 - per meeting. For this experiment, we simulate equation (2-1) over a thousand periods, setting the following values for its parameters: $\Gamma = 0.99$, $\Upsilon = 1$, $\sigma^s = 1$ and $\sigma^r = 1$, which are not far from what is estimated with real data (scales considered).

A first data set is simulated for each member, for both PSI and *dots* values, following the *space-state* model described in (2-1) with the parameters imposed above. The respective paths are shown in Figure 2.1:

Figure 2.1: 3 members simulated as *dots* and *PSI* series.



Our algorithm is then recursively run over these 1000 periods, computing and storing joint and individual probabilities for each stage. The *joint hypothesis probabilities* are shown in Figure 2.2. The notation designs the hypothesis probability for each member (letter) and *dot* (number) that appear consecutively, as assigned by the tracking algorithm - for instance, p_A1B2C3 refers to the following association: $\{$*dot 1 is member A, dot 2 is member B, dot 3 is member C*$\}$ (in this case, the correct assignment, as we defined earlier). We also display the *individual member-to-dot* probability in Figure 2.3, β_j^t - which corresponds to p_tj.

Some desirable features of the tracking algorithm behaviour are worth noticing. First, most of the time it is pointing out the true association as the most likely hypothesis - areas in which the red line is above all other lines. The regions in which it doesn't happen are mostly those in which members B and C get really close from each other, where it was expected to have a greater level of uncertainty. Second, the individual probabilities follow the same logic, specially for member A - which by the simulated paths is further away than the other members.

Figure 2.2: Joint hypothesis probabilities given by tracking algorithm. The red line represents the true association.

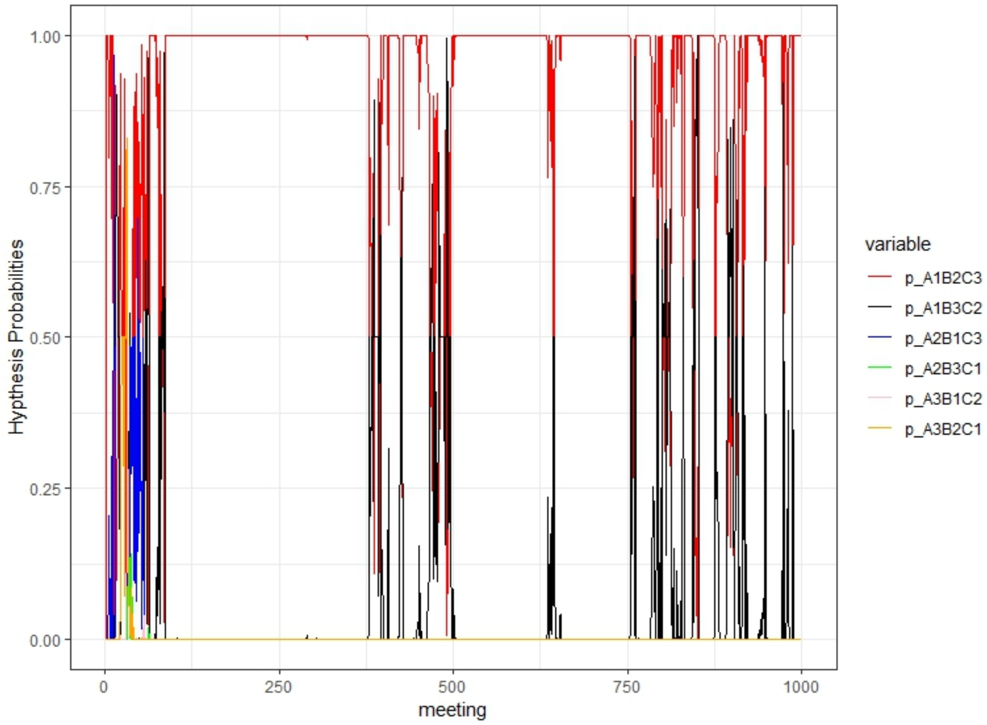
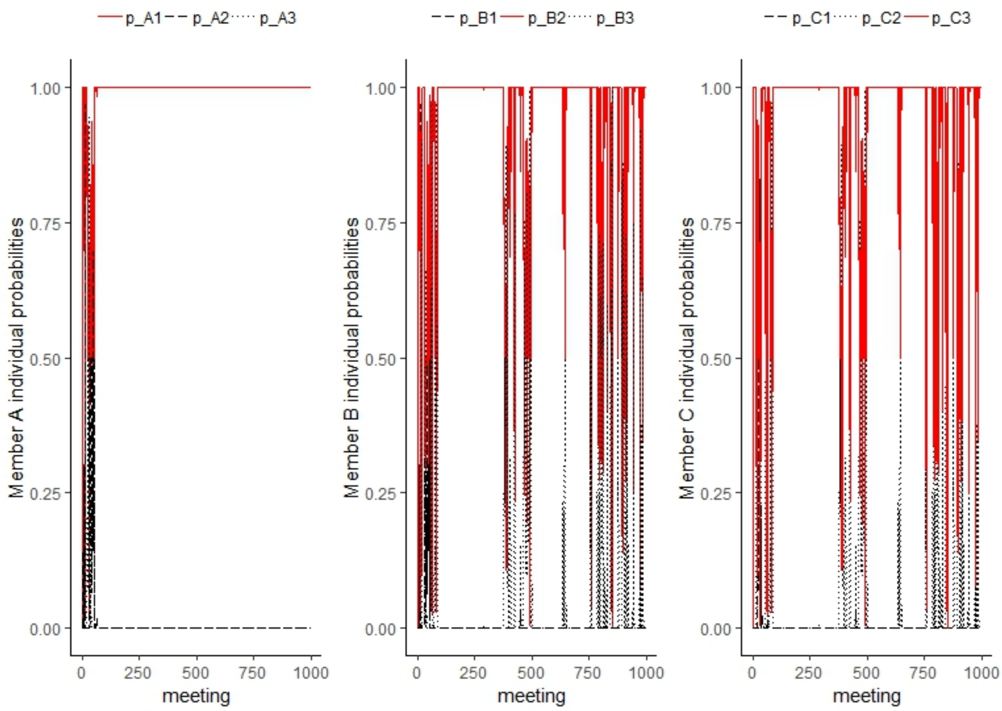


Figure 2.3: Individual hypotheses probabilities given by tracking algorithm. The red line in each plot represents the probability estimated for that member true association.



Still, one may argue that this feature comes more than than the simulated path than from the algorithm capability. To deal with that, we perform the same experiment, but repeated a thousand times (for a thousand different paths for *dots* and PSI), taking the average calculated probabilities. Those are displayed in Figures 2.4 and 2.5.

The *smoothed* simulation over a thousand different paths actually improves the tracking algorithm performance, in a sense that, on average, not only it *always* gets both joint and individual probabilities as the highest one, but it also shows a *learning pattern* - with the correct assignment increasing its probability as time passes by and more information is incorporated. Closer to the last periods, on average, the true link between *dots* and members is considered to be next to the unity.

Some components play an important role in the tracker performance. Specifically, one can point out the *PSI standard deviation* ($\hat{\sigma}_k^r$) and the *initial assignment prior*. Appendix B brings some robustness tests about both these characteristics. The overall algorithm performance falls, as expected, but it is still able to identify the true association with the highest probability throughout the whole period in most cases.

Figure 2.4: 1000 paths average for joint hypotheses probabilities. The red line represents the true association.

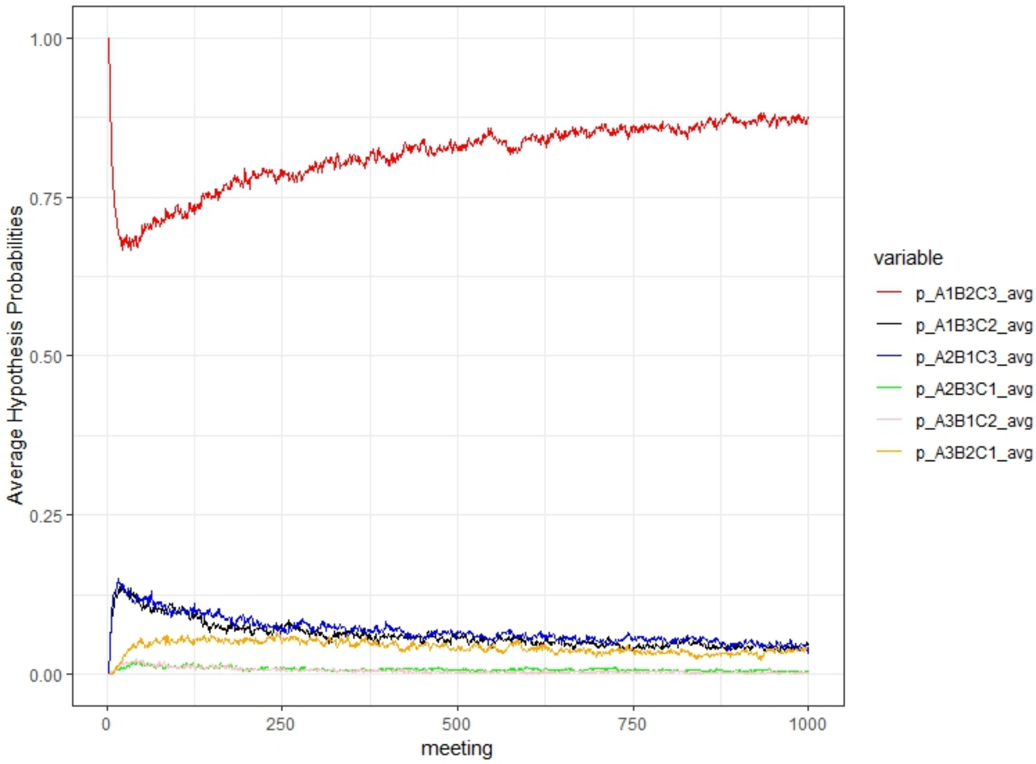
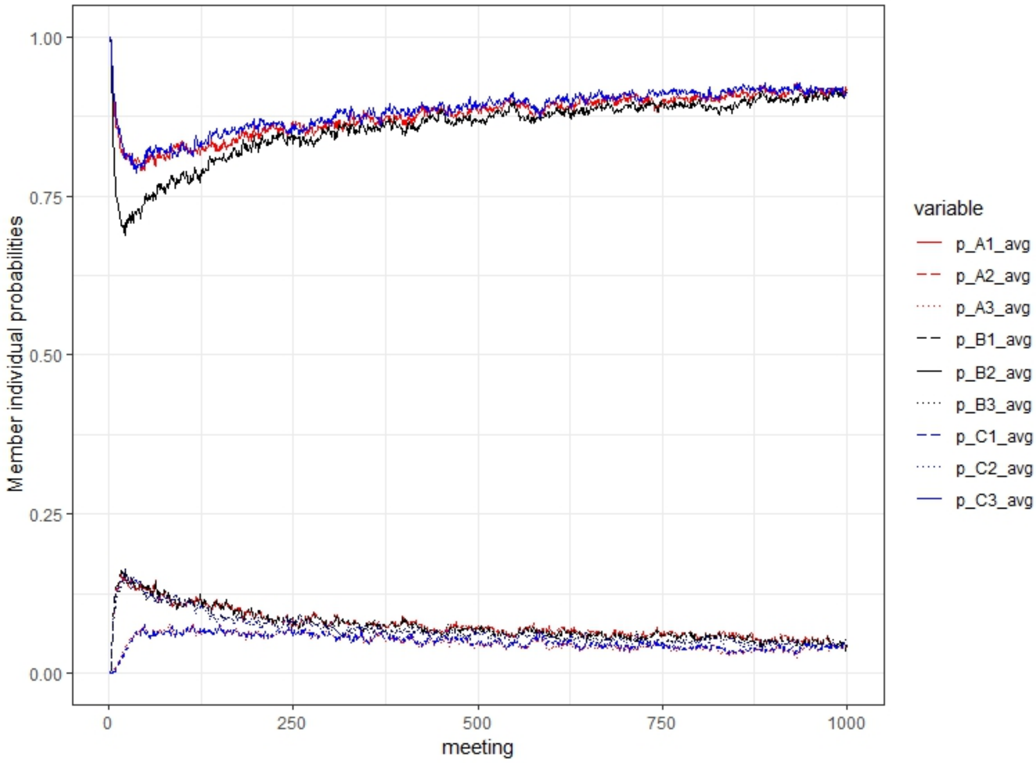


Figure 2.5: 1000 paths average for individual association probabilities. The full-weighted line represents the correct individual associations.



3
Data

3.1
FOMC structure

Before presenting the data itself, it is worth explaining how the monetary policy committee works in practice. Overall, the FOMC is composed by 19 members, divided in those who have the right to vote on the policy decision and those who only take part in the discussion. The 7 members from the *Board of Governors* (including the *president* and *vice-president*), the *Federal Reserve Bank of New York* president and other four *Regional Reserve Bank Presidents* (who act on one-year rotating basis) belong to the first group, while the other five regional presidents do not vote. Nevertheless, even them can provide their individual *dots*.

More than that, each member is *allowed* but not *compelled* to provide this information, in a way it is common to end up with less *dots* than present members: on average there are about 18 present members in each of the 25 meetings covered by this paper, but an average of just 17 *dots* published in them. In only 5 occasions these numbers are equal.

In other words, the *dot plot* presents the whole FOMC *corpus* view, both voting and non-voting members - which is why we found it relevant to know which ones of them represent the decision-making members and, therefore, get a clearer view regarding the FOMC bias towards a tighter or looser monetary policy stance.

Another important mechanic in FOMC functioning regards the possibility of members being present in some meetings and not in others, for several reasons - from *vacancies* opened in a Regional Fed position to particular reasons. This has a direct implication to our tracking algorithm: regard from (2-10) that it updates the first *prior* for period k taking in account the probability-weighted *dot* for member t in period $k-1$ ($\tilde{x}^t_{k-1|k-1}$). But if this member wasn't being tracked in $k-1$, this term doesn't exist *a priori*. To overcome this issue in order to keep our algorithm running, whenever this is the case, we impose our own conservative prior and consider $\tilde{x}^t_{k-1|k-1}$ equal to that meeting median.

3.2 Numeric Data

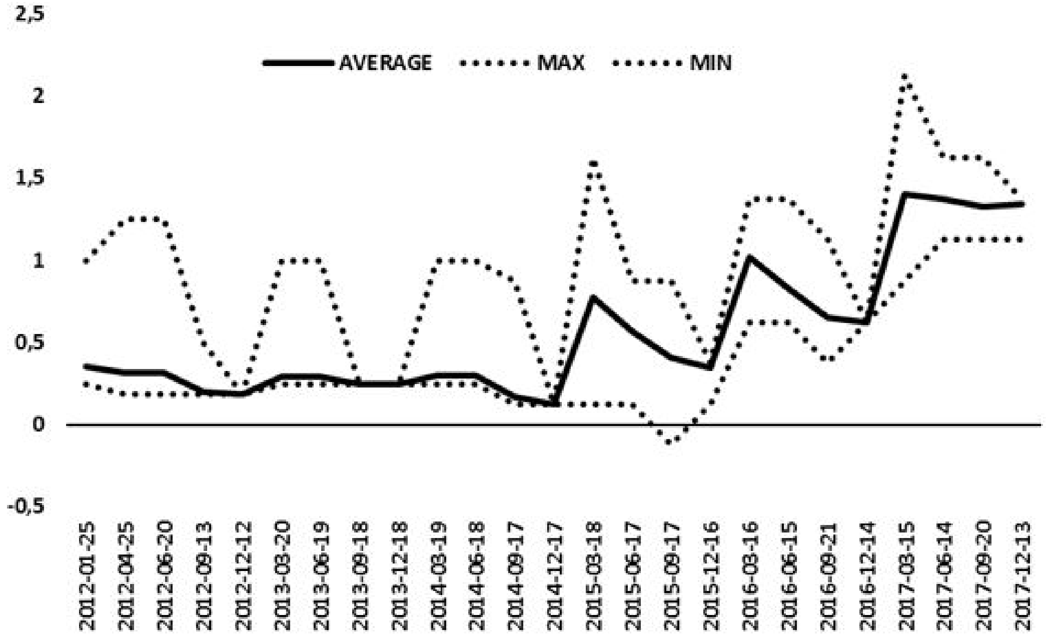
3.2.1 Dot Plot

Beginning in January 2012, the interest rate *dot plot* is published once in every two meetings - a total of 25 times between 2012 and 2017. According the *FOMC projection material*, it is formally defined as it follows: *"This chart is based on policymakers' assessments of appropriate monetary policy, which, by definition, is the future path of policy that each participant deems most likely to foster outcomes for economic activity and inflation that best satisfy his or her interpretation of the Federal Reserve's dual objectives of maximum employment and stable prices. Each shaded circle indicates the value (rounded to the nearest $\frac{1}{8}$ percentage point) of an individual participant's judgment of the midpoint of the appropriate target range for the federal funds rate or the appropriate target level for the federal funds rate at the end of the specified calendar year or over the longer run"*.

The interest rate *dot plot* contain information about the current year, the following two years and the long run. Although we recognize longer term assessments may bring useful information, we opt to work only with the current year-end *dots*, since our text-quantifying strategy is heavily based on language surrounding short-term decisions provided by the FOMC, as it will become clear in chapter 4.

As mentioned earlier, since it is optional for policymakers to provide their individual views, usually there are fewer observed *dots* for a given year-end than the proper number of members. In our sample, this amount ranges from 16 to 19, and whenever this mismatch happens, we have to fill up the *dots* set until m_k is equal to T_k - a necessity required by our algorithm. We proceed conservatively and add "synthetic" dots equal to the median for that meeting until this condition is reached. The path for the average, the maximum and the minimum year-end *dot* in each meeting are shown in Figure 3.1:

Figure 3.1: The average, the maximum and the minimum value for cross-sectional year-end dots in each meeting from our sample.



There is one more adaptation needed to be done in order to ensure the tracking routine to perform. For some meetings, the possible number of association hypothesis is so huge that it becomes computationally infeasible to estimate their probabilities. For a given meeting k, let L be the number of *dots* levels, each one with r^l *dots*. The number of possible associations between *dots* and members, $TotalHyp_k$ is given by the multinomial formula:

$$TotalHyp_k = \frac{T_k!}{r_k^1 \times ... \times r_k^L} \tag{3-1}$$

As an example, consider the March 2017 meeting represented in Figure 1.1. It has 5 interest rate median levels, each one with 1, 4, 9, 1 and 2 *dots*, respectively. There are over 20.420.400 possible associations between dots and members that can configure its composition. Just for this meeting, it would take weeks to attribute individual likelihoods to each one of them and then collect joint hypotheses probabilities for each individual configuration.

We propose a last simplification to solve this computational problem. Instead of working with the whole set of *levels* in each meeting, we work with only three: *consensus*, *above consensus* and *under consensus*. Every *dot* greater than the median value for that meeting is equalized to $\max\{X_k\}$, and the same for *dots* smaller than the median value, but the other way around - all of them are set to $\min\{X_k\}$. Although we may loose some precious information regarding cross-sectional variability, this necessary change allows us at least to still have the full range between the lowest and the highest *dot* in each FOMC in order to create $\tilde{x}^t_{k|k}$.

3.2.2 Official Speeches

As main input, all the official FOMC members speeches issued between October 2011 and December 2017 will be in our starting data set. They can be found at the official FOMC website, as well as at each of the 12 regional Fed web pages. However, not all of them will be used, since some texts deal with subjects not related to the speaker's economic situation assessment: texts regarding regional economy situation or the Economist profession are common, therefore biasing the *Policy Stance Index* estimation.

To overcome that, the final database will contain only the speeches in which at least one of the words *inflation*, *growth*, *employment* and *GDP* comes up at least once. We believe this is enough to filter out texts which would not bring useful information for our purpose. That makes our final speeches data set end up with 1.132 individual speeches.

3.2.3 FOMC Statements and Monetary Policy Decision

At some point during our text-quantifying strategy, the *FOMC monetary policy statements* will be used as important inputs. Issued right after each meeting together with the monetary policy decision, these statements contain summarized information regarding members' discussion about their economic situation assessment and appropriate policy to be adopted. They can be also be found at the FOMC website.

The FOMC statements started to be published regularly after each meeting in May 1999, yielding a time series of 159 observations up to December 2017 (our *dots* sample end) containing both *quantitative* and *qualitative* information regarding each meeting policy decision - the short-term target interest rate and the FOMC *rationale* behind this decision contained in the statement text. Both of these dimensions will be used at the *Policy Stance Index* construction.

Another possibility could be the use of the FOMC minutes instead of monetary policy statements, since they bring a much more detailed discussion about the economic scenario and has a longer time-series. However, at a first moment we choose the statements mainly so we can depart from an already reduced dimension in our information space. As it will become clear in chapter 4, our mechanism for *PSI* generation highly relies on *shrinkage* econometric tools, and too many variables could jeopardize our technique efficiency. Still, Appendix C brings a different specification which uses the minutes as main input, together with its main results, and we find little change in most measures regarding its efficiency.

3.3 Text Pre-Processing

As it is common in natural language studies, it is useful to pre-process textual data by cleaning the noise contained in some words or symbols, improving efficiency of applied quantifying techniques. In this work, we basically follow (23) *pre-processing* steps.

3.3.1 Collocations and Stop Words

First, we identify sequences of two- or three-word terms that have a specific meaning, mostly those which refer to economic concepts, such as *labor market* or *terms of trade*, and replace them for single-word tokens, such as *"labmkt"* or *"termsoftrade"*, respectively.

After identifying collocations, we need to remove English *stop words* - common words that do not bring very useful information, such as "*the*", "*is*" and "*on*". We use the preset words contained in the "tm" [1] package on R software.

3.3.2 Single Characters and Punctuation

Next, we remove all kinds of punctuation and single-letter characters, such as "a", "I" and currency symbols. We also turn all letters to lower case and remove all numeric characters, since in these speeches they are usually used to refer to dates or other kind of information that won't be used in our methodologies. Doing that prevents the algorithm from treating these characters as relevant tokens.

3.3.3 Stemming

"Bank", "Banks" and "Banking", under our lens, have virtually the same meaning. In cases like these, it is interesting to replace these words with their roots, reducing even more the feature space without greater loss of information. In our example, all the three words above are replaced by "Bank". Therefore, we

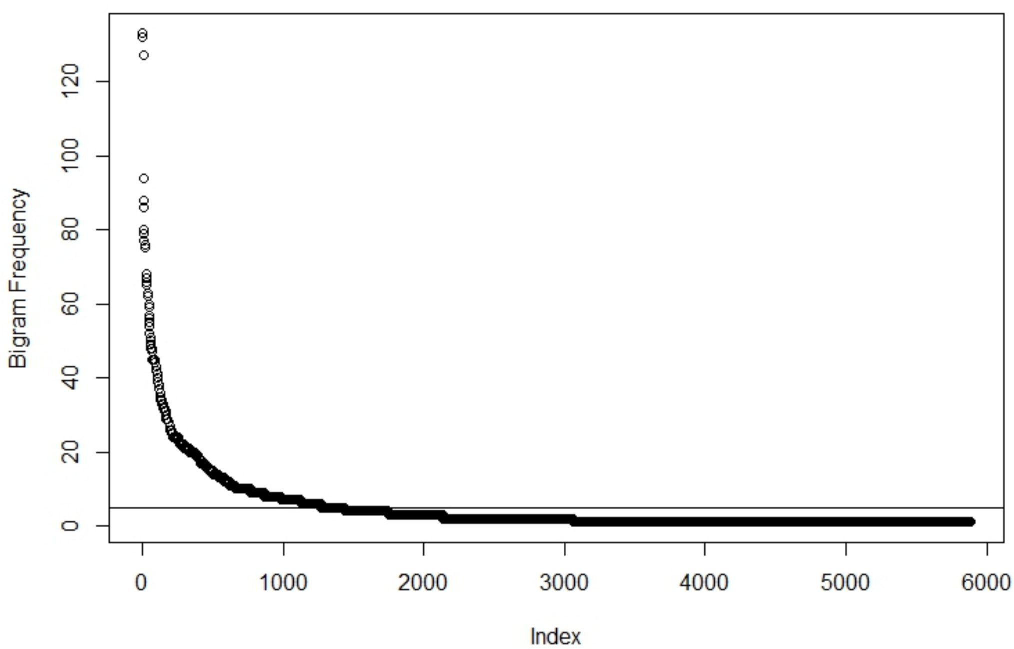
[1] https://cran.r-project.org/web/packages/tm/index.html

will use the *Porter Stemmer*, explained in (39). This approach is also standard in natural language works.

3.3.4 Excluding Common and Rare Words

Finally, we take account of very rare words and also remove them from our corpus. This has proven to be very useful in natural language studies since it drastically reduces problems' dimension by deleting noise that may be read as useful information. Also following (23), we establish a threshold for a minimum number of documents a given word has to appear in our database in order to stay in the text *Corpus* - defined as the collection of all texts composing the database, segregated at the statement level. Specifically, we remove all tokens appearing in less than 5 FOMC statements. Although arbitrary, the results are robust to changes in both these limits.

Figure 3.2: Number of documents in which each bigram appear. Those below the horizontal line (5 documents) are excluded for dimension reduction.



4
Quantifying Speeches

4.1
Numeric Representations

With the processed versions of FOMC statements and speeches in hand, the text quantifying technique will work with two *numeric representations* extracted from them.

The first one will be the *bi-gram* (sequence of two consecutive words) count by document. Terms such *"increasing inflation"* and *"tightening labor market"* (which is translated to *"tight labmkt"* after our pre-processing) are expected to be part of the central banker justification to a given interest rate decision. It follows the rationale behind what is called the *"topic and tone"* approach - first, a subject is identified by a word or a set of them, and then a qualitative information is extracted regarding this subject by analysing other words near it (usuallt adjectives or adverbs). In the *"increasing inflation"* example, the subject would be *"inflation"*, and the tone regarding it would be given by the term *"increasing"*.

It yields a high-dimension matrix, in a sense that the number of covariates exceeds the number of observations in our sample (statements). We still pre-filter this huge matrix by keeping only the bigrams that contain one of the following stemmed terms: *inflat*, *gdp*, *growth* and *empl*, aiming to improve our operators performance by reducing the space dimension. We end up with 159 rows (observations) and 285 columns (unique bi-grams that contain the words above), where the (i, j)th term represent the count of bi-gram j in statement i.

The second numerical representation will be formed by the *latent topic structure* contained in the FOMC texts and member speeches. It is common in natural language studies to think of a text as a set of words which are sequentially drawn according to the following structure: first, a *topic* τ is randomly chosen from $\mathcal{T}$ possible ones; next, since a *topic* is essentially a probability distribution over a set of words, a token is selected from this distribution μ_τ. The process is repeated until the document is filled up.

At a first moment, we will estimate this underlying topic structure using

FOMC statements, and just in a further step we will re-estimate it at the individual speeches level. The reasons for that will soon become clearer.

In order to extract these hidden *topics* from FOMC documents, we use the *Latent Dirichlet Allocation* (LDA) probabilistic topic model - similar to a soft clustering algorithm, it basically groups words in topics throughout density functions, based in the co-occurrence between them. Developed by (19) and with almost 25.000 citations since then, this technique was first applied in Economics in works such (20), (21) and (23). Although here we provide a detailed overview, more information including theory and estimation procedure can be found at (23) online Appendix[1].

Back to the *uni-gram* version of the pre-processed text, we construct the *term-document-matrix* (henceforth *tdm*). Let D be the number of documents (at this step specifically, we follow (21) and group at the *paragraph* level); and V the total count of unique terms in a collection of documents, which will be called *corpus*. The *tdm* consists of a $D \times V$ matrix, where the (d, v)th term represents the number of times word v appears in document d. It again yields us a very high-dimension matrix, as well as a very sparse model, since most of the *tdm* entries will consist of zeroes.

LDA is an unsupervised machine learning algorithm aimed to alleviate precisely these problems. Its ultimate goal is to find $\mathcal{T}$ meaningful word groupings in the data and represent each document in terms of these *topics*, going from the V-dimensional space to the $\mathcal{T} - 1$ simplex, as its outputs are probability distributions of documents over topics. Since it is a mixed-membership model, not only each word can belong to different *topics* with different weights, but also each document can belong to multiple *topics*. In practice, it is not so different than finding latent *factors* in this sparse matrix.

Basically, we feed it with 3 inputs:

- *Documents Corpus:* in our case, the full story of Fed statements, where words are grouped at the paragraph level.
- *Number of Topics ($\mathcal{T}$):* there is not a consensus of how to best choose it since there is a trade-off between the model interpretability and its goodness-of-fit. Therefore, after some value testing and following previous literature, $\mathcal{T}$ is set to 62[2].
- *Parameters:* since the LDA algorithm makes use of *Dirichlet* Distributions as priors, we need to provide the parameters α and η. As in (23), we follow (40) and set $\alpha = \frac{50}{T}$ and $\eta = 0.025$.

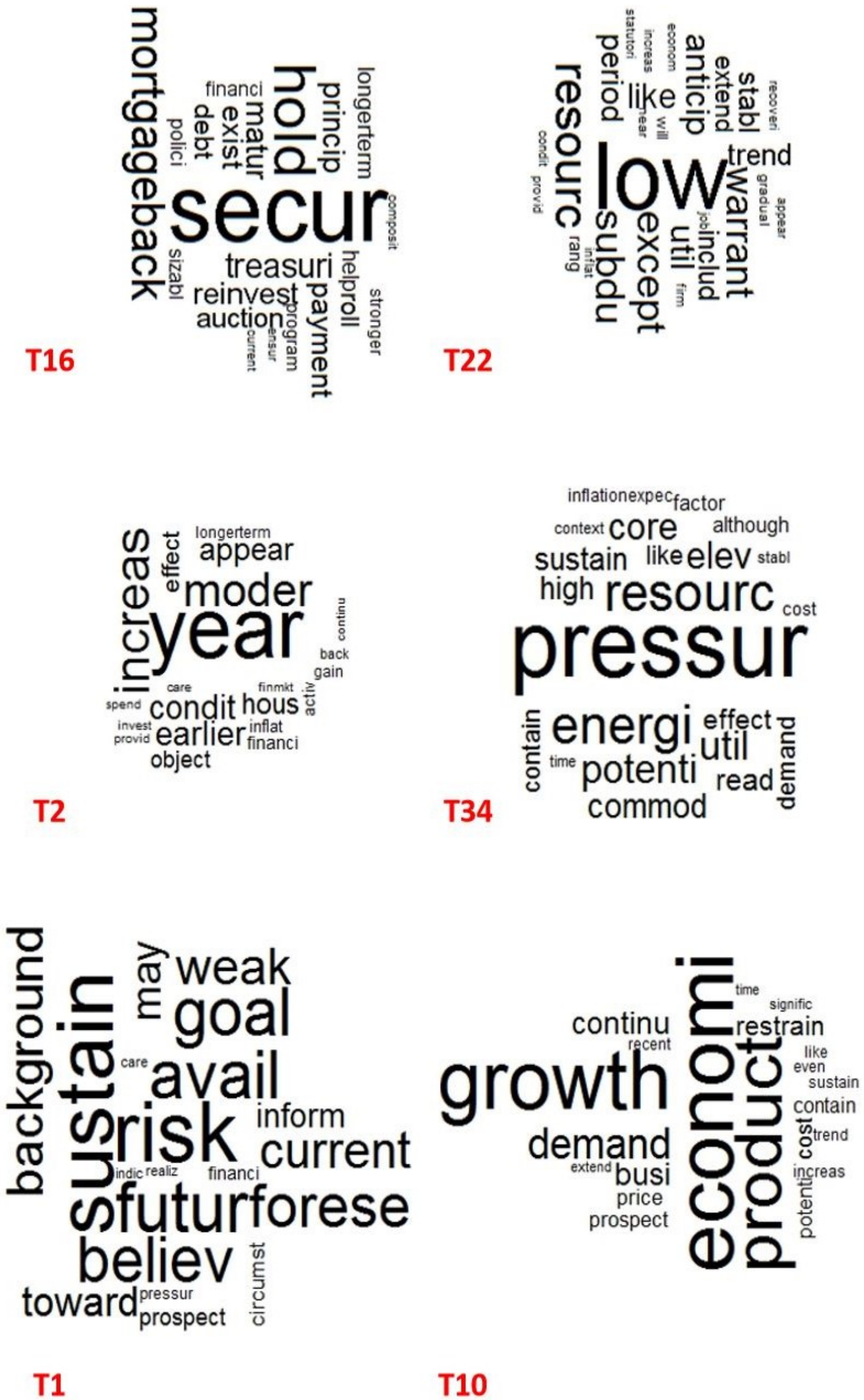
[1] https://sekhansen.github.io/pdf_files/fomc_technical_appendix.pdf
[2] some robustness to that is presented in Appendix C, with an alternate specification.

After providing these inputs, LDA returns us 2 outputs:

- $\mathcal{T}$ topics, which are probability distributions over words: a predicted distribution $\hat{\mu}_{\tau} \in \Delta^{V}$ over the V unique terms.
- Document distributions which capture the fraction of a *document* is devoted to each topic. Formally: a predictive distribution $\hat{\theta}_{\mathbf{d}} \in \Delta^{\mathcal{T}}$ over topics.

It is important to freeze that although LDA *topics* will be estimated at the FOMC minutes *paragraphs* level, we are more interested in the whole minute-level distribution over different topics. To obtain that, we follow the procedure described in (23) with just a few minimal modifications. Figure 4.1 shows some of these *topics* in means of *word clouds*, where the word size is proportional to its weight in this topic. A first note interesting to mention is how they seem related to specific economic matters, even if the algorithm wasn't fed with anything to indicate that.

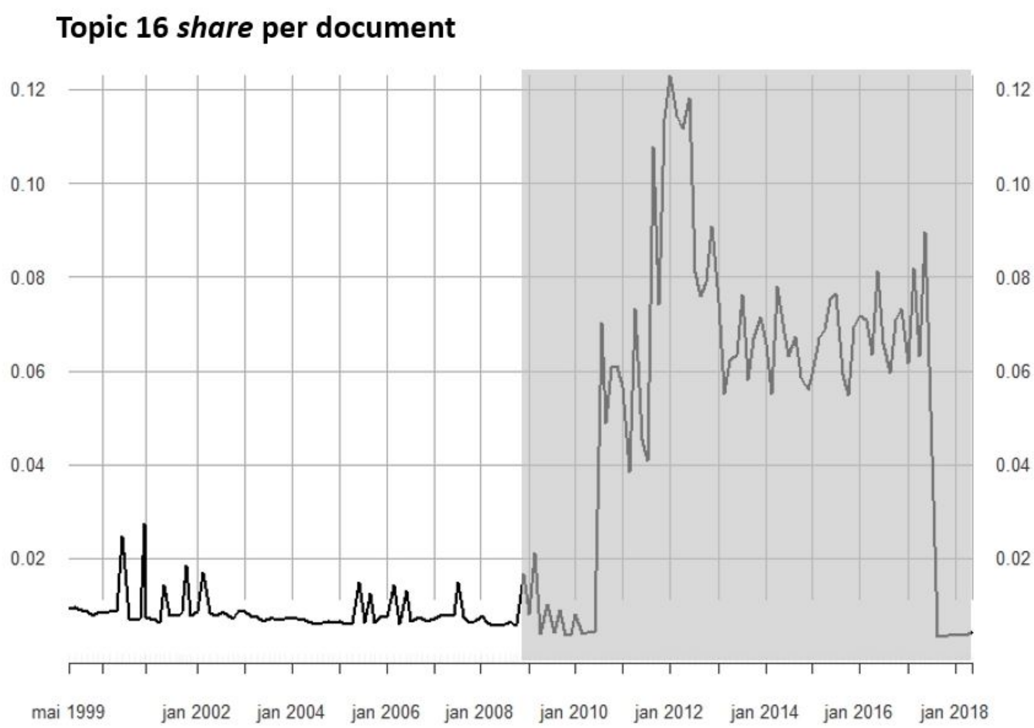
Figure 4.1: 6 out of the 62 topics given by LDA applied on FOMC statements; the word size in the cloud is proportional to its relevance for the given topic.



This kind of representation is complementary to the *bigrams* one, since it tries to summarize the *narrative* content of each statement, in terms of topics covered. If, for instance, the Fed is worried with the economic activity and spend a greater part of its statement talking about it, this topic will have a greater value. This representation also allows a better comparison between different documents.

To illustrate this underlying quantified text structure, Figure 4.2 plots the time series for Topic 16 *share per document* for each statement in our time series. Figure 4.1 indicates that this topic is somewhat related to the Fed balance sheet and financial instruments, thanks to words such as "*securities*", "*reinvestment*", "*maturities*", "*debt*" and many other, a vocabulary also very common when talking about the Fed balance sheet expansion, also known in the literature as *Quantitative Easing*. The grey area starts at the beginning of the first phase ("QE1"). There seems to be a correlation between these two facts, with the topic 16 coverage increasing up to five or six times from one period to another.

Figure 4.2: Topic 16 share per document time-series. Grey area goes from the start of QE1 in January 2009 until the first interest rate hike after the Financial Crisis, in December 2015.



4.2
The Policy Stance Index

With these newly generated numerical series in hand, our ultimate goal is to obtain a value which indicates the current stance of a member regarding a tighter (*hawk*) or looser (*dove*) monetary policy formulation (i.e. higher or lower target nominal interest rate). This procedure will be developed in two steps: first, the correlation between FOMC minutes numerical representations and nominal target *fed funds* will be extracted using *machine learning* tools, revealing those set of bi-grams and *topics* that tend to appear more in periods of higher or lower nominal rates. Next, we apply these correlation coefficients to speeches numerical representations themselves - ending up with sort of a "fitted value" for each individual statement - our *Policy Stance Index* (PSI), as explained in chapter 2. We also point out that this approach was highly based on (24), but with some modifications of our own.

4.2.1
Step I - Estimating Coefficients

Our primary data matrix consists of 347 characteristics (columns) - the aggregation of bigrams count and *topics* shares per document - for 159 points in time (rows), yielding a classical *high dimension* econometric problem: not enough degrees of freedom since the number of regressors exceeds (by a large amount) the number of observations. To deal with this issue, we make use of the *Elastic Net* operator, introduced in (41). In general, a *shrinkage* operator is designed to filter relevant covariates in high dimension environments like this by introducing a "penalty" function $\mathcal{F}(\beta)$ in a least squares operation:

$$\hat{\beta^*} = \arg\min_{\beta^*} \sum_t (\hat{u}_t - \beta^{*\prime} X_t)^2 + \mathcal{F}(\beta^*) \tag{4-1}$$

The *Elastic Net* operator introduces a double-step penalization process. First, for a given set of penalization parameters (λ_1, λ_2), the $\hat{\beta}(naive)$ is estimated:

$$\hat{\beta^*}(naive) = \arg\min_{\beta^*} \sum_t (\hat{u}_t - \beta^{*\prime} X_t)^2 + \lambda_2 \sum_{j=1}^{p} \beta_j^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| \tag{4-2}$$

Then, the final estimator vector $\hat{\beta}$ is normalized by λ_2:

$$\hat{\beta} = (1 + \lambda_2)\hat{\beta}(naive) \tag{4-3}$$

It is common to work with a linear combination of λ_1 and λ_2 in this literature, transforming the problem in:

$$\hat{\beta}^* = \arg\min_{\beta} \sum_t (\hat{u}_t - \beta' X_t)^2 + (1-\nu)\lambda \sum_{j=1}^{p} \beta_j^2 + \nu\lambda \sum_{j=1}^{p} |\beta_j| \tag{4-4}$$

With ν set at 0.5 (conservative approach)[3]. As a model selection tool, the *Elastic Net* operator is expected to produce as output only the coefficients for the most relevant bi-grams and topics used to "justify" such decisions, setting all others to zero.

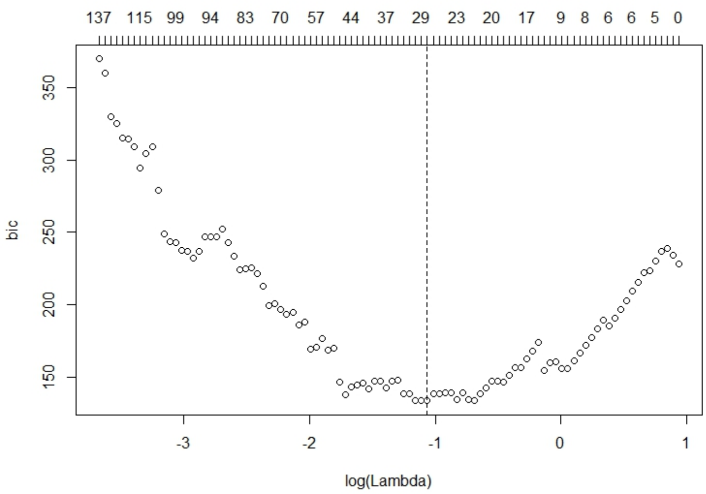
The main equation to be estimated is then:

$$i_t = \alpha_0 + \psi' z_{f,t} + \gamma' \hat{\theta}_{f,t} + \epsilon_t \tag{4-5}$$

Where i_t is the main nominal target interest rate set by the FOMC at meeting t, ψ is the vector of all bi-grams related coefficients, which are contained in $z_{f,t}$; γ is the vector of coefficients related to all *topics distributions* per document, which are contained in $\hat{\theta}_{f,t}$, and ϵ_t is an error term. The subscript f denotes that these covariates come from FOMC statements.

One of its main challenges though is how to find the "optimal" penalty term, λ. There are several techniques proposed in the literature. We work with the *Bayesian Information Criterion* (BIC), which has proven to be very useful in environments with short time series compared to the number of covariates. The final value for λ is set to 0.346, with a total of 29 variables between topics and bigrams chosen as relevant by this machine learning tool, with the constant included. Figure 4.3 plots, for each λ, the corresponding BIC value[4].

Figure 4.3: *Bayesian Information Criterion* (BIC) for different values of λ. The values under x-axis indicate the $\log\lambda$, while the numbers above the graphic indicates the number of variables included for each λ value. The BIC is minimized at $\lambda = 0.346$, at 133.9.



[3]Other specifications were tested in scales of 0.1 but no major improvements were detected
[4]More detailed information and theory in (41)

The main values estimated in this step are the non-zero coefficients $\hat{\psi}$ and $\hat{\gamma}$. The intuition behind them is that they compound an important part of the arguments used to justify decisions for higher or lower interest rates by the monetary policy committee, in terms of sequences of words and time spent covering specific topics. As an example, imagine the Fed thinks it is appropriate to raise interest rates in a given meeting when it sees an "*increasing inflation*" risk in its horizon. With our proposed method, we hope to find relevant coefficients both for this *bigram* and topics related to price level changes.

Here we find important to notice our method advantage over simple *word counting* methods - one of the most common procedures in text mining literature. In general lines, *word counting* techniques follow the *topic and tone* approach - text parts (usually sentences) are pre-selected when they cover a determined *topic*, and then the *tone* is measured by counting the number of times certain words comes up. These words are usually pre-defined and divided in two lists, a *"positive"* and a *"negative"* one, and to them some arbitrary values are assigned - generally, "positive" words get a value of 1, while "negative" ones gets a score of -1. Finally, summing across all relevant sentences *tone* values in a document will result in its *word counting index* or *sentiment index*.

There are two main improvements given by our methodology when compared to *word counting*. First, instead of pre-selecting *"positive"* and *"negative"* words, therefore running the risk of leaving important ones out (or bringing insignificant ones in), we let the algorithm decide which terms will be included and which ones will be left out. Second, while *topic and tone* methods usually apply the same values for different words, our technique attribute different weights for different sentences formulation, which seems a more realistic feature of text processing. For an example, "*slow growth*" and "*plummeting growth*" are both bigrams that communicate a significantly different idea, and probably would get the same value under a *word counting* approach. under these terms, our approach is less *subjective* than other common techniques in the literature.

4.2.2
Step II - Fitting Speeches

Once the coefficients $\hat{\alpha_0}$, $\hat{\psi}$ and $\hat{\gamma}$ are estimated, we build for the *individual speeches* the same data matrix built for the FOMC statements - with the *bigrams* count per document as well as the *topics* coverage per speech, re-estimated exactly like we did to get FOMC minutes distributions from paragraphs distributions. Finally, we run the following vector multiplication:

$$PSI_{m,s} = \hat{\alpha}_0 + \hat{\psi}' z_{m,s} + \hat{\gamma}' \hat{\theta}_{m,s} \tag{4-6}$$

Where $PSI_{m,s}$ denotes the *Policy Stance Index* for member m in the speech s, while $z_{m,s}$ and $\hat{\theta}_{m,s}$ are the same as defined in chapter 4.2.1 (but now for speeches). Assuming that FOMC members follow a somewhat standard procedure when speaking about their views on the economic situation, these newly created PSI scores may satisfactorily capture member's m position about the appropriate policy to be followed at the time discourse s was given.

Finally, there is need to aggregate these PSI values between meetings, to end up with a single y_k value for each member present at given meeting k, as explained in chapter 2. We do that by taking a simple average of all member's m PSI values for speeches issued between meetings $k-1$ and k. If there were no individual statements during this period, we repeat the last observed value. If it is also not available, we take the last observed PSI for the member that preceded the current member's position (common in cases where we find vacancies at regional chairs and the local vice-president assumes for one or two meetings). Lastly, if any of the above applies and there is still a missing value, we complete with the cross-sectional PSI average available at meeting k.

5
Results

5.1
Policy Stance Index

Figure 5.1 shows both the true nominal rate and *Elastic Net* fitted time series. With only 0.21 *out-of-sample* MSE, this specification shows a great fit, and although we have set our *in-sample* data at the 110-th observation, the result is fairly robust when we vary this parameter both towards a smaller and a greater expand.

Figure 5.1: *Elastic Net* operator performance in fitting the target nominal interest rate and our numerical representations, as shown in C-1. Shaded area indicates *out-of-sample* region.

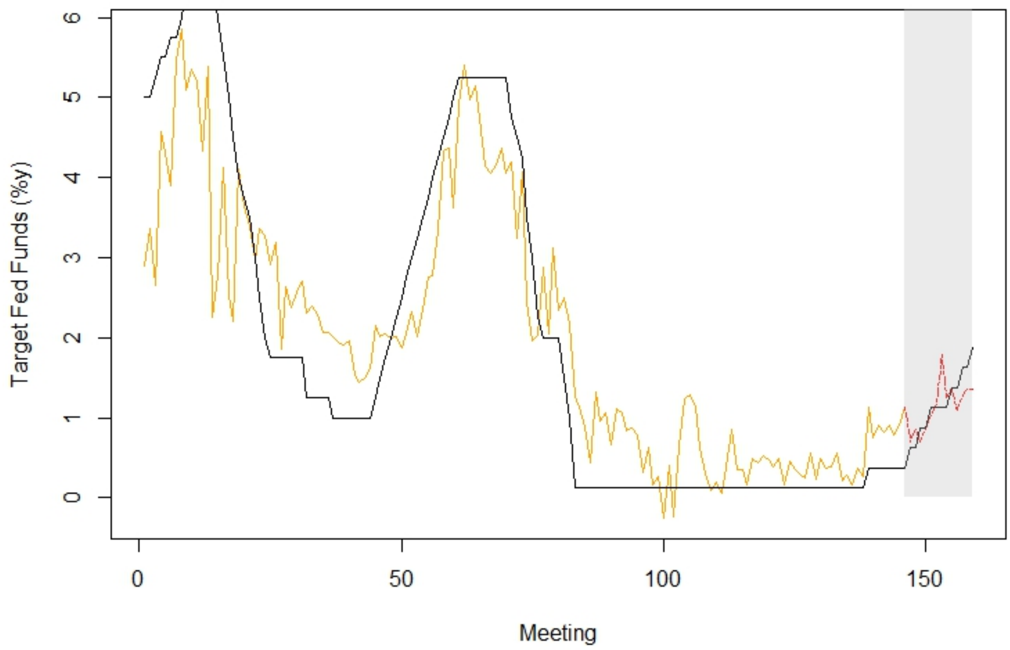


Table 5.1 shows the *Elastic Net* selected *topics* and *bi-grams*, along with their estimated coefficients and the intercept, α_0. The topics are represented by

letter T followed by their respective number - since they come from a random soft clustering device, there is no natural order or labeling for them.

Table 5.1: *Elastic Net* selected *topics* and *bi-grams*, in their stemmed versions. Coefficients set to zero means that the variable was considered as relevant but its magnitude is tiny, in absolute value (below 10^{-15}).

Variable	coef.	Variable	coef.
T16	-4.3214	T34	22.4199
T22	-2.6327	T10	17.4348
employ avail	-0.4943	T2	7.1905
foster maxemploy	-0.4541	(Intercept)	1.45734
longinflatexpect stabl	-0.3941	heighten inflatpressur	0.9421
fed employ	-0.0000	T1	0.6923
T21	0	inflatpressur forese	0.00008
T27	0	T3	0
T47	0	T4	0
T49	0	T5	0
T60	0	T6	0
econgrowth inform	0	T7	0
far growth	0	T8	0
growth next	0	T9	0
recent longinflatexpect	0		

It is worth spending some time analysing Table 5.1. First, consider the *negative* topics, 16 and 22, which *word clouds* are represented in Figure 4.1. Topic 16, the one with the most negative coefficient, is formed by combinations of words like *mortgage*, *holdings*, *securities* and *reinvestment* - a vocabulary widely used around the 2008 financial crisis, and a period during the nominal interest rate reached the *zero lower bound*. Topic 22 also seems to be related with lower rates - it somewhat contain words which relate to a *"low"* state of the economy, with other that indicate a *"gradual"* or *"stable" "trend"* or *"period"*.

Switching to the *positive* topics, the same subjective rationality can be inferred. Topic 34 shows the greater coefficient value, and clearly relates to one of dearest topics for any central banker - *inflation*. Closely behind, topic 10 displays terms such as *"growth"*, *"economic"*, *"production"* and *"demand"*, which can be linked to economic activity matters without greater effort. Together, these 2 topics shows relevant information (with the correct coefficient sign) about the Fed main objectives given by its institutional dual mandate - price stability and maximum employment. Even topic 2, with a relative smaller value, also holds information about something related to higher rates, thanks to the word *"increase"* and other terms such as *"inflation"* or *"financial conditions"*.

Even though few were selected, our relevant *bi-grams* also seems to be in the correct direction. Higher inflation usually asks for a higher interest rate, which are correctly captured by *"heighten inflation pressure"* and *"inflation pressure foreseen"*. On the other hand, *"foster maximum employment"* relates to the need of more activity, which in turn is related to lower nominal rates - a similar thought may be done with *"employment available"* and *"long-term inflation expectations stability"*.

Although there are lots of other *bi-grams* that may represent an important part in policy decision but are not captured individually, it is important pointing out that these words may be as well somewhat present inside our *topics* structure, since they can be interpreted as linear combinations of all tokens in our text *corpus*.

The discrepancy between *topics* and *bi-grams* coefficients magnitude may seem too high at a first glance, but we note that, while *bi-grams* variables takes integer values, since they measure the frequency with which word sequence appears in each document; the *topics* variables take values between 0 and 1, since they represent a "share" of each of the 62 topics contained in a text, as defined in chapter 4.1.

As explained in chapter 4.2.2, these coefficients are then applied to the same variables re-calculated at each individual speech. Table 5.2 shows some descriptive statistics about the PSI for each of the six covered years in our sample, along with the same observations regarding the *dot plot* for current year-end.

Table 5.2: *PSI* and *Dots* descriptive statistics for each year in our sample. For the *average* columns, *PSI* was taken as an average of all observations in each year's meeting, while an average of each meeting median was taken for the dots. As for the *Standard Deviation* columns, the proceeding was the same for the *PSI*, while an average of in-meeting *dots* dispersion across each year was taken for *dots* values.

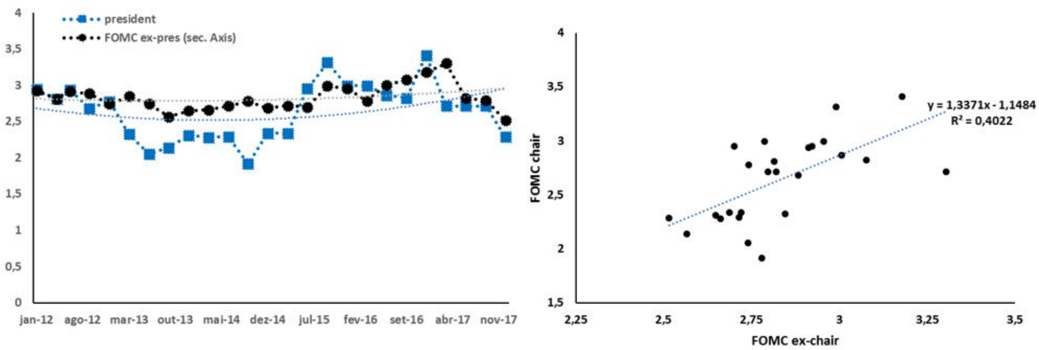
	Average		Std. Dev.	
	PSI	Dots	*PSI*	Dots
Year				
2012	2.853	0.198	0.551	0.192
2013	2.649	0.250	0.459	0.086
2014	2.678	0.187	0.561	0.139
2015	2.845	0.500	0.549	0.251
2016	3.013	0.750	0.629	0.158
2017	2.844	1.375	0.551	0.174
correl.		48,2%		42,2%

It is very hard to assess if we are in the right direction in capturing members *actual* monetary policy stance, since there is no numerical value we can directly compare them to. However, there are some empirical or anecdotal facts we can contrast with to check if our artificial series make some sense, at least.

First, the positive and somewhat high correlation both for average speeches and median *dots* and their cross-sectional standard deviations is cheering. There is a clear upward trend in both average measures, specially between 2013 and 2016 - period in which the U.S. economy gradually departed from the *zero lower bound* constraint. Although we recognize a clear *level* difference between both series, we argue that the *PSI* values contain not only information about the short-run, but also for longer-term monetary policy in speeches given, which may help to partially explain this discrepancy.

To evaluate our results, we can also think of the different approach adopted by any regular Fed member and the FOMC president. As reinforced in works such as (6), while usually FOMC members adopt a *collegial approach*, with relative freedom when talking about their individual views on the current state of the economy, the FOMC chair have a more *institutional* role since he/she speaks in the name of the whole Committee. Therefore, we hope the FOMC president PSI to correlate positively with the rest of the board. Figure 5.2 shows the time-series of both the president (Ben Bernanke until September 2013, followed by Janet Yellen) and an average of all other members present in each FOMC meeting covered in our sample, as well as the same information in the means of a scatter-plot. In fact, there is a clearly high positive correlation between both measures, of around 63%.

Figure 5.2: Correlation between the FOMC chair PSI and other members' average PSI. On the left, the time-series of both variables; on the right, a scatter-plot in which each point represent the chair PSI on the vertical axis and the PSI average of other members in that same meeting on the horizontal axis. Correlation is significantly positive at 63%.



Another clue comes from anecdotal evidence contained in financial blogs or newspaper. It is very common for big newspaper websites or blogs that cover financial markets to have their own views about which active FOMC members are more hawk, which are more dove, and which ones lie somewhere in between. Figures 5.3 and 5.4 are examples of that kind of information, extracted from big media vehicles. The first thing to notice is that, although there is a one-year window between both diagrams, there is little change in members positioning in these *hawk-dove* scales, leaning towards an expected not-swinging profile, specially for this short time scope. We also point out that this ordering is very common across many other similar market media channels, omitted here for concision, specially when it comes to the extremes (hawks and doves).

Figure 5.3: A private market website ordering FOMC members in a subjective *hawk-dove* scale in the beginning of 2017. Source: https://www.businessinsider.com.au/the-essential-guide-to-whos-a-policy-hawk-and-whos-a-dove-at-the-fed-ecb-boe-and-boj-2017-3

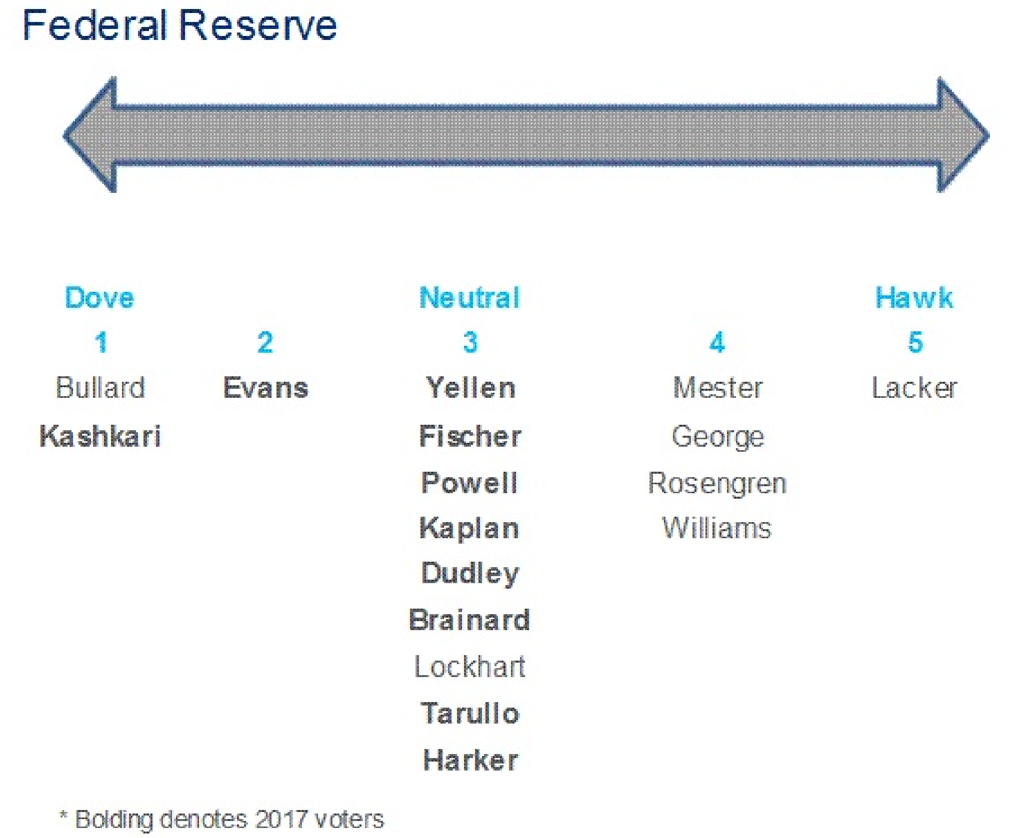
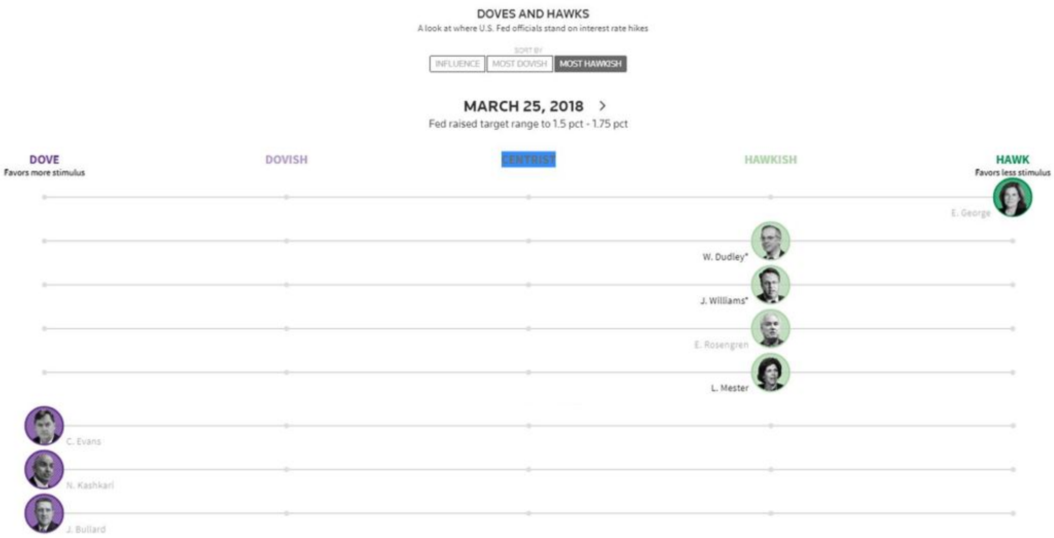


Figure 5.4: Another news website ordering FOMC members in a *hawk-dove* scale, this time for the first meeting of 2018. *Neutral* members omitted for concision. Source: http://fingfx.thomsonreuters.com/gfx/rngs/USA-FED/010030ZL253/index.html



We can compare this "subjective evidence" with our own generated PSI values for these years, for instance. Figure 5.5 compiles the average PSI for each member that took part in FOMC meetings in 2017, ordered from the "hawkest" to the "dovest" one. We paint in red the ones which appears in *hawkish* or *hawk* regions at least one time during this period, and in blue those who do so but in *dovish* or *dove* regions. This comparison also points out to a good performance of our quantifying techniques, specially when we focus on the members more prone to adopt a looser monetary policy stance - the three of them are precisely the ones with the lowest average PSI values the year of 2017.

Figure 5.5: PSI average values for each member that was part of the FOMC in 2017. Blue-colored names represents those which appears as most inclined to a looser monetary policy in media subjective scales, while red-colored ones represents those more inclined to higher rates.

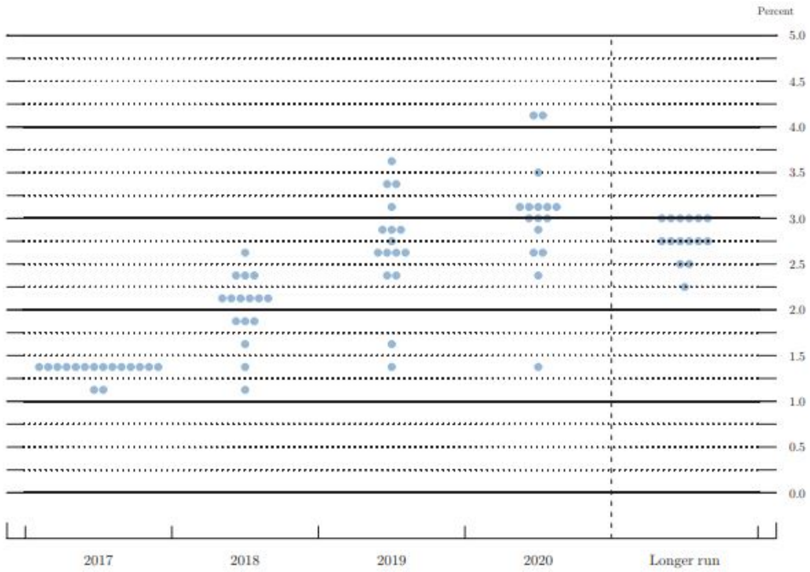
PSI - YEAR AVERAGE	
MEMBER	AVG 2017
tarullo	3,893
kaplan	3,544
williams	3,467
rosengren	3,415
fischer	3,043
mester	3,017
bostic	2,967
powell	2,934
harker	2,891
brainard	2,848
mullinix	2,800
yellen	2,606
george	2,594
dudley	2,588
lacker	2,411
bullard	2,407
evans	2,391
kashkari	1,826

5.2 Member-to-dot Tracking

Even though the true identity behind each *dot* will never become public, there is one meeting in which we know exactly in which interest rate level each member lies - our sample last meeting. Figure 5.6 shows the dot plot for December 2017 meeting, in which the Committee decided to raise the nominal rate range from $(1, 1.125)$ per cent to $(1.125, 1.25)$ per cent.

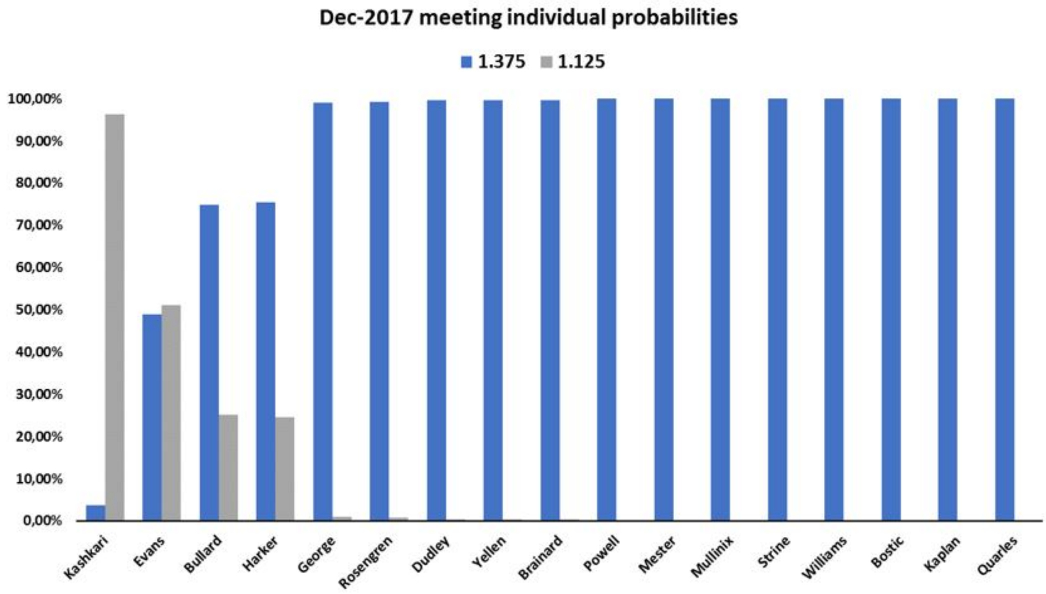
However, this was not an unanimous voting. According to the statement following the decision, there were two members that voted *against* this hike, preferring to maintain the nominal rate at that time current level - Mr. Charles Evans and Mr. Neel Kashkari. Since it was the last meeting of the year, and the dot plot brings information about individual assessments on the appropriate interest rate to be set at each year end, it is almost sure that the two under-consensus dots corresponds, in fact, to these to members.

Figure 5.6: December 2017 meeting interest rate dot plot. For the year-end frame (2017), two members indicated a lower-than-consensus level for the nominal rate, and exactly two members voted against the decision of hiking the interest rate by that time - namely, Mr. Charles Evans and Mr. Neel Kashkari.



With that information, we are now able to at least assess the performance of our algorithm in that specific meeting. If it manages to attribute higher probabilities to these members when labeling the under-consensus dot, we believe it is a good indicative. Figure 5.7 shows these *member-to-dots* probabilities.

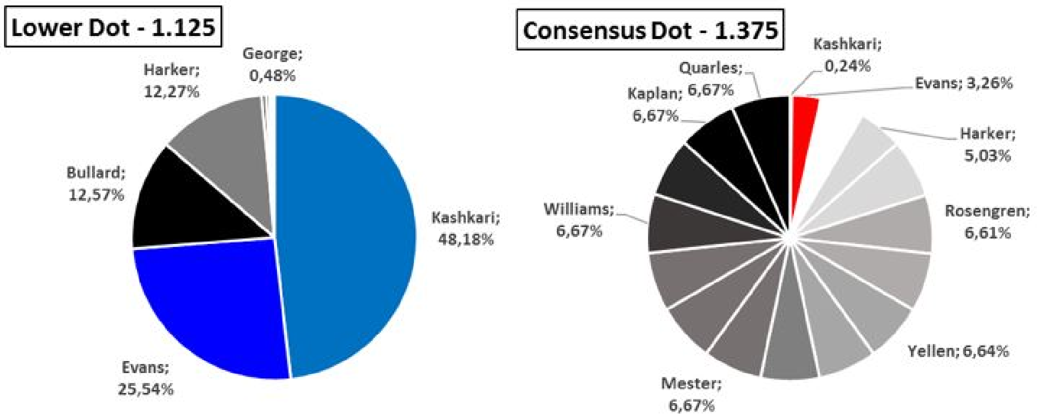
Figure 5.7: *Member-to-dot* individual probabilities - for each member, a double showing with how much chance he/she corresponds to the lowest or the highest December 2017 *dot*.



Taking in account that our algorithm is ran over only 25 meetings, with an average of 19 members per meeting - which sometimes gives us literally millions of possible associations between *dots* and members -, we find our algorithm performance strongly satisfactory under some criteria. It manages to consider the true *under consensus* members - Mr. Kashkari and Mr. Evans - as the most likely "low" *dots*, one of them with almost certainty - our estimated probability for Mr. Kashkari being the responsible for one of these low *dots* is almost 99%.

We also exercise the inverse thinking and present the *dot-to-member* probabilities for December 2017 FOMC below, obtained again by collecting all hypothesis' likelihoods in which each condition is true (i.e. *"dot 1.125 corresponds to member Y "*) and normalizing by the sum of all joint likelihoods. As expected, both members considered as the *true low dots* are in fact those with higher share of probability.

Figure 5.8: *Dot-to-member* individual probabilities. For each dot level, a *pizza-shaped* graphic in which each slice size is proportional to a probability of this dot corresponding to a given member.



The colored areas in each graphic (blue on the left and red on the right) are the two members in which we are most interested in. Now, analysing the *lower dot* probabilities, they configure the second and third most likely members to be positioned that way. As a comparison, the *naive* assignment probability - uniformly attributing equal chance for a dot correspond to each member - would be around 11% for the unnder-consensus region (two dots out of seventeen). Our tracking algorithm enhances this probability for the true answers, Mr. Kashkari and Mr. Evans, by more than four and two times, respectively.

6
Conclusions

"It is increasingly standard for central bankers, financial market participants, and academic researchers to describe management of expectations as central to monetary policy"[1], and under that framework, a decision making committee consensus level plays an important role. This paper addresses this issue by proposing techniques that may reduce uncertainty regarding a relevant source of heterogeneity amongst the FOMC: its composition between inflation fighting hawks and growth promoting doves. These procedures consist of first quantifying individual speeches in this monetary stance *spectrum*, then inputting those in a new target tracking algorithm to estimate full probability distributions over all possible associations between members and anonymous *dots*. In both scales our results point for a solid success under many criteria.

Our methodology leaves the door open for several extensions. Accurate measures for individual views on appropriate policy may help to investigate different impacts on market rates, in the spirit of (5), or even in the committee posterior decisions. Besides, we believe our tracking algorithm is general enough so it can be applied in settings other than the one presented here, such as other anonymous voting schemes. Finally, our proposed method for quantifying text can be extended to a number of questions regarding not only central bank communication and its relation to economic variables or policy signalling, but also to other fields from economics in which there is interest in text mining.

[1](8)

Bibliography

[1] WOODFORD, M.. **Interest and Prices: Foundations of a Theory of Monetary Policy**. Princeton University Press, 2011.

[2] BLINDER, A. S.. **Monetary policy by committee: Why and how?** European Journal of Political Economy, 23(1):106 – 123, 2007. Central bank transparency and central bank communication.

[3] BORDO, M. D.; ISTREFI, K.. **Perceived FOMC: The Making of Hawks, Doves and Swingers**. NBER Working Papers 24650, National Bureau of Economic Research, Inc, May 2018.

[4] MALMENDIER, U. M.; NAGEL, S. ; YAN, Z.. **The Making of Hawks and Doves: Inflation Experiences on the FOMC**. CEPR Discussion Papers 11902, C.E.P.R. Discussion Papers, Mar. 2017.

[5] EHRMANN, M.; FRATZSCHER, M.. **Communication by central bank committee members: Different strategies, same effectiveness?** Journal of Money, Credit and Banking, 39(2/3):509–541, 2007.

[6] BLINDER, A. S.; EHRMANN, M.; FRATZSCHER, M.; DE HAAN, J. ; JANSEN, D.-J.. **Central bank communication and monetary policy: A survey of theory and evidence**. Journal of Economic Literature, 46(4):910–45, December 2008.

[7] BLATTNER, T. S.; CATENARO, M.; EHRMANN, M.; STRAUCH, R. ; TURUNEN, J.. **The predictability of monetary policy**. Occasional Paper Series 83, European Central Bank, Mar. 2008.

[8] KING, R. G.; LU, Y. K. ; PASTÉN, E. S.. **Managing expectations**. Journal of Money, Credit and Banking, 40(8):1625–1666, 2008.

[9] GERAATS, P. M.. **Central Bank Transparency**. Economic Journal, 112(483):532–565, November 2002.

[10] SIBERT, A.. **Is Central Bank Transparency Desirable?** CEPR Discussion Papers 5641, C.E.P.R. Discussion Papers, Apr. 2006.

[11] EHRMANN, M.; FRATZSCHER, M.. **Explaining Monetary Policy in Press Conferences**. International Journal of Central Banking, 5(2):42–84, June 2009.

[12] GÜRKAYNAK, R. S.; SACK, B. ; SWANSON, E.. **Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements**. International Journal of Central Banking, 1(1), May 2005.

[13] ANDERSSON, M.; DILLÉN, H. ; SELLIN, P.. **Monetary policy signaling and movements in the term structure of interest rates**. Journal of Monetary Economics, 53(8):1815 – 1855, 2006.

[14] JANSEN, D.-J.; DE HAAN, J.. **Were verbal efforts to support the euro effective? a high-frequency analysis of ecb statements**. European Journal of Political Economy, 23(1):245 – 259, 2007. Central bank transparency and central bank communication.

[15] ROSA, C.; VERGA, G.. **On the consistency and effectiveness of central bank communication: Evidence from the ecb**. European Journal of Political Economy, 23(1):146 – 175, 2007. Central bank transparency and central bank communication.

[16] STEKLER, H.; SYMINGTON, H.. **Evaluating qualitative forecasts: The fomc minutes, 2006–2010**. International Journal of Forecasting, 32(2):559 – 570, 2016.

[17] LUCCA, D. O.; TREBBI, F.. **Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements**. NBER Working Papers 15367, National Bureau of Economic Research, Inc, Sept. 2009.

[18] BHOLAT, D.; HANS, S.; SANTOS, P. ; SCHONHARDT-BAILEY, C.. **Text mining for central banks**. Número 33 em Handbooks. Centre for Central Banking Studies, Bank of England, 2015.

[19] BLEI, D. M.; NG, A. Y. ; JORDAN, M. I.. **Latent dirichlet allocation**. J. Mach. Learn. Res., 3:993–1022, Mar. 2003.

[20] MONIZ, A.; DE JONG, F.. **Predicting the impact of central bank communications on financial market investors' interest rate expectations**. In: Presutti, V.; Blomqvist, E.; Troncy, R.; Sack, H.; Papadakis, I. ; Tordai, A., editors, THE SEMANTIC WEB: ESWC 2014 SATELLITE EVENTS, p. 144–155, Cham, 2014. Springer International Publishing.

[21] HANSEN, S.; MCMAHON, M.. **Shocking language: Understanding the macroeconomic effects of central bank communication**. Journal of International Economics, 99(S1):S114–S133, 2016.

[22] KAWAMURA, K.; KOBASHI, Y.; SHIZUME, M. ; UEDA, K.. **Strategic central bank communication: Discourse analysis of the bank of japan's monthly report**. Journal of Economic Dynamics and Control, 100:230 – 250, 2019.

[23] HANSEN, S.; MCMAHON, M. ; PRAT, A.. **Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach**. The Quarterly Journal of Economics, 133(2):801–870, 2018.

[24] HANSEN, S.; MCMAHON, M. ; TONG, M.. **The long-run information effect of central bank communication**. CEPR Discussion Papers, (DP13438), 2019.

[25] GENTZKOW, M.; KELLY, B. T. ; TADDY, M.. **Text as data**. NBER Working Papers 23276, National Bureau of Economic Research, Inc, 2017.

[26] EGGERTSSON, G. B.; WOODFORD, M.. **The zero bound on interest rates and optimal monetary policy**. Brookings Papers on Economic Activity, 2003(1):139–211, 2003.

[27] WOODFORD, M.. **Methods of policy accommodation at the interest-rate lower bound**. Proceedings - Economic Policy Symposium - Jackson Hole, p. 185–288, 2012.

[28] MACHADO, O. P.; CARVALHO, C. ; BERRIEL, T.. **Lift-off Uncertainty: What Can We Infer From the FOMC's Summary of Economic Projections?** Technical report, 2015.

[29] BONGARD, M.; GALATI, G.; MOESSNER, R. ; NELSON, W.. **Connecting the dots: market reactions to forecasts of policy rates and forward guidance provided by the Fed**. DNB Working Papers 523, Netherlands Central Bank, Research Department, Oct. 2016.

[30] DETMERS, G.-A.. **Forward Guidance under Disagreement - Evidence from the Fed's Dot Projections**. SFB 649 Discussion Papers SFB649DP2016-041, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, Oct. 2016.

[31] BLINDER, A.. **The Quiet Revolution: Central Banking Goes Modern**. Arthur Okun Memorial Lectures Series. Yale University Press, 2008.

[32] JR., H. W. C.; MCGREGOR, R. R. ; VERMILYEA, T. A.. **Power-Sharing in Monetary Policy Committees: Evidence from the United Kingdom and Sweden**. Journal of Money, Credit and Banking, 46(4):665–692, June 2014.

[33] SMALES, L. A.; APERGIS, N.. **The influence of FOMC member characteristics on the monetary policy decision-making process**. Journal of Banking & Finance, 64(C):216–231, 2016.

[34] LUO, W.; XING, J.; MILAN, A.; ZHANG, X.; LIU, W.; ZHAO, X. ; KIM, T.-K.. **Multiple object tracking: A literature review**. 05 2017.

[35] REID, D. B.. **"an algorithm for tracking multiple targets"**. 24:1202 – 1211, 02 1978.

[36] BAR-SHALOM, Y.; TSE, E.. **Tracking in a cluttered environment with probabilistic data association**. Automatica, 11(5):451 – 460, 1975.

[37] FORTMANN, T.; BAR-SHALOM, Y. ; SCHEFFE, M.. **Sonar tracking of multiple targets using joint probabilistic data association**. IEEE Journal of Oceanic Engineering, 8:173–184, July 1983.

[38] KONSTANTINOVA, P.; UDVAREV, A. ; SEMERDJIEV, T.. **A study of a target tracking algorithm using global nearest neighbor approach**. In: PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE CONFERENCE ON COMPUTER SYSTEMS AND TECHNOLOGIES: E-LEARNING, CompSysTech '03, p. 290–295, New York, NY, USA, 2003. ACM.

[39] PORTER, M. F.. **Readings in information retrieval**. p. 313–316, 1997.

[40] GRIFFITHS, T. L.; STEYVERS, M.. **Finding scientific topics**. Proceedings of the National Academy of Sciences, 101(suppl 1):5228–5235, 2004.

[41] ZOU, H.; HASTIE, T.. **Regularization and variable selection via the elastic net**. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67(2):301–320, 2005.

A
Target Tracking Algorithm: Revisiting the JPDAF

This section digs a bit deeper in the theoretical framework behind the Dot Tracking Algorithm developed in Section 2. For that, we believe it is worth revisiting some concepts from the *Joint Probabilistic Data Association Filter* (JPDAF hereafter) presented in (37) at a first moment to clarify how we get to 2-8, which dictates how to calculate the probability for a given joint association hypothesis.

Bearing in mind the dim difference between what is defined as an event in each framework[1], as well as the possibility of "false measurements" (*clutter*, measurements emitted by a source which is not of tracking interest) in the JPDAF setting, this filter is built over the same hypotheses from the *Probabilistic Data Association Filter* (36):

I The probability density of a measurement conditioned upon past data and given it is correct, $p(y_k^j|\chi_k^j, Y^{k-1}) \equiv f(y_k^j|Y^{k-1})$, is assumed to be available;

II The density of a measurement given that it is incorrect is uniform in the validation region whose volume is denoted by V_k, i.e. $p(y_k^j|\chi_k^j, Y^{k-1}) = V_k^{-1}$;

III No inference can be made on the number of validated returns from past data;

IV The probability of each return being correct, conditioned on past data, is the same, i.e. no target signature information is used.

We shall also define some extra notation. The *measurement association indicator* indicates whether measurement j is associated with any established target in event χ, or if it is assigned as *clutter*:

$$\begin{aligned} \tau_j(\chi) &\equiv 1; \quad \text{if } t_j > 0; \\ \tau_j(\chi) &\equiv 0; \quad \text{if } t_j = 0 \end{aligned} \tag{A-1}$$

[1]In ours, we fully observe measurements time-series and observe not-linked values for states. In the JPDAF, the state is fully latent, while measurements are partially observed through time.

On the other hand, the *target detection indicator* indicates if target t is detected in hypothesis χ:

$$\begin{aligned} \delta_t(\chi) &\equiv 1; \quad \text{if } t_j = t \text{ for some } j; \\ \delta_t(\chi) &\equiv 0; \quad \text{if } t_j \neq t \text{ for all } j; \end{aligned} \tag{A-2}$$

We can now proceed like done in (2-8) and compute the joint probability using Bayes' Rule:

$$\begin{aligned} P\{\chi|Y^k\} &= P\{\chi|y_1, ..., y_m, m_k, Y^{k-1}\} \\ &= p(y_1, ..., y_m|\chi, m_k, Y^{k-1})P\{\chi|m_k, Y^{k-1}\}/c \end{aligned} \tag{A-3}$$

The first factor in A-3 is *joint probabilistic density of the candidate* m *measurements*, conditioned on event χ:

$$p(y_1, ..., y_m|\chi, m_k, Y^{k-1}) = \prod_{j=1}^{m} p(y_j|\chi_j^{t_j}, Y^{k-1}) \tag{A-4}$$

For them, a *Gaussian approximation* is assumed since the computation of the exact state density is a costly sum of Gaussian densities, while a conservative uniform density prior over the surveillance region is assumed for those measurements to which no target is assigned:

$$\begin{aligned} p(y_j|\chi_j^t, Y^{k-1}) &= \mathcal{N}(y_j^t; \hat{y}^t, S_t); \quad \tau_j(\chi) = 1; \\ p(y_j|\chi_j^t, Y^{k-1}) &= V^{-1}; \quad \tau_j(\chi) = 0. \end{aligned} \tag{A-5}$$

The right-hand side term in (A-3) is the prior probability of a joint event up do time k, and it is a function only of the number of targets in period k, m_k, and the number of false measurements assigned to each hypothesis - since the candidate measurements weren't observed yet, there is no further value to make any inference about. (37) shows that this term may be written as it follows:

$$P\{\chi|m_k, Y^{k-1}\} = \frac{\phi(\chi)!}{m_k!} \prod_{t:\delta_t=1} P_D^t \prod_{t:\delta_t=0} (1 - P_D^t) \frac{e^{-CV}(CV)^{\phi(\chi)}}{\phi(\chi)!} \tag{A-6}$$

Where P_D^t denotes the detection probability of target t, the number of false measurements $\phi(\chi) \equiv \sum_{j=1}^{m}[1 - \tau_j(\chi)]$ is assumed Poisson distributed with parameter CV, where C is the density of false measurements and V the total volume searched.

After some algebra, (A-5) and (A-6) together yields:

$$P\{\chi|Y^k\} = \frac{1}{c} \prod_{j:\tau_j=1} \mathcal{N}(y_j^t; \hat{y}^t, S_t) \prod_{t \in D(\chi)} P_D^t \prod_{t \in ND(\chi)} (1 - P_D^t) \tag{A-7}$$

Where c again is a normalizing constant. Finally, We are able to obtain the probability β_j^t that measurement j belongs to target t. We just need to sum over *all* feasible events χ for which this condition is true:

$$\beta_j^t = \sum_{\chi} P\{\chi|Y^k\}\hat{\omega}_j^t(\chi); \quad j = 1, ..., m; \ t = 0, ..., T. \tag{A-8}$$

$$\beta_0^t = 1 - \sum_{j=1}^{m} \beta_j^t \tag{A-9}$$

We can go back to the *dot tracking problem* now. Note that, although very similar and following the same rules of movement described in (2-1), the uncertainty sources between both problems are slightly different. While in our main problem we completely observe measurements (y_k) along time and the not-linked and unmatched states values (x_k), the JPDAF works in a setting in which only unmatched measurements values are observed. In the end, however, this will be of minor importance, since in both cases events are defined in a generic manner. This way, since in our framework an event associates a target to a state (which are observed only at the end of period k), the computation of $P\{\chi|m_k, X^{k-1}, Y^k\}$ follows the same rationale in (A-6). Moreover, most of the terms contained in it are algebraically canceled as showed in (A-10). Replacing assumption (A-5) with (2-7), one gets:

$$P\{\chi|Y^k, X^k\} = \frac{1}{c} \prod_{j:\tau_j=1} \mathcal{N}(x_k^j; \hat{x}_{k|k}^j, \Sigma_k) \prod_{t \in D(\chi)} P_D^t \prod_{t \in ND(\chi)} (1 - P_D^t) \tag{A-10}$$

Finally, given the data treatment explained in chapter 3.2.1, since the number of *dots* and present members will be always the same, the Committee participants will always be "detected". This means the term P_D^t above is equal to 1 for every member t and the set $ND(\chi)$ is empty, therefore leading to equation (2-8), rewritten here for convenience:

$$P\{\chi|Y^k, X^k\} = \frac{1}{c} \prod_{j:\tau_j=1} \mathcal{N}(x_k^j; \hat{x}_{k|k}^j, \Sigma_k)$$

B
Robustness Checks - Target Tracking Algorithm

As mentioned in chapter 2, the target tracking algorithm proposed in this paper is sensitive to some variables which may impact its overall performance in a simulated environment. To shed some light on these questions, this chapter brings some robustness checks. First of all, the PSI variance term σ^r measure the precision of our quantified speech value around the true policy stance of member m. Greater values for σ^r are therefore associated to less precise signals, which tend to jeopardize our tracking performance.

Figures B.1 and B.2 repeats the results shown in Figures 2.4 and 2.5, but now with the term σ^r increased from 1 to 4 - something like 3 or 4 times what is estimated for real data, scales considered. As expected, the overall performance falls, but is still able to catch (on average) the true hypothesis as the most likely one.

Figure B.1: 1000 paths average for joint hypotheses probabilities, $\sigma^r = 4$.

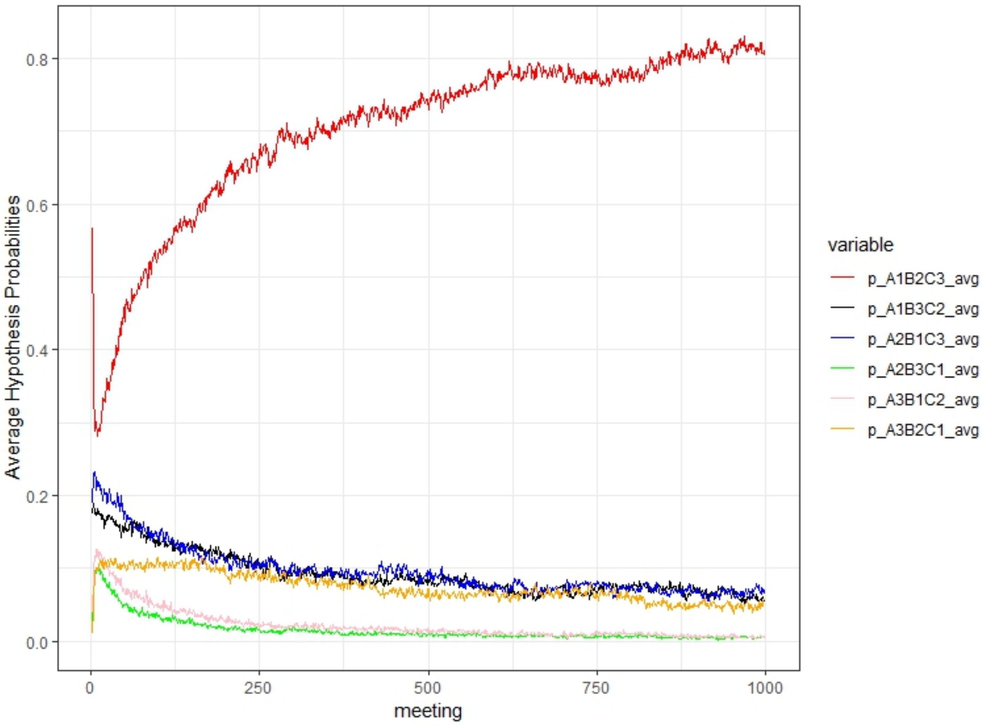
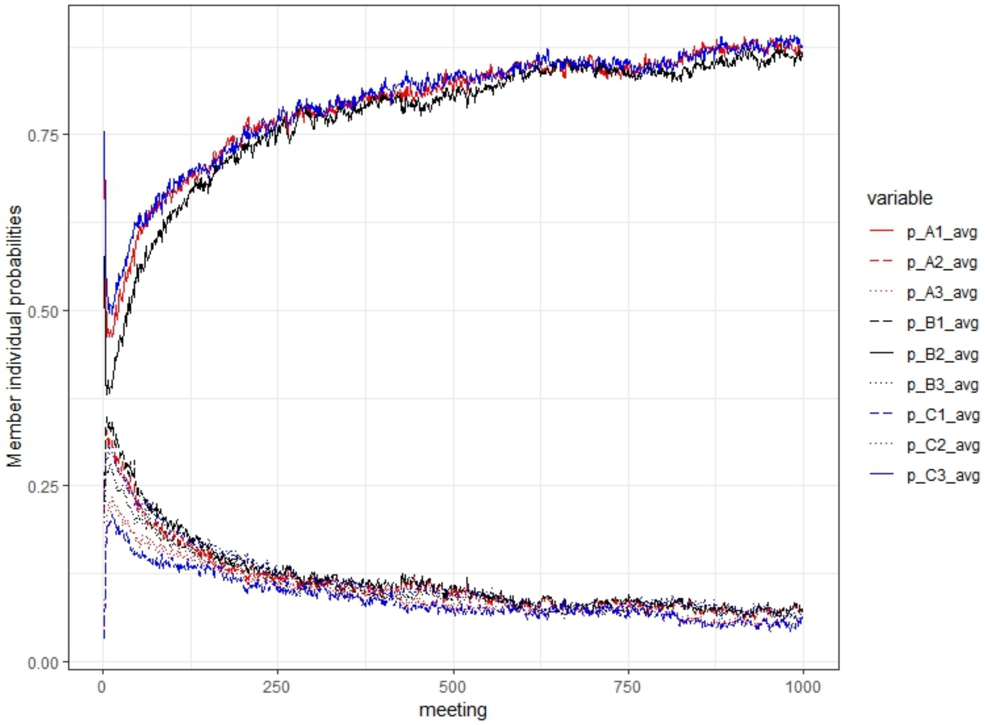


Figure B.2: 1000 paths average for individual association probabilities, $\sigma^r = 4$.



Another source of uncertainty may also come from our arbitrary initial prior on the values for $\tilde{x}_{0|0}$. Again we repeat the previous experiment, but now with totally wrong initial association priors. Figures B.3 and B.4 shows that the experiment results are barely unchanged, except for the very beginning periods.

Figure B.3: 1000 paths average for joint hypotheses probabilities, $\sigma^r = 1$, wrong initial prior.

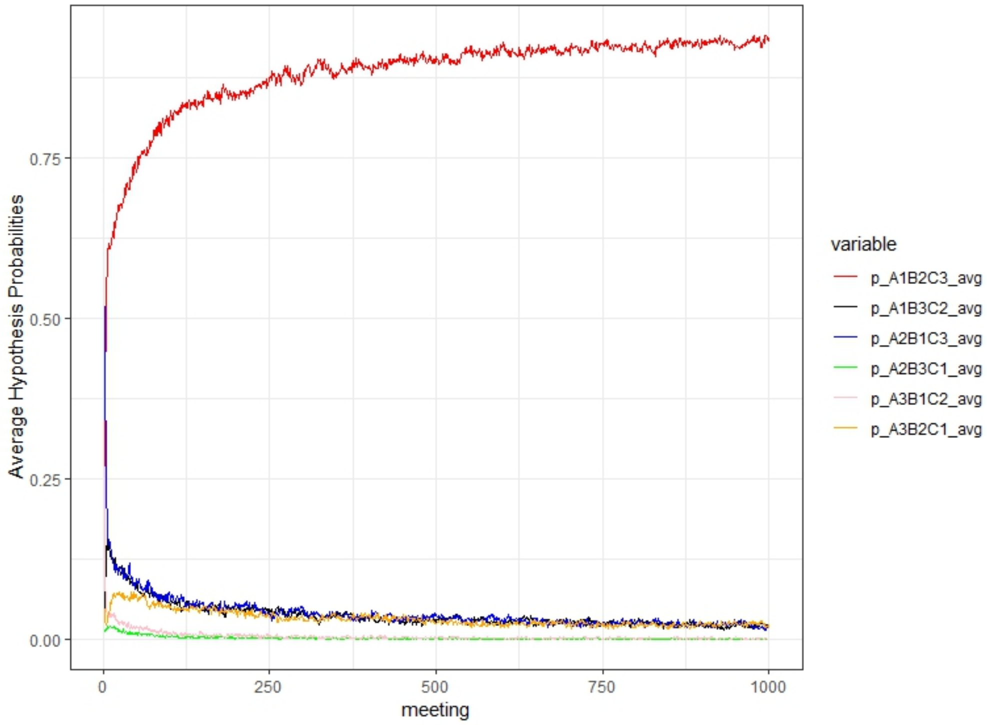
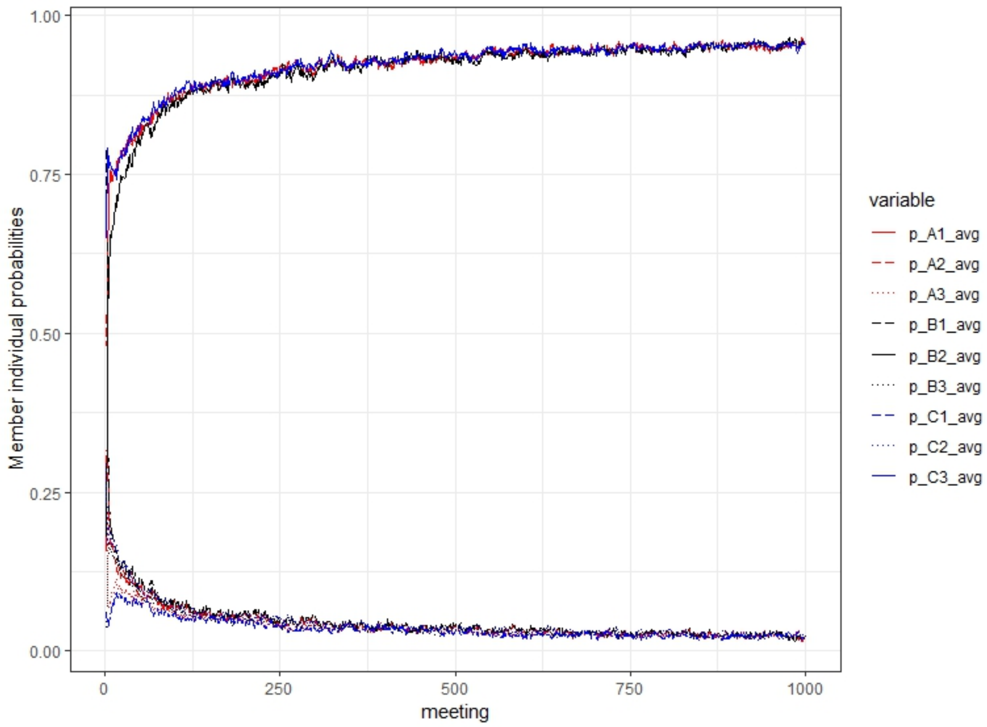


Figure B.4: 1000 paths average for individual association probabilities, $\sigma^r = 1$, wrong initial prior.



Finally, we combine both test above in the same simulations, displaying the results in Figures B.5 and B.6. We recognize the performance fall specially in the first periods can be harmful, due to the short spam we have in our real data exercises. However, we reinforce the argument that the variance term R is much closer to the experiments proposed in chapter 2 than the ones presented here, and also that our routine keeps showing a learning behavior and good results in longer terms.

Figure B.5: 1000 paths average for joint hypotheses probabilities, $\sigma^r = 4$, wrong initial prior.

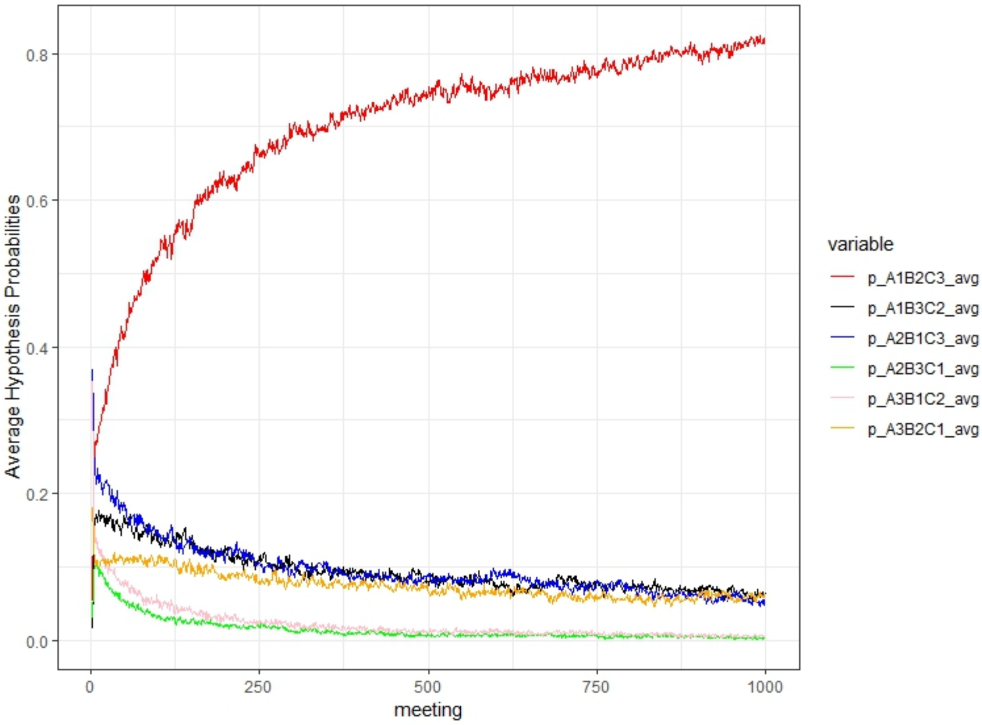
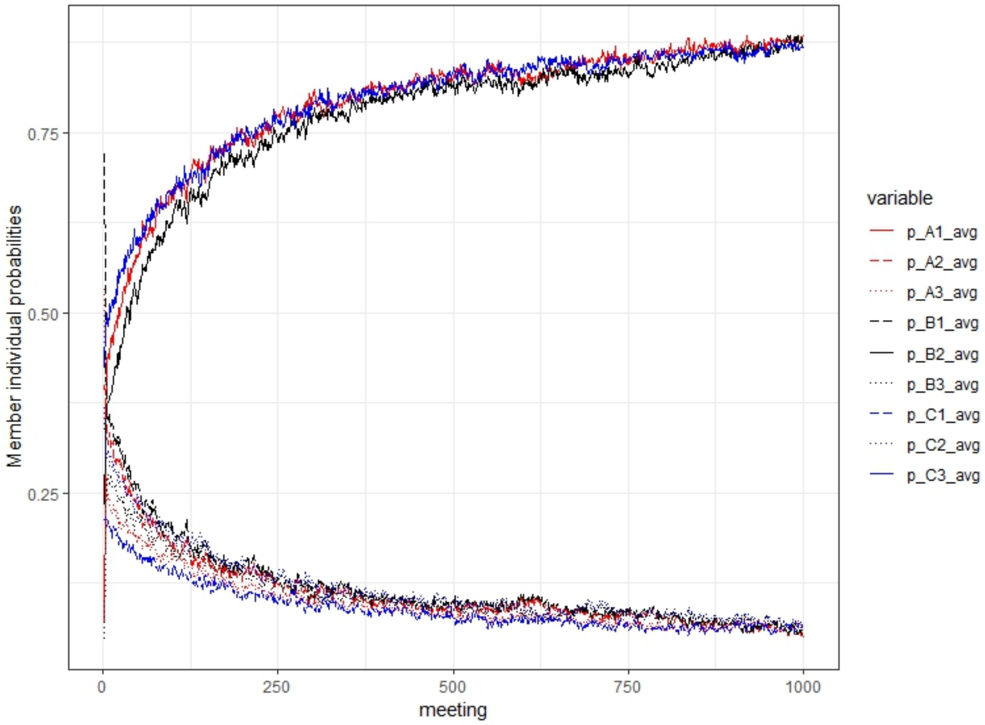


Figure B.6: 1000 paths average for individual association probabilities, $\sigma^r = 4$, wrong initial prior.



C
Alternative Specification

This chapter shows a different specification for our text quantifying technique, which in the end is also able to satisfy some of our desirable criteria, including considering at least one of the right under-consensus member as the most likely one in December 2017 meeting. The discrepancies lie on the source text used in our quantifying first step (FOMC minutes instead of statements), the shrinkage tool used (Ridge operator instead of Elastic Net) and also some parameters from LDA application, like the T number of topics and the threshold for rare words.

The FOMC minutes started to be regularly published after each meeting in May 1995, yielding us 35 more observations to our time series. Moreover, it contains a much more detailed discussion about the economy overview, with few changes in its structure along time. Specifically, we work with the same numeric representations as explained in chapter 4.1 contained between the section "Staff Review of the Economic Situation" and the section which holds each member vote. However, we note our feature space grows exponentially, both for the bi-grams matrix and the *tdm* used in LDA application, in a way even the dimension reduction tools may have trouble in handling such a large amount of data.

We run the LDA over this text corpus (again segregated at the paragraph level) with 30 topics and a threshold of 10 documents as a minimum for each bi-gram to appear to be considered in our used data. We also set an "upper-bound" to exclude very common words, and set this limit at 170 minutes (out of a total of 194). We then repeat chapter 4.2.1, but instead of using the Elastic Net operator, we use the *Ridge* operator, characterized by the following loss function:

$$\hat{\beta}^* = \arg\min_{\beta} \sum_t (\hat{u}_t - \beta' X_t)^2 + \lambda(\beta'\beta) \tag{C-1}$$

With the term λ again calculated by the *Bayesian Information Criterion.* Figure C.1 shows its value for each λ, while Figure C.2 brings the fitted value for the Ridge step applied over the nominal rate set at each meeting, both in-sample and out-of-sample.

Figure C.1: *Bayesian Information Criterion* (BIC) for different values of λ. The values under x-axis indicate the $\log \lambda$, while the numbers above the graphic indicates the number of variables included for each λ value. The BIC is minimized at $\lambda =$.

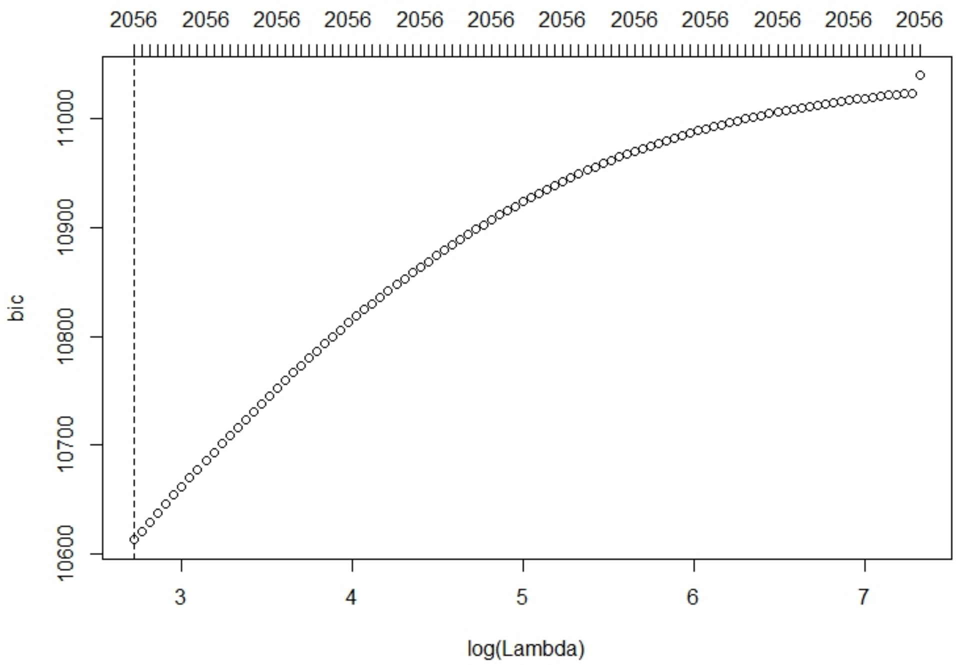


Figure C.2: *Ridge* operator performance in fitting the target nominal interest rate and our numerical representations, as shown in C-1. Shaded area indicates *out-of-sample* region.

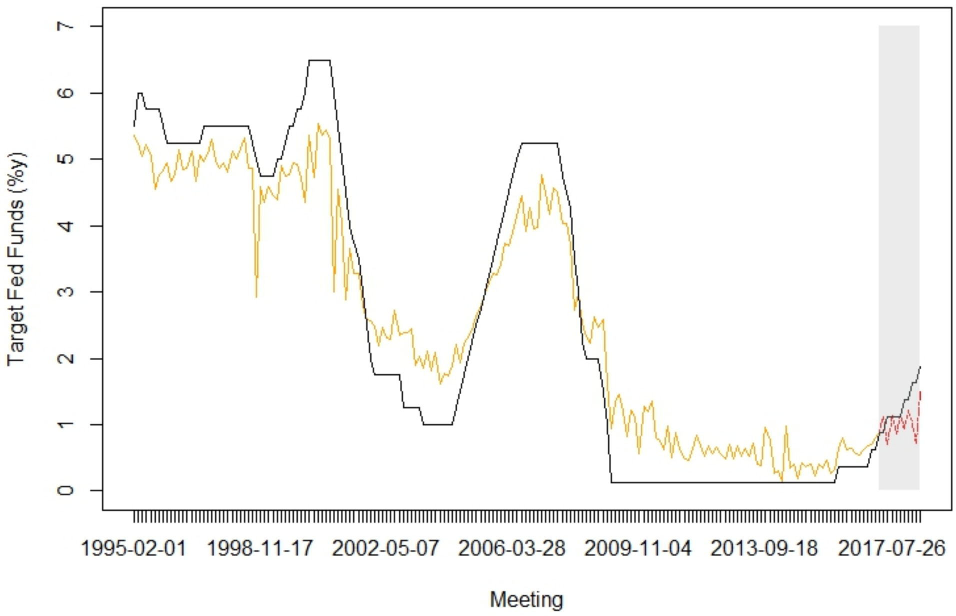


Table C.1 shows the topics and bi-grams considered as relevant by our algorithm. Note that, again, there is a clear intuition behind them, specially when interpreting the selected topics - note that, for instance, topics related to increasing prices and demand expansion are still selected as the most positives. All word clouds are presented below, in Figures C.3 and C.4.

Table C.1: *Ridge* selected *topics* and *bi-grams*, in their stemmed versions. Coefficients below 10^{-15} were omitted for simplicity.

Variable	coef.	Variable	coef.
T25	-17.9650	T16	16.4745
T8	-17.4041	T15	12.6679
T10	-13.7067	T17	9.7100
T4	-13.6423	T30	3.8293
T1	-3.5843	(Intercept)	2.9406
presid inflat	-0.1463	condit contempl	0.3224
inflat inflatexpec	-0.1261	sustain econgrowth	0.2260
subdu inflat	-0.0835	realgdp unemploy	0.2229
ensur inflat	-0.0296		

Figure C.3: *Ridge* selected positive topics in alternative specification; the word size in the cloud is proportional to its relevance for the given topic.

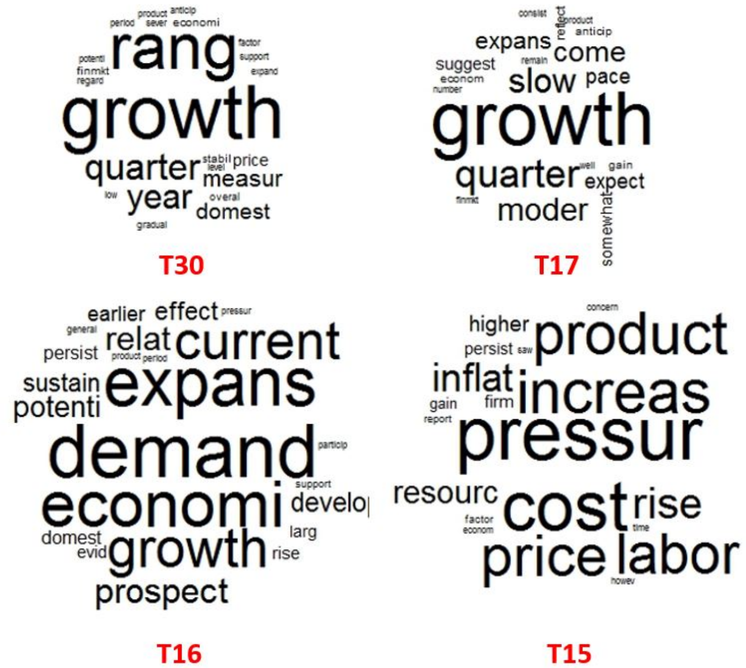
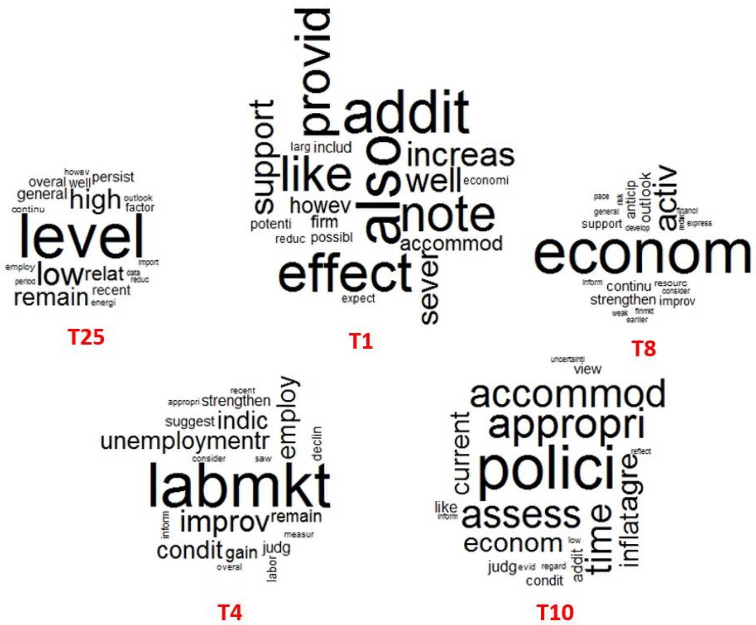


Figure C.4: *Ridge* selected negative topics in alternative specification; the word size in the cloud is proportional to its relevance for the given topic.



Our anecdotal experiment collected from media vehicles opinions is repeated in Figure C.5. In some measures, we note a slight improvement when comparing to our main specification, specially on the *hawk* end of the specter - 2 out of the 3 "hawkest" members identified by our technique were also done so by market analysts. The lower average PSI value of 1.50 for 2017 also suggests this technique is less biased to the upside than the original one.

Figure C.5: PSI average values for each member that was part of the FOMC in 2017. Blue-colored names represents those which appears as most inclined to a looser monetary policy in media subjective scales, while red-colored ones represents those more inclined to higher rates.

PSI - YEAR AVERAGE	
MEMBER	AVG 2017
williams	2,646
tarullo	2,543
lacker	2,227
harker	2,173
powell	2,125
fischer	2,020
mullinix	1,911
bostic	1,828
mester	1,791
rosengren	1,608
dudley	1,562
george	1,482
kaplan	1,143
evans	1,090
bullard	1,042
brainard	0,985
yellen	0,643
kashkari	-0,488

Finally, we present our final results for this alternative specification in Figures C.6 and C.7, like it was done in chapter 5.2. For 2017 last meeting, the aggregated performance falls, in a sense it is not recognizing Mr. Evans as one of the most likely lower dot. Still, it considers Mr. Kashkari with almost certainty, and Mr. Evans still poses as the third most likely member to be in that position.

Figure C.6: *Member-to-dot* individual probabilities - for each member, a double showing with how much chance he/she corresponds to the lowest or the highest December 2017 *dot*.

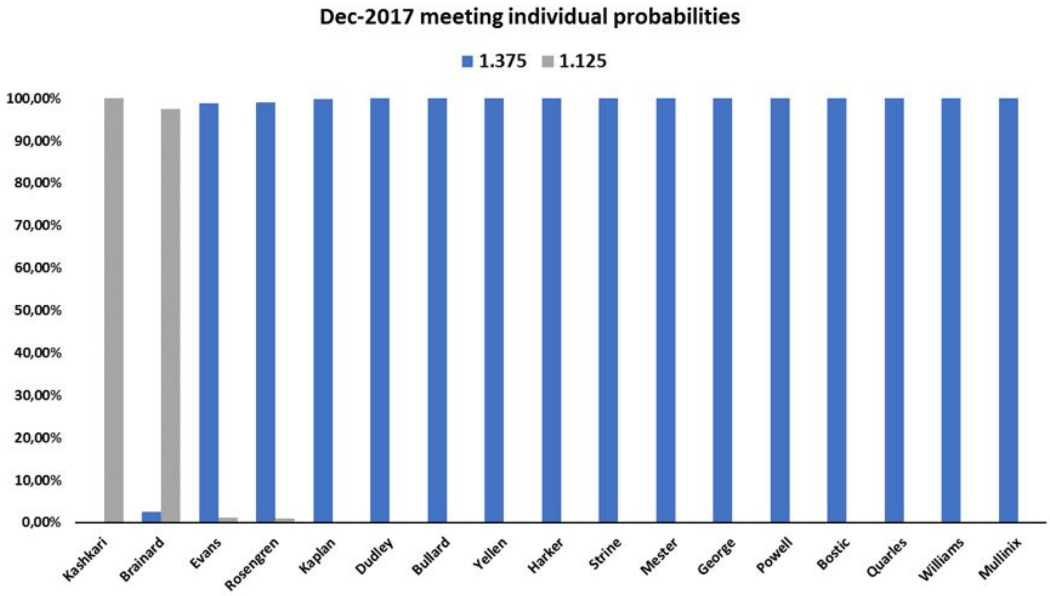


Figure C.7: *Dot-to-member* individual probabilities. For each dot level, a *pizza-shaped* graphic in which each slice size is proportional to a probability of this dot corresponding to a given member.