UTS Business School –
Department of Economics
Dr Chau Chak Wing Building
14 – 28 Ultimo Road, Ultimo NSW 2007

T: +61 2 9514 7713
economics@uts.edu.au

Department of Economics
Working Paper Series

‘Does the Fed say it all? Textual analysis of public communications and private discussions’

Aleksei Chernulich 1
Mengheng Li [2]
Eamon McGinn [3]

[1] New York University Abu Dhabi
[2] The University of Technology Sydney
[3] Australian National University

No. 01 / 2024 **January 2024**

Does the Fed say it all? Textual analysis of public communications and private discussions*

Aleksei Chernulich [a,†], Mengheng Li [b,c,‡], and Eamon McGinn [b,§]

[a] *Division of Social Science, New York University Abu Dhabi*
[b] *Economics Discipline Group, University of Technology Sydney*
[c] *The Centre for Applied Macroeconomic Analysis, Australian National University*

Abstract

Public communications have been the center of studies on the effect of forward guidance. The Fed's public communications are based on private discussions within FOMC. This paper analyses private information revealed in FOMC meeting transcripts. Combining textual analysis and time series techniques, our study finds: (1) private information has an information advantage in predicting future policy rate changes, conditional on public information and economic variables; and (2) public information does not overshadow the predictive power of economic variables. These findings imply that the Fed can be more assertive and transparent in its communication to anchor market expectations on future policy.

Keywords: Central Bank Communication, Textual Analysis, Forecasting, Monetary Policy, Taylor Rule, FOMC, Forward Guidance.
JEL Classification: E44, E52, E58, C53.

*We thank Dick van Dijk, James Morley, Emi Nakamura, Jordan Norris, Jón Steinsson, Etienne Wasmer and participants of NYUAD research seminar and the 2021 Asian Meeting of Econometric Society for useful comments.

† *E-mail:* Aleksei.Chernulich@nyu.edu
‡ *E-mail:* Mengheng.Li@uts.edu.au
§ *E-mail:* Eamon.G.McGinn@uts.edu.au

1 Introduction

The introduction of forward guidance as an alternative policy tool honed by the Fed amid the 2008 global financial crisis has stimulated much research on central bank communication (Blinder et al., 2008). It is recognized that policy deliberation in FOMC meetings results in the Fed's public communication with "controlled" transparency and consistency (Acosta, 2015). On the one hand, with the growing need to maintain market stability and share information among various financial policy committees, the Fed aims to achieve communication transparency to better signal its future policy paths (Reis, 2013; Gilchrist et al., 2015; Hansen et al., 2018). On the other hand, business cycle-induced economic uncertainty means that central banks must remain somewhat ambiguous in their communication, so their credibility is protected and they have the ammunition to shock the market when needed (Stein, 2014; Blinder et al., 2017; Jia and Wu, 2022). While we do not answer if the Fed lacks the ability to keep its public communication fully transparent or it deliberately withholds some private information, we estimate to what extent the Fed under-informs the public about its decisions.

Our motivation comes from the assumption that the Fed utilizes an approximate Taylor-type decision rule (Taylor, 1993) to set the Federal funds target rate based on a set of macroeconomic variables such as GDP and inflation rate (see *e.g.* Woodford, 2001 and Fernandez et al., 2008 for detailed discussions). Such a decision rule can be described as

$$r_t = f(\Omega_{t-1}) + \epsilon_{r_t}, \tag{1}$$

where r_t is the setting of a policy instrument of meeting t, Ω_{t-1} is the information prior to meeting t and enters a linear policy function f, and ϵ_{r_t} is a monetary policy shock orthogonal to Ω_{t-1}.[1] Under this interpretation, the market cannot fully track the decisions of the Fed, despite that Ω_{t-1} is observed. This is because ϵ_{r_t} collects various random

[1]This decision rule can emerge from an infinite-horizon optimal control problem where the monetary authority maximizes the expected value of a quadratic criterion function subject to a set of linear constraints of technology and private agents' decision rules.

factors that affect policy decisions, including the views of the FOMC members that are affected by their understanding of the state of the economy (Gorodnichenko et al., 2023). In fact, a large literature attempts to measure and rationalize ϵ_{r_t} whose magnitude reflects macroeconomic uncertainty and the deliberation of the Fed (Stein, 2014; Hanson and Stein, 2015; Jia and Wu, 2022). As policy shocks manifest in sizable market reactions (Gorodnichenko et al., 2023), it is important to assess how the Fed's information, both publicly communicated and private, which enters ϵ_{r_t} in (1), affects its decision on r_t.

To estimate policy shocks that are traditionally considered unobserved, Romer and Romer (1989), Sims (1992) and Nakamura and Steinsson (2018), among many others, have developed various statistical methods, such as structural VAR models and narrative and high-frequency approaches. More recently, many have pointed out that ϵ_{r_t} is at least partially observable through FOMC's public communication. For example, Romer and Romer (2004) and Lucca and Trebbi (2009) construct sentiment indices from FOMC statements to gauge the directional information on r_t. Similarly, Gürkaynak et al. (2005) and Hansen and McMahon (2016) devise information measures and show how FOMC language shocks explain future policy paths.

We extend this literature that mostly focuses on public communications to the Fed's private discussions.[2] Considering the different informational content they embody, we term FOMC statements and minutes as public information, while FOMC transcripts are private information.[3] The former are published right after a policy meeting in a deliberate and well-structured manner, whereas the latter are unstructured and only accessible with a 5-year lag. To this end, we estimate the effect of the latter on the policy rule (1) and address a simple question: Can private information from the Fed provide additional information on future policy rate changes?

[2] In this way, we also complement the literature that studies the effects of private but informal communications as in Cieslak et al. (2019), Vissing-Jorgensen (2020), and Morse and Vissing-Jorgensen (2020).

[3] In this paper, we focus on textual communication that is readily available on the Fed's official website. One can extend the analysis to include scattered press conferences and other public speeches of the Fed's representatives.

Methodologically, we combine the text-as-data approach from Gentzkow et al. (2019) and standard time series models to analyze the Fed's transcripts, statements, and minutes from 357 FOMC meetings between 1982 and 2016 collected from the Board of Governors of the Federal Reserve System website.[4] Our paper differs from the existing literature that uses transcripts (Hansen et al., 2018; Acosta, 2015; Shapiro and Wilson, 2022) in two main aspects. First, we adopt a time series approach where the effect of private information is estimated conditional on observed economic indicators and public information. Second, we run an extensive forecasting exercise to show that private information lifts the informational limit of public information. Our results imply that incorporating more information in the Fed's public communication can improve the market understanding of its future policy stances, making alternative unconventional tools such as forward guidance more effective.

2 Text as data

Members of the FOMC meet on average 8 times a year to decide on the target rate upon reviewing the present and forecast economic conditions. Each transcript consists of speeches given by the committee members. We use all $96,000$ speeches chronologically from all meetings in our analysis. In the online appendix, we provide some summary statistics of the speech-level data before and after 1994, when the policy decision started to be announced.

We automate the cleaning of the textual data following steps of other computational linguistic studies, such as Hansen et al. (2018) and Gentzkow et al. (2019). This procedure includes conversion to lowercase, removal of punctuation marks and other symbols, trimming excess white space, and removing all common stop words. Common surnames (and some specific surnames of FOMC members) are also removed using the list of the 2000 most common surnames in the US Census. We apply a Porter stemming algorithm

[4]Statements are available from 1998, prior to which we only consider minutes as public communication.

(Hornik, 2007) to reduce inflections and retain the root of each word. After cleaning, we obtain a filtered version of the raw speech data. Statements and minutes are cleaned in the same way.

To enable econometric analysis on texts, we build the *Document Term Matrix* (DTM). Elements in the matrix show whether key bigrams, defined as pairs of consecutive written units (a word after cleaning may only contain its root as a basic written unit), are used at a given speech. Below we describe the process for the case of the transcripts, while for statements and minutes it is analogous.

First, bigrams are generated by grouping together all pairs of written units. This gives us 3.4 million bigrams, of which more than 1 million are unique and appear at least ten times. Second, we identify a manageable set of relevant and informative bigrams by using widely applied *log odds ratio* $\delta_i = \log(p_i + 1) - \log(q_i + 1)$, where p_i (q_i) is the number of times bigram i being used in documents of a meeting that lead to positive (negative) rate changes. So $\delta_i > 0$ ($\delta_i < 0$) indicates that bigram i is more likely associated with a rate increase (decrease). This step leads to around 2400 bigrams with large $|\delta_i|$, where large $|\delta_i|$ is defined as being greater than 2.407 or in the top 4% of bigrams. We also try monograms and trigrams, and use the *frequency-inverse document* of Gentzkow et al. (2019) to determine informative written units. The results show little difference from those presented in this paper and are thus omitted for clarity.

Finally, we define the DTM X with typical binary elements $X_{ti,j}$, which equals 1 if bigram j is used in the i-th speech in meeting t and zero if otherwise. Let K and N denote the numbers of columns and rows of X, respectively. We have $K = 2,438$ bigrams and $N = 96,872$ speeches from transcripts that are arranged chronologically in our analysis. X can be partitioned into $T = 374$ sections, each corresponding to a meeting. Let n_t denote the number of speeches in meeting t, which is also the number of rows in the t-th partitioned section of X. We have $\sum_{t=1}^{T} n_t = N$.

3 Econometric method

We consider three classes of models – Taylor-type regression models, text-based models, and the combination of the two, which we term text-augmented Taylor models. In a nutshell, we conduct a series of forecasting exercises following the decision rule in (1) for the three classes of models. All models are simple linear time series models and specified with autoregressive distributed lag (ARDL) dynamics which is a linear regression with lagged dependent and explanatory variables. The ARDL model estimates the textual effect on the target rate conditional on the history of interest rates and other macroeconomic variables and thus enables us to explore the incremental informational content in the public and the private information of the Fed.

3.1 The time series model

In all models, the dependent variable is the first difference in target rates. There are two reasons for differencing: 1) FOMC meetings decide on changes in target rates instead of the level, and 2) a series of ADF and KPSS tests (Keblowski and Welfe, 2004) strongly suggest the presence of a unit root, and differencing ensures stationarity. We also difference the levels or the logarithms of non-stationary macroeconomic variables used in the Taylor models. We consider a multi-step prediction exercise, although they deviate from the standard decision rule (1). As introduced previously, central bank communication, such as the forward guidance, aims to anchor market opinions on future interest rates, and thus is expected to have multi-step predictive power even when economic variables are not informative.

We propose the following meeting-level h-period ahead predictive regression model

$$\Delta r_t = c + \rho \Delta r_{t-h} + \gamma' z_{t-h} + \chi_{t-h}\beta + \epsilon_t, \tag{2}$$

for a positive integer h and $t = 1, ..., T$, where Δr_t is the policy decision made in meeting

t. z_{t-h} is a vector of stationary economic variables available to the meeting $t-h+1$. For example, if $h=1$, z_{t-1} contains economic variables available to the meeting t; this is the decision rule in (1). Of our interest is χ_{t-h}, a row vector that scores bigrams in meeting $t-h$. Lastly, ϵ_t is an independent and identically distributed error term.

χ_t aggregates textual information in central bank information from the rows in the DTM X that correspond to all speeches in meeting t. Each element in this vector can be interpreted as the importance or weight of a bigram used in that meeting. Define $x_{ti} = (X_{ti,1}, ..., X_{ti,K})$ which is ti-th row of the DTM X and collects dummy variables for bigrams in the i-th speech made in meeting t. Let ω_{ti} denote the length of speech i in meeting t, measured by the number of words in that speech. We define χ_t by

$$\chi_t = \sum_{i=1}^{n_t} \omega_{ti}^* x_{ti}, \quad \text{where} \quad \omega_{ti}^* = \frac{(1+i/n_t)\omega_{ti}}{\sum_{i=1}^{n_t}(1+i/n_t)\omega_{ti}}. \tag{3}$$

The construction of the normalized weight ω_{ti}^* is based on two rationales: 1) The longer a speech is (*i.e.* ω_{ti} is larger), the more influential it is; and 2) the closer to the end of a meeting a speech is (*i.e.* i is close to n_t), the more likely it is to determine the policy outcome.

The length effect captured by ω_{ti} is intuitive. It reflects that a stronger policy signal in the deliberation of a FOMC member tends to strengthen the equilibrium deliberation effort of other members; this resembles the discipline effect in agency theory as documented by Hansen et al. (2018). The simple linear trend in the weighting function controls for the timing effect. In particular, earlier speeches might bring information distortion and lead to antiherding among committee members and exaggeration in their deliberation (Meade and Stasavage, 2008), whereas later speeches (especially those at the end of the meeting) promote herding and conformity, and eventually lead to a policy decision (Hansen et al., 2018; Fehrler and Hughes, 2018). This means that later speeches are "deal closers" that conclude a meeting and thus receive larger weights, while earlier speeches are "ice breakers" that stimulate discussions and open a meeting.[5]

Model (2) boils down to a Taylor-type decision rule if $\beta = 0$ and z_{t-h} contains inflation expectation, output growth (or gap), and inflation rate (or deviation from its target); see *e.g.* Woodford (2001), Siklos and Wohar (2005) and Rühl (2015). The estimated β in the model tells us if there is an information advantage in texts in addition to the economic variables. Reversely, should we impose $\gamma = 0$, we have a purely text-based model with the policy innovation $\Delta r_t - c - \rho\Delta r_{t-h}$ fully decomposed into textual information. How economic information and central bank communication complement each other in determining policy rates is of our interest. We provide some additional remarks on empirical strategy in online appendix.

3.2 Dimensionality reduction

Even the simple ARDL(1,1) model (2) is prohibitively high-dimensional, meaning that the number of bigrams in X far exceeds the number of observations T. To achieve dimension reduction, we opt for a variant of the least absolute shrinkage and selection operator (LASSO): double-selection, or DS-LASSO of Belloni et al. (2014).

Compared to the standard LASSO, DS-LASSO achieves correct model selection (*i.e.* selection of influential bigrams) in the presence of potential omitted variables. We find it attractive because the monetary policy shock ϵ_{r_t} in (1) likely contains determining factors not captured in the regression (2). Furthermore, LASSO pulls elements in β uniformly towards zero, introducing bias and uncertainty. In contrast, the DS-LASSO gets rid of bigrams with a close-to-zero effect on future interest rate moves without mistakenly pulling other coefficients towards zero. This feature allows us to correctly select the most informative bigrams in central bank communications.

[5] This is in line with Alan Greenspan's view when addressing the House Banking Committee:

> *... The prevailing views of many [FOMC meeting] participants change as evidence and insights emerge. This process has proven to be a very effective procedure for gaining a consensus...* (Greenspan, 1993, as reported in Meade and Stasavage, 2008 and Hansen et al., 2018)

Greenspan's comments summarize the consensus decision-making process at FOMC meetings, where it takes time for divergent opinions to eventually converge to a policy outcome.

It is worth noting that LASSO estimators are linear and involve simple convex optimizations. Other variable selection procedures such as regression tree, neural network, subset selection, and random forest can also replace (2) (Chernozhukov et al., 2018). We stick to DS-LASSO, because the Taylor-type policy function (1) is linear, and the identified effects are more interpretable. Also, if the simple LASSO model can already identify the incremental predictive power of private communications, more advanced models are expected to only strengthen our results.

Lastly, LASSO approaches have a tuning parameter that controls the degree of shrinkage and is usually determined by cross-validation. We adopt the method of stratified cross-validation of Li and Chen (2014) that takes into account the time series feature of our dataset by randomly splitting the sample into chronologically ordered subsamples. The optimal tuning parameter is then fixed throughout our forecasting exercise introduced in the next section.

3.3 Predictive analytics

To investigate whether central bank public and private information provides more explanatory power on future policy trajectory than a Taylor-type principle, we compare models with and without textual information using the Diebold-Mariano (DM) test for equal predictive accuracy (Diebold and Mariano, 2002; Giacomini and White, 2006).

The DM test statistically quantifies the distance between functionals of out-of-sample forecast errors of two competing models. We follow the convention and use the difference between squared forecast errors. Diebold and Mariano (2002) gives the asymptotic distribution for the average error differential that serves as the base for the test, and Giacomini and White (2006) derives the test statistic under nonnested models. The DM test statistic is a z-score of error differentials constructed from a series of out-of-sample forecast errors obtained by rolling-window estimations. Suppose that the window size is τ. In the first window, we estimate the model using data from $t = 1$ to $t = \tau$ and make a prediction of $\Delta_{\tau+h}$. In the last window, the model is estimated using data from

$t = T - h - \tau + 1$ to $t = T - h$ and fitted to generate a forecast of Δr_T. This gives us $T - h - \tau + 1$ forecast errors for each model under consideration, which are then used to perform the test. Rejecting the null of equal predictive accuracy, we compare the root mean squared errors, defined as the square root of the sample average of squared forecast errors, of a model with textual information and one without to see if the text-augmented model is superior in prediction.

Bernanke (2015) clarified that Taylor's mechanistic approach could never replace the extensive deliberations of the FOMC. Thus, rolling-window estimation also allows for β in (2) to change over time. Hansen et al. (2018) documents significant behavioral changes in FOMC members in their deliberations before and after the release of transcripts in October 1993. As our sample covers this particular period, rolling windows can capture how this change affects the predictive power of bigrams. In essence, our procedure is similar to Favero (2006), Fernandez et al. (2008), and Greenwood-Nimmo and Shin (2012), who employ rolling windows to examine how the reaction functions of central banks evolve over time. But we run a predictive exercise with central bank public and private information at a meeting-level frequency.

In a predictive framework, Inoue et al. (2017) has derived the optimal window size that asymptotically minimizes the sum of squared forecast errors. Throughout the empirical study, we follow their advice and set the window size to contain 100 meetings (equivalent to 12.5 years on average), but we also perform robustness checks using eight years as in Fernandez et al. (2008) and 20 years as in Greenwood-Nimmo and Shin (2012) and reach largely the same conclusions.

4 Empirical results

This section focuses on forecasting exercises that shed light on the marginal effect of meeting deliberations. It tells us how the Fed trades off between communication transparency, brevity in its forward guidance and intentional vagueness for credibility. Specifically, we explore i) whether or not private central bank information contains textual information on policy rate movements, conditional on public communications and observed economic variables; and ii) how stable is their information advantage over the sample period during which various reforms and events took place.

4.1 Model specifications

We consider three classes of predictive regression models, including 2 Taylor models, 2 text models, and 6 hybrid models. To differentiate model specifications, we use a set of combined abbreviations: "T", "AT", "Pr", and "Pu".

"T" denotes our benchmark Taylor-type forecasting rule where the dependent variable is the change in the federal fund target rate made in meeting t, or Δr_t, while independent variables include Δr_{t-h} and z_{t-h} as in (2), with the latter containing changes in inflation and real GDP realized h meetings before. In this exercise, we use revised macro variables which is impractical in real-time forecasting. But since we focus on the additional predictive power brought by communications, this is not an issue: if revised macro variables cannot dilute information in communications, vintage data can only strengthen our results.

"AT" stands for an augmented Taylor-type forecasting rule. In this specification, z_{t-h} also includes other economic variables that are found to be explanatory about the Fed's reaction function. Specifically, it includes the University of Michigan Consumer Inflation Expectation and forecasts from the Survey of Professional Forecasters. These variables reflect how the Fed understands the forward-looking behavior of economic agents Bernanke (2004); Fernandez et al. (2008). Additionally, it includes a sentiment index derived from FOMC statements, following Romer and Romer (2004) and Lucca and Trebbi (2009). The index is a single quantity that provides compressed directional information on Δr_t. Lastly, "AT" features seven principal components constructed from the FRED-MD dataset of 128 monthly macroeconomic variables (McCracken and Ng, 2016), including the government bond yield curve that reflects interest rate expectations similar to the

Federal Funds Futures. These components are in line with the coincidental factors of Stock and Watson (2002) and span the information set available to the Fed during each FOMC meeting. To mitigate overfitting and reduce forecasting bias brought by correlated predictors, we further shrink the regression coefficients of added economic variables via a ridge-type penalty term. We do not use LASSO to nullify their effects because when the Fed gauges the state of the economy, these variables are inside its information set in spite of possibly small effects.

Both text and hybrid models have "Pr" and "Pu" that stand for central bank private and public information, respectively. The former is taken from the DTM of FOMC meeting transcripts, whereas the latter comes from minutes and statements. Our research question hinges on the incremental information advantage of privation information, conditional on economic variables and public information. For example, we can compare an augmented Taylor model with public information, *i.e.* ATPu, with one with both private and public information, *i.e.* ATPrPu to see if the Fed under-informs the public about its policy rate change. Or, we can compare ATPr with AT to see if extra economic variables dilute the information contained in private information.

4.2 Informative bigrams in explaining policy rate changes

It is of interest to run the most flexible ATPrPu model for the whole sample period and see what the influential bigrams are. Should public information be succinct enough, there should be no textual information extracted by the DS-LASSO predictive regression. In other words, if public information fully summarizes future policy directions and magnitudes, conditional on the large information set used in the augmented Taylor rule, private information should only contain noise.

For this exercise, we consider $h = 1, 2, 3$ and 4 step-ahead forecasts of policy rate change Δr_t following Equation (2), and average over the coefficient estimate β across the four regressions. In addition, we separate regressions for rate increases and rate decreases. Figure 1 shows the word cloud of the most informative bigrams. Font with a larger size

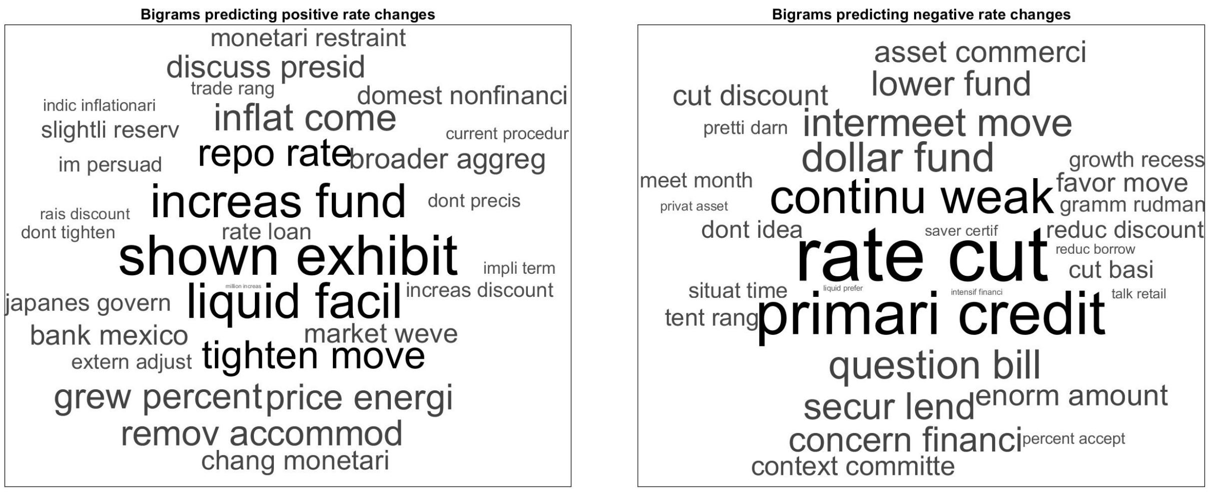

Figure 1: The most informative bigrams that predict positive (left) and negative (right) rate changes. The bigrams are identified by DS-LASSO in the ATPrPu model using the whole sample period. Bigram influence is measured by the coefficient estimate in an absolute value weighted by its normalized R-squared contribution.

indicates a larger average coefficient in absolute value and thus is more influential.

It comes with no surprise that both positive and negative rate changes are associated with bigrams related to the real economy and inflation, broadly in line with Taylor (1993)'s original view. For example, "shown exhibit" and "continu weak" may signal the state of the economy and are associated with positive and negative rate changes, respectively. Also, "infla come" and "lower fund" may imply how inflation status affects FOMC's decision. The financial market, as captured by "primari credit", "secur lend", and "concern financi", tends to affect decisions with negative changes.

For robustness, we also apply the adaptive LASSO Zou (2006) method with stratified cross-validation for time series data. The set of selected bigrams is nearly identical. It is reassuring to see that both LASSO methods pick similar bigrams, as both enjoy the oracle property (*i.e.* asymptotically correct selection) when combined with double selection. Unlike most LASSO methods that do not allow for statistical inference, the employed DS-LASSO allows for both asymptotic and finite-sample inference, enabling us to compute in-sample statistics. For $h = 1$ (h=4), around 1/3 (1/5) of the selected bigrams are statistically significant at 10% level in the ATPrPu model, highlighting the marginal gain from textual information in the augmented Taylor rule. Furthermore, for $h = 1$, the adjusted R-squared ranking T<AT≈ATPu<ATPrPu makes it clear that additional economic variables improve upon the simple Taylor rule, with additional public information unable to further the in-sample fit. Yet, once the model is augmented with textual information extracted from FOMC transcripts, a significantly higher adjusted R-squared is observed. This finding lends evidence on the incremental information that is contained in the private discussions affecting FOMC's decisions, but is largely absent in the Fed's public communications.

4.3 Forecasting with private and public information

Due to the persistence of economic variables, past macroeconomic conditions are expected to be informative about the near-future policy trajectory. This implies that a multi-step-ahead Taylor-type principle is in place.[6] Furthermore, public communications, particularly forward guidance, are meant to anchor future market expectations on policy targets and thus should provide multi-step-ahead predictive power. To see how textual information, emerging from both public and private information, augments the information set used in Taylor-type decision rules, we conduct a multi-step forecasting exercise following the statistical procedure introduced in Section 3.3.

Since the 2008 financial crisis, the US economy has entered a decade-long ultra-loose monetary policy regime. To see if central bank communications are robust in the "zero lower bound" era, we also consider predictive regressions that replace the federal funds target rate (FFTR) with the Wu-Xia shadow federal funds rate (SR) Wu and Xia (2016). The latter is implied by movements in the yield curve caused by non-conventional monetary policies such as asset purchase programs. Therefore, it proxies what the short-term interest rate would be, should it be allowed to go below zero.

From Table 1 that summarizes the RMSE obtained by the proposed rolling-window procedure for different forecast horizons, ATPr and ATPrPu stand out among all model

[6] This can be derived in a model where the monetary authority follows a log-linear Taylor rule with interest rate smoothing.

Table 1: RMSE of Taylor and text models

T	TPu	TPr	TPrPu	AT	ATPu	ATPr	ATPrPu	Pr	Pu
One-step ahead forecast, $h = 1$									
31.2	30.8	28.7	29.0	28.5	29.2	28.1	28.1	33.7	36.3
30.6	29.3	28.2	27.9	28.0	27.8	27.4	27.2	34.5	34.0
Two-step ahead forecast, $h = 2$									
30.7	31.6	29.3	29.6	29.0	28.8	28.3	28.2	32.4	33.5
31.0	31.3	29.4	29.6	28.7	28.4	28.1	28.3	33.2	32.8
Three-step ahead forecast, $h = 3$									
31.6	31.4	30.8	30.6	29.8	30.0	28.7	28.9	35.2	31.0
32.7	32.5	33.1	31.7	30.8	31.8	31.2	32.0	33.3	32.8
Four-step ahead forecast, $h = 4$									
32.5	31.8	31.2	31.0	30.2	29.7	29.2	29.6	31.4	29.9
33.7	33.4	32.5	32.7	31.1	30.8	29.3	30.2	34.5	33.0

Each panel in the table corresponds to a forecast horizon $h = 1, 2, 3$ or 4. The upper (lower) row in each panel indicates the RMSE of a predictive regression with the federal funds target rate (shadow rate) as the dependent variable. "T", "AT", "Pr", and "Pu" indicate Taylor, augmented Taylor, private, and public information, respectively. See Section 4.1 for model specifications.

specifications. This indicates that policy trajectories are determined by central bank information and a broad set of economic variables. In particular, when we look at both T and AT models, it is clear that private information brings predictive power that is captured neither by public information nor by economic variables. The only exception is the SR predictive regression for $h = 3$. Overall, for both FFTR and SR, predictive power extracted from private information improves forecasting performance more than public information.

Furthermore, the addition of a richer set of economic variables does not overshadow the incremental effect of private information: ATPr clearly wins over AT and ATPu. It is interesting to see that even though AT models outperform T models due to a larger information set, a result consistent with Stock and Watson (2002), private information still brings useful information to our predictive regressions.

Figure 2 shows the DM test results for all pairs of competing models for one-step-ahead forecast change in FFTR and SR. Two main observations are made: (i) the Fed

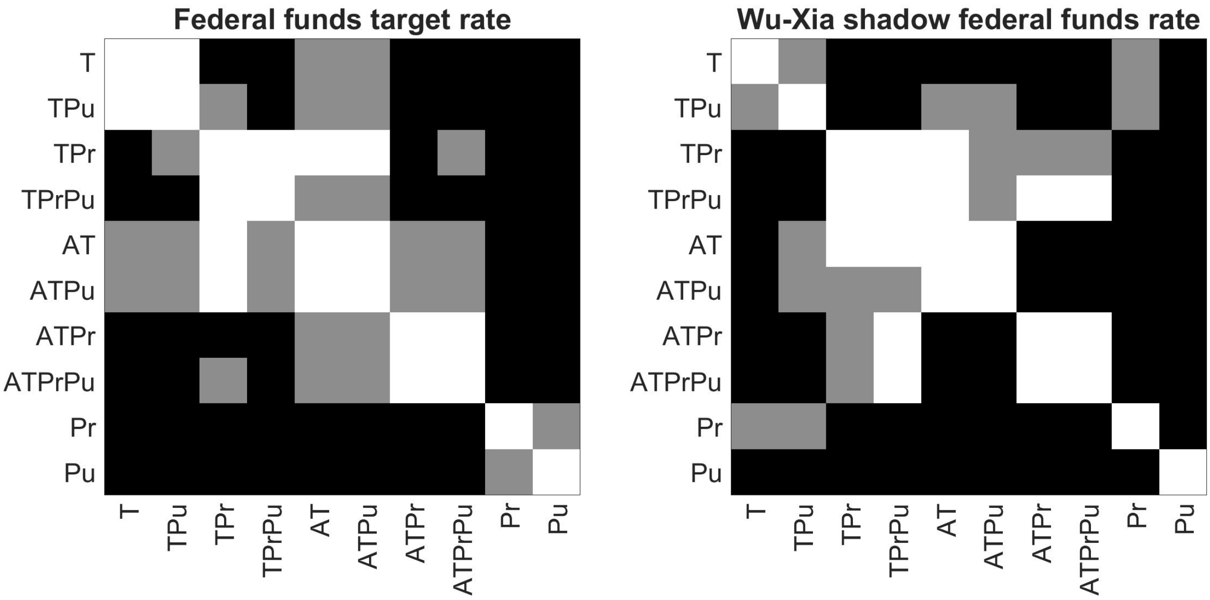

Figure 2: Test for equal predictive accuracy. Pairwise DM test results are shown, using the difference between squared forecast errors from two competing models as the loss differential. Black and grey indicate rejection of the null of equal predictive accuracy at the 5% and 10% levels, respectively. White indicates a failure to reject the null.

under-informs the market on its policy trajectory; (i) public communications do not statistically improve upon a larger set of macroeconomic variables. For $h > 1$, these observations remain and thus are omitted for clarity.

For both FFTR and SR, the TPr model outperforms the T model at a 5% level, implying a significant predictive improvement introduced by textual information in FOMC meeting transcripts over conventional Taylor variables. The improvement in forecast errors is around 5 basis points, which is an economically significant scale considering the target rate movement is usually 25 basis points. With a richer set of economic variables, the ATPr model outperforms the AT model at a 10% level for the FFTR. This means that private information tends to contain more information than economic observables.

Lastly, pure text models perform poorly, as shown in Table 1, statistically dominated by all other models as shown in Figure 2. Econometrically, this is due to the bias caused by the lack of economic variables used in the linear regression model. This means that the Fed's decision rule still follows a Taylor-type principle, albeit not fully.

4.4 Implications and related literature

Testing results from the previous section support the hypothesis that the Fed under-informs the market. This is in agreement with two interpretations in the literature.

First, the way in which the Fed conveys or uses information on the state of the economy may differ from the Taylor principle. As the market usually assumes that the policy rate moves according to a Neo-Fisherian model as in (1), this discrepancy creates an information effect (Nakamura and Steinsson, 2018). Our text-as-data approach thus shows that it is possible to design information measures that predict future policy paths as in Gürkaynak et al. (2005) and Hansen and McMahon (2016). One additional insight from our result is that the information effect is predominantly found when analyzing private rather than public information. Although the TPu model outperforms the T model, ATPu does not outperform AT. Therefore, public information does not provide an information advantage over augmented economic variables. Our result confirms Acosta (2015) that documents the difference between private and public information, but in a predictive regression framework.

Second, as it is not possible to obtain meeting transcripts in real-time, the information advantage of private information means that the Fed only hints at "most likely" future policy movements in its public communications, leaving room for adjustments (Stein, 2014; Blinder et al., 2017). Our result lends evidence to support this intentional uncertainty that effectively preserves the Fed's credibility and accountability (Jia and Wu, 2022).

4.5 Effect of private information over time

The previous section investigates the average forecasting performance of models across the whole $T-h-\tau+1$ testing period introduced in Section 3.3. It is of interest to explore whether, uniformly or episodically, the information advantage from private information improves upon Taylor-type models.

Figure 3 shows the one-step ahead cumulative squared forecast errors of all models relative to the total squared forecast errors of the T model. Results for $h > 1$ are comparable and available upon request. A less than 1 end point indicates the model outperforms the Taylor model in a RMSE sense. The increase of squared errors from all plots exhibits three jumps, indicating major monetary policy shocks that took place. The jump prior to 1995 is associated with a tightening cycle between early 1994 to mid-1995 which naturally makes the prediction of rate changes difficult when those changes are non-zero. The other two jumps are apparently related to economic conditions: the 2001 recession and the 2008 financial crisis.

For FFTR, models with private information are observed to robustly outperform models without, while public information only marginally improves upon the Taylor model from the 2000s. Consistent with our previous findings, augmented with a larger set of economic variables, public information does not seem to generate any information advantage, as seen by the nearly indistinguishable curves of the AT and ATPu models. Private information, on the other hand, still improves upon AT models uniformly. Thus, in terms of short-term interest rates, the Fed tends to under-inform the market in its public communications in a constant manner, not specific to any certain point in time.

The story changes when we turn to SR, which facilitates the zero lower bound and starts to differ from FFTR since the early 2000s. Notice the cumulative squared errors for FFTR level off for all models after 2008. Around that time, policy rates, and thus the dependent variable in rolling-window regressions, are basically zero with little to no variations, leading to near-zero coefficients and forecast errors. This is not the case for SR. The Fed resorted to non-conventional monetary policy with the implied short-term interest rate, or SR, showing large variations due to the uncertain economic conditions at that time. Among T models, private information does not significantly improve upon prediction until the 2000s. As SR is imputed from yield curve movements, this finding lends evidence to Nakamura and Steinsson (2018) who document that central bank information effect becomes effective rather recently. Interestingly, the additional economic

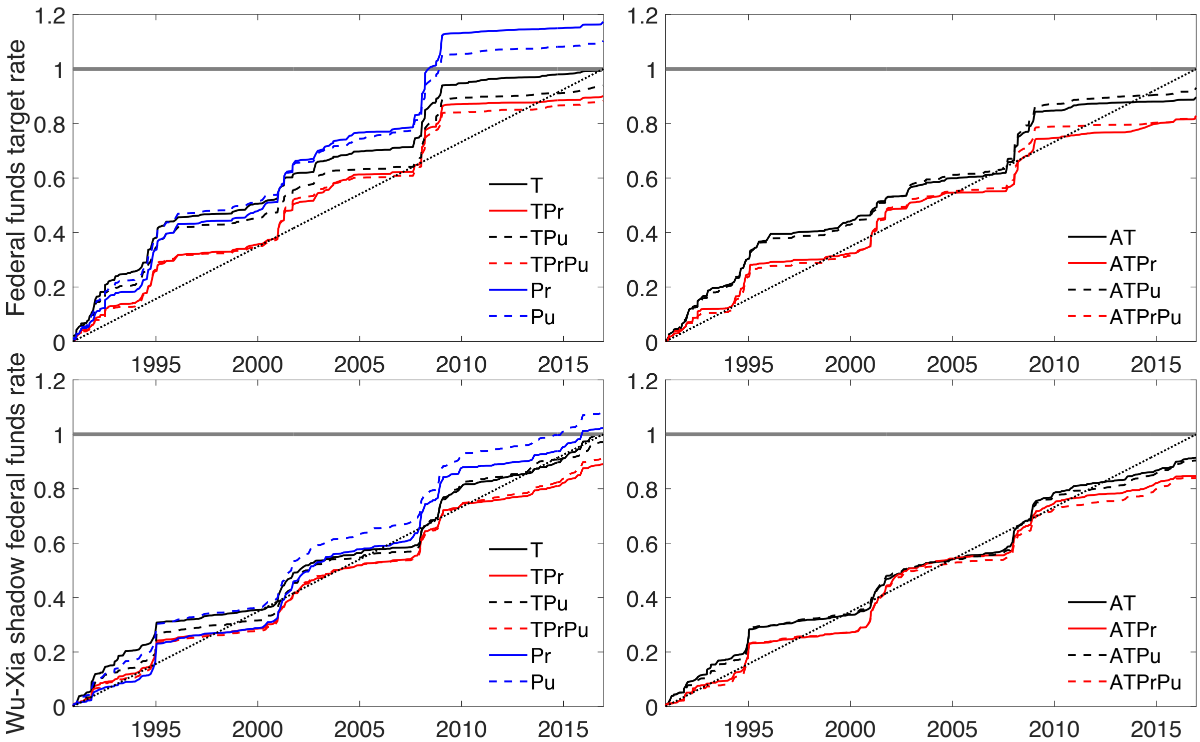

Figure 3: Cumulative squared forecast errors. Forecast errors are computed for each rolling window under all model specifications. Reported is the cumulative squared forecast errors of a model relative to the sum of total squared forecast errors generated by the Taylor model for $h = 1$.

variables seem to help reduce forecast errors prior to the 2008 financial crisis, as seen by the converging squared errors curve between the AT and ATPr models or between the ATPu and ATPrPu models. This may suggest that the final policy decision during that period closely followed a Neo-Fisherian, or augmented Taylor-type, policy rule as in (1), while FOMC members might have expressed some concerns over the prolonged lax lending standards that led to the 2008 housing bubble and financial stress.

Next, we statistically assess whether private information has an information advantage over economic variables and public information used by the market to gauge future policy trajectory. Considering structural breaks and instabilities due to the 2001 and 2008 recessions, we use the fluctuation rationality test of Rossi and Sekhposyan (2016) that is robust to a wide range of model assumptions and forms of structural instability. Following Equation (33) in their paper, test statistics are constructed over time for the later sample

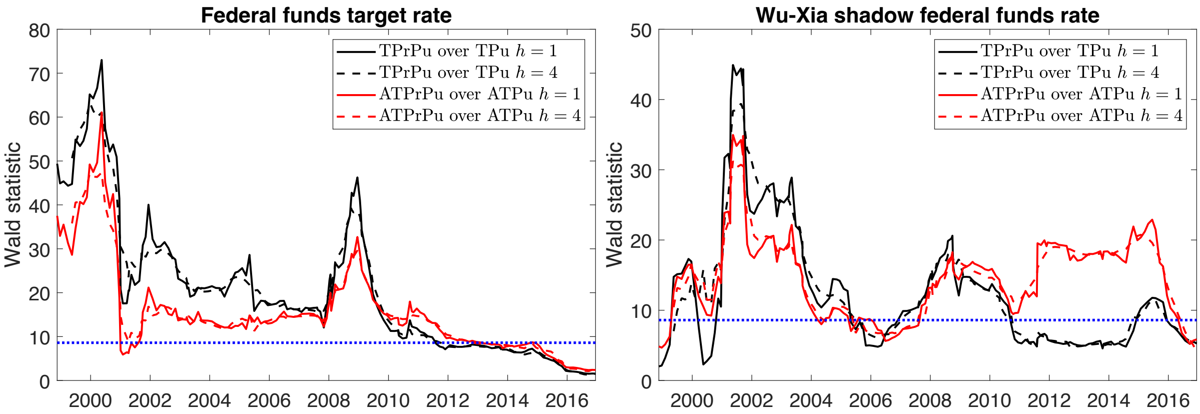

Figure 4: Fluctuation test for model pairs with and without private communications. Stability of one- and four-step-ahead forecasting performance of models with textual information from private communications relative to those without. The constant dashed line indicates the 5% critical value reported by Rossi and Sekhposyan (2016).

period.[7] The results are given in Figure 4.

For both FFTR and SR, we see periods of information advantage. In the former case, the information advantage in private information is predominantly present prior to 2010, especially at the beginning of the 2000s, and when the model is not augmented with additional economic variables. After 2010, the zero lower bound causes the information advantage to diminish, similar to our findings in Figure 3. In the latter case, the peak of information advantage is observed after the 2001 recession when SR started to differ from FFTR. It is worth noting that after entering the zero lower bound era, the information advantage of private information quickly diminishes for models without the larger set of economic variables, whereas AT models with private information produce a significant information advantage over AT models without it. When comparing the two bottom plots in Figure 3, it becomes clear that while meeting deliberations convey non-conventional monetary conducts such as asset purchase programs, private information alone tends to generate inefficient forecasts. In other words, during economic downturns, the Fed utilizes a larger information set when setting its (non-conventional) policy target, and the information advantage is only observed conditional on it. This has implications for

[7] Test statistics are based on sections of forecasting errors, which are themselves obtained from rolling window regressions. Cut-off points are optimally chosen according to Rossi and Sekhposyan (2016), and this leaves us with a time series of test statistics after 1999.

the transparency of central banks' communication during economic downturns, when meeting deliberations can show more divergent opinions.

Lastly, we find that the meeting discussions do effectively inform future policy trajectories, but the Fed fails to fully communicate them to the public. This can be seen from similar evaluations of the fluctuation test statistics for $h = 1$ and $h = 4$, suggesting that the information advantage of private information extends to no less than 4 meetings.

5 Discussion

This paper compares the predictive power of private transcripts and public minutes and statements with respect to actual decisions on the targeted interest rate changes. The analysis indicates that the textual information contained in transcripts is a more powerful instrument for explaining future targeted interest rates. Notably, our empirical analysis is simple by design: (1) it ignores speakers' effects and does not distinguish economy and policy go-rounds; (2) it builds on linear Taylor rules and simple LASSO-type predictive models; and (3) it uses revised macro variables. Improving upon these features either sharpens the edge the private discussions have over publicly available information or leads to a smaller role that economic variables play. Therefore, our findings establish a conservative lower bound on differences in predictive power between public communications and private discussions.

Currently, transcripts are released with a substantial five-year time delay to facilitate efficient information flow and opinion exchange among the board members, whose behavior is affected by various factors, such as career concerns (Hansen et al., 2018). While the extent of transparency in central bank communications is subject to institutional constraints, such as its credibility concern, welfare function, and the format of statements (Jia and Wu, 2022), our study provides direct evidence for the Fed under-informing the market about its future policy trajectory. This can have profound policy implications, because it suggests room for adjusting policy transparency, whenever the Fed chooses to either forward guide or “shock” the market. For example, greater transparency can be achieved if more information from private discussions is communicated to the public via better structured minutes and public speeches by the board members (Bernanke, 2004; Gürkaynak et al., 2005; Nakamura and Steinsson, 2018), or releasing the transcripts with a shorter time lag.

References

Acosta, M. (2015). FOMC responses to calls for transparency. *Finance and Economics Discussion Series 2015-060*, Washington: Board of Governors of the Federal Reserve System.

Andreou, E., E. Ghysels, and A. Kourtellos (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics 158*(2), 246–261.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Bernanke, B. S. (2004). Fedspeak. Remarks at the Meetings of the American Economic Association. San Diego. California. January 3.

Bernanke, B. S. (2015). The Taylor rule: A benchmark for monetary policy?

Blinder, A., M. Ehrmann, J. De Haan, and D.-J. Jansen (2017). Necessity as the mother of invention: Monetary policy after the crisis. *Economic Policy 32*(92), 707–755.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature 46*(4), 910–45.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal 21*(1).

Cieslak, A., A. Morse, and A. Vissing-Jorgensen (2019). Stock returns over the fomc cycle. *The Journal of Finance 74*(5), 2201–2248.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1), 134–144.

Favero, C. A. (2006). Taylor rules and the term structure. *Journal of Monetary Economics 53*(7), 1377–1393.

Fehrler, S. and N. Hughes (2018). How transparency kills information aggregation: theory and experiment. *American Economic Journal: Microeconomics 10*(1), 181–209.

Fernandez, A., E. F. Koenig, A. Nikolsko-Rzhevskyy, et al. (2008). The relative performance of alternative taylor rule specifications. *Staff Papers* (Jun).

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Ghysels, E. and R. Valkanov (2009). Granger causality tests with mixed data frequencies. *Kenan-Flagler Business School Working Paper*.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gilchrist, S., D. López-Salido, and E. Zakrajšek (2015). Monetary policy and real borrowing costs at the zero lower bound. *American Economic Journal: Macroeconomics 7*(1), 77–109.

Gorodnichenko, Y., T. Pham, and O. Talavera (2023). The voice of monetary policy. *The American Economic Review 113*(2).

Greenspan, A. (1993). Comments on FOMC communication. in u.s. house of representatives, hearing before the committee on banking, finance and urban affairs, 103rd congress, october 19, 1993.

Greenwood-Nimmo, M. and Y. Shin (2012). Shifting preferences at the fed: Evidence from rolling dynamic multipliers and impulse response analysis. *Available at SSRN 1810643 2012*.

Gürkaynak, R., B. Sack, and E. Swanson (2005). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *International Journal of Central Banking 1*(1), 55–93.

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

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Hanson, S. G. and J. C. Stein (2015). Monetary policy and long-term real rates. *Journal of Financial Economics 115*(3), 429–448.

Hornik, K. (2007). *Snowball: Snowball Stemmer*. R package version 0.0-1.

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*(1), 55–67.

Jia, C. and J. C. Wu (2022). Average inflation targeting: Time inconsistency and intentional ambiguity. Technical report, National Bureau of Economic Research.

Keblowski, P. and A. Welfe (2004). The ADF–KPSS test of the joint confirmation hypothesis of unit autoregressive root. *Economics Letters 85*(2), 257–263.

Li, J. and W. Chen (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting 30*(4), 996–1015.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance 66*(1), 35–65.

Lucca, D. O. and F. Trebbi (2009). Measuring central bank communication: an automated approach with application to fomc statements. Technical report, National Bureau of Economic Research.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Meade, E. E. and D. Stasavage (2008). Publicity of debate and the incentive to dissent: Evidence from the us federal reserve. *The Economic Journal 118*(528), 695–717.

Morse, A. and A. Vissing-Jorgensen (2020). Information transmission from the federal reserve to the stock market: Evidence from governors' calendars. *Unpublished Working Paper*.

Nakamura, E. and J. Steinsson (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics 133*(3), 1283–1330.

NBER (2020). Us business cycle expansions and contractions. https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions. (Accessed on 10/30/2020).

Reis, R. (2013). Central bank design. *Journal of Economic Perspectives 27*(4), 17–44.

Romer, C. D. and D. H. Romer (1989). Does monetary policy matter? A new test in the spirit of friedman and schwartz. *NBER Macroeconomics Annual 4*, 121–170.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review 94*(4), 1055–1084.

Rossi, B. and T. Sekhposyan (2016). Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics 31*(3), 507–532.

Rühl, T. R. (2015). Taylor rules revisited: ECB and Bundesbank in comparison. *Empirical Economics 48*(3), 951–967.

Shapiro, A. H. and D. J. Wilson (2022). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. *The Review of Economic Studies 89*(5), 2768–2805.

Siklos, P. L. and M. E. Wohar (2005). Estimating Taylor-type rules: An unbalanced regression? *Econometric Analysis of Financial and Economic Time Series 20*, 239–276.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review 36*(5), 975–1000.

Stein, J. C. (2014). Challenges for Monetary Policy Communication. *speech delivered at the Money Marketers of New York University, New York, May 6*.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, Volume 39, pp. 195–214. Elsevier.

Vissing-Jorgensen, A. (2020). Informal central bank communication. Technical report, National Bureau of Economic Research.

Woodford, M. (2001). The Taylor rule and optimal monetary policy. *American Economic Review 91*(2), 232–237.

Wu, J. C. and F. D. Xia (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking 48*(2-3), 253–291.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

ONLINE APPENDIX

A Additional remarks on empirical strategy

Below are some remarks we make about the empirical strategy introduced in Section 3.1.

Remark 1. One may consider other weighting schemes as opposed to (3). For example, speech timings can be weighted equally or with a bell-shaped function. In our empirical study, we find that neither outperforms a simple linear trend as in (3) in terms of out-of-sample forecasting performance. Furthermore, one can ignore the length and the timing effects altogether and assign equal weights to all speeches in a meeting. In such a case, χ_{t-h} in (2) is given by $\chi_{t-h} = (\sum_{i=1}^{n_{t-h}} x_{(t-h)i})/n_{t-h}$ – the average frequency counts of bigrams used in meeting $t-h$. Alternatively, since speeches are ordered chronologically, one may use the Mixed Data Sampling (MIDAS) of Andreou et al. (2010) to summarise meeting-level speeches. In this case, a weighted average replaces χ_{t-h} in (2) by $\sum_{i=1}^{n_{t-h}} m_i x_{(t-h)i}$ with the MIDAS weights m_i implied by a smoothing function whose shape depends on some estimable parameters. However, neither approach is favored because they ignore inter-speech variation in deliberations, potentially leading to information loss. In the MIDAS case, the weighting function needs to be estimated by maximum likelihood, which cannot be implemented in our case due to the high-dimensional text variable. Also, the fact that the DTM X contains only dummy variables renders the smoothing effect of the MIDAS weighting scheme ineffective and uninterpretable.

Remark 2. Model (2) is an ARDL model with order $(1, 1)$; namely there is one past dependent variable and one vector of past explanatory variables. The generalization to ARDL(p, q) is straightforward and thus omitted. Moreover, the regression model controls for past policy instruments Δr_{t-h} and economic history z_{t-h} on the right-hand side and thus ensures that β captures additional predictive power of textual information. Equivalently, our model gives a Granger causality interpretation (Ghysels and Valkanov, 2009) to the role of texts.

Remark 3. A cut or a hike in the target rates usually equals to few multiples of 25 basis points,[8] suggesting that the decision rule (1) might be better described by a discrete choice model, say an ordered probit model. However, as policy rate changes are usually small, a linear specification is not a key issue and can serve as a first-order approximation where estimated β closely follows the average marginal effect obtained from the ordered

[8]Since December 2008, the Fed has formulated its policy target in terms of a 25-basis-point interval. But changes of the lower and the upper bounds of the interval are the same and equal to a few multiples of 25 basis points.

probit model.

B Descriptive statistics

Table 2: Macroeconomic Variables Summary Statistics

	Min	Median	Mean	Max	SD	Description
Change	-1.25	0	-0.03	1.12	0.25	Change in target rate
Recession	0	0	0.14	1	0.35	Indicator variable, based on NBER (2020)
GDPC1	$6,794.88$	$10,575.10$	$11,530.63$	$16,663.65$	$3,019.97$	Real Gross Domestic Product
GDPPOT	$7,168.14$	$10,710.13$	$11,748.89$	$17,145.98$	$3,087.61$	Real Potential Gross Domestic Product
GAP	-7.36	-1.62	-1.88	2.31	2.02	GDPC1 - GDPPOT
CPILFESL	93.67	160.73	164.02	235.35	41.07	Consumer Price Index for All Urban Consumers
DFF	0.07	5.24	4.86	14.51	3.14	Effective Federal Funds Rate
MICH	1.07	3.07	3.15	5.03	0.59	University of Michigan: Inflation Expectation
PHIL.CPI	-0.92	3.08	3.11	6.23	1.12	Philadelphia Fed: Survey of Professional Forecasters, Mean CPI Level
ShadowRate	-1.99	5.07	4.71	14.92	3.48	Wu and Xia (2016)
ln(CPI)	4.54	5.08	5.07	5.46	0.26	ln(CPILFESL)
ln(DFF)	-2.60	1.66	1.08	2.68	1.36	ln(DFF)
ln(MICH)	0.06	1.12	1.13	1.62	0.19	ln(MICH)
ln(PHIL.CPI)	-0.05	1.13	1.09	1.83	0.34	ln(PHIL.CPI)
ln(ShadowRate)	-1.52	1.67	1.49	2.70	0.74	ln(ShadowRate)
Δln(CPI)	0.06	2.72	3.05	7.33	1.33	[ln(CPILFESL$_t$) - ln(CPILFESL$_{t-1}$)]*400
Δln(DFF)	-538.23	-3.86	-21.82	151.93	83.95	[ln(DFF$_t$) - ln(DFF$_{t-1}$)]*400
Δln(MICH)	-371.49	0	-6.66	295.58	68.85	[ln(MICH$_t$) - ln(MICH$_{t-1}$)]*400
Δln(PHIL.CPI)	-233.18	-6.55	-7.71	262.45	72.24	[ln(PHIL.CPI$_t$) - ln(PHIL.CPI$_{t-1}$)]*400
Δln(ShadowRate)	-491.52	-13.73	-24.18	170.17	85.02	[ln(ShadowRate$_t$) - ln(ShadowRate$_{t-1}$)]*400

Note: When ShadowRate is negative, the observations for ln(ShadowRate) and Δln(ShadowRate) are omitted.

Table 3: Speech Level Summary

	All Data		Pre Forward Guidance		Post Forward Guidance		Description
	Mean	SD	Mean	SD	Mean	SD	
Change	-0.02	0.29	-0.04	0.35	-0.01	0.22	Change in target rate
Post Forward Guidance	0.51	0.50	0	0	1	0	Indicator for post forward guidance policy (i.e. 1994-02-04)
Recession	0.10	0.31	0.08	0.27	0.13	0.33	Indicator variable, based on NBER (2020)
Loughran McDonald Count	3.16	9.29	1.55	4.22	4.67	12.08	Number of words present in Loughran McDonald lexicon (2011)
Loughran McDonald Score	-1.09	4.32	-0.54	2.11	-1.60	5.61	Total sentiment score for Loughran McDonald lexicon (2011)
Word Count	107.76	277.87	56.86	125.94	155.71	360.93	Word count of speech