

HOW NATURAL LANGUAGE PROCESSING WILL IMPROVE CENTRAL BANK ACCOUNTABILITY AND POLICY

Charles W. Calomiris and Harry Mamaysky

In the movie “True Lies,” Arnold Schwarzenegger plays the role of a skilled spy. In one scene, he is injected with truth serum in preparation for interrogation. Under the influence of the truth serum, he reveals how he plans to kill his captors and escape and then applies his skills and weapons precisely as he said he would.

Imagine if central bankers similarly were required to tell us, in a clear, credible, and comprehensible way, how they will follow a systematic strategy to use their knowledge, skills, and policy instruments to react to the economic data they will face. Many observers have argued that this would substantially improve monetary policy by eliminating unnecessary uncertainty associated with the difficulty of forecasting a central bank’s reaction function—the mapping from publicly observable data to publicly observable actions by the central bank (Calomiris 2018).

Dincer, Eichengreen, and Geraats (2019) identify self-discipline, predictability, and credibility as the fruits of central bank transparency. Central bankers following a disclosed systematic strategy of monetary policy clearly would achieve the highest marks on all three dimensions. They would display perfect self-discipline by

Cato Journal, Vol. 40, No. 2 (Spring/Summer 2020). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.40.2.11.

Charles W. Calomiris is the Henry Kaufman Professor of Financial Institutions at Columbia Business School and a Visiting Fellow at the Hoover Institution. Harry Mamaysky is Associate Professor of Professional Practice at Columbia Business School, where he serves as the Director of the Program for Financial Studies.

forcing themselves as a group to formulate a strategy to guide their actions, as economists have shown is highly desirable (e.g., Kydland and Prescott 1977). Because their actions would be predictable reactions to changing circumstances, they would create no avoidable uncertainty. They would enjoy perfect credibility, proving with each relevant data release and every action that they were reliable public servants.

Of course, there would still be the unavoidable long-term uncertainty associated with the necessary changes over time in the central bank's agreed reaction function. But, with the exception of reactions to severe crises, those changes would be announced in advance.¹ The unpredictability arising from changes in an evolving framework for the central bank's reaction function is necessary because the structure of the economy is not static, because frameworks make use of different tools under different circumstances (e.g., at very low interest rates, quantitative easing [QE] actions may replace interest rate actions), and because humans' understanding of that structure is always imperfect.

Unfortunately, however, many commentators on central bank transparency take too narrow a view of the topic, and consequently, they fail to consider as part of transparency, the most important part: formulating and disclosing the strategy for central bank policy. Central bankers often receive very high grades for "transparency" despite the absence of any articulation of a systematic framework. Dincer, Eichengreen, and Geraats (2019) offer a careful, comprehensive, and valuable cross-country comparison of central banks' practices, and their measure takes into account not only data about central banks' speeches and minutes, but also the extent to which the central bank provides useful information about its policy intentions. Their measure of transparency highlights important differences across countries. But their definition, which scores the Fed as one of the most transparent central banks in the world (exceeded only by two countries), may not penalize central bankers sufficiently for failing to agree upon and announce a *systematic* framework for policy. Announcing a specific policy framework is one of the criteria in their scoring model, but they may give too much weight to other

¹Most advocates of this view also recognize that central banks could also preserve the option to deviate from their announced strategy in response to an economic or financial system emergency.

factors—whether central bankers give public speeches and testimony, disclose the minutes of their meetings fully and promptly, and share some of their forecasting opinions with the public.

Although it is desirable for central bankers to disclose the content of their speeches, testimony, committee deliberations, and forecasts to the public, the act of providing those words and facts to the public may not be very useful if a central bank avoids clearly stating a systematic framework for policy. True transparency requires central banks to provide understandable and reliable information to the public that makes it possible for the public and their delegated representatives to actually hold central bankers accountable for their statements and behavior, and that is not possible in the absence of a detailed, quantitative description of central bank strategy, embodied in a reaction function.

Meltzer (2014) provides a brief summary of the history of policy errors by the Fed, and shows how adherence to systematic policy would have prevented many of them. Systematic policy should not be seen as a constraint on what the Fed is able to do, either in the long run or the short run (Meltzer, along with all serious advocates of systematic policy, understood the necessity of revisions in the framework over time, and the desirability of short-run deviations from it during financial crises). Meltzer argued, however, that the point of systematic policy was to make the Fed accountable. Meltzer recognized that transparency only in the sense of mechanical disclosure of facts, without a commitment to systematic policy, is incapable of producing accountability.

Without a systematic policy framework, neither Congress (which is responsible for overseeing the Fed under Article I, Section 8, of the Constitution) nor the public can exert any effective oversight of the Fed. Requiring the Fed to announce a systematic policy allows others to evaluate that procedure, to verify whether the Fed is following it, and to ask specific questions about observable deviations. Without a systematic policy, questions to Fed officials during congressional hearings produce vague and lengthy responses, a form of double talk that Fed officials are expert at delivering, precisely because doing so avoids accountability. Alan Greenspan is said to have once quipped, “If you understood me, then I misspoke.” The joke is funny for a reason.

Fed governors are also expert at orchestrating their recorded Federal Open Market Committee (FOMC) deliberations to ensure

that potentially embarrassing real disagreements are negotiated outside the meeting rather than reflected in conflicting votes. This has led to a severe secular decline in recent years in the public FOMC voting disagreements among Fed governors (Thornton and Wheelock 2014). From 1957 to 2013, Federal Reserve bank presidents dissented 241 times, and Fed governors dissented 208, but from 1994 to 2013, governors effectively stopped dissenting (presidents account for 72 of 76 dissents during that time frame). Presidents also tend to dissent much more when they are in favor of tighter policy—78 percent of their dissents were for tighter policy, in contrast to governors for whom only 28 percent of dissents were for tighter policy.

Similarly, although Hansen, McMahon, and Prat's (2018) analysis of FOMC transcripts finds that greater disclosure of Fed discussions are helpful in encouraging FOMC members to inform themselves more, they also confirm earlier findings that the greater transparency produced by the publication of FOMC minutes promotes greater apparent conformity of expressed opinions, as policymakers are less willing to express differences publicly. Cannon (2015) shows that this change has been particularly pronounced for Fed governors. These findings confirm the early research by Meade and Stasavage (2008), who find that the decision to disclose FOMC transcripts reduced dissent and encouraged greater use of prepared comments rather than unscripted ones. Acosta (2015) also finds evidence of greater conformity in word choice after the increase in disclosure. These findings reinforce the distinction between disclosure and true transparency. Indeed, it appears that, in the absence of an articulation of systematic policy framework, greater disclosure may have made the disclosed discussions less meaningful to the public.

What about the Fed's commitment to a long-run 2 percent inflation target? What effect does that have on improving transparency? The 2 percent commitment was a major step in the direction of making policy goals clear, but it said little about short-run strategy to achieve that goal. And even that commitment is now uncertain. The 2 percent inflation goal is not a matter of law, merely of Fed officials' voluntary commitment, and the Fed recently decided to reconsider that goal (see Thomson Reuters 2018). When your long-run goals are open to revision, and your short-term strategy is either secret or nonexistent, how can you be considered a model of self-discipline, predictability, and credibility?

But perhaps this is all about to change. What if central bankers found it in their own interest to create and announce a systematic framework for monetary policy? Why would that happen? Returning to the “True Lies” analogy, if central bankers knew that they would be forced to speak under the influence of truth serum, then they might decide that it is in their best interest to adopt and announce a systematic monetary policy. Central bankers who know that their thoughts are being revealed publicly—and then analyzed systematically—would no longer be able to use the absence of a systematic framework to avoid accountability for clear mistakes in their thinking, or for duplicity or opacity in their policy releases, or inconsistency over time in their reactions to information. Avoiding a systematic policy, therefore, would cease to be attractive to Fed officials as a way to skirt accountability. Adopting a systematic approach to monetary policy that could be defended on its merits would allow them to avoid the embarrassment of revealed errors, dishonesty, and inconsistency.²

But how, you may be wondering, do we get them to take the truth serum? Maybe we don’t have to. Instead, we can subject their words to natural language processing (NLP). Like mind reading and lie detecting, NLP allows researchers to extract the underlying meanings of word flow, even when they have been shrouded by vague or misleading statements by the author of those words. Because it subjects the entirety of central bank communications to in-depth, real-time analysis, NLP shines new light on what central bankers know, believe, and do, and that new light will make it easier for the public to hold central bankers to account for missing, withholding, or distorting information, and this will provide a strong incentive for them to construct a systematic, transparent approach to policy.

Natural Language Processing Is Decoding Central Banking Words and Actions

In the past several years, NLP has been revolutionizing the way social scientists measure the flow of information. NLP has proven to be effective in supplementing traditional data news (TDN), by which

²Hansen, McMahon, and Prat (2018) find a similar “discipline” effect from the disclosure of FOMC transcripts: Fed officials did their homework more when they realized that they were accountable for the quality of their analysis. Similarly, holding them accountable for inconsistent, incompetent, or deceptive actions should encourage them to improve the policy process by making it more systematic.

we mean the familiar flow of quantitative information about market prices and the occasional data releases that measure GDP growth, employment, inflation, industrial production, housing sales, or other relevant economic concepts.

NLP is a systematic approach to reading the flow of text and distilling from that flow measures that capture useful aspects of its meaning. There are many approaches to NLP measurement that have proven useful. By useful we mean that the quantitative measures of the flow of text yield information not already contained in TDN. For example, a massive new literature shows that NLP news can predict high-frequency and low-frequency market price movements and future economic data releases.

NLP has been shown to be useful for measuring the meaning of the words used by central bankers. We do not provide a comprehensive review of that literature here, but instead, a selective review that focuses on how NLP can promote improvements in monetary policy. This has several aspects, and each of them creates real transparency, which should promote accountability, which in turn should encourage monetary authorities to adopt more systematic approaches to policy. The outcomes of NLP research for central bankers can be divided into three categories of influence: (1) NLP makes it easier for the public to tell when central bankers are failing to pursue transparent, systematic policies, or failing to disclose their true beliefs; (2) NLP allows the public to understand which economic phenomena central bank policies are actually responding to; and (3) NLP creates new publicly available information resources that central bankers will increasingly use as inputs into policymaking, which will reduce (although not eliminate) the information asymmetry between central bankers and the public about the state of the economy. In the remainder of this article, we provide examples of each of these three categories.

NLP's Role in Measuring How Nonsystematic and Nontransparent Policy Is

In the absence of a systematic approach to monetary policy that is clearly communicated to the public and that the public can use to interpret clearly the facts disclosed by the central bank, more disclosure by a central bank, per se, does not necessarily produce more useful information about monetary policy. Indeed, there is evidence

that it can have the opposite effect. Davis and Wynne (2016) count the number of words in each FOMC statement and find that this grew from 200–300 words circa 2007 to 800–900 words by the end of 2013. The word count for the policy statements of the Bank of Canada also grew over time, but by less. Lange (2019) finds that the length of Fed statements fell dramatically after 2015 and now averages roughly the same length as prior to the crisis.

Davis and Wynne (2016) also use the Flesch-Kincaid score to examine the comprehensibility of Fed communications, based on a reading of the postmeeting FOMC statement. The Flesch-Kincaid score, developed by the U.S. military in the 1970s, measures the education level needed to comprehend text. Sentence length and the number of syllables per word affect this score. These aspects of language are sometimes viewed as indicative of an intent of deception, but they can also measure the complexity of the policy problem the central bank is grappling with. Davis and Wynne (2016) find substantial variation in the Flesch-Kincaid score for the postmeeting statement, which peaked above 20 in late 2013, implying a fifth-year PhD student level. They find no change in the Flesch-Kincaid score for the Bank of Canada over time. Lange (2019) finds that in 2017–18, under Chair Powell's leadership, the Flesch-Kincaid score for the Fed fell from 14 to 10, but with the Fed's change in policy stance away from tightening at the end of 2018, it returned to 14.

In summary, there has been dramatic variation over time in the Fed's FOMC communications, both in terms of their length and their linguistic complexity. One interpretation of these facts is that there is inherent variation in the complexity of the economy, which necessitates variation in the length and complexity of Fed statements. Another interpretation is that the absence of a systematic framework for policy means that the Fed's policies cannot be explained as the application of preexisting rules; ad hoc policies require lengthy explanation. A third explanation is that difficult circumstances produce intentional obfuscation by policymakers.³

³In an international context, Correa et al. (2017) show that central banks from stable countries (e.g., Germany, Canada, and the Netherlands) produce financial stability reports with much more negative tone than those produced by less stable countries (e.g., Turkey, Argentina, Spain, and Portugal) suggesting obfuscation, as well as window dressing.

Davis and Wynne (2016) also examine market reactions to the postmeeting statement and the relationship between the word count and the grade level of Fed disclosures and the magnitude of market reactions. They find that both greater complexity and greater word count are associated with important and statistically significant increases in the size of market reactions to Fed policy statements, but that magnification effect is only important at times when the Fed is communicating a policy change (defined as a surprising change in policy compared to prior market expectations).

Davis and Wynne (2016) interpret this as evidence of the usefulness of Fed explanations in helping the market understand complex policy changes, but there is a deeper point that they do not make. The magnitude of market reactions to Fed statements indicate the extent to which markets are surprised by the statement, which reflects the opacity of the preexisting policy framework in which those statements are made. The size of the market's reaction measures, at least in part, the extent to which policy is not operating in a systematic and predictable way.⁴ If monetary policy were a perfectly predictable consequence of changes in the economic environment, then market responses to Fed meetings would be nil. From that perspective, the Davis and Wynne findings can be seen as indicative of a lack of transparency when it is needed most. The large market reactions that occur when policy changes are accompanied by the use of more words and greater complexity captures the costs of pursuing confusing, non-systematic, ad hoc policies.

The fact that greater length and complexity are associated with larger market reactions does, however, indicate that these aspects of Fed disclosure are not simply a means of obfuscation (i.e., a systematic version of Greenspan's aforementioned quip). If they were merely indicative of obfuscation, then more words

⁴Another possibility is that the time variation in market reactions may reflect variation over time in the extent of the information about the economy that the Fed reveals by virtue of its statements or actions. The Fed has superior information about the economy, and therefore, its actions can be revealing of that information. Even if the Fed were totally rule based, but the rule depended on better information, policy changes could still cause market reactions as they contain new information. And it is conceivable that those changes could vary over time.

and more complex words would diminish rather than magnify market reactions.

But there is evidence that other aspects of Fed communication may be intended to obfuscate for self-serving purposes. We already noted the evidence in Meade and Stasavage's (2008) and Hansen, McMahon, and Prat's (2018) studies of FOMC transcripts that greater disclosures produce more conformity in the opinions expressed, which is a form of obfuscation to avoid perceived costs of publicly displaying differences of opinion. It remains true, however, that despite the reduced quality of FOMC discussions after 1993, NLP analysis of FOMC minutes shows systematic links between the word choices of FOMC members and the future state of the economy, as shown by Stekler and Symington (2016) and Ericsson (2016), and by Correa et al. (2017) with regard to financial stability reports of global central banks.

Cannon (2015) provides additional evidence about how the Fed's decision to disclose FOMC minutes after 1993 produced changes in Fed officials' behavior. She finds that the 1993 change resulted in fewer comments per FOMC member and longer comment length. These changes, however, were much more pronounced for Fed governors than for Fed presidents. Presidents did not reduce their comments by as much as governors. Cannon (2015) also finds that pre-1993 Fed presidents' use of language exhibited tone that was much more predictive far in advance of the state of the economy, while governors' tone was more of a coincident indicator. After 1993, presidents' speech continued to be positively associated with the state of the economy and predictive of that state, but the correlation between the language of governors and the state of the economy reversed.

These findings are consistent with the view that after the 1993 change, governors actively sought to reduce the meaningfulness of their expressions of opinions, but the presidents did not. Governors apparently are more susceptible to the conformity-inducing effects of transparency. As we already noted, governors were much less likely to dissent than previously, especially in a tightening direction, but dissents did not change noticeably for presidents after 1993 (Thornton and Wheelock 2014). The pre-1993 findings about the predictive power of their word tone for the economy also show that Fed presidents appear to be better informed earlier and more candid about

changes in economic trends, which is consistent with the motivations that gave rise to the Fed's decentralized structure.⁵ In summary, Cannon's findings not only reinforce prior research about the chilling effects of disclosure on the meaningfulness of FOMC discussions, they also show that the decentralized governance structure of the Fed has succeeded in fostering independent sources of opinion.

Similar to potential obfuscation by individual Fed members, there is also evidence that the Fed *as an institution* may distort the information it releases for self-serving purposes. The Fed produces forecasts of the economy, which are included in its Green Book and Teal Book and released to the public with a lag. Sharpe, Sinha, and Hollrah (2018) find that the text produced by Fed economists to accompany these quantitative forecasts contains important information that can be used to improve the accuracy of the published forecasts. In other words, the people who construct the forecasts seem to know more than what they reveal in the actual forecasts. Researchers we have spoken with at the Fed all agree that this is the result of the Fed wanting to avoid any admissions of having missed important indicators in its prior analysis. To avoid appearing to make large mistakes, the Fed avoids sudden shifts in its forecasts, which means that the published forecasts do not change as rapidly as the opinions of Fed economists.⁶

All of these studies illustrate the power of NLP in holding the Fed accountable. The fact that the Fed varies the length and complexity of its statements so much over time, and the high market impact of lengthy and complex statements, reveals the extent to which the Fed avoids systematic, predictable policy reactions to publicly observable information. The increase in FOMC conformity that accompanied the greater disclosure of transcripts, as well as the failure of quantitative forecasts to reflect all new qualitative information in the text of the Green Book or Teal Book, reveal that greater

⁵This historical intent had two aspects. First, regional Feds were expected to be more knowledgeable about their regions, as the result of geographic proximity and their close relationships with bankers in those regions. Second, because Fed presidents are appointed by local business and banking interests, not by the president of the United States, they were expected to be more independent of politics (see Meltzer 2003).

⁶This effect has been termed conservatism—a tendency to underreact to new information that differs from one's priors—and has been identified in other economic contexts (see, e.g., Barberis, Shleifer, and Vishny 1998).

central bank disclosure affects the behavior of central bankers by making them more likely to reduce the value of the information being disclosed. Finally, evidence about systematic differences in the responses of Fed governors and presidents to changes in incentives to be forthcoming have important lessons about the desirability of the decentralized structure of the Fed. By revealing these shortcomings or advantages of the various aspects of the Fed's policy process, disclosures, and structure, NLP provides a basis for preserving desirable features and encouraging improvements in undesirable ones.⁷

NLP as a Mind Reader of Monetary Policymakers

What are central bankers reacting to, and what effects do those reactions have on the economy? Do the answers to those questions comport with what central bankers say they are doing? It is difficult to gauge the answers to these questions using TDN. Clearly identifiable policy actions, such as changes in the target interest rate or the QE policy stance of the central bank, are rare events that coincide with many other changes in relevant data. If they are stable over sufficiently long periods of time, it is possible, using time series analysis, to identify central bank reaction functions that link those identifiable policy actions with data on relevant variables, such as inflation and unemployment. But that is a big if. Central banks' policy reaction functions are nearly impossible to identify reliably from statistical analysis if those reaction functions are subject to frequent change, and here, frequent enough change to render identification impossible may mean once or twice per decade. That would be enough to undermine the power of time series estimation of the parameters of the reaction function.

More fundamentally, changes in interest rates or central bank balance sheets are not the only, or perhaps even the primary, tools of central banks, and that is especially true in recent years. Recently, central bankers have made use of jawboning ("forward guidance") to influence market perceptions of what they might do under certain circumstances, which they hope will affect interest rates and other

⁷We leave as an open question whether, as more systematic analysis is applied to the study of the Fed's textual communications, the possibility arises that the Fed may be able to "game" the algorithms used to analyze its statements. We discuss this further in the conclusion.

asset prices. Ben Bernanke did this as a Fed governor (prior to becoming chairman) in his influential speeches about the bond market in 2003. Alan Greenspan was criticized for not jawboning sufficiently about the frothiness of the housing market in 2005–2007. He has also been criticized for providing a “Greenspan put” to the stock market—effectively supporting stock prices by being perceived as providing an insurance policy, promising to intervene as needed to keep stock prices from falling dramatically.⁸ Such promises, if credible, can affect market behavior. The belief that such talk can magnify the power of monetary policy underlay the use of forward guidance as a regular feature of Fed policy after the crisis.

If monetary policy must be construed more broadly to encompass the effects of Fed statements on financial markets, how do we measure policy, and how do we gauge its effects? One answer is to apply NLP to Fed speeches, postmeeting official statements, and other press releases in order to measure monetary policy as a continuous variable. With that measure of policy in hand, one can then ask what variables influence policy and what variables are influenced by policy.

Measuring monetary policy is a mouthful. How can one measure policy change without deciding in advance what constitutes policy? Along what dimension(s) should policy change be defined? There are many approaches to answering these questions, but the simplest approach is to collapse policy into a single dimension (a continuum of policy defined alternatively as relatively hawkish or relatively Dovish), and to define policy actions as combinations of words that tend to have meaning for market prices. In other words, central banking policy is as central banking policy does.

Perhaps the most advanced approach to constructing such a measure of monetary policy is provided by Prattle, a private data service provider (see Schnidman and MacMillan 2016). Prattle covers all the central banks in major developed and many developing countries. It tracks all of the most relevant publicly available verbal output produced by central banks (official statements, committee minutes, officials’ speeches, press releases). It considers, for each central bank, all the combinations of words or strings of words that appear in this verbal output and identifies which of those word combinations tend to affect various market prices positively or negatively. It uses

⁸Cieslak and Vissing-Jorgensen (2018) show that federal fund rates react much more to negative than to positive stock market returns.

this “training sample” to identify indicators of hawkish or dovish statements, which are then applied out-of-sample to construct a continuous measure over time (measured for each day) of the hawkish/dovish policy stance of the central bank.

Calomiris and Mamaysky (2019a) and Calomiris et al. (2020) analyze these Prattle scores of the major central banks, including the Fed. They validate Prattle scores as a useful measure by showing that they capture both plausible observable influences on monetary policy and the consequences of changes in monetary policy. For example, major QE announcements are associated with large dovish changes in Prattle scores, which result in declines in interest rates and depreciation of the currencies of the dovish central banks.

More interestingly, Calomiris and Mamaysky (2019a) show that monetary policy, as indicated by the Prattle score, responds strongly to variables that are not generally included in Taylor Rules, such as the VIX (Chicago Board Options Exchange [CBOE] Volatility Index), and changes in international capital and reserve levels. They also find that monetary policy matters, not only for interest rates, exchange rates, and the macroeconomy, but also for financial market risk indicators such as the VIX. Their evidence is consistent with what one would expect if a central bank were providing a put option to the stock market. Such a central bank would respond to positive VIX shocks by becoming more dovish. Furthermore, given that policy stance, markets should respond to unexpected dovishness by negatively updating their beliefs about future economic and market conditions (thereby resulting in an increasing VIX). Both conjectures are confirmed by the data for the Fed since the mid-1990s and for the ECB and Bank of England since 2008.

Calomiris et al. (2020) find that their method for measuring the flow of news about the economy (developed in Calomiris and Mamaysky 2019b) also explains a large proportion of future Prattle score changes several months in advance. In other words, central banks respond to economic news that is captured by NLP with a lag, which Calomiris et al. (2020) interpret as reflecting a recognition lag: central banks mainly respond to TDN, and much of the news that is apparent in NLP predicts changes several months later in TDN measures.

By creating a continuous, objective measure of monetary policy, NLP subjects central bankers to a new form of oversight that increases central bank accountability. For example, central bankers

cannot plausibly deny that they are providing put options to the stock market if one can show that their words indicate that they do precisely that. Furthermore, if their policy stance does not respond consistently to macroeconomic variables such as unemployment and inflation, they can be faulted by advocates of such a policy for failing to implement an identifiable, predictable rule.

NLP as an Information Leveler

Central banks invest substantial resources in data collection and are among the best informed parties in the world about the current state of the economy. The fact that central banks possess unique access to timely private information makes it harder to hold them to account for deviating from a systematic policy framework because it may not be possible to distinguish inconsistent policy actions from consistent policy actions that result from access to private information.

But what if much of the best and most timely information were publicly available? In that case, it would be embarrassing if central banks were found to be ignoring such information, or relying on inferior private information in the place of better public information.⁹

There is evidence that the Fed is doing precisely that: failing to make use of timely public information while devoting substantial resources to processing inferior private information. As already noted, Calomiris et al. (2020) find that a parsimonious vector of NLP measures of the news flow contained in Thomson Reuters predicts about a fifth of the variation in the Fed's Prattle score as much as a year into the future. Calomiris and Mamaysky (2019b) find that this same vector of NLP measures predicts stock returns a year into the future for a large sample of countries. Thorsrud (2017) finds that employing a somewhat similar approach to measuring news from a Norwegian business newspaper, he is able to use the news available at the time of the first forecast of quarterly GDP growth to correct about a fifth of the error in the initial GDP growth forecast (that is,

⁹We are not saying that all of the best and timely information on which to base monetary policy is necessarily accessible through NLP. The Fed has other private sources of information in addition to NLP, which may have incremental value in forecasting. For example, FedEx supplies the Fed with real-time information about FedEx delivery volumes that are informative about the state of the economy and that are not publicly available.

about a fifth of the difference between the initial forecast and the final version of the forecast produced 18 months later).

While failing to systematically “nowcast” the economy with concurrent news flow, the Fed maintains a Beige Book procedure involving private interviews with a small number of business leaders in each Fed district, which are assembled and reported to the FOMC. This approach focuses on a narrow aspect of private information (the conscious, qualitative opinions of a small number of people). Newspapers, Twitter feed, and other textual sources likely would add greatly to the nowcasting information contained in the Beige Book.¹⁰ The fact that Prattle scores are forecastable on the basis of lagged news measures used by Calomiris et al. (2020) shows that the Fed’s reliance on the stale news contained in TDN and Beige Book surveys means that it is missing important information that is publicly available through natural language processing.

We do not mean that observation as a criticism of the Fed. NLP is a new field and the power of NLP of news is only now becoming apparent. Our point is that soon it will be impossible for the Fed to pretend that monetary policy can be made without systematic analysis of the news as an important input into nowcasting. It seems likely that soon Fed officials will learn how to make systematic use of news flow—perhaps from the local business newspapers produced in each large MSA—to optimize their forecasts. Once they, and others, do so, because news flow is likely correlated with some of the Fed’s private information, it is unlikely that the Fed will retain as large an information advantage in nowcasting. The implication is that the public will be able to apply new discipline to the Fed because the information gap about the state of the economy between the Fed and public forecasters will be substantially reduced.

Conclusion

In spite of the fact that Fed disclosure standards have improved markedly over the past two decades (e.g., the release of detailed minutes and the release of FOMC members’ interest rate forecasts), and in spite of the fact that the Fed has made use sometimes of explicit forward guidance about the likely future path of interest rates over

¹⁰Some in the popular press often accuse the Fed of being out of touch with real-time business conditions. See, for example, CNBC (2018).

the near term, U.S. monetary policy is much less systematic and intelligible than it was two decades ago. It used to be possible to describe Fed policy reasonably accurately by referring to a Taylor Rule relating inflation and unemployment (or the GDP gap) to the fed funds rate. Since about 2002, that relationship changed as monetary policy began to react to other influences in unpredictable ways. And since the 2007–2009 crisis, the toolkit of monetary policy has expanded dramatically, the goals of monetary policy have been broadened, and the variables to which monetary policy reacts have changed. The Fed maintains (for now) a long-run 2 percent inflation target and continues to adhere to a vague dual mandate, but it does not even attempt to describe its actions via any systematic mapping from observables to outcomes.

Despite the high marks the Fed earns for disclosure, therefore, the true transparency and accountability of U.S. monetary policy are quite low. Indeed, there is evidence that increased disclosure has reduced transparency by leading some Fed officials—especially Fed governors—to make their intentions less comprehensible.

We have shown, however, that the same incentives to avoid accountability that have produced the decline in systematic policy could have the reverse effect as the result of the new influence of NLP as a source of discipline on policymakers. NLP has allowed researchers to measure many aspects of policy that are embarrassing to the Fed, including Fed governors' attempts to hide their beliefs in response to disclosure, the Fed's apparent manipulation of its Green Book forecasts to avoid embarrassment for prior forecasting inaccuracy, and the increasingly unsystematic nature of policy (as shown by the growing impacts of monetary policy news and its relation to the words and complexity of FOMC statements).

NLP analysis of central bankers, using measures such as Prattle scores, allows us to construct a continuous measure of monetary policy stance, which can be validated by connecting it to other observables. This permits us to observe the true influences on Fed behavior and the effects of Fed policy. Such an analysis confirms that the Fed responds to a wide range of variables and has been providing a put option to the stock market for at least the past two decades. This new ability to measure policy and connect it to its influences and consequences provides another source of discipline on Fed actions because it permits the public to better judge the actions of policymakers.

Finally, NLP has also identified new opportunities for analyzing news flow to gain information about the economy and has also shown that the Fed is not yet harnessing those opportunities optimally, which means it is relying unnecessarily on stale news. Eventually, the Fed will have to recognize this reality and improve its information collection practices. Nevertheless, the important long-run change will be a narrowing of the information gap about the economy between the Fed and the public. This, too, will be a source of new discipline on the Fed.

NLP is still in its infancy, and it will grow in importance as a source of monetary policy discipline as techniques for measuring Fed officials' statements, intent, and actions improve over time. NLP will produce new incentives for Fed leaders to return to a more systematic and transparent approach to monetary policy, if only to preserve their own reputations as consistent, predictable, and effective policymakers.

It may be objected that an alternative way for policymakers to react to the NLP revolution is by reducing the content of their language in speeches, testimony, policy statements, and minutes. This is unlikely for two reasons. First, as Ben Bernanke recently remarked, monetary policy is "98 percent talk and only two percent action."¹¹ The impact of talk is clearly visible in, for example, the impact of Fed speeches on the VIX and exchange rate returns (Calomiris and Mamaysky 2019a). The Fed uses communication as its main way of affecting the economy and so choosing not to communicate is not an attractive option. Second, any attempt to alter language use to manipulate the inferences NLP algorithms would derive from Fed communications would be detectable. It is important to remember that NLP is capable of identifying both a retreat away from content toward vacuous policy statements (indeed, it has already done so in its analysis of post-1993 Fed governors' behavior), as well as manipulative and untrue language.¹² An observable retreat away from content should not be sustainable in a democracy where Fed officials are forced to defend their actions and statements. The Fed cannot exercise policy under its current strategy without including significant and straight-forward content in policymakers' speeches, and therefore, the Fed cannot eliminate content from its communications.

¹¹This is from Bernanke's blog on March 20, 2015, available at <https://www.brookings.edu/blog/ben-bernanke/2015/03/30/inaugurating-a-new-blog/>.

¹²See, for example, Kogan, Moskowitz, and Niessner (2019).

References

- Acosta, M. (2015) "FOMC Responses to Calls for Transparency." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2015–60.
- Barberis, N.; Shleifer, A.; and Vishny, R. (1998) "A Model of Investor Sentiment." *Journal of Financial Economics* 40: 307–43.
- Calomiris, C. W. (2018) "Reforming the Rules That Govern the Fed." *Cato Journal* 38 (1): 109–38.
- Calomiris, C. W.; Harris, J.; Mamaysky, H.; and Tessari, C. (2020) "Fed-Implied Market Conditions." Working Paper, Columbia Business School.
- Calomiris, C. W., and Mamaysky, H. (2019a) "Monetary Policy and Exchange Rate Returns: Time-Varying Risk Regimes." Working Paper, Columbia Business School.
- _____ (2019b) "How News and Its Context Drive Risk and Return around the World." *Journal of Financial Economics* 133 (2): 299–336.
- Cannon, S. (2015) "Sentiment of the FOMC: Unscripted." Federal Reserve Bank of Kansas City *Economic Review* (Fourth Quarter): 5–31.
- Cieslak, A., and Vissing-Jorgensen, A. (2018) "The Economics of the Fed Put." Working Paper, University of California at Berkeley, Haas School of Business. Available at https://faculty.haas.berkeley.edu/vissing/cieslak_vissingjorgensen.pdf.
- CNBC (2018) "Cramer: Fed Didn't Do 'Its Homework' and Is Now Forced to Hike Rates This Month." CNBC (December 10).
- Correa, R.; Garud, K.; Londono, J.; and Mislav, N. (2017) "Sentiment in Central Banks' Financial Stability Reports." International Finance Discussion Papers No. 1203. Washington: Board of Governors of the Federal Reserve System (March).
- Davis, J. S., and Wynne, M. A. (2016) "Central Bank Communications: A Case Study." Working Paper No. 283, Globalization and Monetary Policy Institute.
- Dincer, N.; Eichengreen, B.; and Geraats, P. (2019) "Transparency of Monetary Policy in the Postcrisis World." In D. G. Mayes, P. L. Siklos, and J. E. Sturm (eds.), *The Oxford Handbook of the Economics of Central Banking*, chap. 10. New York: Oxford University Press.
- Ericsson, N. R. (2016) "Eliciting GDP Forecasts from the FOMC Minutes around the Financial Crisis." *International Journal of Forecasting* 32: 571–83.

- Hansen, S.; McMahon, M.; and Prat, A. (2018) "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach." *Quarterly Journal of Economics* 133 (2): 801–70.
- Kogan, S.; Moskowitz, T.; and Niessner, M. (2019) "Fake News: Evidence from Financial Markets." Available at <https://ssrn.com/abstract=3237763> (April 15).
- Kydland, F. E., and Prescott, E. C. (1977) "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85: 473–91.
- Lange, J. (2019) "You Don't Need a PhD Anymore to Read Fed Statements." Available at www.reuters.com/article/us-usa-fed-communications-analysis/you-dont-need-a-phd-anymore-to-read-feds-statements-idUSKCN1QG0JE (February 27).
- Meade, E. E., and Stasavage, D. (2008) "Publicity of Debate and the Incentive to Dissent: Evidence from the U.S. Federal Reserve." *Economic Journal* 118: 695–717.
- Meltzer, A. H. (2003) *A History of the Federal Reserve, Vol. I, 1913–1951*. Chicago: University of Chicago Press.
- (2014) "Current Lessons from the Past: How the Fed Repeats Its History." *Cato Journal* 34: 519–39.
- Schmidman, E., and MacMillan, W. (2016) "The Prattle Machine Learning Algorithm: How and Why." Prattle White Paper.
- Sharpe, S.; Sinha, N. R.; and Hollrah, C. (2018) "What's the Story? A New Perspective on the Value of Fed Forecasts." Working Paper, Federal Reserve Board of Governors (March).
- Stekler, H. O., and Symington, H. (2016) "Evaluating Qualitative Forecasts: The FOMC Minutes, 2006–2010." *International Journal of Forecasting* 32: 571–83.
- Thomson Reuters (2018) "Fed's Bullard: It's Time to Reconsider 2 Percent Inflation Target." Thomson Reuters (May 17). Available at www.newsmax.com/finance/economy/fed-bullard-inflation-framework/2018/05/17/id/860916.
- Thornton, D. L., and Wheelock, D. C. (2014) "Making Sense of Dissents: A History of FOMC Dissents." Federal Reserve Bank of St. Louis *Review* 96 (3): 213–27.
- Thorsrud, L. A. (2017) "The Value of News." Working Paper, Norges Bank.