Mark My Words: The Transmission of Central Bank Communication to the General Public via the Print Media*

Tim Munday[†], James Brookes[‡]

Abstract

Central banks need to influence wage and price-setters' expectations to fulfill their objectives. Despite its importance, communication to the general public is far less studied than communication to financial markets. This paper posits that a key channel through which the general public receives central bank communication is through the print media. We examine which features of central bank text are associated with increased newspaper reporting of central bank communication. We write down a model of news production and consumption in which news generation is endogenous. We use our model to show that standard econometric techniques will likely (i) provide biased estimates and (ii) fail to deal with the high-dimensionality of the estimation problem. We use computational linguistics to measure the the extent of news coverage a central bank communication receives, and the textual features that might cause a communication to be more (or less) likely to be considered newsworthy. We apply our model to the case of the Bank of England, and utilise machine learning techniques designed for high-dimensional equations to estimate the relationship between news coverage and central bank communication. We find that the interaction between the state of the economy and the way in which the Bank of England writes its communication is important for determining news coverage. Content and the state of the economy on their own do not seem to have an effect on news coverage. We provide concrete suggestions for ways in which central bank communication can increase its news coverage by improving readability in line with our results.

*We are extremely grateful to Michael McMahon for his guidance and support with this project. We also thank our discussant Diego Calderon for insightful feedback. This paper benefited from comments from John Muellbaeur, David Bholat, Julian Aswhin, Max-Sebastian Dovi, and seminar participants at the Bank of England, University of Oxford, and the Warwick 2020 PhD Conference.

[†]University of Oxford, Bank of England

[‡]Bank of England

1. Introduction

Central Bank communication to the general public is important. The control of household and firm expectations, particularly when operating monetary policy at or close to the effective lower bound, can in theory be a powerful tool for modern central bankers. In addition, central banks also have a democratic obligation to speak to the general public (Binder 2017a). Central bank power, legitimacy, and in many cases independence, are granted by the public under the tacit agreement that the central bank remains accountable. As a result, communication to the public has been part of a growing body of research since Blinder, Ehrmann, Fratzscher, De Haan, and Jansen (2008) and Blinder (2009).

The provision of information to households and firms can have real effects on the economy. Randomised Controlled Trials (RCTs) where a treatment group is given information about an economic variable (typically inflation), show significant effects in the survey respondents expectations and subsequent actions in households (Coibion, Georgarakos, Gorodnichenko, and Van Rooij 2019; Kryvtsov and Petersen 2020) and firms (Coibion, Gorodnichenko, and Kumar 2018; Coibion, Gorodnichenko, and Ropele 2020).

However despite the RCT evidence that communication can have substantial effects on the economy as a policy tool, direct communication from central banks to the public has had an "abysmal track record" of influencing expectations (Coibion, Gorodnichenko, Kumar, and Pedemonte 2020) despite central banks' best efforts. Consumers and firms (i) know little about the central bank and their objectives (Van der Cruijsen, Jansen, De Haan, et al. 2015; Haldane and McMahon 2018), (ii) pay little attention to inflation dynamics in low-inflation environments (Candia, Coibion, and Gorodnichenko 2020; Cavallo, Cruces, and Perez-Truglia 2017), (iii) don't react to monetary policy announcements (D'Acunto, Hoang, and Weber 2020), and (iv) rarely if ever read monetary policy reports or other forms of direct communication (Kumar, Afrouzi, Coibion, and Gorodnichenko 2015).

Nonetheless, survey evidence (Haldane and McMahon 2018; Bholat, Broughton, Ter Meer, and Walczak 2019), and theoretical analysis (Haldane, Macaulay, and McMahon 2020) has suggested that altering aspects of communication can overcome some of these problems of communication to the general public.

Given the lack of *direct* engagement with central bank communication, reaching the public via the print media has become a route of interest for central banks looking to communicate. There is evidence that the media's interpretation of central bank communication can move financial markets (Hayo and Neuenkirch 2012; Hendry 2012), and that professional forecasters predominantly rely on media reports to process central bank news (Hayo and Neuenkirch 2015). Turning to communication with the general public, Blinder and Krueger (2004) find that TV and newspapers are the two top sources for economic information for the general public, and Larsen, Thorsrud, and Zhulanova (2020) provide evidence that news topics are good predictors of households inflation expectations. But the survey evidence is somewhat mixed as to how effective central bank communication is at altering consumer beliefs when filtered through the news media (Coibion, Gorodnichenko, and Weber 2019; Lamla and Vinogradov 2019).

We ask: what features of central bank text are associated with greater reporting of that text in the news? To the extent that we can, we try to design our approach to make these associations have a causal interpretation.

We develop a model in which the central bank, the newspapers and agents in the economy produce and consume news (respectively). Our model does not presume that text generation by the central bank is exogenous. The central bank is forward looking and anticipates the effect on the news cycle of its communication. In this sense, we draw on the work of Gentzkow and Shapiro (2010), who find that the demand of consumers drives media slant. In our model the state of the economy also affects consumers preferences for news. The model is sequential and is solved under perfect and complete information.

The model serves to illustrate that estimating the relationship between central bank communications and their coverage in newspapers is a high-dimensional inference problem, compounded by highly complicated relationships between variables. This motivates our estimation procedure.

Before we can estimate any relationships between communication and the news coverage it receives, we need to measure the variables in our model. We measure: the proportion of a newspaper article that is paraphrased or strongly influenced by central bank text, and a vector of textual features of central bank communication.

We measure the first of these using an event-study methodology. More specifically, we take every newspaper article in two approximately one and half day windows around 1211 Bank of England communication events, that contain the words 'Bank of England'. We then calculate the weighted change in the document-term matrices of the news between the two windows where the weights are related to how similar the newspaper articles are to the Bank of England's own communication. This provides us with a measure of how much the news flow has changed as a result of a central bank communication. To calculate this weighted similarity measure, we use a combination of word2vec (Mikolov, Sutskever, Chen, G. S. Corrado, and Dean 2013) — an embedding based approach based on a shallow neural network — and soft cosine similarity (Sidorov, Gelbukh, Gómez-Adorno, and Pinto 2014).

We measure the vector of textual features of central bank communication that could potentially alter its news coverage using a pipeline of computational linguistics. Our vector of features has three main components: topics, linguistic processing features, and newsworthiness features. Topics are measured using a dictionary method derived from the tags given by Guardian journalists to articles in the business section of their newspaper. Linguistic processing features and newsworthiness features are measured using a large range of different annotation methods, primarily based on sophisticated computational linguistic toolkits.

Our paper creates a comprehensive range of measures (351 in total), drawing on literature from journalism studies, psycholinguistics, computational linguistics, and economics, to determine what makes events receive of news coverage.

The model solution implies a extremely high number of features to estimate parameters for — 4695 in total. Indeed this high dimensionality problem is common to many textual analyses. To deal with the dimensionality issue and (approximately) retain the unbiased estimates that traditional econometric estimation methods have, we use the desparsified LASSO (Van de Geer, Bühlmann, Ritov, Dezeure, et al. 2014; Adamek, Smeekes, and Wilms 2020). This allows us to design an estimation procedure that, to the extent that we can, permits us to treat the associations that we uncover between the textual features of central bank communication, and the news coverage that central bank communication receives, as causal, whilst simultaneously performing feature selection.

We find that the *interactions* between the state of the economy all three categories of textual feature (content, linguistic processing, and newsworthiness) are important for explaining the pass through of central bank communication to the mainstream media. We find that the state of the economy on its own has no significant impact on the news coverage of central bank communication, and that only a small number of textual features on their own are significant.

Furthermore, we find that it is the variance of economic variables that, when interacted with textual features of central bank communication, influence the news coverage of central bank communication. It seems that the second moment of the economy drives how the content and style of central bank communication affects news coverage.

On the textual side, after performing feature selection on our 351 possible drivers of newsworthiness, we find five main categories of features are significant in explaining the news coverage that central bank communication receives, and derive five policy implications from them. These are:

1. Keep things simple. Our results show that one should avoid introducing embedded clauses and particle verb structures.
2. Personalize the text. Use we/us/you to engage the reader.
3. Write in short sentences. Long dependence arcs reduce the likelihood of newspaper coverage.
4. Summarise your message in the first sentence.
5. Use facts and figures.

Each of these recommendations is produced by accurately measuring the relevant features of the Bank of England communications and their media coverage. We believe that applying the above suggestions to central bank communication will improve news coverage and ultimately help central banks reach a wider audience.

Our paper predominantly builds on two different literatures. The first examines the role of central bank communication using text analysis. Much of this literature centers around how central bank communication, appropriately measured via various natural language processing techniques, affects financial market variables and the real economy. S. Hansen and McMahon (2016) examine how the FOMC statements impact the US economy through a FAVAR, S. Hansen, McMahon, and Tong (2019) analyse how Bank of England inflation report topics are related to high-frequency movements in financial markets, Hendry and Madeley (2010) investigate the impact of Bank of Canada statements on returns and volatility, Ehrmann and Talmi (2020) suggest that Bank of Canada statements that are less similar to previous releases cause more market volatility, and Born, Ehrmann, and Fratzscher (2014) examine central bank communication on financial stability and its effect on the stock market. A smaller literature examines central bank communication in the media. For example, Hendry (2012) and Hayo and Neuenkirch (2012) supplement the text of the Bank of Canada with subsequent market news reports to determine which basic features of both pieces of text move financial markets.

Other papers examine the role of the media in transmitting central bank communication rather than purely financial market implications, but generally stay within the bounds of whether or not the central bank is perceived positively or negatively in the media. Berger, Ehrmann, and Fratzscher (2011) use a manually labelled dataset from ECB staff of how favourably the media report ECB monetary policy decisions, and find that decisions that have large informational content and those that have been preceded by large numbers of statements gain the most favourable coverage.[1] Lamla and Sturm (2013) perform a

[1] Applied to the case of South Africa in Reid, Du Plessis, et al. (2011)

similar analysis (using a dataset labelled manually by a private company) to investigate how expectations of future monetary policy decisions portrayed in the media are affected by interest rate decisions. Rybinski et al. (2019) use dictionary methods to generalise these manual approaches and apply a similar analysis to the case of the Polish central bank. Binder (2017b) is an exception to the focus on favourability and uses a manually coded dataset from PEW to determine whether communication events influence the prominence of the Federal Reserve (and its chair) in the news. Our paper goes much further than the current literature, explicitly modelling the news coverage process, creating a far richer measure of news coverage of a communication using an embedding based approach, and finally forming over 4000 features (351 textual and 11 economic and various polynomials and interactions between them) which could potentially cause a communication to be newsworthy and examining which features are important.[2]

The second strand of literature analyses features that may influence news coverage of an event. Galtung and Ruge (1965)'s seminal paper examining news articles on crises in four Norwegian newspapers spawned a subfield of journalism studies aimed at developing taxonomies of so-called "news values" (e.g. Bednarek and Caple 2017; Harcup and O'neill 2017). Through detailed qualitative textual analysis, this field of research has shown that certain properties of an event can make it more or less newsworthy. For instance, bad-news events that affect many people close to home, such as a natural disaster or a terrorist attack in a country's capital city, are more likely to make the news than other types of story. On the other hand, however, there has been far less research using computational tools to study newsworthiness. A notable exception is the study of Piotrkowicz (2017), which computationally operationalises 6 aspects of newsworthiness to predict readership engagement with news headlines taken from The Guardian and New York Times. Building on this literature, inspired in particular by Piotrkowicz (2017)'s study, we develop annotation schemes for 9 dimensions of newsworthiness.

Intimately interwoven with the inherent newsworthiness of an event is the extent to which the text describing it can be successfully comprehended by the reader. Comprehensibility has recently been of interest to scholars in a wide range of economics sub-fields. For instance, Guay, Samuels, and D. Taylor (2016) analyse the obfuscating effects of financial statement complexity; Amadxarif, Brookes, Garbarino, Patel, and Walczak (2019) analyse the linguistic complexity of prudential regulation; Fullwood (2016) compares the readability of central bank communications with other genres of text. However, in general, researchers analysing economic texts for their readability have far too often used vastly simplified models of language comprehension. For example, it has been assumed that classical readability metrics—comprising minimal features such as word length and sentence length—can adequately model text complexity. But decades of research in the cognitive psychology of language has demonstrated that language processing is far more complex (see, for instance, the contributions in Rueschemeyer and Gaskell 2018). At the same time, research in computational linguistics has developed superior tools and techniques through which we might computationally model language comprehension (e.g., Gonzalez-Garduno and Søgaard 2017; Howcroft and Demberg 2017). Building on this recent research in computational readability and drawing on decades of research into the cognitive psychology of language, we design a suite of novel linguistic features across three core levels of linguistic comprehension — word access, sentence parsing, and discourse

[2] An interesting, although largely unrelated paper to ours, that uses news text around monetary policy meetings is Ellen, Larsen, and Thorsrud (2019), in which the authors compare the restricted document term matrix, projected down using SVD, between Norges Bank communications and the preceding articles, and use this as a measure of narrative monetary policy shocks.

integration — to better capture the reading experience.

Our paper makes three main contributions to the literature: (i) it is the first to structurally model central bank communication to the public via the print media, (ii) it develops a comprehensive feature set of potential newsworthy features that go well beyond naive approaches used before (e.g. Flesch-Kincaid scores, Haldane and McMahon (2018)), and (iii) we perform inference on our model that allows us to make policy recommendations for how central banks should alter their communication in order to reach a greater share of the population.

The paper is structured as follows. Section 2 outlines the model framework. Section 3 discusses the data sources. Section 4 details the computational linguistic approaches we use to measure the degree of news coverage a central bank communication receives. Section 5 details the features of central bank communication that may influence its newsworthiness, and how we measure them. Section 6 contains the estimation procedure used to create our estimates for what influences newsworthiness, and discusses the results and their policy implications. Section 7 concludes.

2. Framework

The framework presented below is a simple model of news production and consumption in a three agent world under perfect and complete information. The model contains three agents. The central bank, a representative newspaper, and a continuum of consumers.

Our model serves to illustrate that if one wants to estimate which textual features of the central bank's communication cause increased newspaper reporting that: (i) any reasonable way of measuring the textual features of the news text and the central bank text quickly leads to high a dimensionality issue in which the number of parameters to estimate exceeds the sample size, and (ii) the content produced by the central bank is a function of the desired news characteristics of consumers, the central bank's own objectives regarding communication to the public, and the state of the economy.

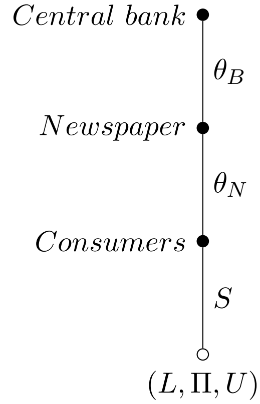
These two points are a result of the endogenous nature of central bank text production. The first leads us to use shrinkage methods to perform feature selection. The second means that when we do feature selection we will create omitted variable bias unless we explicitly account for the relationship between featurs. Consequently, we estimate the relationship given by our model equation using the de-sparsified LASSO (Van de Geer, Bühlmann, Ritov, Dezeure, et al. 2014). See Section 6 for the estimation procedure.

This paper explicitly recognises that central bank communication is endogenous and designed to reach consumers. Thus, central bank's must balance their desired message — which, for example, may be a complicated and heavily caveated one — with the desires of consumers for simpler messages.

Our model is parsimonious, and aims to accurately capture the salient features of central bank text production and consumption. It is not a model of information being released by a central bank, after which agents engage in updating their beliefs in order to fulfil some objective (Morris and Shin 2002). Rather it takes the literature on news consumption as its starting point (Gentzkow and Shapiro 2010), and then builds the role of the central bank and the newspaper to come to a full fledged model.

News is produced and consumed within the same period. There are three stages in the model. First, the central bank publishes content with characteristics described by vector θ_B. The representative newspaper produces news with characteristics θ_N comprised of a combination of central bank content and other news. Finally, share S of consumers decide to buy the newspaper if buying it would give them positive utility. All agents are rational and there is perfect and complete information. Since all information sets are singletons, we solve for a sequentially rational nash equilibrium in pure strategies using backward induction.

Fig. 1. Structure



2.1. Consumers

The consumer side of the model is influenced by Gentzkow and Shapiro (2010). Every day, a continuum of consumers choose whether or not to read the news. Gentzkow and Shapiro are concerned with one possible characteristic of the news: political slant. We want to study multiple characteristics of the news, so we extend their utility function such that θ is a vector of characteristics. A consumer, indexed c, has utility function:

$$U_c = \bar{u}_c - \gamma(\theta_N - \theta^*)^T W(\theta_N - \theta^*) + \epsilon_c \tag{1}$$

Where $\bar{u}_c$ represents some exogenous taste for news, θ_N is a measure of characteristics of the news, θ^* is a measure of the desired characteristics of the news (common to all consumers), W is a diagonal weight matrix (common to all consumers), and ϵ_c is an idiosyncratic taste shock.

Furthermore, we conjecture that consumers' desired characteristics of the news, θ^*, are dependent on the state of the economy, which we denote by vector z. Thus we can write $\theta^* = \theta^*(z)$. For ease of notation we will mainly use the simpler notation of θ^*, but it is worth remembering that these desired characteristics are not completely exogenous or time-invariant.

Household c consumes the news on a given day if $U_c \geq 0$. We assume, as in Gentzkow and Shapiro (2010), that ϵ_c is distributed iid uniform across households on an interval that includes the maximum and minimum values of $-\gamma(\theta_N - \theta^*)W(\theta_N - \theta^*)^T + \epsilon$. As a result, we can write the share of households consuming the news, S, as:

$$S = \delta - \gamma(\theta_N - \theta^*)^T W(\theta_N - \theta^*) + \epsilon \tag{2}$$

Where δ is a constant, and γ has been rescaled after having been integrated over ϵ_c.

2.2. Newspapers

Newspapers act in a competitive market, and are profit maximisers. We consider a representative newspaper. Furthermore, since we are only considering news related to monetary policy, we can think of the newspaper's problem as more accurately being the journalist's problem.

The journalist tasked with economic reporting has one choice variable: k. k represents the fraction of an article that directly paraphrases the central bank. The journalist faces a trade-off. Writing news that satisfies consumer desires will sell more papers, but requires effort. Paraphrasing the central bank may not align with consumer desires, but is costless.

An article produced by the newspaper is constructed as follows. Proportion k of the article is paraphrased central bank communication, and has characteristics equal to that of the central bank's communication (θ_B). Proportion $1-k$ of the article is created by the newspaper. Since the newspaper is profit maximising, proportion $1-k$ of the article will be exactly aligned with consumer desires (θ^*). [3]

The characteristics of the news content produced by the newspaper, θ_N, can be written:

$$\theta_N = k\theta_B + (1-k)\theta^* \tag{3}$$

The newspaper maximises profit, which can be written as:

$$\Pi = \lambda S - C(1-k) \tag{4}$$

Where C is a cost function which we assume to be quadratic: $C(1-k) \equiv \alpha(1-k)^2$. Subbing in for S from the consumer demand equation gives:

$$\Pi = \lambda(\sigma - \gamma(k\theta_B + (1-k)\theta^* - \theta^*)^T W(k\theta_B + (1-k)\theta^* - \theta^*) + \epsilon) - \alpha(1-k)^2 \tag{5}$$

Newspapers optimally choose k. Taking the first order condition with respect to k gives:

$$\frac{\partial \Pi}{\partial k} = -2\lambda\gamma k(\theta_B - \theta^*)^T W(\theta_B - \theta^*) + 2\alpha(1-k) = 0 \tag{6}$$

Rearranging for k gives a key model equation:

$$k = \alpha\left(\alpha + \lambda\gamma(\theta_B - \theta^*)^T W(\theta_B - \theta^*)\right)^{-1} \tag{7}$$

This equation states that the proportion of newspaper text directly paraphrasing the central bank communications, k, is high when the central bank releases text with characteristics (θ_B) that are close to consumer desires (θ^*).

2.3. Central bank

To incorporate that the central bank's production of content is not exogenous and may take into account consumer preferences, we model the central bank's problem as follows.

The central bank has some over-arching set of objectives that relate to monetary policy communication to the public. First amongst these objectives is the anchoring of inflation expectations at target. Other objectives may include civic engagement or influencing household and firm expectations of variables other than inflation.

Let's suppose that the central bank has a (weighted) quadratic loss function, L, over the deviations

[3] Note we don't allow the characteristics to offset.The newspaper cannot set the characteristics of the article in the part which they create as to cancel out the paraphrased part, and arrive at a news article that has characteristics exactly in line with consumer desires.

from this small vector of objectives, y, from their target values $\bar{y}$:

$$L = (y - \bar{y})^T H (y - \bar{y}) \tag{8}$$

Where H is a positive definite diagonal weight matrix.

The central bank's only instrument to achieve these particular objectives (it still retains its usual monetary policy tools to achieve other objectives) is the characteristics of the text it produces, θ_B. The objectives, y, are a function of the news the consumer receives, θ_N.

The central bank takes the gradient vector of its loss function with respect to its instrument.

$$\nabla L(\theta_B) = 2(y(\theta_N) - \bar{y})^T H J_{\theta_B}(y) = \vec{0} \tag{9}$$

Where $J_{\theta_B}(y)$ denotes the Jacobian matrix of y with respect to θ_B.

There are possibly many solutions to this set of equations. Our aim is only to convince the reader that the solution is that the cenrtal bank sets θ_B as a function of its objectives, the state of the economy, and consumer desired characteristics.

If one assumes that the central bank can achieve the global minimum of its loss function, i.e. it sets $y(\theta_N) = \bar{y}$ through its manipulation of its instrument θ_B, then the global minimum solution is detailed as follows. Other solutions that satisfy the first order conditions, but are not global minima are detailed in Appendix Section 8.1.

Assuming that the function which maps the news consumers recieve to the central bank's objectives is invertible we can write:

$$\begin{aligned} y(\theta_N) - \bar{y} &= 0 \\ \theta_N &= y^{-1}(\bar{y}) \end{aligned} \tag{10}$$

For ease of notation, we denote $y^{-1}(\bar{y})$ as θ_B^*, to represent the central bank's desired characteristics for the elements of θ_N.

$$\begin{aligned} \theta_N - \theta_B^* &= 0 \\ &= k\theta_B + (1-k)\theta^* - \theta_B^* \\ &= \alpha\theta_B + \lambda\gamma(\theta_B - \theta^*)^T W(\theta_B - \theta^*)\theta^* - \theta_B^* \end{aligned} \tag{11}$$

This is a quadratic equation in the vector θ_B. If it has a solution, the solution(s) for a given element of θ_B are:

$$\theta_{B,j}^{opt} = \theta_j^* + \frac{(\theta_{B,j}^* - \theta_j^*)\left(\alpha + 2\lambda\gamma w_j(\theta_{B,j}^* - \theta_j^*)^2 \pm \sqrt{\alpha^2 - 4\lambda\gamma(\alpha + \lambda\gamma w_j(\theta_{B,j}^* - \theta_j^*)^2)\sum_{i\neq j}(\theta_{B,i}^* - \theta_i^*)^2}\right)}{2\gamma\lambda\sum_i(\theta_{B,i}^* - \theta_i^*)^2} \tag{12}$$

This is a non-linear equation in which θ_B is a function of consumers' desired characteristics, θ^* and the central bank's desired characteristics, θ_B^*.[4]

[4]Clearly if Equation 12 holds $\forall j$, then $\theta_N = \theta_B^*$, and so the solutions are global minima of the loss function because they set $L = 0$. As a result, there is no need to check the second order condition, the only condition that must be met is that the

Intuitively, the optimal choice of characteristic j by the central bank, $\theta_{B,j}^{opt}$, is a weighted average between the central bank's own desired characteristic $\theta_{B,j}^{*}$ and the consumer's desired characteristic θ_{j}^{*}. The central bank must balance communicating exactly what it desires, and communicating in a way that will reach consumers. This is the fundamental trade-off of our model.

For notation, we rewrite the solution in Equation 12, or indeed the approximate solution if there is no exact one:

$$\theta_{B,j}^{opt} = \theta_{B,j}\left(\theta^{*}, \theta_{B}^{*}\right) \tag{14}$$

We can rewrite Equation 7 as:

$$k = \alpha\left(\alpha + \lambda\gamma\left(\theta_{B}\left(\theta^{*}, \theta_{B}^{*}\right) - \theta^{*}\right)^{T} W\left(\theta_{B}\left(\theta^{*}, \theta_{B}^{*}\right) - \theta^{*}\right)\right)^{-1} \tag{15}$$

Note that θ_B is observed by us, the researchers, but preferences θ^{*} and θ_{B}^{*} are not.

2.4. Desired Characteristics

The key model equation is Equation 15. We can rearrange this into an equation which can be estimated:

$$\begin{aligned} k &= \frac{\alpha}{\alpha + \lambda\gamma\sum_i w_i\left(\theta_{B,i}\left(\theta^{*}(z), \theta_{B}^{*}(z)\right) - \theta_{i}^{*}(z)\right)^{2}} \\ \frac{1-k}{k}\frac{\alpha}{\lambda\gamma} &= \sum_i w_i\left(\theta_{B,i}\left(\theta^{*}(z), \theta_{B}^{*}(z)\right) - \theta_{i}^{*}(z)\right)^{2} \end{aligned} \tag{16}$$

Where z is a vector of economic variables that affect the desired characteristics for economic news.

This is an equation which is linear in the w_i's: the diagonal elements of W. We assume that consumer desired characteristics θ^{*} for a given feature i, have the following relationship:

$$\theta_{i,t}^{*} = \bar{\theta}_{i}^{*} + \pi_{i}^{T} z_t \tag{17}$$

Where π_i is a $(M \times 1)$ coefficient vector which maps the state of the economy z to the desire for feature i.

A simple rational inattention model in which the consumer demand for news is isomorphic to their 'attention', would suggest the consumer desires in our model should vary linearly with the inverse prior variance of the variable in question (Sims 2003).

The elements of z used in estimation are specified in Section 3 and include both the level and variance of many characteristics of the economy.

Now our equation becomes:

---

solutions exist. i.e. that:

$$\alpha^{2} - 4\lambda\gamma(\alpha - \lambda\gamma w_j(\theta_{B,j}^{*} - \theta_{j}^{*})^{2})\sum_{i\neq j}(\theta_{B,i}^{*} - \theta_{i}^{*})^{2} > 0 \quad \forall j \tag{13}$$

Intuitively, this restriction can be understood as not allowing the preferences of the central bank, θ_{B}^{*} to stray too far from the preferences of the consumer, θ^{*}. Since we posit that both are driven by the state of the economy, z, this does not seem too far fetched.

$$
\begin{aligned}
\frac{1-k}{k}\frac{\alpha}{\lambda\gamma} &= \sum_i w_i \left(\theta_{B,i} - \bar{\theta}_i^* - \pi_i^T z\right)^2 \\
&= \sum_i w_i \left(\theta_{B,i}^2 + (\bar{\theta}_i^*)^2 + (\pi_i^T z)^2 - 2\theta_{B,i}\bar{\theta}_i^* - 2\theta_{B,i}(\pi_i^T z) + 2\bar{\theta}_i^*(\pi_i^T z)\right)
\end{aligned} \tag{18}
$$

In terms of observables, this is:

$$
\frac{1-k}{k} = \beta_0 + \beta_1^T(\theta_B^T\theta_B) + \beta_2^T\theta_B + \beta_3^T(z \otimes \theta_B) + \beta_4^T z + \beta_5^T(z \otimes z) \tag{19}
$$

Where we have defined:

$$
\begin{aligned}
\beta_0 &= \frac{\alpha}{\lambda\gamma}\sum_i w_i(\theta_i^*)^2 \\
\beta_1 &= \frac{\alpha}{\lambda\gamma}[w_0, w_1, ..., w_N] \\
\beta_2 &= -2\frac{\alpha}{\lambda\gamma}[w_0\bar{\theta}_0^*, w_1\bar{\theta}_1^*, ..., w_N\bar{\theta}_N^*] \\
\beta_3 &= -2\frac{\alpha}{\lambda\gamma}[w_0\pi_0^T, w_1\pi_1^T, ..., w_N\pi_N^T] \\
\beta_4 &= 2\frac{\alpha}{\lambda\gamma}[w_0\bar{\theta}_0^*\pi_0, w_1\bar{\theta}_1^*\pi_1, ..., w_N\bar{\theta}_N^*\pi_N] \\
\beta_5 &= \frac{\alpha}{\lambda\gamma}[w_0(\pi_0 \otimes \pi_0), w_1(\pi_1 \otimes \pi_1), ..., w_N(\pi_N \otimes \pi_N)]
\end{aligned} \tag{20}
$$

Adding in time subscripts gives the equation we wish to estimate:

$$
\frac{1-k_t}{k_t} = \beta_0 + \beta_1^T(\theta_{B,t}^T\theta_{B,t}) + \beta_2^T\theta_{B,t} + \beta_3^T(z_t \otimes \theta_{B,t}) + \beta_4^T z_t + \beta_5^T(z_t \otimes z_t) \tag{21}
$$

This equation relates the observable features of the text the central bank produces (θ_B) and the state of the economy (z), to the degree of reporting that central bank communication receives (k).

The inclusion of kroeneker products in Equation 21 causes the dimensionality to become unmanageable for standard econometrics once we start to include the full set of features that comprise θ_B and controls that comprise z: point (i) made at the beginning of this section. Point (ii), that the textual features of the central bank's communication are themselves functions of the state of the economy, can be deduced from Equation 12.

This model posited that the central bank was forward looking and thus its communication is a function of (i) consumer desired characteristics for news, (ii) the state of the economy via the central bank's own objectives, and (iii) the state of the economy via its impact on consumer demand for news. The model results in Equation 21; a linear equation between news coverage, textual features and the state of the economy.

What if news production has no demand side? Our model incorporates the demands of consumers for news as an important part of the central bank's (and the newspaper's) problem in determining the production of news. As a result of this, the central bank's communication is a combination of what the central bank wants to publish (θ_B^*) and what consumers want to read (θ^*). If consumers always bought the newspaper regardless of its content, then W becomes a matrix of zeros, k is always 1, the journalist does no work (she simply copies the central bank's communications into the newspaper), and the central bank can print its exact desired message (θ_B^*) knowing that it will reach all consumers verbatim. Patently, this is not how central bank communication works. Central banks have large public communication departments, research how to increase the penetration of their communication to the public through altering its form (Haldane and McMahon 2018), and think deeply about how to balance their desired communication with its palatability to a general readers.

Now we use the result of the framework just outlined, that news coverage of central bank communication (k) is a function of the communication itself (θ_B) and the state of the economy (z) through an equation such as Equation 21, to answer the question as to *which* features of communication or the economy matter for news coverage

The data and the techniques applied to that data to create the variables needed are detailed in Sections 3 and 4 and 5. The methodology used to estimate the β coefficients of Equation 21 is detailed in Section 6.

3. Data

3.1. Bank of England communications data

We study the communication of the Bank of England. The Bank of England communications data comes from a number of sources. Text data on the Introductory Statements and Inflation Reports are from S. Hansen, McMahon, and Tong (2019), the Q and A text data and Inflation Reports past 2015 are from Munday (2019), and the speech and minutes data were provided by Al Firrel at the Bank of England.

The speech data originally contained 771 speeches. After cleaning out speeches for which the text data was garbled or non-existent, we have 654 speeches.

The length of the series and number of communication instances are described in Table 1.

Table 1: Bank of England Communication Data

	Inflation Report	Q and A	Introductory Statement	Minutes	Speeches
First Observation	1998-02-11	2007-05-16	2001-02-14	1997-07-16	1997-06-12
Last Observation	2018-08-02	2018-08-02	2018-08-02	2019-05-02	2019-05-30
No. of observations	83	45	71	253	654
Total No. of words	1,682,165	311,511	29,301	1,475,554	2,386,576
No. of unique words	10,835	7,369	3,780	9,743	30,650

The intra-day timing of Bank of England communications varies over the sample. Figure 2 plots the intra-day publication times of Bank of England communications in our sample against the dates they were published. All timing data are from Bloomberg. We do not use intra-day timing data for speeches, namely because the time at which a speech is scheduled is often (i) inaccurate to when the speech was actually given, and (ii) not the same as when the speech text was released to a wider audience.

Both Table 1 and Figure 2 show how a large share of the communications corpus is comprised of the speeches and minutes. The other salient feature of Figure 2 is the move to “Super Thursday” in August 2015. This represents a stark break in Bank of England publication practices. Prior to August 2015, the Inflation Report, Q and A, Introductory Statement and Minutes did not coincide with the interest rate decision. From mid-2015 onwards, every interest rate decision is accompanied by publication of the Minutes, and every other interest rate decision is accompanied by the publication of both the Minutes and the Inflation report, with the Introductory Statement and Q and A following shortly afterwards.

Fig. 2. Timing of Bank of England communications

We analyse the Speeches, Minutes, Introductory Statements and Q and A's as separate communication events. For the Inflation Report, we analyse each section of the Inflation Report separately. We do this because (i) it reduces the length of the text of each chunk of communication text to a similar size, (ii) it allows us to determine which, if any, sections of the Inflation Report are influencing the news, (iii) the sections of the Inflation Report are labelled by content, and have changed over time (as explained below), analysing them individually accounts for these changes.

Table 2: Inflation Report Section Key

Section code	Feb 1998- Aug 2002	Nov 2002- Aug 2005	Nov 2005 - May 2015	Aug 2015 - Aug 2018
0	Overview	Overview	Overview	
1	Money and asset prices	Money and asset prices	Money and asset prices	
2	Demand and output	Demand	Demand	Demand and output
3	The labour market	Output and supply	Output and supply	Supply and the labour market
4	Costs and prices	Costs and prices	Costs and prices	Costs and prices
5	Mon. pol. since the prev. report	Mon. pol. since the prev. report		
6	Prospects for inflation	Prospects for inflation	Prospects for inflation	Prospects for inflation
7				Global economic and fin. developments
8				Monetary policy summary

Table 2 shows how the sections of the Inflation Report have evolved over time. In some cases, when the content is similar between sections despite the name of the section changing, we treat them as identical sections for our analysis (e.g. Demand vs Demand and Output). When the content is considerably different, we analyse a new section as a separate type of communication (e.g. Overview vs Monetary Policy Summary).

3.2. Newspaper data

The newspaper data are provided by Dow Jones. The data cover five major British daily newspapers: The Daily Mail, The Daily Mirror, The Guardian, The Sun, and The Times.[5] [6] Collectively these papers have a monthly physical circulation of 3,420,888, and account for 42% of the total circulation of the top 16 non-Sunday papers recognised by the Audit Bureau of Circulations and reported on by the Press Gazette.[7]. However, physical circulation does not account for (i) reaching consumers via online platforms, (ii) the proportion of the physical printed copies that are actually bought and read. Table 3 shows estimates by PAMCo of the total reach of the newspapers in our dataset. Combined, the newspapers in our sample have an estimated total monthly reach of 115,000 people, which is 47% of the estimated total monthly reach of all UK non-Sunday non-regional newspapers covered by PAMCo. [8]

[5]These data do not include the sister Sunday papers

[6]The dataset does not include papers that are aimed at a predominantly financial audience, such as City AM or the Financial Times. These papers have lower circulation, and do not transmit news to the general public in the way that we are concerned with studying here.

[7]Data from ABC. Accessed on 11/11/2019 https://www.pressgazette.co.uk/national-newspaper-abcs-guardian-sees-smallest-circulation-decline-for-july-2019/

[8]The data is constructed by combining a face-to-face survey of 35,000 adults aged 15 and above, direct measures of online audiences from comScore, and adjusting for duplicate readers who consume the news through both physical and digital mediums via a digital panel.

Table 3: Circulation of British Newspapers within our dataset

	Monthly Estimated Audience (000s)				
	Total	Phone	Tablet	Desktop	Print
The Daily Mail	24,775	17,026	2,415	4,123	6,398
The Daily Mirror	27,045	21,948	2,798	2,662	3,142
The Guardian	23,810	17,525	2,703	6,377	2,755
The Sun	32,438	25,399	3,276	3,413	7,135
The Times	7,629	3,793	709	1,102	3,285

Source: PAMCo 3 2019: Jul 18 Jun 19 (June 19 Comscore data).

3.3. Economic data

We include a broad range of economic variables to try and capture any economic state variables that affect preferences for economic news. The unemployment rate, the 10 year government bond rate and the CPI inflation rate are included to control for variables that may influence consumer demand for economic news. The the 1-year OIS rate and its daily change are included to control for news that is written in reaction to market movements. Appendix Section 8.4 shows regressions that suggest that news coverage is related to high frequency financial market variables, and so they are variables one would want to control for.

In addition we include the inverse variance (calculated over the previous year) of all of these variables, to control for preferences for news that take the forms predicted by the rational inattention literature (Sims 2003).

Data for all these variables except the 1 year OIS rate is taken from FRED. The 1 year OIS rate is taken from Bloomberg. The 1 year OIS rate only starts in the year 2000. So whilst the training of the word vectors can take place on the Bank's corpus extending back to 1997, the analysis performed on news coverage in Section 6 is based on the 2000-2018 period.

4. Measurement of media coverage

We want to estimate Equation 12. In this section we discuss the construction of k, a variable which measures news coverage of a central bank communication. This involves a novel event-study methodology that leverages natural language processing tools. In Section 5 we discuss how we measure the features of the text contained in θ_B and the elements of z.

4.1. Event study methodology

In our framework, k is the proportion of the news that is directly paraphrasing Bank of England communication. We could proxy this by manually coding a dummy variable as to whether a given central bank communication received news coverage. However, on the scale we wish to perform analysis, this is not possible. We have over 1000 communication events, and every newspaper article of five British newspapers (online and print) since 1997. To determine whether a given inflation report received news coverage, let alone whether or not it was the Inflation Report as opposed to the subsequent Q & A, would be very labour intensive. Thus we seek an automated approach. Not only does this save time, it allows us a more precise measurement of k, and permits our methodology to be easily applied to other research questions.

Using Natural Language Processing, we could measure k using the similarity between the news that reports on the Bank of England's communication, and said communication. However, this approach has two problems. Firstly, any external ongoing economic events that both newspapers and the Bank of England comment on will be picked up as newspapers reporting Bank of England content. Secondly, central bank communication that is written in a more journalistic style will again be picked up as newspapers reporting Bank of England content.

To remove these confounding factors, we calculate a measure of the communication *surprise* imparted by the Bank of England for each communication event, and use this as a proxy for k. In practice, this is the *change* in similarity between (i) the news the day before and the central bank communication, and (ii) the news the day after and the central bank communication.

A simple example follows to illustrate this point.

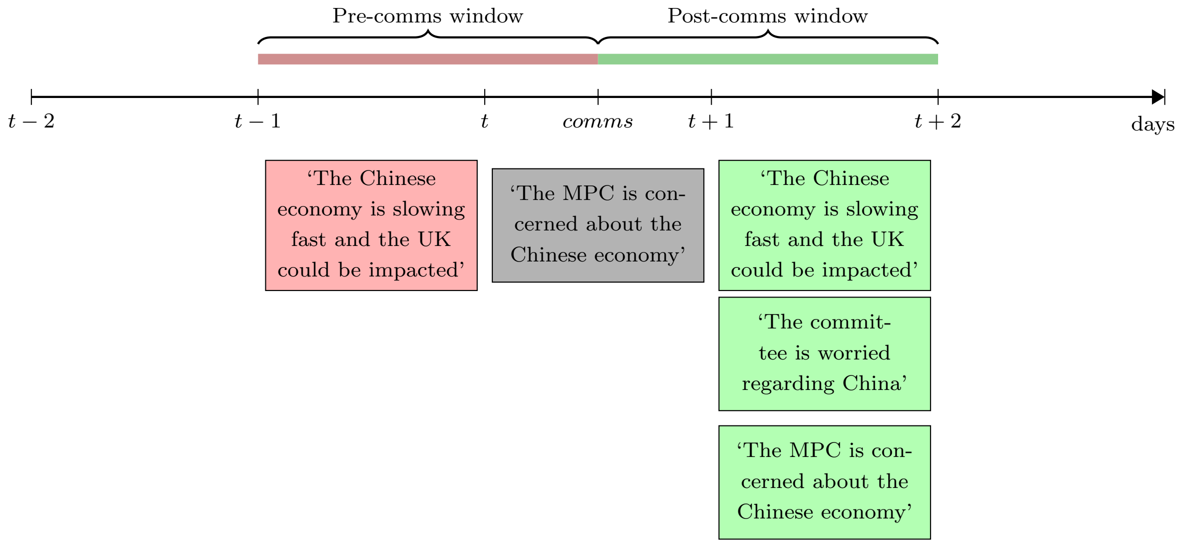
Suppose that the Chinese economy suddenly slows down. The Bank of England communicates its concern to wage and price setters (i.e. the public). It releases a statement to relay its message. It's statement reads "The MPC is concerned about the Chinese economy'". It releases this statement on day t at 12 noon exactly.

An article was published the day before (day $t-1$) stating that "The Chinese economy is slowing considerably and the UK could be impacted". Furthermore, an identical article was published the day afterwards (day $t+1$) along with two additional articles that reported on the Bank's communication. See Figure 3.

One way to measure k would be to calculate the similarity (defined more precisely in the following subsection) between the communication and the post-communication articles. However, a method that did not account for the articles prior to the Bank's communication would be likely to count the article on the day afterwards as being influenced by the Bank's communication since they both contain the words "Chinese economy".

To guard against this we take the *change* between the average similarity of articles in the post-communication and the communication and articles in the pre-communication window and the communication.

Fig. 3. Timeline of example communications event



4.2. Defining similarity

It is fair to say that much of the press coverage following Central Bank communication is an interpretation of the Bank's words. To try and capture to what extent the *message* of the central bank transmits into the news media we use a technique from Natural Language Processing, called word embeddings (Mikolov, Chen, G. Corrado, and Dean 2013) combined with soft-cosine similarity (Sidorov, Gelbukh, Gómez-Adorno, and Pinto 2014) to measure the similarity between central bank communication and news articles.[9]

It is worth noting that this measure combines two separate sources of vectors. Word embeddings are word-specific vectors, in our case of length 100, which are the result of a supervised machine learning algorithm. Term-frequency vectors are document-specific vectors which map words from a dictionary to their frequency in a document.

Following Mikolov, Chen, G. Corrado, and Dean (2013) and Mikolov, Sutskever, Chen, G. S. Corrado, and Dean (2013), word embeddings have become a popular way of representing individual words as vectors, whilst retaining desirable features of the words. Famously, word embeddings retain the semantic relationships between words, insofar as — if trained on the appropriate corpus — the vector for 'King' minus the vector for 'Man' plus the vector for 'Woman' yields a vector similar to that of 'Queen'.

In our case we use the word2vec Continuous Bag of Words implementation. This implementation takes a word within a sentence as the variable that a shallow neural network is asked to predict. We then provide the words surrounding the word in question (the "context") to the neural network, and ask it to predict the missing word. We train on the entire corpus of Bank communications and news articles.

[9]Doc2vec, the analagous form of word2vec for document embeddings is another possible way to get at a measure of similarity that we want, but we have reason to doubt its accuracy on a corpus as small as our own.

We use a total window size of 10 — so any word within five words either side of the word we want to predict is included as an input. The neural network then "learns" to predict words based on their context. Or, it maximises the probability of the correct word, given the context. The word embeddings we extract are the weights the network eventually uses to perform its predictive task. A more detailed explanation is found in Appendix Section 8.2.

We pre-process our data in this case by by removing words less than two letters, removing punctuation, removing all stopwords (words like 'and' and 'the', that add more noise than signal to the data), and converting all uppercase letters to lowercase.

Once each word in the dictionary has been assigned a word2vec vector, we move on to calculating the soft-cosine similarity between two documents.

Cosine similarity is the use of the cosine of an angle between two vectors as a measure of how similar the vectors are. In our case the vectors in question are unigram term frequency vectors. If one was to use pure cosine similarity on these vectors, only words that co-occur in both texts (i.e. the news article and the Bank of England communication) would translate into a higher cosine similarity. Soft cosine similarity, a measure that has been show to outperform many methods in text-similarity competitions (Charlet and Damnati 2017), uses the embeddings derived from the word2vec procedure to weight the cosine similarity measure.

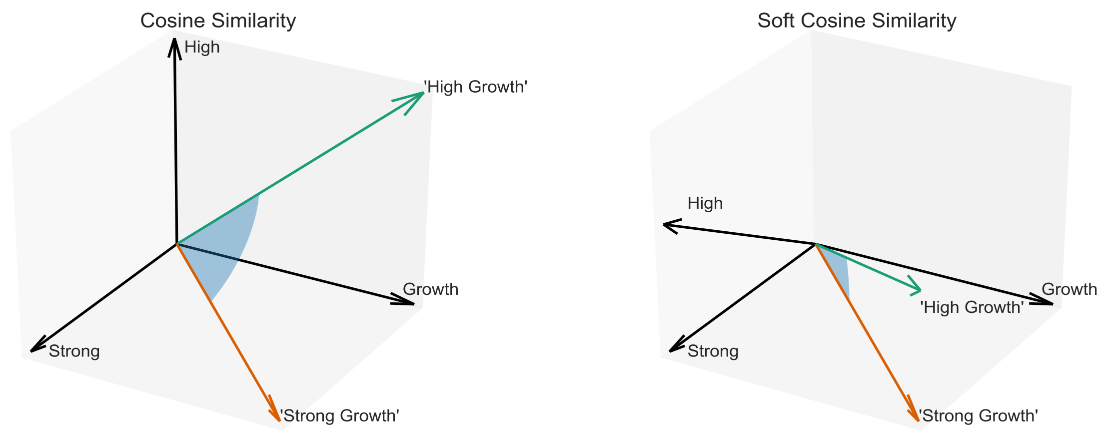
It is instructive to consider an example.[10] Suppose that we want to compare the similarity of the Bank of England's two word communication 'strong growth', with the news article published the following day of 'high growth'. Since there are only three words in the dictionary — {'strong', 'growth', 'high'} — we can visualise the term frequency vectors on a 3d chart (Figure 4). Calculating the cosine similarity between the term frequency vectors (1, 1, 0) and (0, 1, 1) is simple and yields a similarity of $\frac{1}{2}$. However this calculation ignores the fact that 'high' and 'strong' are used synonymously in this case. Indeed, the vector-space representations of 'high' and 'strong' are orthogonal to one another.

Soft cosine similarity uses the cosine similarity between the word embedding vectors (the set of word specific vectors) to weight the cosine similarity calculation between the term frequency vectors (the set of document vectors).

This is illustrated in Figure 4. The word embeddings for 'high' and 'strong' have high cosine similarity, so the term-frequency representations of these words cease to be orthogonal. Calculating the cosine similarity on the non-orthogonal vectors yields the soft-cosine similarity, which is higher (i.e. the angle between the vectors is smaller) than without the word-embedding weights.

[10]The idea for this example comes from the python package gensim's documentation. See https://github.com/RaRe-Technologies/gensim/blob/develop/docs/notebooks/soft_cosine_tutorial.ipynb

Fig. 4. Cosine similarity versus soft-cosine similarity



Mathematically, soft cosine similarity adds an extra weighting term to the cosine similarity formula.

$$CosineSim(a,b) = \frac{\sum_{i=1}^{N} a_i b_i}{\sqrt{\sum_{i=1}^{N} a_i^2}\sqrt{\sum_{i=1}^{N} b_i^2}} \tag{22}$$

$$SoftCosineSim(a,b) = \frac{\sum\sum_{i,j}^{N} s_{i,j} a_i b_j}{\sqrt{\sum\sum_{i,j}^{N} s_{i,j} a_i a_j}\sqrt{\sum\sum_{i,j}^{N} s_{i,j} b_i b_j}} \tag{23}$$

Where $s_{i,j}$ is the similarity between word i and word j as measured by the *cosine* similarity of their word2vec vector representations.

Having calculated the soft-cosine similarity for all articles with respect to the Bank's communication, we can take the change in the average between the windows as our semantic measure of k.

4.3. Implementation

We take all the relevant news articles that are published in a window before a Bank communication event, and all the relevant news articles that occur in a window afterwards. For our analysis, relevant is defined as containing the words "Bank of England".

When performing the analysis on our dataset the event windows vary slightly based on the form of communication. For Inflation Reports, Introductory Statements, and Minutes, we take our pre-announcement window as any articles published before the time of publication on the same day or the day before, and our post-announcement window as any articles published after publication on the day of the announcement or on the day afterwards. For the Q and A, we take the same approach to the pre-announcement window, but allow a gap of 2 hours before the post-announcement window begins to allow the Q and A to have finished before we collect the news articles. For speeches, we do not use the time at which the speech was delivered or published to the public, so our windows exclude any articles published on the same day, and only cover articles published on the day before or after the speech took place.

Our measure has two important features. Firstly, it is a measure of how much the content of the news has changed as a result of an official communication event. Indeed, if the news content is the same in both windows, the measure will return a value of 0. Secondly, it is adjusted for how much of the change in news content can be ascribed to the Bank's communication. The largest readings will take place when the news in the pre-announcement window is unrelated to the Bank's communication - i.e. it was not trailed in the press beforehand, or indeed forecasted by press articles leading up to the communication event - and when the news in the post-announcement window is closely related to the Bank's own message.

One caveat to our approach is that similarity does not capture the accuracy of reporting of central bank communication — indeed it is possible to write news of central bank communication that is subtly inaccurate, yet very similar textually to the original communication.

Returning to the original example, the same diagram with the similarity measures now imposed can be seen in Figure 5 and the results of the similarity measure calculations are displayed in Table 4.

Table 4 shows that the post-communication article which is an exact copy of the pre-communication article has an identical similarity score. When we take the difference in some moment between the two distributions of pre and post articles (in our case, the mean), we control for the confounding effect of news stories that were already in the press before the Bank of England made a communication.

Furthermore, Figure 5 highlights the fact that post-communication article number two, whilst having a meaning that is very similar to the Bank's communication, would score zero on a cosine similarity measure (as there are no identical common words that aren't stopwords), but recieves a positive soft-cosine similarity score because it uses synonyms.

Our measure, as explained above, uses a similarity matrix between words to help weight similarity queries between documents. In the example, one of the synonyms used in post-communication article number 2 is 'worried' in place of 'concerned' which the Bank of England communication uses. The word2vec model tells us that the similarity ($s_{i,j}$) between these two words is 0.85. As a result, when computing the soft-cosine similarity between post-communication document 2 and the Bank's communication, the words 'worried' and 'concerned' do not return a contribution of zero to the similarity score, as would be the case with normal cosine similarity.

We can see in Figure 5 that the measure gives post-communication article number 2 a score of 0.35, higher than post-communication article 1. Finally we take the difference in the average scores between the two windows to arrive at our measure. Our example receives a k of 0.22. This is a relatively high score, which should be expected: there was little trailing of the Bank's message in the press, and one article reported what the Bank said verbatim.

We perform this exercise for all of the Bank of England's communications. Figure 6 shows the time series of our measures of k.

In Appendix Section 8.3, as a robustness check, we show that the kernel density estimates for the soft cosine similarity scores of articles in the pre-communication window are shifted substantially closer to zero versus those in the post-communication window.

Fig. 5. Timeline of example communications event

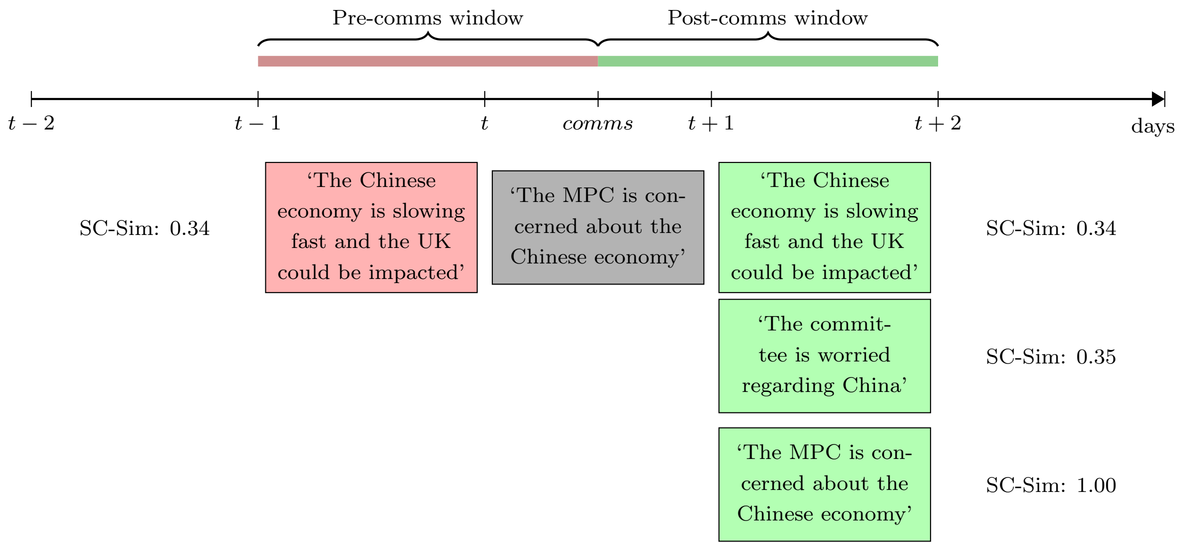
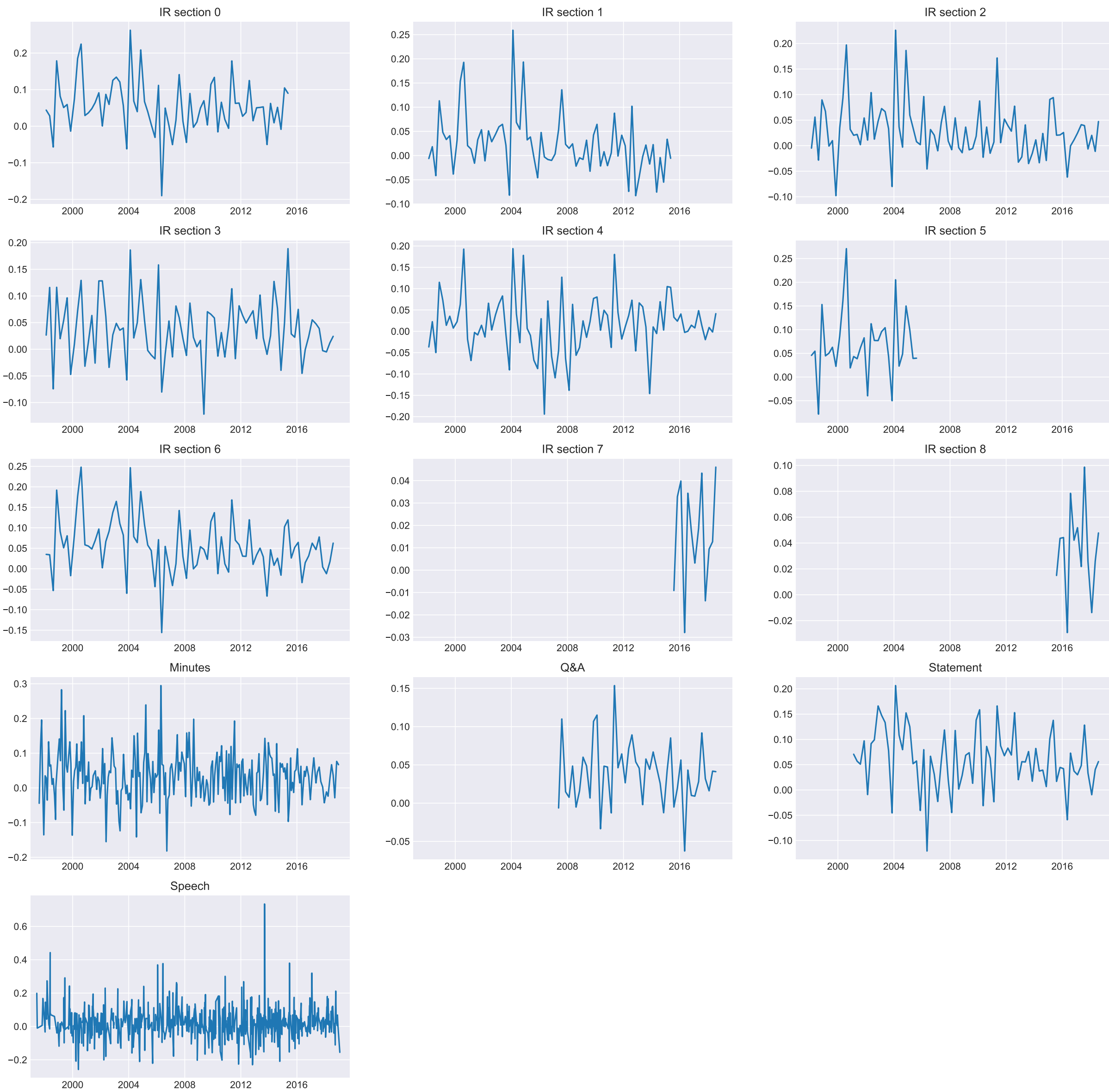


Table 4: Calculating the measures of an example communications event

	Pre-communication		Post-communication		Difference in averages
	Per-article	Average	Per-article	Average	
Measure	0.34	0.34	0.34	0.56	0.22
			0.35		
			1		

Fig. 6. Semantic similarity measure time series



5. Measurement of text features

What factors might influence whether an event is picked up in the news? In this section, we discuss the set of characteristics of central bank text, θ_B, that we surmise may be associated with our outcome variable k, the proportion of newspaper text that paraphrases Bank of England communication events.

In total we measure 351 different variables that make up the vector θ_B. The explanation of how we measure these variables is the only issue dealt with in this section.

First we include dummy variables for all the different types of communication we study: 9 sections of the Inflation Report, the minutes, the Q & A, and speeches. We also include a dummy variable as to whether there was a monetary policy decision made on the same day as the communication.

To motivate our textual feature sets, we draw on scholarly investigations the cognitive psychology of language processing and crucially news values and news discourse (e.g. Galtung and Ruge (1965), Bednarek and Caple (2017), and Harcup and O'neill (2017)).

Drawing on the above-mentioned literature, for some central bank communication to be picked up by the media, we assume that it has to be of topical interest (TOPIC), quickly and efficiently processed (LINGUISTIC PROCESSING), and contain certain characteristics that journalists value (NEWS VALUES). Thus, we decompose θ_B into these three components, the last two with three and ten sub-components respectively:[11]

1. TOPIC: θ_B^{TP}
2. LINGUISTIC PROCESSING: θ_B^{LP}
   (a) lexical access
   (b) syntactic processing
   (c) discourse processing
3. NEWS VALUES: θ_B^{NV}
   (a) size
   (b) impact
   (c) sentiment
   (d) personalization
   (e) proximity
   (f) facticity
   (g) uncertainty
   (h) prominence
   (i) timeliness
   (j) novelty

Below we enumerate and discuss these features

[11] In addition, journalism scholars suggest that news media pick-up is also influenced by other factors, such as other stories competing for space, reporter availability, proximity of the communication event to a given deadline, etc. Bednarek and Caple (2017) term these NEWS SELECTION FACTORS. For us, these are features that are excluded by design or are unmeasurable, and as such are caught up in the error term ϵ, or are else captured by our controls.

5.1. Topic

We wish to measure the the extent to which Bank communication touches on topics that consumers want to read about — i.e. that are contained in their preference vector θ^*. We measure 49 different topics using simple dictionary methods. To find these topics, we first obtain guardian articles containing the word 'economy' since January 1st 2000 until the present day, a total of 13203 articles. We then store the tags that these articles are assigned. Tags are attached manually by guardian journalists. There are over 50,000 distinct tags across the guardian's text corpus.

Tags have two 'levels', an upper and a lower level. The upper level represents a broader category than the lower level. For example, a 2014 article titled "Recycling, saving energy, reducing waste: how is it going for you?" is tagged on the upper level as 'environment' and on the lower level has three tags of: 'recycling', 'plasticbags', and 'energyefficiency'. For our purposes we only consider tags with the upper level tag of 'business' which encompasses all economics reporting from the guardian. This is in total 886 tags.

We split the string of the lower level tag into the most likely set of words using a probabilistic model based on zipf's law.[12] In the above example 'energyefficiency' is split into 'energy' and 'efficiency'.

Then, we use the word-embeddings trained on the entire news and Bank communication corpus that we constructed during our measurement of k to assign each lower level tag an average word embedding (obviously if the lower level tag is just one word, then the average word embedding is just the embedding of that word). The tag 'energyefficiency' will be assigned a word embedding of length 100 that is the average of the embeddings for 'energy' and 'efficiency'.

We remove tags that have been used less than 100 times. We then use a k-means clustering algorithm to group the tags into distinct groups. The optimal number of clusters, 49, is determined by the silhouette score across a grid search. These 49 clusters form the topics of content that we wish to measure.

Once the tags are clustered, we take the centroids of the clusters, and take the ten words — excluding numbers and words that are clearly typos[13] — that are closest to the centroid from our word embeddings. These ten words form a dictionary for each topic that is used to measure the extent to which that topic is discussed by the Bank of England. More specifically, our measure for each topic is the total sum of the occurences of the words in the topic dictionary for a given communication, divided by the length of the communication.

Dictionary methods — counting certain words relating to a topic of interest — are a very common and simple method for measuring content. They have been applied most notably to measurement of uncertainty in text (S. R. Baker, Bloom, and Davis (2016), Manela and Moreira (2017), Soto (2019)).

We could have opted for an unsupervised approach to content modelling, such as Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003). This would have involved estimating a generative model of text production on some external corpus (e.g. the guardian articles), and then querying the Bank of England communications to determine each communication's distribution over the estimated topics. However, for our purposes there are a number of issues with this approach. Since words are not unique to topics, it is difficult to justify ex-post labelling of topics as being related to certain content features. This is

[12] see python package wordninja for more details

[13] The reason for this being that typos and numbers are likely to have vectors associated with them that are close to the random vector assigned at the beginning of the word2vec training. The fact that they are close to our centroids is just random chance.

particularly the case if the random seed used in LDA changes — potentially altering the topics. Using word embeddings escapes this problem because each of our topics is simply a cluster around a *specific* word embedding, and so the closest word to that embedding can be said to be the label of the topic. Indeed, cases in which LDA has been used the authors in the field generally shy away from ex-post labelling which we explicitly want to do here (S. Hansen, McMahon, and Tong (2019), Munday (2019)), and use LDA as a dimensionality reduction tool only.

The list of words for each topic, and the tags associated with them are detailed in Appendix Section 8.5. It's worth noting that some of the topics look to be simply noise (topics 0 and 29), and we would expect these topics to return insignificant results when regressed against k. Some of the topics are clearly topics that consumers are interested in, but the Bank is unlikely to comment on (e.g. Topic 38, which is to do with retail supermarkets), but this is by design! We want to include topics that consumers care about but that the Bank may regard as unimportant — how to trade off talking about topics close to consumers interests versus those close to the Bank's is the exact problem we outlined mathematically in Section 2.

5.2. Linguistic Processing

Our linguistic processing features relate to three main goals in interpreting an incoming (written) linguistic signal that will influence whether a central bank communication is picked up by the media. One can think of our linguistic processing features as capturing how easy it is to read a given central bank communication.[14]

- We (i.e, language users) have to identify that a string of text is a potential word, and if so, access information about it from our mental lexicon—for instance, its meaning(s), part-of-speech, phonological make-up, etc. We also need to resolve ambiguities around meaning. The point at which we have successfully accessed a word and correct information about it is called *lexical access.*
- Next, we need to fit each word together to arrive at a representation of the meaning of the sentence. This is *syntactic processing.*
- Finally, we need to integrate the sentence just processed into the prior discourse (and potentially world knowledge). This is called *discourse processing.*

5.2.1. Lexical Access and Processing

We begin our discussion of language processing with lexical access and processing features, which deal with dimensions relating to:

- an individual's exposure to a particular word (in general, exposure to a word increases its activation, and thus ease of processing);
- formal linguistic properties of the word;
- semantic features of the word;
- and neighborhood effects.

[14] Note that we have obviously simplified things here. For instance, we completely ignore individual differences that may bear on an individual's ability to comprehend a text. For a full and complete treatment, we direct the interested reader to the very accessible introductory text of Warren (2013).

Usage Rates and Exposure

Frequency One of the earliest and most robust findings in psycholinguistics that has been replicated across experimental paradigms is that frequency of usage plays a central role in accessing words from the mental lexicon (Howes and Solomon 1951; Forster and Chambers 1973; Whaley 1978). In language comprehension and production, high-frequency words are accessed more rapidly and more accurately than low-frequency words. In particular, the effect is not binary, i.e. only between low versus high frequency words, but persists throughout the frequency range (Embick, Hackl, Schaeffer, Kelepir, and Marantz 2001).

To operationalize this feature, we drew on frequency information from the SUBTLEX-UK database of Van Heuven, Pawel Mandera, Keuleers, and Brysbaert (2014). This database contains frequency information for over $160,000$ word types from subtitles of BBC programs. Frequencies derived from this database have been shown to better predict individuals' word processing performance than those derived from other databases.

We extracted word tokens and word lemmas from each document using the spaCy package for Python (Honnibal and Montani 2017). For each word token within each document, we then measured its type frequency and lemma frequency in SUBTLEX-UK.[15] This results in two vectors of length n_d, where n_d is the number of words in document d: a type frequency vector $type$ and lemma frequency vector $lemma$. To provide summary measures for each document, we computed the mean from each of these vectors to yield two frequency attributes per document.

Contextual Diversity A relatively recent finding in psycholinguistic research is the importance of a word's CONTEXTUAL DIVERSITY—the number of contexts in which an individual has experience of a word (Adelman, Brown, and Quesada 2006; Plummer, Perea, and Rayner 2014). In the presence of contextual diversity, the above-mentioned strong effect of usage frequency in isolation seems to be attenuated.

To build this feature, we took type-based contextual diversity scores from the subtitle corpus of Van Heuven, Pawel Mandera, Keuleers, and Brysbaert 2014.[16] Following the above procedure, we built one contextual diversity feature (mean vector).

Age of Acquisition Words acquired early in life are processed faster than words acquired later, even when other variables are controlled for (see Johnston and Barry 2006 for an overview).

We obtained type-based age-of-acquisition information from the dataset of Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012), and again derived one summary feature based on the mean.

Prevalence Brysbaert, Paweł Mandera, McCormick, and Keuleers 2019 find that *word prevalence*, "the percentage of people who indicate they know the word", explains an additional 3.6% of variance in word-processing studies.

We operationalize this feature using the dataset of Brysbaert, Paweł Mandera, McCormick, and Keuleers 2019, which along with overall prevalence scores also contains scores split by respondent's gen-

[15]Type frequency is the frequency of a word form in the text, such as *banks* or *strengthening*. Lemma frequency is the frequency of a word's dictionary entry form, such as *bank* or *strengthen*.
[16]Lemma-based contextual diversity is not available.

der (male or female). As sociological information may be important in explaining word access across individuals, we include this information to derive $p = 4$ prevalence features: (1) overall prevalence scores, (2) prevalence scores for females, (3) prevalence scores for males, and (4) the difference of the last two mentioned scores. Again, this is computed first at the word token level, and then summarized at the document level by taking the mean.

Repetition Priming Recent prior exposure to a word facilitates its re-access (D. L. Scarborough, Cortese, and H. S. Scarborough 1977). This is called REPETITION PRIMING. The prior mention of the token is termed the *prime* and the re-accessed token is termed the *target*.

We use three programs. Our first program checks whether a word occurs in the prior context and fires a boolean.

Second, given that memory decays with distance and can thus impact on word retrieval, we used a second program that first checks whether a word occurs in the prior context; if it does, we take the reciprocal of the distance (in tokens) between the prime and the target; if it does not occur, we give it a score of zero.

Third, given that human memory may decay logarithmically rather than linearly (see Singh, Tiganj, and Howard 2018), we also took the natural log of the sum of the second measure and unit constant.

We measured the above three variables at the word-type and word-token level, to derive six features summarized by the mean.

Expectancy in the Sentential or Discourse Context Lexical access is also facilitated (pre-empted) by the sentential context (e.g., Schuberth and Eimas 1977; Kutas and Hillyard 1984).[17] This suggests that during comprehension language users are predicting the upcoming context. In other words, upcoming words are already being accessed from the mental lexicon ahead of their being read. When a word is read that is not expected, we have to retrieve that unexpected word from the mental lexicon, and this causes a processing difficulty. For example, in the sentence below the word *roses* fits the context perfectly. Words such as *tulips*, even though it is from the same semantic field, are unexpected in the context and cause a processing delay.

(1) The gardener really impressed his wife on Valentine's Day. To surprise her, he had secretly grown some {<u>roses</u>, tulips}.

We operationalized this feature by using spaCy's word vector engine to return the similarity score between a target word (e.g. *roses*) and the prior context (e.g. *...his wife on Valentine's day...*) Thus, according to our measure, for the example above, *roses* receives a fit with the context of 0.38, while *tulips* has a lower score of 0.22. We have our one summary feature for this dimension of lexical access.

Formal Word Properties

[17] For instance, using an electrophysiological paradigm, Federmeier and Kutas 1999 showed that words that are unexpected within the sentence or discourse context induce larger N400 amplitudes than words that fit the context perfectly (N400 is a negative-going potential peaking around 400 ms after the onset of a stimulus).

Word Status Psycholinguists have also found some evidence for the differential processing of CONTENT words (words with rich semantic content) versus FUNCTION words (words with grammatical functions and minimal semantic content) (Pulvermüller 1999). For instance, Pulvermüller, Lutzenberger, and Birbaumer 1995 showed that the processing of function words is localized in the left-hemisphere of the brain, whereas content words are processed bilaterally.

We first used spaCy's part-of-speech (PoS) tagger to perform annotation. Then, for each relevant token (i.e. excluding symbols, punctuation, and whitespace), we fired a binary variable for whether the token was a content word (i.e. adjective, adverb, interjection, noun, proper noun, verb) or a function word (i.e. adposition, auxiliary, coordinating conjunction, determiner, numeral, particle, pronoun, subordinating conjunction). Then, to derive a single measure for at the document level, we took the ratio of the number of content words to the number of content words and function words combined—namely:

$$\frac{Count(content)}{Count(content) + Count(function)}.$$

Grammatical Category Relatedly, specific parts-of-speech may be processed differently (West and Stanovich 1986).

Using information from spaCy's PoS (Part of Speech) tagger, we extracted the relative frequency for each broad part of speech in the document (as listed in Table 5; Nivre et al. 2016). To provide more nuanced information, we also extracted relative frequencies for fine-grained PoS tags (Penn Treebank: Marcus, Santorini, and Marcinkiewicz 1993).

We derived 14 broad PoS features and 33 fine-grained PoS features. (Recall that these are proportions, so we simply have a single measure for each PoS-type.) The following two tables detail the PoS tags that we consider.

Table 5: Broad Part-of-Speech tagset with description and examples

Part-of-Speech Tag	Description	Examples
ADJ	adjective	big, old, green, incomprehensible, first
ADP	adposition	in, to, during
ADV	adverb	very, tomorrow, down, where, there
AUX	auxiliary	is, has (done), will (do), should (do)
CCONJ	coordinating conjunction	and, or, but
DET	determiner	a, an, the
NOUN	noun	girl, cat, tree, air, beauty
NUM	numeral	1, 2017, one, seventy-seven, IV, MMXIV
PART	particle	s, not,
PRON	pronoun	I, you, he, she, myself, themselves, somebody
PROPN	proper noun	Mary, John, London, NATO, HBO
PUNCT	punctuation	., (,), ?
SCONJ	subordinating conjunction	if, while, that
SYM	symbol	$, %, , , +, , , , =, :)
VERB	verb	run, runs, running, eat, ate, eating

Drawn from `https://spacy.io/api/annotation#pos-tagging`

Table 6: Narrow Part-of-Speech tagset with description and examples

Part-of-Speech Tag	Description
CC	conjunction, coordinating
CD	cardinal number
DT	determiner
EX	existential there
IN	conjunction, subordinating or preposition
JJ	adjective
JJR	adjective, comparative
JJS	adjective, superlative
MD	verb, modal auxiliary
NN	noun, singular or mass
NNP	noun, proper singular
NNPS	noun, proper plural
NNS	noun, plural
PDT	predeterminer
POS	possessive ending
PRP	pronoun, personal
PRP$	pronoun, possessive
RB	adverb
RBR	adverb, comparative
RBS	adverb, superlative
RP	adverb, particle
TO	infinitival to
UH	interjection
VB	verb, base form
VBD	verb, past tense
VBG	verb, gerund or present participle
VBN	verb, past participle
VBP	verb, non-3rd person singular present
VBZ	verb, 3rd person singular present
WDT	wh-determiner
WP	wh-pronoun, personal
WP$	wh-pronoun, possessive
WRB	wh-adverb

Drawn from `https://spacy.io/api/annotation#pos-tagging`

Word Length Effects Word length (or 'bulk') is another pervasive factor in word recognition performance, with the simplest measures (e.g. number of characters) already incorporated in earlier computational measures of text complexity.

We used four different dimensions of word length: (1) number of characters, (2) number of phonemes (that is, the number of units of sound), (3) number of syllables, and (4) number of morphemes (e.g. *organiz-ation-s* has three morphmes, *walk-ed* has two). With 1 summary measure for each dimension based on the document mean, we derived four features of word bulk.

Semantics

Concreteness Concreteness refers to the extent to which the concept of a given lexical item can be perceived by one of the five senses or not. Thus, *money* is concrete, and *inflation* is not.[18]

We used concreteness ratings from the experimental study of Brysbaert, Warriner, and Kuperman (2014), in which almost $40,000$ word lemmas were rated 1 (abstract) through 5 (concrete) across over 4,000 participants, to derive one concreteness feature based on the document mean.

Emotionality Another dimension of semantics that has recently come to the fore is effect of word emotionality in word processing.[19] Words with emotional feature specifications are typically processed faster than those which are more neutral (Scott, O'Donnell, and Sereno 2012).

We drew on the database of Warriner, Kuperman, and Brysbaert 2013 to develop a total of 3 features of emotional *valence* (word pleasantness), emotional *arousal* (the intensity of emotion provoked by the word), and emotional *dominance* (degree of control).

Lexical Ambiguity When we read a word such as *bank* we gain access to its multiple meanings, e.g. the word's meaning as financial institution and its meaning as place alongside a river. Even in contexts in which the other meaning makes complete nonsense, we still access and consider as a potential the other meaning (Swinney 1979). This choice of meanings causes a processing disruption, though we are seldom aware of it. This is called *lexical ambiguity.*

We measured the degree of lexical ambiguity in a document in two ways. First, we extracted the number of meanings for each word in the document using WordNet (Miller 1995). Second, we measured the semantic diversity of a word—that is, the degree to which the contexts in which a given word occurs are similar in meaning overall (Hoffman, Ralph, and Rogers 2013). For each measure, we computed our usual mean vector.

Neighborhood Effects

Orthographic Neighborhood A target word a has an orthographic neighbor b if one can create b from a by changing a single letter in one of the word's positions (Coltheart, Davelaar, Jonasson, Besner, and Dornic 1977). Thus, some orthographic neighbors of the word *bank* are *balk*, *bane*, *lank*. The size of a word's neighborhood affects its access: the larger the neighborhood of a word, the faster its access (Andrews 1989).

[18] According to one theory, words denoting concrete concepts activate both a language (verbal) system and an imagistic (nonverbal) system whereas words denoting abstract concepts activate only the language system. This 'dual-coding' (activity in two interconnected systems) affords processing advantages to concrete words (Paivio 2013).

[19] Emotionality has drawn attention in monetary economics too, for instance Tuckett (2011).

We collected orthographic neighborhood statistics for each word in a given document from the dataset of Balota et al. (2007), and from these derived the mean vector.

Phonological Neighborhood Similarly, a word's phonological neighborhood size refers to the number of words that can be formed from the original word by a single phoneme substitution, addition or deletion. For example, *sort* has *thought* (substitution), *sorts* (addition), and *ought* (deletion) as phonological neighbors. Mulatti, Reynolds, and Derek Besner (2006) demonstrate that phonological neighborhood size is a stronger predictor of lexical access than orthographic neighborhood size in contexts in which words are read aloud.

As for orthographic neighborhood, we used the dataset of Balota et al. (2007) to derive our mean summary variable for this feature.

5.2.2. Syntactic Processing

We detail next features that are intended to capture the processing costs associated with the comprehension of syntax (sentence structure) and its interface with meaning (semantics).

Drawing on and adapting Bornkessel-Schlesewsky and Schlesewsky (2009, p. 90)'s list of requirements of a syntactic processor, we aim to featurize five aspects of sentence parsing:

1. formal structure building;
2. grammatical dependency relation linking;
3. working memory and storage limitations;
4. expectation;
5. ambiguity processing and conflict resolution.

Syntactic Structure Building

Constituency Types As soon as we encounter textual material, we need to impose structure upon it and build out the individual words into larger constituent units. This forms the basis for subsequent interpretation. We call this *constituency parsing*.

Using the spacy add-on component benepar (Kitaev, Cao, and Klein 2019; Kitaev and Klein 2018), we parsed each sentence in each document into its constituents. An example of a sentence's constituent parse can be seen in Figure 7.

Table 7: Constituency parse labels

Label	Description	Label	Description
S	main clause declarative	NX	head of noun phrase in complex NPs
SBAR	subordinate clause	PP	prepositional phrase
SBARQ	direct question	PRN	parenthetical
SINV	inverted declarative	PRT	particle
SQ	inverted yes/no question	QP	quantifier phrase
ADJP	adjective phrase	RRC	reduced relative clause
ADVP	adverbial phrase	UCP	unlike coordinated phrase
CONJP	conjunction phrase	VP	verb phrase
FRAG	fragment	WHADJP	*wh*-adjectival phrase
LST	list marker	WHADVP	*wh*-adverbial phrase
NAC	not a constituent	WHNP	*wh*-noun phrase
NP	noun phrase	WHPP	*wh*-adjectival phrase

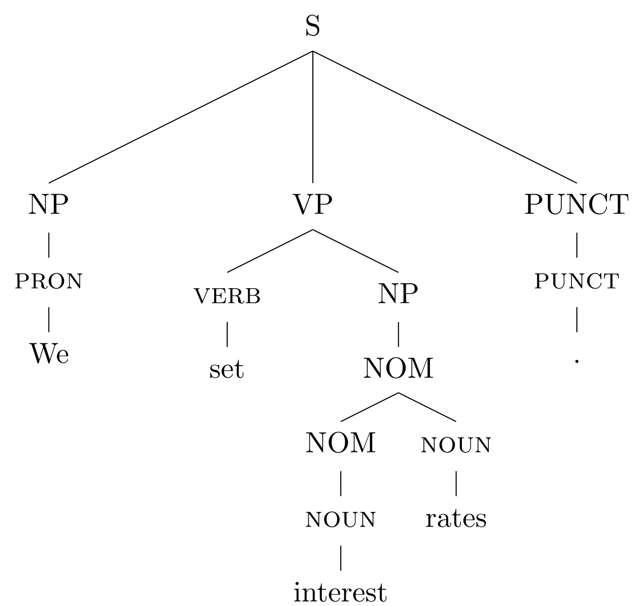


Fig. 7. Syntactic constituency parse for an example sentence: “We set interest rates.”

For each syntactic constituent type, listed in Table 7, we calculated its mean sentence rate per document.

Dependency Relation Building

Dependency Types As we build constituency structure, we need to link each syntactic constituent with a grammatical role. For example, consider again the following simple sentence:

(2) We set interest rates.

Upon identifying the noun phrase constituent *We* we need to realise that it is the grammatical subject of the sentence; when we encounter the verb *set* we need to realise that it is the verbal root of the sentence, upon which *We* depends; when we encounter the noun phrase *interest rates*, we determine that it is the grammatical object, dependent upon *set*. We thus map the constituency built in (e.g.) Figure 7 onto a grammatical dependency parse, which links each word to its relational parent. The dependency parse equivalent of Figure 7 is given in Figure (2).

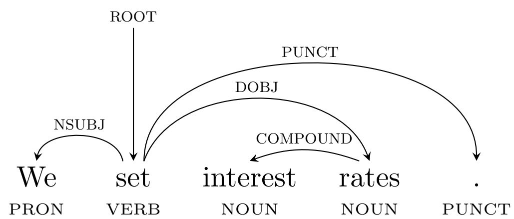


Fig. 8. Syntactic dependency parse for an example sentence *We set interest rates.* The arcs above the text denote the grammatical relations. We also show part-of-speech information below the text. This shows that *We* has involved in a nominal subject relation with respect to the root *set*, and *interest rates* is involved in a direct object relation with respect to the root. The root *set* is associated with 3 dependencies in this graph. The nominal subject *We*, which is here instantiated by a pronoun (PRON) .

We used spacy to build dependency parses for each sentence within each document. Then, for each of the 48 syntactic dependency types, listed in Table 8, we calculated its mean per sentence rate.

Table 8: Dependency parses

Label	Description	Label	Description	Label	Description
acl	adjectival clause	csubj	clausal subject	nummod	numeric modifier
acomp	adjectival complement	csubjpass	clausal subject (passive)	oprd	object predicate
advcl	adverbial clause modifier	dative	dative	obj	object
advmod	adverbial modifier	dep	unclassified dependent	obl	oblique nominal
agent	agent	det	determiner	parataxis	parataxis
amod	adjectival modifier	dobj	direct object	pcomp	complement of preposition
appos	appositional modifier	expl	expletive	pobj	object of preposition
attr	attribute	intj	interjection	poss	possession modifier
aux	auxiliary	mark	marker	preconj	pre-correlative conjunction
auxpass	auxiliary (passive)	meta	meta modifier	prep	prepositional modifier
case	case marking	neg	negation modifier	prt	particle
cc	coordinating conjunction	nn	noun compound modifier	punct	punctuation
ccomp	clausal complement	nounmod	modifier of nominal	quantmod	modifier of quantifier
compound	compound	npmod	noun phrase as adverbial modifier	relcl	relative clause modifier
conj	conjunct	nsubj	nominal subject	root	root
cop	copula	nsubjpass	nominal subject (passive)	xcomp	open clausal complement

Root Type Some root types are more "canonical" than others; for instance, a verbal root might be considered more basic than a nominal root. Following related work in computational readability (e.g. Brunato, De Mattei, Dell'Orletta, Iavarone, and Venturi 2018), we computed the proportion of sentences within a document for which the sentence's root was instantiated by a verb.

Working Memory and Storage Limitations A considerable body of research has investigated the role of *working memory* and *storage limitations* in sentence processing (Gibson 1998; Gibson 2000). When reading a sentence, we process each word incrementally over time, integrating each word one by one into the structure being built. As the sentence unfolds, it is necessary to retrieve information that has gone before and link current information with it. This burdens the sentence processor, because linguistic material has to be held in memory until it can be fully integrated.

We take as an example two sentences from Jaeger and Tily (2011). The first sentence (3-a) is relatively easy to process, while (3-b) for most readers is almost impossible (although it does actually make perfect sense).

(3) a. This is the malt that was eaten by the rat that was killed by the cat.
b. This is the malt that the rat that the cat killed ate.

The reason (3-a) is easier to process than (3-b) is because in the former the dependency relations between the individuals words are fairly *local*. In (3-b), by contrast, both *that* and *the rat* have to be stored in working memory until the verb *ate* is encountered (they are the object and subject of *ate*, respectively). These *long-distance* or *non-adjacent* dependency overtax memory resources and result in processing difficulty.

We operationalize various features that can be assumed to relate to storage and integration costs. Specifically, we featurized the following.

- dependency arc lengths
  - mean dependency arc length per sentence
  - the mean maximum dependency arc length per sentence
- number of dependencies
  - mean number of dependencies per sentence
  - mean number of dependencies per root
  - mean number of dependencies per subject
- dependency location
  - mean number of left-edge dependencies per root
  - mean number of left-edge dependencies per subject
  - mean number of right-edge dependencies per root
  - mean number of right-edge dependencies per subject
  - ratio of left-edge and right-edge dependencies per root
  - ratio of left-edge and right-edge dependencies per subject
- offset distances
  - offset distance of subject
  - offset distance of root
- mean number of leaves (terminal nodes, i.e. words) per sentence
- mean number of non-binary branching constituents
- mean number of non-terminal nodes
- mean parse tree height
- mean number of words per syntactic phrase
- ratio between the length of the first syntactic phrase (usually the subject noun phrase) and the second syntactic phrase (usually the verb phrase)

Structural Expectation and Priming

Structural Expectation Other researchers in the psychology of language have focused on processing difficulty/ease as associated with the likelihood of syntactic structures in the discourse (Demberg and Keller 2008; Levy 2008). In general, structures that are more frequently encountered in a language user's experience are preferred over those that are less frequent.

We operationalized this feature by extracting the mean sentence surprisal score for each document. Specificallty, this is defined as the Shannon information content of the sentence's best constituency parse, i.e. $Surprisal(parse_1) = IC(parse_1) = \log_2 P(\frac{1}{parse_1})$.

Structural Priming We have already discussed lexical priming, whereby a word is more easily accessed if it has been used already in the discourse. Psychologists of language have researched similar effects on the syntax plane. Specifically, syntactic structures that have previously been used in the discourse are easier to build than those which are encountered for the first time.[20]

We featurized this aspect of sentence process in various ways:

- dependency type type-token ratio ($2 - 6$ grams)
- part-of-speech type-token ratio ($2 - 6$ grams)
- syntactic production similarity

Structural Ambiguity Processing and Conflict Resolution

Sentence Ambiguity Score Like lexical ambiguities, structural ambiguities permeate natural language. A robust finding in the experimental literature is that such structural ambiguities result in processing difficulty and delay (for an overview see e.g. Van Gompel and Pickering 2007). A famous example of *local syntactic ambiguity* is given in (4)

(4) The horse raced past the barn fell.

As we process the above sentence incrementally, we encounter the word *raced* and analyse it as the main verb of the sentence. But when we reach *fell*, we realize we've made a mistake because there's no place in the structure to attach it. Subsequently, we have to reanalyse the sentence with *raced* as a participle introducing a reduced relative clause (i.e. *(that was) raced...*) and *fell* as the main verb. This kind reanalysis causes a processing slow-down. It is called a local ambiguity because the ambiguity is resolved locally within the sentence being processed.

Other sentences exhibit *global syntactic ambiguity*, such as that in (5).

(5) The girl saw the boy with the binoculars.

The attachment of the prepositional phrase *with the binoculars* is ambiguous: it can modify the seeing event or the noun phrase *the boy.* This is a global ambiguity because the ambiguity still has to be resolved at the end of the sentence. Again, such ambiguities complicate the parsing process.

[20] See e.g. Pinker (2015): "A bare syntactic tree, minus the words at the tips of its branches, lingers in memory for a few seconds after the words are gone, and during that time it is available as a template for the reader to use in parsing the next phrase. If the new phrase has the same structure as the preceding one, its words can be slotted into the waiting tree, and the reader will absorb it effortlessly."

Given the importance of structural ambiguity processing in the psychological literature, we decided to construct a feature operationalizing it. Specifically, for each sentence in each document, we used a K-best parser to extract parsing surprisal scores for the two best parses. We then took the ratio of the surprisal scores, and averaged at the document level.

Explicit Structure Marking The grammar of English has available a suite of alternate ways of saying the same thing. Oftentimes, one variant is more explicit in some way than another. For instance, grammatical negation can be contracted, where *not* is the explicit variant and *-n't* is the contracted variant. We can optionally omit clausal indicators; for instance in (5) the relativizer *that* is omitted. In addition, in certain types of noun phrase, the determiner *the* is variably realized or omitted.

There is some evidence that explicit alternants can support comprehension, as their presence can reduce ambiguity and help build the correct syntactic parse (Race and MacDonald 2003; Warren 2013; Pinker 2015).

We built features for a variety of fairly frequent syntactic constructions in which one alternant can be considered more explicit than another:

- grammatical negation: *not* vs. *-n't*
- complementizer omission: *that* vs. ∅
- relative pronoun omission: *who/which* vs. ∅
- dative realization: (e.g.) *gave a boost to the economy* vs. *gave the economy a boost*
- infinitival-*to* omission: (e.g.) *help the economy to recover* vs. *help the economy recover*
- comparative choice: (e.g.) vs. *the economy is more healthy* vs. *the economy is healthier*
- *the*-omission: *the* vs. ∅
- auxiliary contraction: (e.g.) *have* vs. *'ve*
- genitive realization: (e.g) *the economy of the UK* vs. *the UK's economy* vs. *the UK economy.*

For each of these constructions, we computed the proportion of explicit realization. For instance, for *the*-realization, we simply computed the number of times a noun phrase began with *the* in a given document divided by the total number of noun phrases in a given document. If the construction happened not to occur in a given document, we include boolean features flagging this.

5.2.3. Discourse Processing

Having accessed words and parsed the incoming linguistic input text into its constituent parts, the language comprehender next needs to construct a mental representation of the text. Below we featurize four main aspects of discourse processing:

- identifying the topic of the discourse;
- constructing propositions and representations for new discourse entities;
- determining how each sentence is connected to other sentences;
- and identifying referents for linguistic expressions.

Topic Identification In order to be able to start building a mental representation of the text, the reader has to quickly identify what it is about. A series of experiments have shown that when the context is given – whether that is a picture, a title, a summary first sentence – readers are better able to recall the contents of the text (Bransford and Johnson 1972). We conjectured, therefore, that if the first sentence in the text effectively summarises the main content, the text will be more readily understood and recalled. We operationalized this aspect of discourse processing by using a doc2vec (Le and Mikolov 2014) algorithm in which we assessed the textual similarity of first sentence to the rest of the text. The more similar the first sentence is to the rest of the text, the more effective a summary it can be considered to be of the text as a whole.

Constructing Propositions and Representations for New Discourse Entities In constructing a mental representation of the text, comprehenders need to be able to extract relevant PROPOSITIONS from the incoming signal of the surface syntax and fix ENTITIES into memory. Propositions are the "smallest units of meaning that can be assigned a truth value", abstracted from their linguistic realization (Traxler 2011). We define entities loosely as people, places, time, things, concepts, etc., which are arguments in propositions. Consider the following propositional representation of the sentence *"Global activity has strengthened over recent months"*:

(6) strengthen(`theme` = 'global activity', `path` = 'recent months', `tense` = present, `aspect` = perfect)

Research has shown that reading time and recall typically depends on the number of propositions that make up the text (Ratcliff and McKoon 1978).

Entities come in two flavours. They are either GIVEN, in which case a comprehender merely has to reactivate an existing mental representation, or NEW in which case the comprehender has to build an entirely new mental representation. There are substantial computational costs associated with constructing mental representations for new discourse referents (Haviland and H. H. Clark 1974; Gibson 1998).

We designed a suite of features intended to capture aspects of proposition construction and discourse entity representation:

- number of words
- number of sentences
- number of noun phrases
- number of entities (defined here as noun phrases tagged as a named entity)
- number of entities normalized by the number of noun phrases
- number of entities normalized by the number of tokens
- number of entities normalized by the number of sentences
- proportion of noun phrases that are contextually given
- proportion of noun phrases that are indefinite
- number of adverbials about the discourse itself ('as stated above' etc)

Coherence A text is coherent if the propositions that are extracted from the text can be easily connected in some way. In a classic experiment, Thorndyke (1977) showed that participants who were shown a jumbled up narrative recalled fewer ideas than participants who were shown the same text in a coherent order. We decided to operationalize features relating to i) temporal cohesion, ii) lexico-semantic cohesion, iii) referential cohesion, and iii) discourse relations.

- TEMPORAL COHESION
  - proportion of present$\rightarrow$present tense sequences
  - proportion of present$\rightarrow$past tense sequences
  - proportion of past$\rightarrow$past tense sequences
  - proportion of past$\rightarrow$ present tense sequences
  - temporal homogeneity of the document
  - temporal sequence homogeneity of the document
- LEXICO-SEMANTIC COHESION
  - proportion of noun phrases with a lexical chain
  - mean lexical chain span
  - mean lexical chain length
  - proportion of lexical chains spanning over half the document
  - proportion of sentences with at least one overlapping lemma
  - mean number of overlapping word per sentence
  - mean similarity between each successive pair of sentences
  - difference in mean similarity between each successive pairs versus shuffled pairs of sentences
- REFERENTIAL COHESION
  - entity graph coherence score (unweighted)
  - entity graph coherence score (weighted by number of entities)
  - entity graph coherence score (weighted by distance between mentions)
  - proportion subject $\rightarrow$ subject sequence
  - proportion subject$\rightarrow$ object sequence
  - proportion subject$\rightarrow$other sequence
  - proportion subject$\rightarrow$none sequence
  - proportion object$\rightarrow$subject sequence
  - proportion object$\rightarrow$object sequence
  - proportion object$\rightarrow$other sequence
  - proportion object$\rightarrow$none sequence
  - proportion other$\rightarrow$subject sequence
  - proportion other$\rightarrow$object sequence
  - proportion other$\rightarrow$other sequence
  - proportion other$\rightarrow$none sequence
  - proportion none$\rightarrow$subject sequence
  - proportion none$\rightarrow$object sequence
  - proportion none$\rightarrow$other sequence
  - proportion none$\rightarrow$none sequence
- DISCOURSE RELATIONS

- comparison connective rate per sentence
- contingency connective rate per sentence
- expansion connective rate per sentence
- temporal connective rate per sentence
- mean number of connectives per sentences

Coreference Resolution Texts are full of linguistic expressions that refer to the same entity, which we call COREFERENTS. For example, consider the short excerpt of text below in (7).

(7) [The MPC] is [committed to clear, transparent communication]. [The Monetary Policy Report] is a key part of [that]. [It] allows [the group] to share [its] thinking and explain the reasons for [its] decisions.

To introduce more terminology, *The Monetary Policy Report* in the second sentence is termed an ANTECEDENT and the *it*, which refers back to it, is called an ANAPHOR. When a reader encounters a pronominal anaphor like *it* or a noun phrase anaphor *the group*, they need to be able to rapidly identify the correct antecedent. If they match an anaphor to the wrong antecedent, discourse processing breaks down and the text seems incoherent. The process by which readers do this is variously called COREFERENCE RESOLUTION, ANTECEDENT SEARCH, or ANAPHOR RESOLUTION.

Psycholinguists have studied how language comprehenders resolve anaphors, focusing on the factors that facilitate or hinder the process. In particular, there is an effect of distance, with longer distances between antecedent and anaphor causing processing disruption (O'Brien, Raney, Albrecht, and Rayner 1997).[21]

We utilized neuralcoref, a state-of-the-art coreference resolution module for spacy (K. Clark and Manning 2016), and engineered the following features intended to capture aspects of coreference processing:

- ANAPHOR AMBIGUITY
  - number of coreferences per coreference chain
  - mean likelihood of the coreference
  - mean coreference ambiguity score
- DISTANCE
  - mean distance (in words) between each coreferenced entity
- OTHER/GENERAL
  - number of coreference chains in the document

5.3. Newsworthiness

We now move on from the processing of linguistic units to motivate our third main dimension of features—NEWSWORTHINESS—namely, there are certain characteristics of any published news article that

[21] A greater number of competing possibilities for the antecedent results in the ambiguity. In English, but especially morphologically rich languages, language users make use of grammatical information encoded on the anaphor, such as gender and number (Arnold, Eisenband, Brown-Schmidt, and Trueswell 2000). We prefer to match anaphors with antecedents that are in the same grammatical position (Grober, Beardsley, and Caramazza 1978). And we prefer to match anaphors with antecedents that are highly salient or foregrounded in the discourse (Almor and Eimas 2008).

have made it 'newsworthy', i.e. "worthy of being published as news" (Caple 2018). We have drawn on academic journalism research since the 1960s, from Galtung and Ruge (1965) through Bednarek and Caple (2017), to identify 9 news values: (1) SIZE, (2) IMPACT, (3) SENTIMENT, (4) PERSONALIZATION, (5) PROXIMITY, (6) FACTICITY, (7) UNCERTAINTY, (8) PROMINENCE, and (9) NOVELTY. In the following we discuss the specific granular features that make up these 10 news values.

5.3.1. Size

For events to get picked up, they need to be sizeable—that is, they need to be "of a scale large enough to warrant attention" (Montgomery 2007, p. 6). Event size (or scale) can be linguistically encoded in a number of ways, which we draw upon to derive feature sets for this dimension of news values.

- We took the document relative frequencies of **comparative or superlative modifiers** (e.g. *worse/worst, better/best, easier/easiest*). Thus is easily operationalized by checking if a word's fine-grained part-of-speech tag $PoS(w)$ is in the set of comparative or superlative tags, i.e. $PoS(w) \in \{JJR, JJS, RBR, RBS\}$ as defined in Table 6.
- We took the document relative frequencies of **numerals** (and other number terms), which was operationalized by checking if a word's high-level part-of-speech was a numeral, i.e. $PoS(w) == NUM$.
- Similarly, we took the document relative frequencies of **symbols** (e.g. %, £, $, etc.), which was operationalized by checking if a word's high-level part-of-speech was a symbol, i.e. $PoS(w) == SYM$.
- We used regular expressions on the raw text to count the relative occurrence of **intensifiers**—i.e., terms such as *extremely, exceedingly, in all respects, maximally, profoundly.* For this, we drew on Piotrkowicz (2017)'s list of such terms.
- We used regular expressions to count the relative occurrence of **quantifiers** (and other size terms)—e.g., terms such as *plethora, numerous, abundance, myriad, substantial.* Quantifier terms were taken to be those lemmas in the QUANTITY, QUANTIFIED_MASS, SIZE frames in FrameNet (C. F. Baker, Fillmore, and Lowe 1998).
- We computed the relative frequency of **predicates of scalar position**—e.g. *appreciate, diminish, double, dwindle, escalate, expand, fall, gain, grow,* etc. These terms were taken to be those lemmas in CAUSE_CHANGE_OF_POSITION_ON_A_SCALE, CHANGE_POSITION_ON_A_SCALE, CAUSE_EXPANSION, EXPANSION, CAUSE_PROLIFERATION_IN_NUMBER, PROLIFERATING_IN_NUMBER frames in FrameNet (C. F. Baker, Fillmore, and Lowe 1998) and those in lemmas in CALIBRATABLE_COS-45.6 verb class in VerbNet (Schuler 2005).

5.3.2. Impact

Similar to size is the IMPACT of an event. For an item to be newsworthy, it has to be of "considerable significance for large numbers of people" (Golding and Elliott 1979, p. 117). For this dimension, we looked at five features.

- The relative occurrence of synonyms of ***impact*** and ***significance***.

- The relative occurrence of **resultative conjunctions**, i.e. those conjunctions that explicitly index a result discourse relation (e.g. *with the result that, so that, consequently*, etc.). We took such terms from the grammars of Quirk (2010) and Huddleston, Pullum, et al. (2002).
- The relative occurrence of **result-state predicates**. These are predicates that have a result state encoded in their lexical semantics—e.g. *die, build*.
- We also examined **potentially result-state predicates**. These are predicates that do not inherently encode a result state component, but can do in combination with other linguistic material, e.g. *run*, as in *run the economy*, is inherently an activity predicate, but can have a result state e.g. *run the economy dry*.
- We include the proportion of **perfect aspect verb constructions** (*has/had/having* —) as these denote past events with present consequences.

5.3.3. Sentiment

A wealth of research on news discourse has shown that SENTIMENT (viz. positivity, negativity, conflict) is a critical factor in news selection (e.g. Johnson-Cartee 2004; Harcup and O'neill 2017; Bednarek and Caple 2017). We evaluated 6 linguistic markers of event sentiment.

- The relative occurrence of **positive** labeled words in Loughran and B. McDonald (2015)'s dictionary of financial sentiment terms.
- The relative occurrence of **negative** labeled words in Loughran and B. McDonald (2015)'s dictionary.
- In addition to the above, we measured **overall subjectivity**—i.e., the relative occurrence of both negative-labeled words and positive-labeled words in Loughran and B. McDonald (2015)'s dictionary, and **overall sentiment**, i.e. $(Count(positive) - Count(negative))/n$.
- We supplemented traditional features drawn from sentiment lexicons, with two features relating to conflict and contrast(ing views). Specifically, we computed the relative frequency of **contrastive predicates**, i.e. predicates that require two (and potentially in-conflict) agents, e.g. *collide, dissent, disagree, fight, negotiate*, etc. These predicates were taken to be verbs from those verb classes in VerbNet that take agent and co-agent arguments.
- The second new feature is the relative occurrence **adversative conjunctions**—i.e., terms that explicitly indicate a discourse contrast, e.g. *but, however, on the other hand.* We took such terms from the grammars of Quirk (2010) and Huddleston, Pullum, et al. (2002).

5.3.4. Personalization

As Johnson-Cartee (2004) puts it, "people identify with other people, and they are more able to understand and remember stories that are concretized by such examples than those that are not." We derived the following features to operationalize this dimension.

- We used named entity recognition (NER) models to identify the relative frequency of entities tagged as **PERSON**.
- We used a dictionary approach to compute the relative frequency of **personal pronouns that linguistically encode an animate entity**—e.g., *he, him, she, her*, etc.

- Another important aspect of personalization is the degree to which speakers (writers) and interlocuters (audience) are involved in the overall narrative. Accordingly, we computed the relative frequency of **local (speaker/addressee) personal pronouns**: *I, me, my, mine, you, your, yours*, etc.
- Using information from VerbNet, we computed the relative frequency of predicates that require an **animate agentive subject**, e.g. *[Mark Carney] <u>increased</u> interest rates.*
- Also using information from VerbNet, we computed the relative frequency of predicates that require an **animate experiencer subject or object**, e.g. *[Mark Carney] <u>feels</u> that the economy is recovering*, *The downturn has <u>frightened</u> [people] into drawing out their deposits.*.
- Again, using information from VerbNet, we computed the number and proportion of predicates that require an **animate patient subject or object**, e.g. *[Many people] have <u>died</u>*, *The coronavirus crisis has <u>killed</u> [many people] directly or indirectly.*
- The relative frequency of the words ***people*** and ***person***.

5.3.5. Proximity

Consumers typically prefer news relating to events that have happened closer to them in some sense—usually taken to be geographically or culturally (e.g. Galtung and Ruge 1965). We operationalized this dimension of newsworthiness as follows.

- First, to measure **geographic proximity** we took the relative occurrence of terms such as *UK, English, Scottish, British,* etc., in the document.
- Second, we chose to operationalize **cultural proximity** by measuring how close the text is to British English (versus, say, American English). For each word in the text, we measured its relative frequency in two corpora—a British English corpus and an American English corpus. We then used the log-likelihood ratio of the word, i.e. $\log(P(w_{BrE})/P(w_{AmE}))$ as a cultural proximity score for that word. We then took the mean over the document to produce an overall cultural proximity score for the document. Higher values indicate closer cultural proximity of the text to British English.

5.3.6. Facticity

The famous ex-BBC reporter Martin Bell (Bell 1991) remarks that the newsworthiness of a story partly depends on "the degree to which a story contains the kinds of facts and figures on which hard news thrives: locations, names, sums of money, numbers of all kinds." To measure this news value, we measured the proportion of each **entity type** in Table 9. (Note that we exclude PERSON here, because this is treated under PERSONALIZATION.)

Table 9: Named Entity Types

Tag	Description
NORP	Nationalities or religious or political groups.
FAC	Buildings, airports, highways, bridges, etc.
ORG	Companies, agencies, institutions, etc.
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water.
PRODUCT	Objects, vehicles, foods, etc. (Not services.)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART	Titles of books, songs, etc.
LAW	Named documents made into laws.
LANGUAGE	Any named language.
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day.
PERCENT	Percentage, including %.
MONEY	Monetary values, including unit.
QUANTITY	Measurements, as of weight or distance.
ORDINAL	first, second, etc.
CARDINAL	Numerals that do not fall under another type.

5.3.7. Uncertainty

There is a preference for events that are certain and unambiguous. For instance, in their seminal article, Galtung and Ruge (1965, p. 66) note that "an event with a clear interpretation, free from ambiguities in its meaning, is preferred to the highly ambiguous event from which many and inconsistent implications can and will be made". We operationalized this news value by annotating for two features.

- We measured the proportion of words in a document that are in Loughran and B. McDonald (2015)'s **uncertainty lexicon**.
- We measured the proportion of words that are **modal verbs** in the document (e.g. *may, might,*, etc.).

5.3.8. Prominence

Events that involve prominent individuals and organizations are ripe for reportage. For example, Golding and Elliott (1979, p. 122) remark that "big names are better news than nobodies, major personalities of more interest than ordinary folk". Although there might be more sophisticated ways to measure this news value,[22] we chose to operationalize it by counting the **number of references to BoE governors** (normalized for document length).

[22] For instance, we refer the interested reader to the features discussed in Piotrkowicz (2017).

5.3.9. Novelty

News *needs* to be novel in order to become picked up. For example, Dijk (1988) observes, '[t]he requirement that news should in principle be about new events is fundamental.' We chose to operationalize novelty in two ways.

- First, we computed the relative frequency of clauses introduced by existential-*there* (such clauses typically introduce new discourse entities onto the scene).
- Second, we evaluated the textual (dis)similarity between the target document and all other documents published in the prior 30 days before the target document's publication.

Altogether, θ_B comprises of a total of 351 features that we chose to measure based on an extensive review of the literature.

In this way our approach differs from other studies of central bank text that primarily use NLP to reduce dimensionality (S. Hansen, McMahon, and Tong 2019; Larsen, Thorsrud, and Zhulanova 2020; Munday 2019). We measure *specific* features of said communication in order to investigate their relationship to an outcome variable (in our case reporting in the media).

6. Estimation

We have discussed how the variables contained in the model are measured. We now want to estimate the equation given by the structural model (Equation 21) to determine which features of central bank communication and which features of the state of the economy are associated with increased news coverage.

6.1. Method

Our framework, detailed in Section 2, delivered a model equation: Equation 21. Appending an approximation error to this equation gives:

$$\frac{1-k_t}{k_t} = \beta_0 + \beta_1^T(\theta_{B,t}^T\theta_{B,t}) + \beta_2^T\theta_{B,t} + \beta_3^T(z_t \otimes \theta_{B,t}) + \beta_4^T z_t + \beta_5^T(z_t \otimes z_t) + \epsilon_t \tag{24}$$

Equation 24 suffers from a dimensionality problem. The kroeneker product terms cause the number of coefficients we want to estimate to be substantially larger than the number of observations. For example, if one only wanted to measure 100 features of the text, and 10 features of the economy, that would result in 1310 composite features in total, close to the total number of observations.

If the goal was to *predict* $\frac{1-k}{k}$ then we could apply an approximately sparse regression model (e.g. a LASSO model a la Robert Tibshirani (1996)) that implemented a regularisation approach to shrink the dimension to an appropriate size.

Our goal, however, is perform causal inference on the parameters. And methods that perform well at prediction often achieve that predictive ability at the cost of biased or non-consistent coefficient estimates (Leeb and Pötscher 2008a; Leeb and Pötscher 2008b). If one tried, for example, to select a subset of variables that were important from Equation 24 using a LASSO model, and then do causal inference by performing OLS on the selected variables, then substantial omitted variable bias would be likely to occur. For example, suppose a member of θ and a member of z are highly correlated with each other. From the perspective of prediction, including both is inefficient, and so one variable will likely be dropped in the process of regularisation — suppose for arguments sake it is the member of z. But now we have excluded a variable that is highly correlated with a variable of interest, the member of θ, leading to significant omitted variable bias.

This problem is not just a possibility in our case. The solution to the central bank's problem, Equation 12, suggests that the correlation between the controls (z) and the variables of interest (θ) is likely to be strong. Dropping variables that are not predictive of $\frac{1-k}{k}$ in order to overcome the dimensionality issue will prevent us from performing inference.

We turn to the de-sparsified LASSO of Van de Geer, Bühlmann, Ritov, Dezeure, et al. (2014) to perform estimation. The de-sparsified LASSO (sometimes called the de-biased LASSO) is a semi-parametric method which allows the researcher to perform inference on a subset of parameters in a high-dimensional model. We follow the treatment of Adamek, Smeekes, and Wilms (2020) who establish the uniform asymptotic normality of the de-sparsified LASSO in the time series case under Near-Epoch Dependence.

The de-sparisifed LASSO is a shrinkage method. Shrinkage methods apply structure to the parameter vector that the researcher wants to perform inference on in order to circumvent the problem of high-dimensionality. In our case we apply an assumption of weak sparsity to the parameter vector

$[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5]$: within the true structural parameter vector, there are only a few entries that are not exactly or close to zero.

The de-sparsified LASSO applies shrinkage to the parameter vector, thus performing variable selection. However, this variable selection results in the exact post-selection inference issue outlined before (Leeb and Pötscher 2008a). To correct for this the de-sparsified LASSO uses node-wise regressions (regressions of each variable on the right hand side on all other regressors) to de-bias the estimates of the parameter vector.

Alternative methods for performing valid inference in high-dimensional settings include those that use selective inference — i.e. performing inference conditional on a model selected via shrinkage (Tian and J. Taylor 2017; J. Taylor and Robert Tibshirani 2018; Fithian, Sun, and J. Taylor 2014; R. J. Tibshirani, Rinaldo, Rob Tibshirani, Wasserman, et al. 2018) typically under the assumption of IID data; and orthogonalizing the parameter of interest to the estimation of the other parameters using double selection (sometimes called double machine learning) (Belloni, Chernozhukov, and C. Hansen 2014; Chernozhukov, Chetverikov, et al. 2018), a method that has been extended to various time series cases in Chernozhukov, Härdle, Huang, and Wang (2019), Hecq, Margaritella, and Smeekes (2019), and Babii, Ghysels, and Striaukas (2020).

The de-sparsified LASSO of Van de Geer, Bühlmann, Ritov, Dezeure, et al. (2014) provides estimates of the parameter vector in the form:

$$\hat{b} = \hat{\beta} + \frac{\hat{\Theta} X^T (y - X\hat{\beta})}{T}$$

Where $\hat{\beta}$ is the biased parameter vector from a typical LASSO regression of y on X, $\hat{b}$ are the "corrected" estimates of the parameter vector, and $\hat{\Theta}$ is a matrix constructed from the set of node-wise regressions.

More specifically, the lasso estimates from the nodewise regressions yield the parameters:

$$\hat{\gamma}_j := argmin\left(\frac{||x_j - X_{-j}\gamma_j||_2^2}{T} + 2\lambda_j ||\gamma_j||_1\right) \tag{25}$$

where X_{-j} is X with x_j removed.

We can also extract the estimated loss from the loss functions of the nodewise regressions:

$$\hat{\tau}_j^2 := \frac{1}{T}||x_j - X_{-j}\hat{\gamma}_j||_2^2 + 2\lambda_j ||\hat{\gamma}_j||_1 \tag{26}$$

Defining $\hat{\Gamma}$ as the stacked matrix of the parameter vectors $\hat{\gamma}$ with ones along the diagonal, and $\hat{\Upsilon}^{-2} := diag(\frac{1}{\hat{\tau}_1^2}, ..., \frac{1}{\hat{\tau}_N^2})$ then we can write that $\hat{\Theta} := \hat{\Upsilon}^{-2}\hat{\Gamma}$.

Adamek, Smeekes, and Wilms (2020) show that under the assumption of weak sparsity of the parameter vector, and other general conditions that the desparsified LASSO is asymptotically normal, including where inference is performed on weakly dependent data and where the "errors may exhibit serial dependence, heteroskedasticity and fat tails". This allows us to perform valid inference on the estimated parameter vector in the case this paper describes.

Furthermore, one should note that our identification scheme is based on (i) the event study methodology outlined in Section 4 removing confounding factors, and (ii) a selection on observables approach through controlling for the state of the economy and a large number of textual features. There is no natural experiment for us to exploit, and so whilst — to the extent to which we can — we aim for unbiased estimates of the effects we document, we treat these primarily as associations between variables rather than truly causal links.

In total we estimate all 4695 parameters of Equation 24 using 1211 instances of Bank of England communication and their corresponding news coverage.

6.2. Results

Before we perform inference on elements of the parameter vector $[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5]$, we ask a broader question. Is the information in the right hand side variables of Equation 21 significant in explaining the variance in the left hand side variable? Or in other words, is there information in textual features and the state of the economy that explains the extent to which central bank communication is reported on in the news?

In a non-high-dimensional setting the answer could be found by performing an F-test on the entire set of regressors. Unfortunately, owing to the number of regressors in our model, this is not possible. We follow the approach in Bühlmann et al. (2013) and Van de Geer, Bühlmann, Ritov, Dezeure, et al. (2014) in which a test statistic regarding the significance of a group of variables can be constructed using the maximum individual test statistic of the group when performing inference via the de-sparsified LASSO.[23] The p-values for these tests are displayed in Table 10.

Firstly, we find that textual features matter. The p-value for whether θ_B and $\theta_B^T\theta_B$ matter, i.e. whether one can set $\beta_1 = \beta_2 = 0$ to zero, is zero to three decimal places. Secondly, we find that the *interaction* between the state of the economy and textual features is also important: the p-value that $\beta_3 = 0$ is also zero to three decimal places. We find that the state of the economy on its own is not significant in explaining news coverage. The p-value for whether $\beta_4 = \beta_5 = 0$ is not below the critical value of 0.05, so we cannot reject this null hypothesis. Finally, overall, the regressors in Equation 21 explain a significant proportion of the variance in the dependent variable (see the final row of Table 10 — with the most useful variables for explaining news coverage either textual features or their interactions with the economy.

Table 10: Significance of groups of variables

H0	No. of coefficients	p-value
$\beta_1 = \beta_2 = \vec{0}$	752	0.000
$\beta_3 = \vec{0}$	3861	0.000
$\beta_4 = \beta_5 = \vec{0}$	132	0.301
$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \vec{0}$	4695	0.000

[23] Specifically, as in (Van de Geer, Bühlmann, Ritov, Dezeure, et al. 2014), for any fixed group G, conditionally on the features X, the asymptotic distribution of

$$\max_{j \in G} n|\hat{b_j}|^2 / \sigma_\epsilon^2 \hat{\Omega_{j,j}}$$

, where $\hat{b}$ are the desparsified coefficient estimates, σ^2 is a consistent estimate for the error variance, and $\hat{\Omega} = \hat{\Theta}\hat{\Sigma}\hat{\Theta}^T$, under the null hypothesis that $\beta_j = 0 \quad \forall j \in G$ is asymptotically equal to the maximum of dependent $\chi^2(1)$ variables whose distribution can be simulated.

Turning to inference on individual parameters we can ask the question: which variables are significant in explaining news coverage? Inference on individual parameters raises another issue related to the high-dimensionality of our equation — namely that of multiple comparisons. When testing 4695 coefficients for significance, as we are in our case, a regression that comprised purely of white noise variables as both dependent and independent variables would result in approximately 5% of coefficients returning as "significant" under a critical p-value of 0.05. To adjust for this, we use the Benjamini-Hochberg adjustment (Benjamini and Hochberg 1995) to control the False Discovery Rate (the expected proportion of false discoveries amongst the rejected hypotheses). Table 11 shows the coefficients that are significant at the 5% global False Discovery Rate. One can see that the corrected p-values decline by around two orders of magnitude compared to the unadjusted p-values, once we correct for the multiple comparisons problem.

It is important to note that our explanatory variables are correlated with one another. As a result, the significance of one variable may substitute for the significance of another. This issue is, of course, one that all multiple regression analyses face. For example, it could be that the Inflation Report gets much more news coverage than the Q & A, but that is not reflected in a significant coefficient for an Inflation Report dummy variable, rather it is picked up as a significant coefficient on the topic of 'Inflation' which occurs more often in the Inflation Report than in the Q & A.

Table 11: Significant Coefficients

Definition	$\hat{\beta}$	Class	p-value	Corrected p-value
Topic 40 ('Bonds') × IP growth inv. var	-162.62	Topics	1.67e-07	7.10e-05
proportion VB × IP growth inv. var	283.10	Lexical processing	1.58e-07	7.10e-05
proportion WP × IP growth inv. var	219.75	Lexical processing	1.35e-07	7.10e-05
proportion TO × infl inv. var	207.70	Lexical processing	1.81e-07	7.10e-05
mean number of content words in doc × IP growth inv. var	153.21	Lexical processing	2.29e-04	4.14e-02
proportion PRP × IP growth inv. var	-142.50	Lexical processing	2.80e-04	4.69e-02
proportion particle (universal) × IP growth inv. var	-160.40	Lexical processing	4.48e-05	1.24e-02
proportion PRP × unemp rate	-172.61	Lexical processing	2.26e-04	4.14e-02
mean contextual expectancy for word × IP growth inv. var	-184.13	Lexical processing	1.70e-04	3.46e-02
proportion particle (universal) × infl inv. var	-186.44	Lexical processing	1.15e-06	4.15e-04
proportion RP × infl inv. var	-187.84	Lexical processing	6.07e-05	1.58e-02
proportion PRP squared	-207.55	Lexical processing	1.23e-04	2.75e-02
proportion PRP × IP growth inv. var	-253.89	Lexical processing	1.77e-07	7.10e-05
proportion RP × IP growth inv. var	-285.00	Lexical processing	8.12e-10	7.62e-07
degree to which the first sentence is a 'headline' squared	-89.32	Discourse Processing	1.43e-05	4.46e-03
mean number of full auxiliaries per sentence × infl inv. var	300.74	Syntax Processing	1.73e-12	2.02e-09
mean dependency arc lengths per sentence × IP growth inv. var	230.11	Syntax Processing	6.81e-06	2.28e-03
prt rate per sentence × IP growth inv. var	188.53	Syntax Processing	8.94e-05	2.10e-02
prt rate per sentence × infl inv. var	180.94	Syntax Processing	2.53e-04	4.40e-02
prt rate per sentence squared	132.73	Syntax Processing	1.99e-04	3.90e-02
number CONJP squared	-106.12	Syntax Processing	6.67e-05	1.65e-02
number SQ × infl inv. var	-214.36	Syntax Processing	1.23e-07	7.10e-05
proportion resultative conjucts × infl inv. var	142.84	Newsworthiness	1.35e-04	2.87e-02
proportion MONEY × infl	-244.92	Newsworthiness	3.75e-08	2.94e-05
proportion MONEY squared	-261.06	Newsworthiness	3.61e-17	8.48e-14
proportion MONEY × IP growth inv. var	-297.52	Newsworthiness	3.78e-21	1.78e-17
proportion governor × infl inv. var	-340.46	Newsworthiness	7.56e-14	1.18e-10
Q & A Dummy × infl inv. var	157.17	Other	2.93e-05	8.60e-03

Before we analyse the significant coefficients in Table 11 it is worth briefly discussing what is *not* significant. The dummy variable that denoted that a monetary policy decision took place on the same day as a communication was insignificant. This is corroborated by the fact that the state of the economy z and its kroeneker product $z \otimes z$ are both jointly not significant, as shown in Table 10, and have no single variable that is significant at the 5% level in Table 11. This not only suggests that slow moving variables that capture the state of the economy (GDP, the unemployment rate, CPI Inflation) do not — on their own — influence news coverage of the Bank of England's communication, but that changes in monetary policy stance captured by the one-day difference in the 1-year OIS rate also do not influence news coverage of the Bank' communication.

There is one main broad conclusion one can draw from the significant coefficients in Table 11. The interaction between textual features and the state of the economy is extremely important. Of the 28 significant coefficients, all but five are interaction terms. How you write your communication matters — but how much it matters is state-dependent to a large degree. More surprisingly, the state-dependence of the impact of textual features on news coverage largely owes to the inverse variance of state variables.

Of the 23 significant coefficients that are interaction terms between textual features and state variables, 21 include the inverse variance of a state variable. What does this tell us? That is not the "cycle" that economists traditionally think of that is important in determining news coverage, rather it is the second moment that matters. For example, information from the Q and A gets more news coverage when the inverse variance of CPI inflation is low. Or in other words, the Q and A is more likely to be reported on when the variance of inflation is high, and potential comments on its direction and central bank reactions to it are at their most newsworthy.

We now turn to a discussion of the significant coefficients for the textual variables in Table 11. Before we do so, we make a number of preliminary comments. First, we wish to stress that we must be careful not to interpret the individual effects at too fine-grained a level. The feature space is highly correlated, as happens in observational linguistic designs. To highlight one example, word frequency and word bulk are famously related (for the precise mathematical details, see e.g. Kornai (2007) for discussion).

Second, as noted, the important textual features are interactions with features relating to the state of the economy. In what follows, we concentrate exclusively on the textual part of the interaction, but the reader should always bear in mind that the effect is conditional on the sate of the economy.

Finally, we constructed our model with $\frac{1-k}{k}$ as the response. Thus, a *negative* sign on a coefficient indicates that for an increase in the variable of interest, news coverage *increases*. On the other hand, a *positive* sign on the coefficient indicates that for an increase in the variable of interest, news coverage *decreases*.

Topic Conditional on the state of the economy, text that discusses the bond market receives more news coverage in the popular press. There are potentially many reasons for this, foremost amongst them the prominence of QE as a tool in the period studied, and its effect on bond yields. Interestingly, the effect of discussing other topics is statistically insignificant, including inflation and the labour market. This may be due to the (relative) stability of these markets in the face of economic crises when compared to earlier periods in economic history. Or it could be that enough noise has been introduced to our estimation procedure as to attenuate some of our estimated coefficients towards zero. Future work could utilise evidence from consumer surveys of which topics people view as important to place priors over the coefficients and proceed in a bayesian manner.

Linguistic Processing Our linguistic processing features concern lexical processing, sentence processing, and discourse processing.

Lexical Processing Dealing with lexical processing first, we find that increased textual amounts of base forms of verbs, infinitival-*to*, *wh*-pronouns, and content words result in less text being picked up by the press (conditional on the state of the economy). On the other hand, increased amounts of personal pronouns, adverbs, particles, and greater word contextual expectancy result in more text being picked up by the press (conditional on the state of the economy).

Some of these effects can be explained by reference to processing complexity, that is, keeping things simple so that the consumer can easily understand the message. For instance, base forms of verbs together with *to*-infinitives introduce non-finite embedded clauses and *wh*-pronouns introduce relative clauses, which are another type of embedded clause. Particular types of clausal embedding, such as instances in which the relative pronoun has an object rather than a subject grammatical function, involve higher processing costs than simpler structures (see e.g. Bornkessel-Schlesewsky and Schlesewsky 2009, , particularly Chapter 10). Further, *wh*-clause processing also involves coreference resolution—that is, the referent of the *wh*-expression also needs to be established in order to fully process the clause. Similarly, the greater the expectancy of a word in the discourse context makes it easier to retrieve words from the mental lexicon, integrate and encode in the discourse, and recall.[24]

We interpret the negative coefficient on the proportion of personal pronouns to relate to the newsworthiness feature of *personalization* and the related notion *popularization*, the latter being an aspect of news discourse we have hitherto not discussed in this paper. Popularization refers to the use in journalistic prose of a more colloquial, involved style designed to engage the audience in the narrative and attract a wider lay readership (Douglas 2003; Biber and Gray 2012). The extensive use of personal pronouns, such as *we, us, you*, marks an interpersonal focus that characterizes this style of discourse.[25]

Syntax Processing Conditional on the state of the economy, the use of fully realized auxiliaries, constructions involved long dependency arcs, and particle-verb constructions result in decreased news pick-up for texts. On the other hand, the use of conjunction phrases and main clause interrogatives result increased news pick-ip.

We surmise that the coefficients on the dependency arc length variable and the particle-verb variable reflect the choice of newspaper editors to avoid copying complex language. Recall from the discussion in Section 5 that distances between need to be minimized to facilitate processing. Particle verb structures, which involve a lexical verb and a prepositional or adverbial particle (for instance *pick... up, slow ... down*, are typically discontinuous: the verb has to maintained in working memory until it can be integrated with the particle. Too many of such discontinuities in a given sentence will induce increased processing costs.[26]

Discourse Processing The LASSO model selects only one variable from the set of discourse processing features, namely a text's headlining score. Recall that this feature indicates the degree to which the first sentence in a text summarizes (i.e. serves as a headline for) the main content, and was oper-

[24] Prima facie, it is somewhat curious as to why a larger relative frequency of content words in a document results in a text getting less news coverage, given that content words are by definition more informative than function words (grammatical markers and such). However, content words are generally less frequent and thus harder to process than function words, and it may be that this feature is masking the more basic effect of word frequency.

[25] Biber and Gray 2012 refer to an 'information explosion' in the news press recent years, with "the associated need for economy of expression as there is more information to be communicated". Stories are abundant and page space is limited. As such, while specificity is required, the information needs to be communicated in a compressed, compact style. Contentful adverbial expressions add detail and specific information to the eventualities being described, whereas particles such as the genitive marker *-'s* encode information more economically than other types of phrase (Hinrichs and Szmrecsanyi 2007). Note that the coefficients on *to*-infinitives and *wh*-pronouns could be similarly interpreted (in addition to the processing interpretation discussed above). For example, *to*-infinitives can often be omitted and relative clauses introduced by *wh*-pronouns can be more economically encoded by null pronouns or more compactly realized as a preverbal adjectival clause rather than a postverbal relative clause.

[26] Economization and popularization, mentioned above, may be helpful in understanding the effect of other variables. Full auxiliaries (e.g. *will, is, have*) have informationally more compact contracted variants (*-'ll, -'s. -'ve*) (Krug 1994). The same explanation may also explain the dispreference in news reportage for particle-verbs, which are textually less economical than single-word verbs. In addition, contracted variants occur relatively more frequently in speech and interactive writing styles (Biber 1991), putatively making the text seem less formal and more accessible. We may also appeal to popularization to explain the significant effect seen for — they make the text lively, engaging, and interactive.

ationalized using the doc2vec similarity between the first sentence and the rest of the text. The sign on the coefficient for this feature indicates that documents in which the first sentence more successfully summarizes the main content are result in more news coverage.

Why should this feature matter? It has been consistently shown that individuals process passages better when they are given context, whether that is a picture, a title, or a summary first sentence introducing the topic, before reading the passage (Dooling and Lachman 1971; Bransford and Johnson 1972; Cirilo and Foss 1980; Haberlandt, Berian, and Sandson 1980; Kieras 1980). This context provides an effective frame within which the main material that follows is understood ("global cohesion").

Note in addition that this feature is relevant regardless of the state of the econonomy.

Newsworthiness Finally, we see that several features pertaining to a story's newsworthiness are relevant. Specifically, increases in the relative frequency of money mentions in the text result in greater news reportage. One reason for this is that news editors and journalists prize a story's *facticity*, the extent to which it includes facts and figures. In a central banking context, references to money are facts and figures par excellence.

We also see the importance of the *prominence* aspect of newsworthiness in the sign on the coefficient for textual mentions to the governor.

All in all we provide five concrete proposals for improving the likelihood of central bank communication being reported on in the news, and therefore reaching its intended audience:

1. Keep things simple. Our results show that one should avoid introducing embedded clauses and particle verb structures.
2. Personalize the text. Use we/us/you to engage the reader.
3. Write in short sentences. Long dependence arcs reduce the likelihood of newspaper coverage.
4. Summarise your message in the first sentence.
5. Use facts and figures.

Central banks should be aware that the effectiveness of these measures is dependent on the state of the economy. We found that altering the style of text often had differing effects on news coverage depending on the volatility of the economy. Nonetheless, we believe that applying the above suggestions to central bank communication will improve news coverage and ultimately help central banks reach a wider audience.

7. Conclusion

We have examined which features of central bank communication are important for whether that communication reaches the general public via the print media. To do so, we wrote a model of news production and consumption. We measured the variables in our model using a series of computational linguistic techniques, including a comprehensive set of features that could matter for whether a communication is reported on. We estimated our model using machine learning techniques, and found that it's not only *what* you say that matters, but also *how* you say it. Our paper has concrete policy suggestions for how central banks should mould their communication if they want it be newsworthy.

References

Adamek, Robert, Stephan Smeekes, and Ines Wilms (2020). “LASSO inference for high-dimensional time series”. In: *arXiv preprint arXiv:2007.10952*.

Adelman, James S, Gordon DA Brown, and José F Quesada (2006). “Contextual diversity, not word frequency, determines word-naming and lexical decision times”. In: *Psychological science* 17.9, pp. 814–823.

Almor, Amit and Peter D Eimas (2008). “Focus and noun phrase anaphors in spoken language comprehension”. In: *Language and cognitive processes* 23.2, pp. 201–225.

Amadxarif, Zahid, James Brookes, Nicola Garbarino, Rajan Patel, and Eryk Walczak (2019). “The language of rules: textual complexity in banking reforms”. In:

Andrews, Sally (1989). “Frequency and neighborhood effects on lexical access: Activation or search?” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15.5, p. 802.

Arnold, Jennifer E, Janet G Eisenband, Sarah Brown-Schmidt, and John C Trueswell (2000). “The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking”. In: *Cognition* 76.1, B13–B26.

Babii, Andrii, Eric Ghysels, and Jonas Striaukas (2020). “Inference for High-Dimensional Regressions With Heteroskedasticity and Auto-correlation”. In: *Available at SSRN 3615718*.

Baker, Collin F, Charles J Fillmore, and John B Lowe (1998). “The berkeley framenet project”. In: *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pp. 86–90.

Baker, Scott R, Nicholas Bloom, and Steven J Davis (2016). “Measuring economic policy uncertainty”. In: *The quarterly journal of economics* 131.4, pp. 1593–1636.

Balota, David A et al. (2007). “The English lexicon project”. In: *Behavior research methods* 39.3, pp. 445–459.

Bednarek, Monika and Helen Caple (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford University Press.

Bell, Allan (1991). *The language of news media*. Blackwell Oxford.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). “Inference on treatment effects after selection among high-dimensional controls”. In: *The Review of Economic Studies* 81.2, pp. 608–650.

Benjamini, Yoav and Yosef Hochberg (1995). “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *Journal of the Royal statistical society: series B (Methodological)* 57.1, pp. 289–300.

Berger, Helge, Michael Ehrmann, and Marcel Fratzscher (2011). “Monetary policy in the media”. In: *Journal of Money, Credit and Banking* 43.4, pp. 689–709.

Bholat, David, Nida Broughton, Janna Ter Meer, and Eryk Walczak (2019). “Enhancing central bank communications using simple and relatable information”. In: *Journal of Monetary Economics* 108, pp. 1–15.

Biber, Douglas (1991). *Variation across speech and writing*. Cambridge University Press.

Biber, Douglas and Bethany Gray (2012). "THE COMPETING DEMANDS OF POPULARIZATION VS. ECONOMY". In: *The Oxford handbook of the history of English*, p. 314.

Binder, Carola (2017a). "Fed speak on main street: Central bank communication and household expectations". In: *Journal of Macroeconomics* 52, pp. 238–251.

— (2017b). "Federal reserve communication and the media". In: *Journal of Media Economics* 30.4, pp. 191–214.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Blinder, Alan S (2009). "Talking about monetary policy: the virtues (and vice?) of central bank communication". In:

Blinder, Alan S, Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen (2008). "Central bank communication and monetary policy: A survey of theory and evidence". In: *Journal of Economic Literature* 46.4, pp. 910–45.

Blinder, Alan S and Alan B Krueger (2004). *What does the public know about economic policy, and how does it know it?* Tech. rep. National Bureau of Economic Research.

Born, Benjamin, Michael Ehrmann, and Marcel Fratzscher (2014). "Central bank communication on financial stability". In: *The Economic Journal* 124.577, pp. 701–734.

Bornkessel-Schlesewsky, Ina and Matthias Schlesewsky (2009). *Processing syntax and morphology: A neurocognitive perspective*. Vol. 6. Oxford University Press.

Bransford, John D and Marcia K Johnson (1972). "Contextual prerequisites for understanding: Some investigations of comprehension and recall". In: *Journal of verbal learning and verbal behavior* 11.6, pp. 717–726.

Brunato, Dominique, Lorenzo De Mattei, Felice Dell'Orletta, Benedetta Iavarone, and Giulia Venturi (2018). "Is this Sentence Difficult? Do you Agree?" In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2690–2699.

Brysbaert, Marc, Paweł Mandera, Samantha F McCormick, and Emmanuel Keuleers (2019). "Word prevalence norms for 62,000 English lemmas". In: *Behavior research methods* 51.2, pp. 467–479.

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014). "Concreteness ratings for 40 thousand generally known English word lemmas". In: *Behavior research methods* 46.3, pp. 904–911.

Bühlmann, Peter et al. (2013). "Statistical significance in high-dimensional linear models". In: *Bernoulli* 19.4, pp. 1212–1242.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko (2020). *Communication and the Beliefs of Economic Agents*. Tech. rep. National Bureau of Economic Research.

Caple, Helen (2018). "News values and newsworthiness". In: *Oxford Research Encyclopedia of Communication*.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia (2017). "Inflation expectations, learning, and supermarket prices: Evidence from survey experiments". In: *American Economic Journal: Macroeconomics* 9.3, pp. 1–35.

Charlet, Delphine and Geraldine Damnati (2017). "Simbow at semeval-2017 task 3: Soft-cosine semantic similarity between questions for community question answering". In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 315–319.

Chernozhukov, Victor, Denis Chetverikov, et al. (2018). “Double/debiased machine learning for treatment and structural parameters”. In: *The Econometrics Journal* 21.1, pp. C1–C68.

Chernozhukov, Victor, Wolfgang K Härdle, Chen Huang, and Weining Wang (2019). “LASSO-driven inference in time and space”. In: *Available at SSRN 3188362*.

Cirilo, Randolph K and Donald J Foss (1980). “Text structure and reading time for sentences”. In: *Journal of Verbal Learning and Verbal Behavior* 19.1, pp. 96–109.

Clark, Kevin and Christopher D Manning (2016). “Deep reinforcement learning for mention-ranking coreference models”. In: *arXiv preprint arXiv:1609.08667*.

Cochrane, John H and Monika Piazzesi (2002). “The fed and interest rates-a high-frequency identification”. In: *American Economic Review* 92.2, pp. 90–95.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten Van Rooij (2019). *How does consumption respond to news about inflation? Field evidence from a randomized control trial*. Tech. rep. National Bureau of Economic Research.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar (2018). “How do firms form their expectations? new survey evidence”. In: *American Economic Review* 108.9, pp. 2671–2713.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar, and Mathieu Pedemonte (2020). “Inflation expectations as a policy tool?” In: *Journal of International Economics*, p. 103297.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele (2020). “Inflation expectations and firm decisions: New causal evidence”. In: *The Quarterly Journal of Economics* 135.1, pp. 165–219.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber (2019). *Monetary policy communications and their effects on household inflation expectations*. Tech. rep. National Bureau of Economic Research.

Coltheart, M, E Davelaar, JT Jonasson, D Besner, and S Dornic (1977). *Vol. vol. 6*.

D’Acunto, Francesco, Daniel Hoang, and Michael Weber (2020). *Managing Households’ Expectations with Unconventional Policies*. Tech. rep. National Bureau of Economic Research.

Demberg, Vera and Frank Keller (2008). “Data from eye-tracking corpora as evidence for theories of syntactic processing complexity”. In: *Cognition* 109.2, pp. 193–210.

Dijk, Teun A van (1988). *News as discourse*. Hillsdale: Lawrence Erlbaum.

Dooling, D James and Roy Lachman (1971). “Effects of comprehension on retention of prose.” In: *Journal of experimental psychology* 88.2, p. 216.

Douglas, Biber (2003). “Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs economy”. In: *New media language*, pp. 169–81.

Ehrmann, Michael and Jonathan Talmi (2020). “Starting from a blank page? Semantic similarity in central bank communication and market volatility”. In: *Journal of Monetary Economics* 111, pp. 48–62.

Ellen, Saskia ter, Vegard H Larsen, and Leif Anders Thorsrud (2019). “Narrative monetary policy surprises and the media”. In:

Embick, David, Martin Hackl, Jeannette Schaeffer, Melltem Kelepir, and Alec Marantz (2001). “A magnetoencephalographic component whose latency reflects lexical frequency”. In: *Cognitive Brain Research* 10.3, pp. 345–348.

Federmeier, Kara D and Marta Kutas (1999). “A rose by any other name: Long-term memory structure and sentence processing”. In: *Journal of memory and Language* 41.4, pp. 469–495.

Fithian, William, Dennis Sun, and Jonathan Taylor (2014). "Optimal inference after model selection". In: *arXiv preprint arXiv:1410.2597*.

Forster, Kenneth I and Susan M Chambers (1973). "Lexical access and naming time". In: *Journal of verbal learning and verbal behavior* 12.6, pp. 627–635.

Fullwood, Jon (2016). "A cat, a hat and a simple measure of gobbledygook: how readable is your writing". In: *Bank Underground*. URL: `https://bankunderground.co.uk/2016/10/04/a-cat-a-hat-and-asimple-measure-of-gobbledygook-how-readable-is-your-writing/`.

Galtung, Johan and Mari Holmboe Ruge (1965). "The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers". In: *Journal of peace research* 2.1, pp. 64–90.

Gentzkow, Matthew and Jesse M Shapiro (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71.

Gertler, Mark and Peter Karadi (2015). "Monetary policy surprises, credit costs, and economic activity". In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.

Gibson, Edward (1998). "Linguistic complexity: Locality of syntactic dependencies". In: *Cognition* 68.1, pp. 1–76.

— (2000). "The dependency locality theory: A distance-based theory of linguistic complexity". In: *Image, language, brain* 2000, pp. 95–126.

Golding, Peter and Philip Ross Courtney Elliott (1979). *Making the news*. Longman Publishing Group.

Gonzalez-Garduno, Ana Valeria and Anders Søgaard (2017). "Using gaze to predict text readability". In: *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 438–443.

Grober, Ellen H, William Beardsley, and Alfonso Caramazza (1978). "Parallel function strategy in pronoun assignment". In: *Cognition* 6.2, pp. 117–133.

Guay, Wayne, Delphine Samuels, and Daniel Taylor (2016). "Guiding through the fog: Financial statement complexity and voluntary disclosure". In: *Journal of Accounting and Economics* 62.2-3, pp. 234–269.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson (2005). "The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models". In: *American economic review* 95.1, pp. 425–436.

Haberlandt, Karl, Claire Berian, and Jennifer Sandson (1980). "The episode schema in story processing". In: *Journal of Verbal Learning and Verbal Behavior* 19.6, pp. 635–650.

Haldane, Andrew, Alistair Macaulay, and Michael McMahon (2020). "The 3 E's of central bank communication with the public". In:

Haldane, Andrew and Michael McMahon (2018). "Central bank communications and the general public". In: *AEA Papers and Proceedings*. Vol. 108, pp. 578–83.

Hansen, Stephen and Michael McMahon (2016). "Shocking language: Understanding the macroeconomic effects of central bank communication". In: *Journal of International Economics* 99, S114–S133.

Hansen, Stephen, Michael McMahon, and Matthew Tong (2019). "The long-run information effect of central bank communication". In: *Journal of Monetary Economics* 108, pp. 185–202.

Harcup, Tony and Deirdre O'neill (2017). "What is news? News values revisited (again)". In: *Journalism studies* 18.12, pp. 1470–1488.

Haviland, Susan E and Herbert H Clark (1974). “What’s new? Acquiring new information as a process in comprehension”. In: *Journal of verbal learning and verbal behavior* 13.5, pp. 512–521.

Hayo, Bernd and Matthias Neuenkirch (2012). “Bank of Canada communication, media coverage, and financial market reactions”. In: *Economics Letters* 115.3, pp. 369–372.

— (2015). “Self-monitoring or reliance on media reporting: How do financial market participants process central bank news?” In: *Journal of Banking & Finance* 59, pp. 27–37.

Hecq, Alain, Luca Margaritella, and Stephan Smeekes (2019). “Granger causality testing in high-dimensional VARs: a post-double-selection procedure”. In: *arXiv preprint arXiv:1902.10991*.

Hendry, Scott (2012). *Central bank communication or the media’s interpretation: what moves markets?* Tech. rep. Bank of Canada.

Hendry, Scott and Alison Madeley (2010). “Text mining and the information content of Bank of Canada communications”. In: *Available at SSRN 1722829*.

Hinrichs, Lars and Benedikt Szmrecsanyi (2007). “Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora”. In: *English Language and Linguistics* 11.3, p. 437.

Hoffman, Paul, Matthew A Lambon Ralph, and Timothy T Rogers (2013). “Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words”. In: *Behavior research methods* 45.3, pp. 718–730.

Honnibal, Matthew and Ines Montani (2017). “spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing”. To appear.

Howcroft, David M and Vera Demberg (2017). “Psycholinguistic models of sentence processing improve sentence readability ranking”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 958–968.

Howes, Davis H and Richard L Solomon (1951). “Visual duration threshold as a function of word-probability.” In: *Journal of experimental psychology* 41.6, p. 401.

Huddleston, Rodney, Geoffrey K Pullum, et al. (2002). “The cambridge grammar of english”. In: *Language. Cambridge: Cambridge University Press* 1, p. 23.

Jaeger, T Florian and Harry Tily (2011). “On language ‘utility’: Processing complexity and communicative efficiency”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 2.3, pp. 323–335.

Jarociński, Marek and Peter Karadi (2020). “Deconstructing monetary policy surprises-the role of information shocks”. In: *American Economic Journal: Macroeconomics* 12.2, pp. 1–43.

Johnson-Cartee, Karen S (2004). *News narratives and news framing: Constructing political reality*. Rowman & Littlefield Publishers.

Johnston, Robert A and Christopher Barry (2006). “Age of acquisition and lexical processing”. In: *Visual cognition* 13.7-8, pp. 789–845.

Kieras, David E (1980). “Initial mention as a signal to thematic content in technical passages”. In: *Memory & Cognition* 8.4, pp. 345–353.

Kitaev, Nikita, Steven Cao, and Dan Klein (2019). “Multilingual Constituency Parsing with Self-Attention and Pre-Training”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, pp. 3499–3505. DOI: `10.18653/v1/P19-1340`. URL: `https://www.aclweb.org/anthology/P19-1340`.

Kitaev, Nikita and Dan Klein (2018). "Constituency Parsing with a Self-Attentive Encoder". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, pp. 2676–2686. DOI: `10.18653/v1/P18-1249`. URL: `https://www.aclweb.org/anthology/P18-1249`.

Kornai, András (2007). *Mathematical linguistics*. Springer Science & Business Media.

Krug, Manfred (1994). "Contractions in spoken and written English: A corpus-based study of brachychronic language change". MA thesis. Exeter.

Kryvtsov, Oleksiy and Luba Petersen (2020). "Central bank communication that works: Lessons from lab experiments". In: *Journal of Monetary Economics*.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko (2015). *Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand*. Tech. rep. National Bureau of Economic Research.

Kuperman, Victor, Hans Stadthagen-Gonzalez, and Marc Brysbaert (2012). "Age-of-acquisition ratings for 30,000 English words". In: *Behavior research methods* 44.4, pp. 978–990.

Kutas, Marta and Steven A Hillyard (1984). "Brain potentials during reading reflect word expectancy and semantic association". In: *Nature* 307.5947, pp. 161–163.

Kuttner, Kenneth N (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of monetary economics* 47.3, pp. 523–544.

Lamla, Michael J and Jan-Egbert Sturm (2013). "Interest rate expectations in the media and central bank communication". In: *Central Bank Communication, Decision Making, and Governance: Issues, Challenges, and Case Studies*, pp. 101–111.

Lamla, Michael J and Dmitri V Vinogradov (2019). "Central bank announcements: Big news for little people?" In: *Journal of Monetary Economics*.

Larsen, Vegard H, Leif Anders Thorsrud, and Julia Zhulanova (2020). "News-driven inflation expectations and information rigidities". In: *Journal of Monetary Economics*.

Le, Quoc and Tomas Mikolov (2014). "Distributed representations of sentences and documents". In: *International conference on machine learning*. PMLR, pp. 1188–1196.

Leeb, Hannes and Benedikt M Pötscher (2008a). "Can one estimate the unconditional distribution of post-model-selection estimators?" In: *Econometric Theory* 24.2, pp. 338–376.

— (2008b). "Recent Developments in Model Selection and Related Areas". In: *Econometric Theory* 24.2, pp. 319–322.

Levy, Roger (2008). "Expectation-based syntactic comprehension". In: *Cognition* 106.3, pp. 1126–1177.

Lloyd, Simon P (2020). "Overnight indexed swap-implied interest rate expectations". In: *Finance Research Letters*, p. 101430.

Loughran, Tim and Bill McDonald (2015). "The use of word lists in textual analysis". In: *Journal of Behavioral Finance* 16.1, pp. 1–11.

Manela, Asaf and Alan Moreira (2017). "News implied volatility and disaster concerns". In: *Journal of Financial Economics* 123.1, pp. 137–162.

Marcus, Mitchell, Beatrice Santorini, and Mary Ann Marcinkiewicz (1993). "Building a large annotated corpus of English: The Penn Treebank". In:

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). “Efficient estimation of word representations in vector space”. In: *arXiv preprint arXiv:1301.3781*.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeffrey Dean (2013). “Distributed representations of words and phrases and their compositionality”. In: *Advances in neural information processing systems*, pp. 3111–3119.

Miller, George A (1995). “WordNet: a lexical database for English”. In: *Communications of the ACM* 38.11, pp. 39–41.

Miranda-Agrippino, Silvia and Giovanni Ricco (2018). “The transmission of monetary policy shocks”. In:

Montgomery, Martin (2007). *The discourse of broadcast news: A linguistic approach*. Routledge.

Morris, Stephen and Hyun Song Shin (2002). “Social value of public information”. In: *american economic review* 92.5, pp. 1521–1534.

Mulatti, Claudio, Michael G Reynolds, and Derek Besner (2006). “Neighborhood effects in reading aloud: New findings and new challenges for computational models.” In: *Journal of Experimental Psychology: Human Perception and Performance* 32.4, p. 799.

Munday, Timothy (2019). “Central Bank Communications and Higher Moments”. In:

Nakamura, Emi and Jón Steinsson (2018). “High-frequency identification of monetary non-neutrality: the information effect”. In: *The Quarterly Journal of Economics* 133.3, pp. 1283–1330.

Nivre, Joakim et al. (2016). “Universal dependencies v1: A multilingual treebank collection”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pp. 1659–1666.

O’Brien, Edward J, Gary E Raney, Jason E Albrecht, and Keith Rayner (1997). “Processes involved in the resolution of explicit anaphors”. In: *Discourse Processes* 23.1, pp. 1–24.

Paivio, Allan (2013). *Imagery and verbal processes*. Psychology Press.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). “Glove: Global vectors for word representation”. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Pinker, Steven (2015). *The sense of style: The thinking person’s guide to writing in the 21st century*. Penguin Books.

Piotrkowicz, Alicja (2017). “Modelling Social Media Popularity of News Articles Using Headline Text”. PhD thesis. University of Leeds.

Plummer, Patrick, Manuel Perea, and Keith Rayner (2014). “The influence of contextual diversity on eye movements in reading.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40.1, p. 275.

Pulvermüller, Friedemann (1999). “Words in the brain’s language”. In: *Behavioral and brain sciences* 22.2, pp. 253–279.

Pulvermüller, Friedemann, Werner Lutzenberger, and Niels Birbaumer (1995). “Electrocortical distinction of vocabulary types”. In: *Electroencephalography and clinical Neurophysiology* 94.5, pp. 357–370.

Quirk, Randolph (2010). *A comprehensive grammar of the English language*. Pearson Education India.

Race, David S and Maryellen C MacDonald (2003). “The use of “that” in the production and comprehension of object relative clauses”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 25. 25.

Ratcliff, Roger and Gail McKoon (1978). "Priming in item recognition: Evidence for the propositional structure of sentences". In: *Journal of verbal learning and verbal behavior* 17.4, pp. 403–417.

Reid, Monique, Stan Du Plessis, et al. (2011). "Talking to the inattentive Public: How the media translates the Reserve Bank's communications". In: *Department of Economics and the Bureau for Economic Research at the University of Stellenbosch*.

Rong, Xin (2014). "word2vec parameter learning explained". In: *arXiv preprint arXiv:1411.2738*.

Rueschemeyer, Shirley-Ann and M. Gareth Gaskell (2018). *The Oxford Handbook of Psycholinguistics*. eng. 2nd. Oxford University Press. ISBN: 9780198786825. URL: `https://doi.org/10.1093/oxfordhb/9780198786825.001.0001`.

Rybinski, Krzysztof et al. (2019). "A machine learning framework for automated analysis of central bank communication and media discourse. The case of Narodowy Bank Polski". In: *Bank i Kredyt* 50.1, pp. 1–20.

Scarborough, Don L, Charles Cortese, and Hollis S Scarborough (1977). "Frequency and repetition effects in lexical memory." In: *Journal of Experimental Psychology: Human perception and performance* 3.1, p. 1.

Schuberth, Richard E and Peter D Eimas (1977). "Effects of context on the classification of words and nonwords." In: *Journal of Experimental Psychology: Human Perception and Performance* 3.1, p. 27.

Schuler, Karin Kipper (2005). "VerbNet: A broad-coverage, comprehensive verb lexicon". In:

Scott, Graham G, Patrick J O'Donnell, and Sara C Sereno (2012). "Emotion words affect eye fixations during reading." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38.3, p. 783.

Sidorov, Grigori, Alexander Gelbukh, Helena Gómez-Adorno, and David Pinto (2014). "Soft similarity and soft cosine measure: Similarity of features in vector space model". In: *Computación y Sistemas* 18.3, pp. 491–504.

Sims, Christopher A (2003). "Implications of rational inattention". In: *Journal of monetary Economics* 50.3, pp. 665–690.

Singh, Inder, Zoran Tiganj, and Marc W Howard (2018). "Is working memory stored along a logarithmic timeline? Converging evidence from neuroscience, behavior and models". In: *Neurobiology of learning and memory* 153, pp. 104–110.

Soto, Paul E (2019). "Breaking the Word Bank: Effects of Verbal Uncertainty on Bank Behavior". In: *FDIC Center for Financial Research Paper* 2019-01.

Swinney, David A (1979). "Lexical access during sentence comprehension:(Re) consideration of context effects". In: *Journal of verbal learning and verbal behavior* 18.6, pp. 645–659.

Taylor, Jonathan and Robert Tibshirani (2018). "Post-selection inference for-penalized likelihood models". In: *Canadian Journal of Statistics* 46.1, pp. 41–61.

Thorndyke, Perry W (1977). "Cognitive structures in comprehension and memory of narrative discourse". In: *Cognitive psychology* 9.1, pp. 77–110.

Tian, Xiaoying and Jonathan Taylor (2017). "Asymptotics of selective inference". In: *Scandinavian Journal of Statistics* 44.2, pp. 480–499.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Tibshirani, Ryan J, Alessandro Rinaldo, Rob Tibshirani, Larry Wasserman, et al. (2018). “Uniform asymptotic inference and the bootstrap after model selection”. In: *Annals of Statistics* 46.3, pp. 1255–1287.

Traxler, Matthew J (2011). “Introduction to psycholinguistics: Understanding language science”. In:

Tuckett, David (2011). *Minding the markets: An emotional finance view of financial instability*. Springer.

Van de Geer, Sara, Peter Bühlmann, Ya'acov Ritov, Ruben Dezeure, et al. (2014). “On asymptotically optimal confidence regions and tests for high-dimensional models”. In: *Annals of Statistics* 42.3, pp. 1166–1202.

Van der Cruijsen, Carin, David-Jan Jansen, Jakob De Haan, et al. (2015). “How much does the public know about the ECB's monetary policy? Evidence from a survey of Dutch households”. In: *International Journal of Central Banking* 11.4, pp. 169–218.

Van Gompel, Roger PG and Martin J Pickering (2007). “Syntactic parsing”. In: *The Oxford handbook of psycholinguistics*, pp. 289–307.

Van Heuven, Walter JB, Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert (2014). “SUBTLEX-UK: A new and improved word frequency database for British English”. In: *The Quarterly Journal of Experimental Psychology* 67.6, pp. 1176–1190.

Warren, Paul (2013). *Introducing psycholinguistics*. Cambridge University Press.

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert (2013). “Norms of valence, arousal, and dominance for 13,915 English lemmas”. In: *Behavior research methods* 45.4, pp. 1191–1207.

West, Richard F and Keith E Stanovich (1986). “Robust effects of syntactic structure on visual word processing”. In: *Memory & Cognition* 14.2, pp. 104–112.

Whaley, Charles P (1978). “Word-nonword classification time”. In: *Journal of Verbal learning and Verbal behavior* 17.2, pp. 143–154.

8. Appendix

8.1. Alternative solutions to the central bank's problem

The central bank's problem was to satisfy it's first order condition:

$$\nabla L(\theta_B) = 2(y(\theta_N) - \bar{y})^T H J_{\theta_B}(y) = \vec{0} \tag{27}$$

In the main text, we assumed that the central bank found a global minimum, and set $y = \bar{y}$ through manipulating θ_B.

If we index the vector of objectives and the diagonal elements of H by i and the vector of textual features by j, then a solution to the above first order condition requires that:

$$\sum_i (y_i - \bar{y}_i) h_i \frac{\partial y_i}{\partial \theta_{B,j}} = 0 \quad \forall j \tag{28}$$

Since we have not specified a function that maps θ_B to y the partial derivative is left unevaluated. As a result, there may be many other solutions other than the global minimum assumed in the main text or, indeed, none at all.

If the length of y is greater than one, then there are potentially many solutions to the first order condition. If y is scalar, then we either $y = \bar{y}$ — which is the global minimum solution dealt with in the main text — or $\frac{\partial y}{\partial \theta_{B,j}} = 0 \quad \forall j$. In fact the latter of these conditions leads to a set of linear solutions detailed below. Nonetheless, the possibility of non-linear solutions (or indeed a combination of linear and non-linear solutions) motivates our flexible approach in Section 6.

The central bank's first order conditions under the assumption of a scalar y can be written:

$$\frac{\partial y}{\partial \theta_{B,j}} = 0 \quad \forall j \tag{29}$$

Using the chain rule gives:

$$\nabla y(\theta_N) \frac{\partial \theta_N}{\partial \theta_{B,j}} = 0 \quad \forall j \tag{30}$$

Where $\nabla y(\theta_N)$ denotes the gradient vector of y with respect to θ_N.

We know the partial derivative of θ_N with respect to θ_B, subbing this in gives a system of equations of the form:

$$\nabla_y(\theta_N) \left(\mathbf{1}_j - (\theta_B - \theta^*) \left(\frac{k}{\alpha} 2\gamma\lambda w_j (\theta_{B,j} - \theta_j^*) \right) \right) = 0 \quad \forall j \tag{31}$$

Where $\mathbf{1}_j$ denotes a vector of zeros except for element j which is one.

This system of equations can be manipulated such that any element j can be represented in terms of another element i:

$$\theta_{B,j} = \theta_j^* + \frac{w_i}{w_j} (\theta_{B,i} - \theta_i^*) \left(\frac{\frac{\partial y}{\partial \theta_{N,j}}}{\frac{\partial y}{\partial \theta_{N,i}}} \right) \tag{32}$$

Subbing this in gives

$$\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2 = 2\frac{k}{\alpha}\lambda\gamma w_j(\theta_{B,j}-\theta_j^*)^2\left(\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2 + \sum_{i\neq j}\left(\frac{\partial y}{\partial \theta_{N,i}}\right)^2\frac{w_i}{w_j}\right) \tag{33}$$

Subbing in for k gives

$$\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2 = \frac{2\lambda\gamma w_j(\theta_{B,j}-\theta_j^*)^2\left(\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2 + \sum_{i\neq j}\left(\frac{\partial y}{\partial \theta_{N,i}}\right)^2\frac{w_i}{w_j}\right)}{\alpha+\lambda\gamma\left(w_j(\theta_{B_j}-\theta_j^*)^2+\sum_{i\neq j}w_j\left(\frac{w_i(\theta_{B,j}-\theta_j^*)\frac{\partial y}{\partial \theta_{N,j}}}{w_j\frac{\partial y}{\partial \theta_{N,i}}}\right)^2\right)} \tag{34}$$

Rearranging gives the linear solutions:

$$\theta_{B,j} = \theta_j^* + \frac{\alpha\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2}{2\lambda\gamma w_j\left(\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2+\sum_{i\neq j}\left(\frac{\partial y}{\partial \theta_{N,i}}\right)^2\frac{w_i}{w_j}\right)-\left(\left(\frac{\partial y}{\partial \theta_{N,j}}\right)^2\left(\lambda\gamma w_j+\lambda\gamma\sum_{i\neq j}w_j\left(\frac{w_i(\theta_{B,j}-\theta_j^*)\frac{\partial y}{\partial \theta_{N,j}}}{w_j\frac{\partial y}{\partial \theta_{N,i}}}\right)^2\right)\right)} \tag{35}$$

So there is a set of linear solution that do not achieve the global minimum. However, since we place very little external structure in our estimation procedure, a linear solution is not excluded from being found in the data. The important point is that we use a non-parametric method that allows for the significant possibility for highly non-linear functions between variables.

8.2. Word2Vec

Word2Vec (Mikolov, Sutskever, Chen, G. S. Corrado, and Dean 2013; Mikolov, Chen, G. Corrado, and Dean 2013) is a popular method for transforming words into vectors using their context within a corpus.

Word2Vec uses a shallow (two layer) neural network to produce the vectors. There are two implementation methodologies: Continuous Bag of Words (CBOW) or Skip-Gram. CBOW asks a neural network to predict a target word, given the context of the word (i.e. the words found in a small window around the target word). Skip-gram does the opposite, it asks a neural network to predict the context, given a target word. We use the CBOW implementation.

After training the network, the weights corresponding to the hidden layer are extracted and used as the vector representations for the dictionary of words found within the corpus. These weights are denoted W in the explanation below.

8.2.1. Architecture

The entire corpus has V unique words. For each word in the corpus, we take the context words — words within a window of length C from the target word — and use them to target the word in question. For example, in the sentence "the cat sat on the mat", if we were targeting "sat" and the (symmetric) window length, C, was 4, then we would try to predict "sat" based on the inputs "the", "cat", "on", "the".

For each word w_i, denote x_{w_i} as the unique one-hot encoded length V vector for that word. Upon being inputted to the network, each contextual word input vector, x_{w_i}, is multiplied by a V by N weight matrix (W), where N is the number of nodes in the hidden layer. This is the weight matrix we will eventually extract and call our "word embeddings". The mean weighted vector of the C input vectors is then fed to the hidden layer of neurons. Mathematically, the input to the hidden layer is:

$$h = \frac{1}{C} W^T (x_1 + x_2 + ... + x_C) \tag{36}$$

The hidden layer is fully connected and has a linear activation function. It passes the weighted sum of its inputs to the subsequent N by V weight matrix, W'. The subsequent weighted output is a V length vector, that we denote U:

$$U = hW' \tag{37}$$

Finally, we use softmax to the posterior distribution of words, conditional on their context:

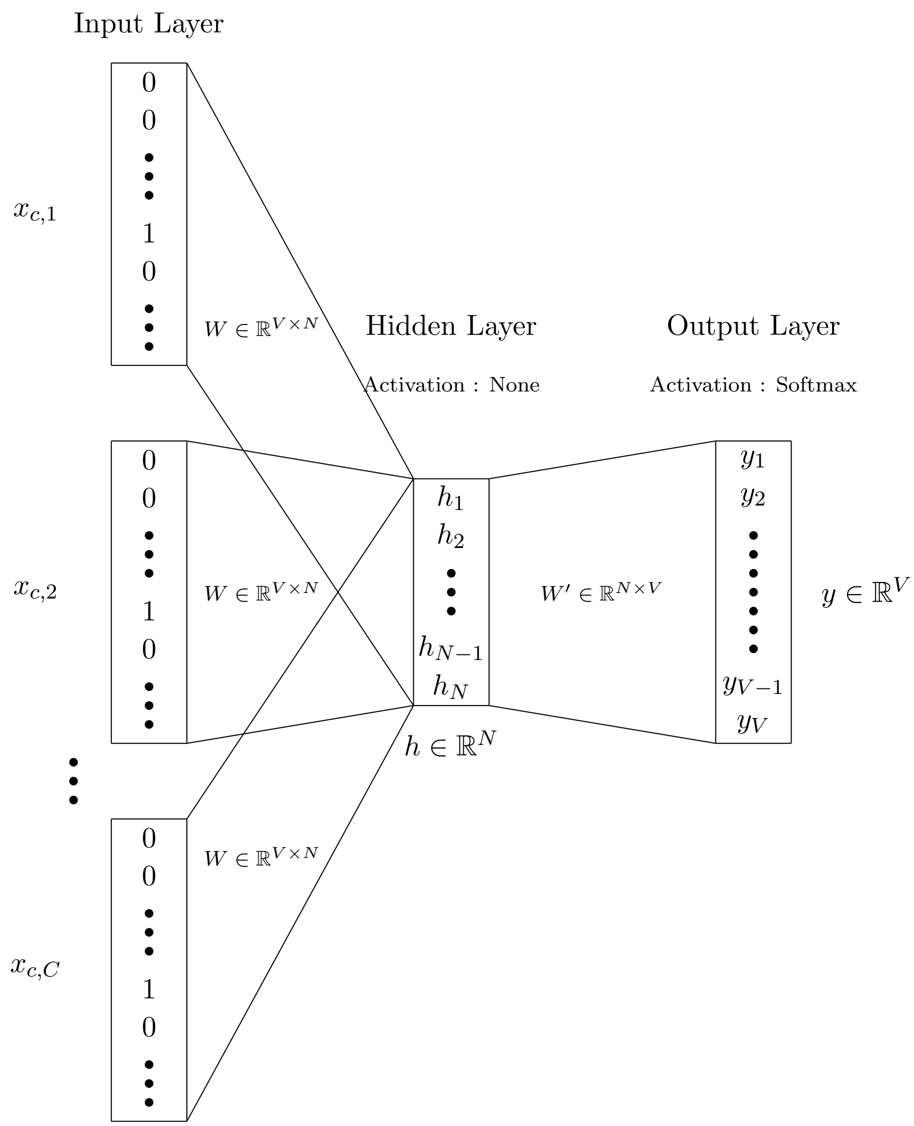
$$p(w_j | w_{c,1}, ..., w_{c,C}) = \frac{exp(u_j)}{\sum_{i=1}^{V} exp(u_i)} = y_j \tag{38}$$

Where u_j is the jth element of the vector U. Denoting the posterior distribution vector across the vocabulary y, we can then compare the prediction of the network, y, with the true result: a one-hot encoded vector of the target word of length V.

The problem is then one of supervised learning. We train the network to minimise the error between the posterior distribution y and the true result.

A sketch of the network architecture is shown below:

Fig. 9. Word2Vec neural network architecture



8.2.2. *Training*

The objective is to maximise the conditional posterior probability of observing the true target word, given the context words. We can write the loss function as:

$$\begin{aligned} L &= -log(y_j) \\ &= -u_j + log\sum_{i=1}^{V} exp(u_i) \end{aligned} \tag{39}$$

Both weight matrices, W and W', are updated during training. Updating is done via a form of gradient descent called backpropagation. For each training example (a word, in a sentence, in a document), one can calculate the error of the neural network, and use that to update the weight matrices. Denote the row of W that refers to word w_i as $v_{w_i}^T$ and the equivalent row of W' as $v_{w_i}'^T$. The updating procedures for each matrix are as follows for a given training instance. For a full derivation of these see Rong (2014).

$$\begin{aligned} v_{w_{c,k}} - \frac{1}{C} \cdot \eta \cdot \frac{\partial L}{\partial h_i} \rightarrow v_{w_{c,k}} \quad for\ k = 1, 2, ..., C \\ v'_{w_j} - \eta \cdot e_j \cdot h \rightarrow v'_{w_j} \quad for\ j = 1, 2, ..., V \end{aligned} \tag{40}$$

Where e_j is the prediction error of the output layer (i.e. $y_j - t_j$ where t_j is a one-hot encoded vector of the target word), and η is the learning rate.

Unfortunately, training the network using the exact process described by the above equations is computationally infeasible. For each training instance, to update v' one has to iterate over every word in the vocabulary of size V and calculate the prediction errors.

We use negative sampling to optimize the computation of training the network. Instead of iterating over every word in V, we only update based on the true output word, and G instances of 0's in the one-hot encoded vector t_j. The noise distribution for this sampling process is as in Mikolov, Sutskever, Chen, G. S. Corrado, and Dean (2013): a uniform distribution raised to the power of 0.75. As a result, the training objective is also modified to that of Mikolov, Sutskever, Chen, G. S. Corrado, and Dean (2013):

$$L = -log\left(\sigma(v_{w_j}'^T h)\right) - \sum_{w_i \in S_{neg}} log\left(\sigma(-v_{w_i}'^T h)\right) \tag{41}$$

Where σ denotes the sigmoid function, and S_{neg} denotes the negative subsample. Consequently, the updating equation for v' becomes:

$$v'_{w_j} - \eta\left(\sigma(v_{w_j}'^T h) - t_j\right) h \rightarrow v'_{w_j} \quad for\ w_j \in S \tag{42}$$

Where S denotes the full subsample, i.e. the negative subsample and the true target word.

8.2.3. *Parameterisation*

We set the context window, C to 10, i.e. 5 words either side of the target word. The hidden layer size N is set to 100. The number of negative samples to draw G is set to 5. The initial learning parameter, η is set to 0.025. We train over 20 epochs.

8.2.4. *Performance relative to pre-trained vectors*

We train our word vectors on the entire Bank of England communication corpus as detailed in Table 1. One can, of course, use pre-trained word vectors from larger corpuses and use these instead. For example one can use the Word2Vec vectors trained on about 100 billion words from Google News (Mikolov, Sutskever, Chen, G. S. Corrado, and Dean 2013; Mikolov, Chen, G. Corrado, and Dean 2013), or the GloVe vectors trained on Wikipedia (6 Billion tokens) or Twitter (2 Billion tweets, 27 Billion tokens) (Pennington, Socher, and Manning 2014). The advantage of using these pre-trained vectors is that they cover a much wider scope of vocabulary, having been trained on a larger corpus. The disadvantage is that they do not capture domain specific knowledge that one gets if one trains on the Bank of England corpus.

Table 12 shows the most similar words to the word 'economy' according to the pre-trained vectors just mentioned, and according to our own vectors. Since 'economy' is not too specific, all the models seem to output sensible similar words. Table 13 shows the most similar words to the word 'cpi'. The pre-trained models, having not seen the word 'cpi' in the context it is used by the Bank of England have a hard time producing similar words. Our model, on the other hand, performs much better, outputting words such as 'rpi' and 'inflation'. Since our strategy to measure news coverage of Bank of England communication is based on the similarity of word vectors between the Bank's communication and the news, the performance of our internally trained word vectors in capturing domain specific knowledge illustrates the benefit of training the vectors ourselves rather than relying on pre-trained models.

Table 12: Most similar words to the word 'economy' across models

GloVe Wikipedia	GloVe Twitter	Word2Vec Google News	This paper
economic	economic	economic	economies
growth	growth	econ_omy	economic
recession	government	economies	recovery
economies	recession	theeconomy	growth
recovery	markets	ecomony	demand
downturn	debt	recession	eurozone

Table 13: Most similar words to the word 'cpi' across models

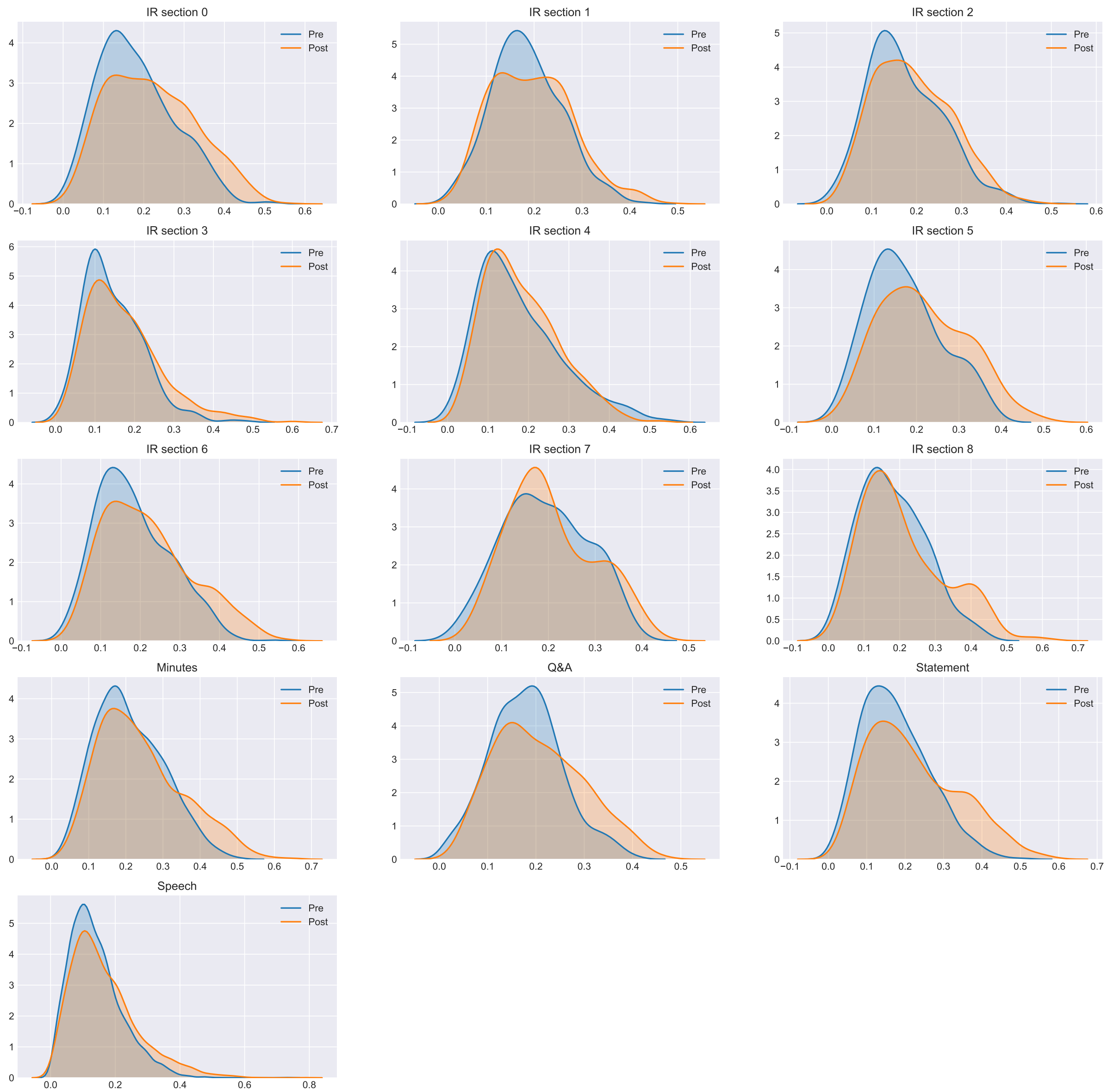
GloVe Wikipedia	GloVe Twitter	Word2Vec Google News	This paper
ppi	cpmi	==_null	rpix
0.1	europeia	1pp	rpi
gdp	stf	infla	hicp
inflation	petrobras	cmp	rpiy
0.2	reduo	MBytes	headline
0.3	cachoeira	idx	inflation

8.3. Visualisation

8.3.1. Visualising k

Firstly, we plot the kernel densities of the similarity of all articles in the post-communication window and in the pre-communication window for the measure defined previously (Figure 10). The distributions are skewed positive in the post-communication window compared to the pre-communication window. This is a useful sanity check. Articles that occur after the communication are more likely to either quote or be semantically similar to the Bank of England communication compared to articles published before the communication. This suggests that (i) articles in the press often draw on Bank communication for their content, and (ii) Bank communication is not simply reacting to the news cycle.

Fig. 10. Semantic similarity measure Kernel Densities in pre and post windows



8.3.2. Visualising topics

Figures 11, 12, 13 show the content elements of θ_B for the minutes. Each chart shows the number of instances of a word in a given topic dictionary divided by the length of the minutes, for all minutes since 1998.

Some topics are extremely rarely commented on in the minutes: Topic 17 — which is concerned with the FTSE100 — and Topic 26 — which is concerned with international economic organisations — are basically never mentioned. Other topics show distinct low frequency movements over time: Topic 34 — which is concerned with fiscal austerity — spikes under the period of fiscal tightening following the 2010 election. Other topics exhibit little trend over time, but significant meeting to meeting variation: Topic 14 — which is concerned with the economy — is one such topic.

Fig. 11. Content measures for Topics 0 to 17 in the minutes

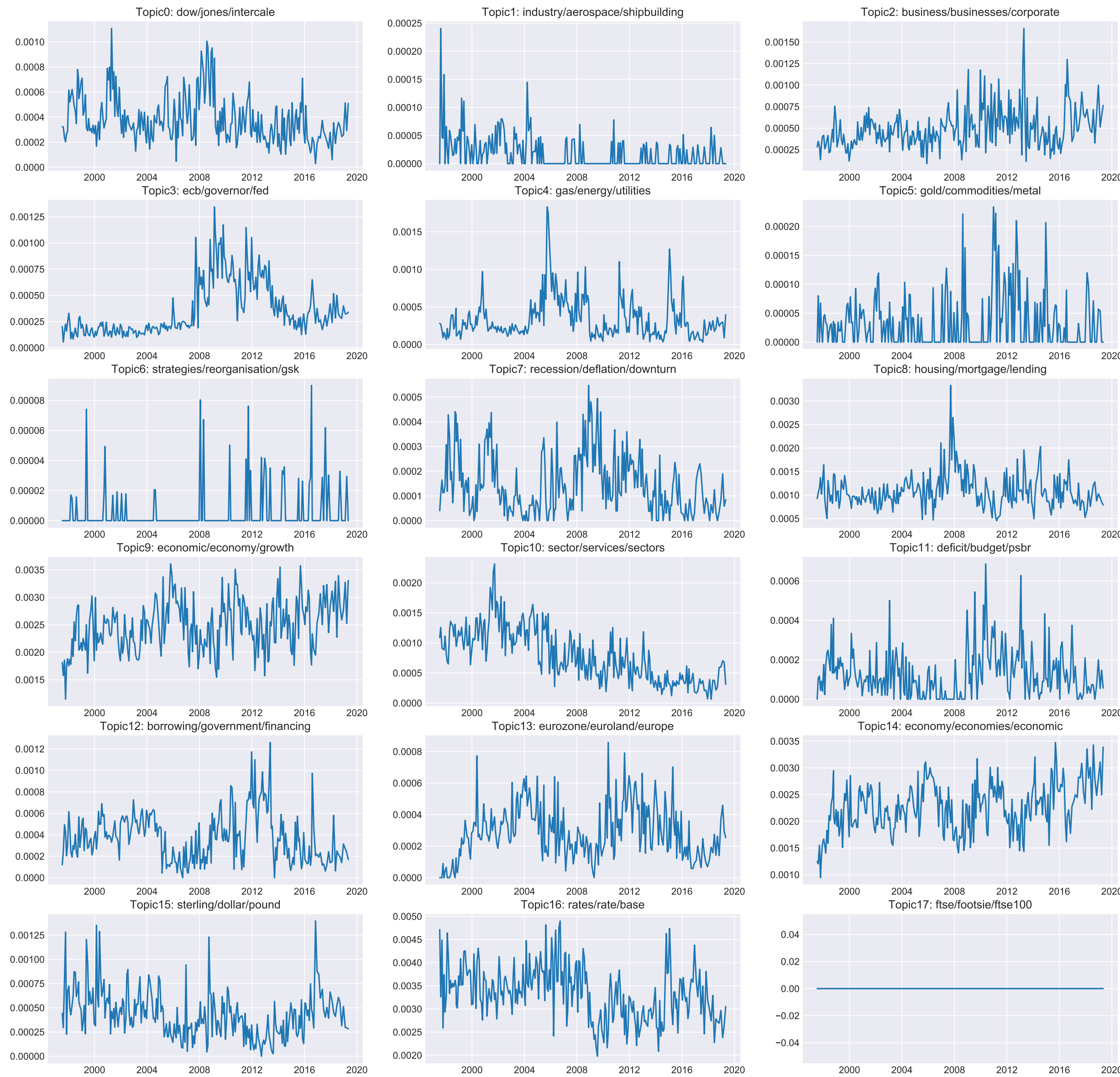


Fig. 12. Content measures for Topics 18 to 35 in the minutes

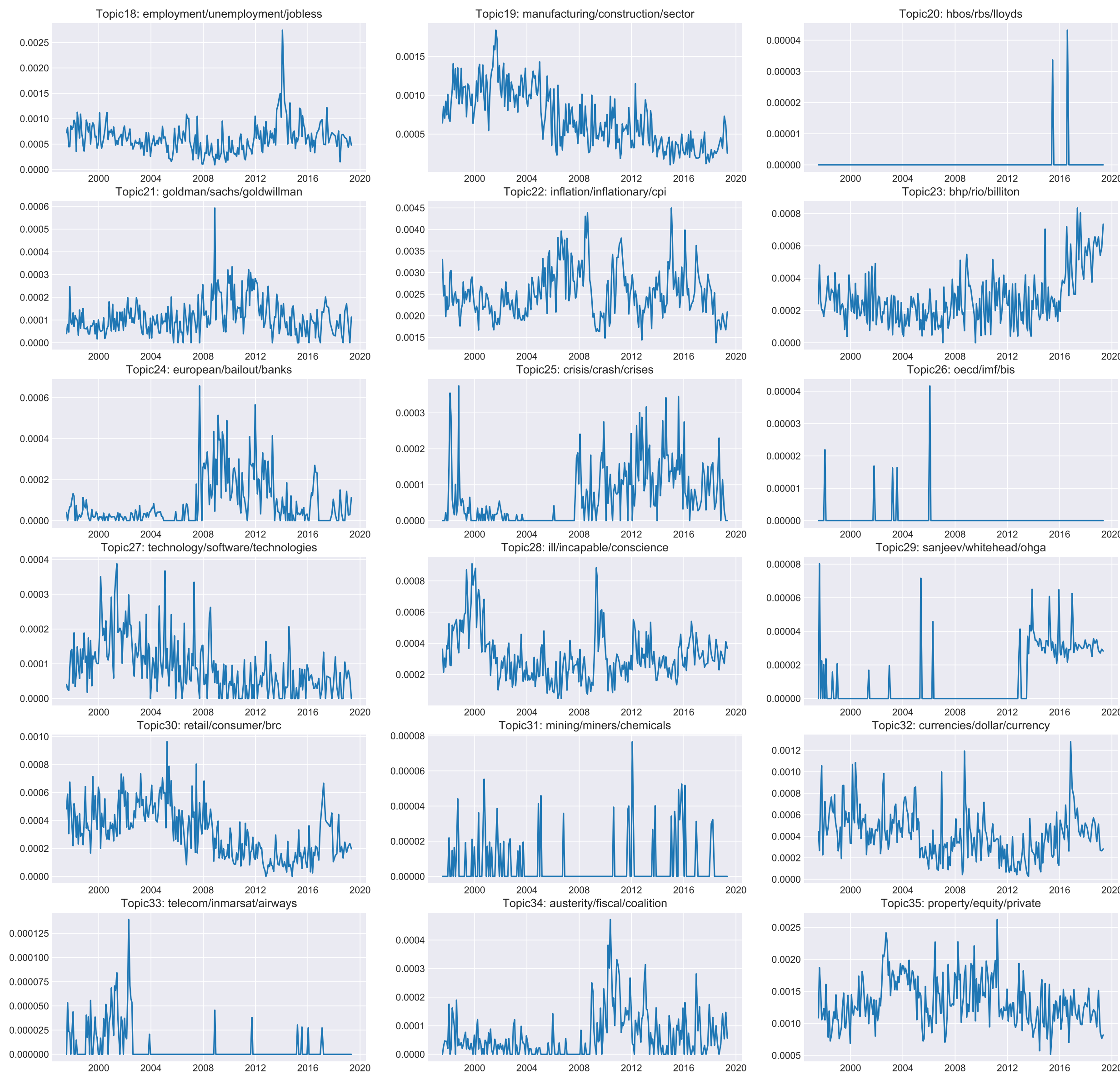
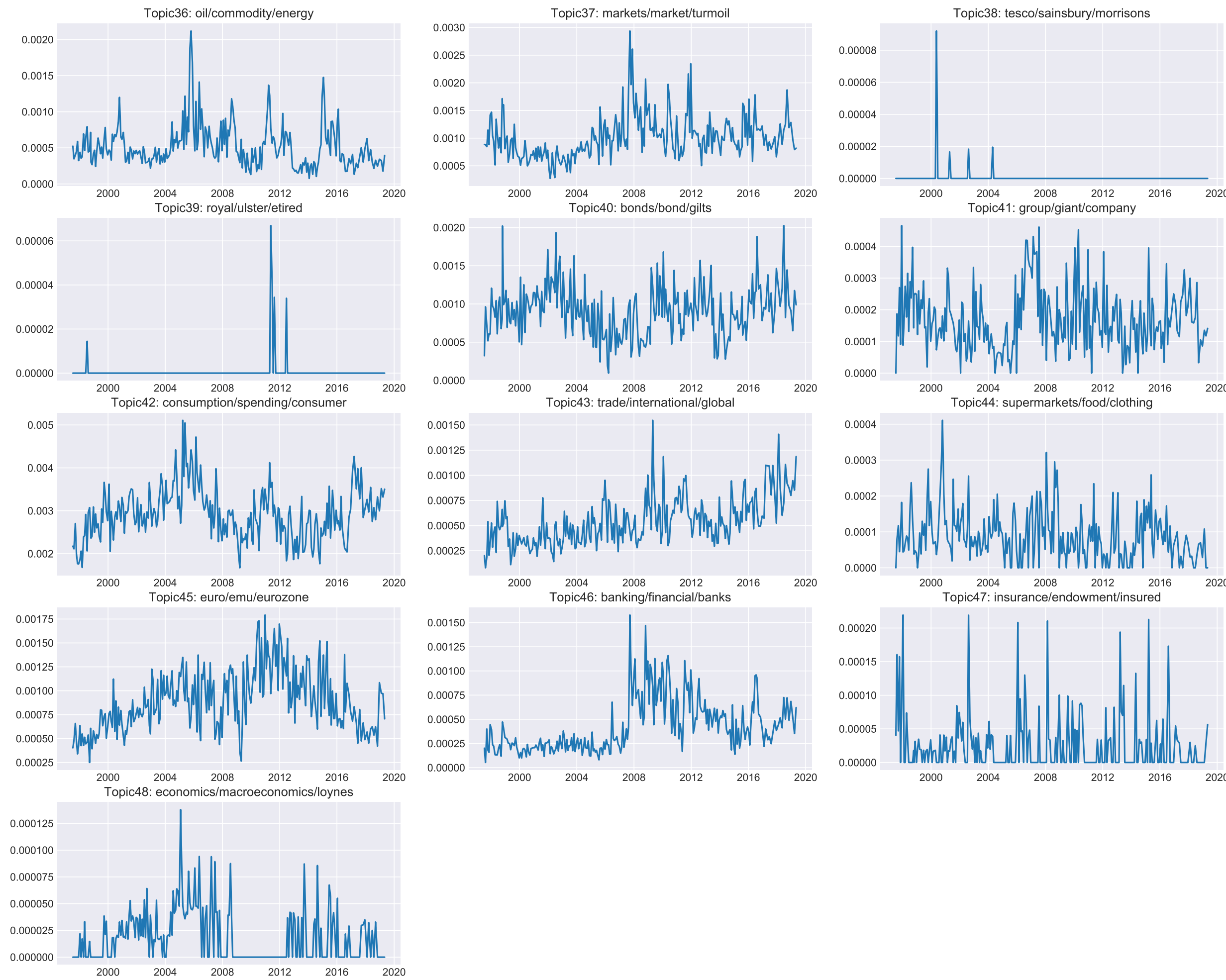


Fig. 13. Content measures for Topics 36 to 48 in the minutes



8.4. Monetary policy surprises as a control

One issue is whether to include a measure of monetary policy surprise as a control in the vector z.

Monetary policy surprises, as measured by the change in financial market prices around a monetary policy event, are often used in identification methods for calculating monetary policy shocks (Kuttner (2001), Cochrane and Piazzesi (2002), Gürkaynak, Sack, and Swanson (2005),Gertler and Karadi (2015), Nakamura and Steinsson (2018), Miranda-Agrippino and Ricco (2018), Jarociński and Karadi (2020)).

In the monetary policy events studied in this paper, only after August 2015 were some communication events by the Bank of England also accompanied by monetary policy decisions. The previous literature on monetary policy shocks has been primarily focused on measuring the surprises that occur around the decisions, with the notable exception of S. Hansen, McMahon, and Tong (2019).

Nonetheless, the question still arises: when financial markets move in response to central bank actions (be they releasing text or otherwise), is the impact on the news flow larger if the financial market move is larger? And - consequently - is a measure of the financial market surprise a relevant variable to include in z?

Table 14 shows the coefficients from simple linear regressions of the measures of the impulse to the news flow (k) for each section of text released against the absolute daily change in the one year OIS rate.

Normally, monetary policy 'surprises' are measured using changes in short term interest rates around a monetary policy decision. In our case, the vast majority of events in which communication is imparted do not occur on the same day as monetary policy decisions — and so short term rates, such as the overnight index swap rate — are unlikely to show any change. A longer term rate that includes investor expectations of future Bank of England decisions (Lloyd 2020), such as the one year rate, is more fitting for our study.

There are several positive and significant coefficients, primarily for sections of the Inflation Report. This is in line with previous research (S. Hansen, McMahon, and Tong (2019), and Munday (2019)), that suggests that the Inflation Report conveys important information to financial markets regarding uncertainty.

That said, the evidence presented in Tables 14is only suggestive. It is not possible to determine the direction of causality from these regressions. Only that a greater impulse to the news flow is associated with larger concurrent moves in financial markets. This could be because if the Bank of England releases information that radically alters the outlook of future interest rate changes, this is likely to be picked up by the press, and will also move financial markets.

Nonetheless, the regressions provide *a priori* evidence that the change in the swap rate is a relevant control variable.

That said, once we add in our textual features and other control variables, the daily change in the swap rate is not significant at the 5% level in our main analysis (Table 11). This suggests that in the naive regressions in Table 14 the change in the OIS rate is proxying for other variables, such as the textual features of the Bank's communication.

Table 14: Coefficients from linear regressions of Semantic shock on the change in the 1-year OIS

	Semantic Shock												
	irdf0	irdf1	irdf2	irdf3	irdf4	irdf5	irdf6	irdf7	irdf8	mindf	qadf	statdf	speechdf
1998-2018	6.70	10.58^{**}	5.82^{*}	6.44^{*}	-0.79	1.65	8.82^{**}	23.61	34.79	-0.96	7.99	6.14	0.36
	(0.19)	(0.01)	(0.10)	(0.10)	(0.87)	(0.78)	(0.04)	(0.17)	(0.17)	(0.62)	(0.19)	(0.14)	(0.82)
1998-2015			5.01	5.63	-0.86		8.00^{*}			-1.43	6.43	4.67	
			(0.20)	(0.19)	(0.88)		(0.10)			(0.49)	(0.34)	(0.30)	

p-values to two decimal places in parentheses

8.5. Content measuring tables

The following tables show the dictionaries for the Topics detailed in Section 5. The first ten words in each column are the words used to create the dictionary. The words in the second part of each column are the tags from the guardian that gave us the centroids of each topic.

Topic0	Topic1	Topic2	Topic3	Topic4
dow	industry	business	ecb	gas
jones	aerospace	businesses	governor	energy
intercale	shipbuilding	corporate	fed	utilities
dickins	volkswagen	enterprise	carney	electricity
nikkei	automotive	sme	mpc	utility
vinnie	steel	enterprises	policymakers	suppliers
dicky	tata	industry	boe	coal
burrill	engineering	innovative	bank	oil
corning	industries	firms	england	installers
bootmaker	bmw	organisational	trichet	petrol
dowjones	theairlineindustry	business	bankofenglandgovernor	energy-industry
	automotive-industry	sustainable-business	european-central-bank	oilandgascompanies
	britishairways	small-business	quantitative-easing	utilities
	pharmaceuticals-industry	corporate-governance	mark-carney	gas
	musicindustry	ethicalbusiness	federal-reserve	
	steel-industry	social-enterprise	monetary-policy-committee	
	tata	avivabusiness	mervyn-king	
			andy-haldane	

Table 15:

Topic5	Topic6	Topic7	Topic8	Topic9
gold	strategies	recession	housing	economic
commodities	reorganisation	deflation	mortgage	economy
metal	gsk	downturn	lending	growth
copper	efficiencies	depression	mortgages	recovery
bullion	licensing	stagnation	remortgaging	macroeconomic
nickel	structures	slump	affordability	upswing
metals	ccps	stagflation	property	eurozone
titanium	logistics	deflationary	lenders	economies
zinc	infrastructure	slowdown	loans	expansion
mineral	systems	contraction	market	demand
commodities	executive-pay-bonuses	recession	housingmarket	economicgrowth
randgoldresources	taxavoidance	globalrecession	mortgage-lending-figures	useconomicgrowth
gold	job-losses	deflation		economic-recovery
vedantaresources	travelleisure			
	davos			
	investing			
	office-for-budget-responsibility			
	entrepreneurs			
	rating-agencies			
	mergers-and-acquisitions			

Table 16:

Topic10	Topic11	Topic12	Topic13	Topic14	Topic15	Topic16
sector	deficit	borrowing	eurozone	economy	sterling	rates
services	budget	government	euroland	economies	dollar	rate
sectors	psbr	financing	europe	economic	pound	base
industries	surplus	lending	euro	eurozone	greenback	inflation
sectors	deficits	funding	bloc	global	yen	svrs
subsectors	surpluses	debt	greece	euroland	sterlings	trichets
areas	fiscal	servicing	continent	recovery	currencies	yields
service	shortfall	borrowing	italy	china	sterlings	borrowing
corporations	obr	borrowings	periphery	world	yuan	messel
intermediation	headroom	governments	germany	growth	rouble	costs
services-sector	budget-deficit	government-borrowing	eurozone	useconomy	sterling	interest-rates
				global-economy	dollar	interest-rates-us
				australia-economy		
				chinese-economy		
				the-gig-economy		
				worldbank		

Topic17	Topic18	Topic19	Topic20
ftse	employment	manufacturing	hbos
footsie	unemployment	construction	rbs
ftse100	jobless	sector	lloyds
techmark	inactivity	sectors	barclays
indexaveraged	joblessness	industries	regulators
shotton	productivity	cips	britannia
nasdaq	vacancies	presumptions	tsb
fste	participation	pmi	subsidiary
miners	workforce	competition	bailed
dax	migration	industrial	fsa
ftse	unemployment-and-employment-statistics	financial-sector	royalbankofscotlandgroup
	usemployment	manufacturing-sector	lloyds-banking-group
	uk-unemployment-and-employment-statistics	construction	regulators
	us-unemployment-and-employment-statistics	manufacturingdata	hsbcholdings
			northern-rock
			financial-services-authority-fsa
			hbos
			armholdings

Topic21	Topic22	Topic23	Topic24	Topic25	Topic26	Topic27
goldman	inflation	bhp	european	crisis	oecd	technology
sachs	inflationary	rio	bailout	crash	imf	software
goldwillman	cpi	billiton	banks	crises	bis	technologies
eshan	unemployment	tinto	bail	crunch	niesr	tech
ubs	prices	xstrata	greece	meltdown	thinktank	ict
goldmans	2pc	miner	ireland	turmoil	cooperation	biotechnology
prefessor	deflation	kazakhmys	counterparties	strains	imfs	telecoms
sachs	persistently	lonmin	cyprus	contagion	studies	technological
nomura	rates	10bhp	eurozone	distress	obr	equipment
morgan	wages	vale	rescue	defaults	lagarde	fintech
goldmansachs	inflation	rio-tinto	europeanbanks	debt-crisis	imf	technology
		bhpbilliton	ireland-bailout	financial-crisis	oecd	
				credit-crunch	institute-for-fiscal-studies	
				subprimecrisis		

Topic28	Topic29	Topic30	Topic31	Topic32	Topic33	Topic34
ill	sanjeev	retail	mining	currencies	telecom	austerity
incapable	whitehead	consumer	miners	dollar	inmarsat	fiscal
conscience	ohga	brc	chemicals	currency	airways	coalition
poison	westec	bumpf	exploration	renminbi	telecoms	budgetary
hindsight	arne	intermediate	mineral	greenback	vita	socialist
conformists	ibstock	grocery	pharmaceutical	sterling	energis	populist
gloomily	easyjet	brcs	wolseley	yen	biotech	coalitions
blimpish	jenning	retailers	antofagasta	yuan	freeserve	protectionism
idiotic	todd	retailing	platinum	pound	vivendi	budgets
careless	pirie	disapponting	alcoa	franc	aerospace	poverty
fresnillo	barclay	retail	mining	currencies	telecoms	austerity
	anglo-american					
	marksspencer					
	ben-bernanke					
	antofagasta					
	kazakhmys					
	lehmanbrothers					
	janet-yellen					
	astrazeneca					
	johnlewis					

Topic35	Topic36	Topic37	Topic38	Topic39	Topic40
property	oil	markets	tesco	royal	bonds
equity	commodity	market	sainsbury	ulster	bond
private	energy	turmoil	morrisons	etired	gilts
rental	crude	rout	safeway	curated	assets
residential	gasoline	crisis	asda	lesney	gilt
cre	petrol	worldscope	waitrose	southerly	ious
estate	import	stockmarkets	somerfield	dutch	treasuries
corporate	wheat	contagion	debenhams	yoko	tranches
rented	copper	stockmarket	grocer	rosyth	iou
landlord	commodities	nervousness	kingfisher	shell	bunds
realestate	oil	stock-markets	tesco	royaldutchshell	bonds
privateequity		marketturmoil	j-sainsbury	royal-mail	
			morrisons		

Topic41	Topic42	Topic43	Topic44	Topic45	Topic46	Topic47	Topic48
group	consumption	trade	supermarkets	euro	banking	insurance	economics
giant	spending	international	food	emu	financial	endowment	macroeconomics
company	consumer	global	clothing	eurozone	banks	insured	loynes
division	household	exporting	supermarket	erm	ambrosianos	insurers	economist
firm	demand	bilateral	meat	currency	regulatory	assurance	economics
conglomerate	expenditure	globally	furniture	euroland	microprudential	families	lbs
subsidiary	growth	external	grocery	ttwa	ccps	debtline	stansfield
specialist	consumers	wto	clothes	shaded	bankers	pension	tombs
operator	concertina	china	footwear	bloc	supervise	flowergram	bowmark
groups	activity	internationa	petrol	single	reformed	grid	disadvantageously
vodafonegroup	consumerspending	internationaltrade	fooddrinks	euro	banking	insurance	economics
btgroup			supermarkets	emu	banking-reform		
co-operative-group							