A machine learning approach to identifying different types of uncertainty

Bennett Saltzman, Julieta Yung*

Economics Department, Bates College, 4 Andrews Rd, Lewiston, ME 04240, United States

HIGHLIGHTS

- We construct a measure of uncertainty from Federal Reserve Beige Books (1970–2018).
- We distinguish uncertainty by type and find that not all uncertainty is the same.
- *Politics & Government Uncertainty* does not relate to changes in the economy.
- *Economic & Business Uncertainty* is associated with weakness in the economy.

ARTICLE INFO

ABSTRACT

We implement natural language processing techniques to extract uncertainty measures from Federal Reserve Beige Books between 1970 and 2018. Business and economic related uncertainty is associated with future weakness in output, higher unemployment, and elevated term premia. On the other hand, political and government uncertainty, while high during recent times, has no statistically significant impact on the economy.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Uncertainty plays a significant a role in shaping investment decisions, household spending patterns, and labor market outcomes, with the potential for large cross-border spillover effects (e.g., Klößner and Sekkel 2014). Unfortunately, very few concepts are simultaneously as important and as difficult to measure, leading to significant efforts to quantify uncertainty through survey-based measures of disagreement among professional forecasters (e.g., Bachmann et al. 2013); option-implied measures of volatility around future asset prices (e.g., VIX); and news-based measures of reported uncertainty (e.g., Baker et al. 2016).

2. Measuring uncertainty

We develop a natural language processing model (combining a trained support vector machine classifier, deterministic algorithms, and manually created graph pattern rules) using *Amenity Analytics*'s text mining engine to extract uncertainty measures from Federal Reserve Beige Books, published between 1970 and 2018.[1]

---

[1] Our method builds from other papers that have taken advantage of machine learning (e.g., Azqueta-Gavaldón 2017) by using natural language processing techniques with deep learning beyond topic modeling, similar to the algorithm in Boudoukh et al. (2016). There is also a growing literature capitalizing on computational linguistics algorithms to analyze FOMC transcripts and other forms of central bank communication (e.g., Hansen et al. 2017).

---

* Corresponding author.
E-mail addresses: jsaltzma@bates.edu (B. Saltzman), jyung@bates.edu (J. Yung).

The outlook among contacts was cautiously optimistic, with significant uncertainty related to domestic politics.

Fig. 1. Example of Identified Uncertainty from January 18, 2017 Beige Book, p. A-2.

Our proposed measure counts every time the concept of "uncertainty" is mentioned in the text. Importantly, our model recognizes negation links that separate instances of uncertainty reduction from negative mentions of uncertainty in the economy. For example, the sentence "there is uncertainty in the economy" would increase the count by one. If it instead read, "there is *not* uncertainty in the economy", it would reduce the count by one. This capability is unique in the literature, as common algorithms treat words as features, and search for set combinations in text without being able to account for context. Moreover, the multi-functional extraction capabilities of the engine, also allow us to extract slotted terms in addition to sentiment, such that we not only identify the word "uncertainty" in a sentence, but we are also able to contextually classify the *type* of uncertainty being extracted. Fig. 1 shows a sentence with slots around the two areas identified: "domestic" and "politics" that relate to "uncertainty".

We define the uncertainty count for each Beige Book $i = 1, 2, \ldots, I$, as the difference between the negative and positive uncertainty mentions divided by the word count for each Beige Book, to control for the tendency of documents to become lengthier over time:

$$v_i = \left[\frac{\begin{array}{c}\text{(number of negative uncertainty extractions in Beige Book } i) \\ -\text{(number of positive uncertainty extractions in Beige Book } i)\end{array}}{\text{(word count in Beige Book } i)} \right]$$
$$\times\ 10,000. \tag{1}$$

The aggregate quarterly measure of uncertainty is the sum of the scaled count of each Beige Book in a given quarter, t,

$$U_t = 100 + \sum_{i \in \text{quarter } t} v_i. \tag{2}$$

3. Quantifying different types of uncertainty

Beyond the main index of uncertainty, we use Eqs. (1) and (2) to distinguish among 13 different areas of uncertainty. We apply principal component analysis and find that the areas can be classified into two broader categories (see Fig. 2).

The areas in groups 1 and 4 both contribute heavily to the second principal component ($> |0.5|$). We combine these groups (e.g., "fiscal", "regulation", etc.) into *Politics & Government (P&G) Uncertainty*. In contrast, groups 2 and 3 are mostly explained by the first component and combined into *Business & Economics (B&E) Uncertainty*. Importantly, given the low factor contribution of the components (0.22 and 0.13), we identify these two categories to be quite different, suggesting that uncertainty differs by type and a simple word count would miss this important distinction.[2]

Fig. 3 depicts the overall index of *Uncertainty* and its two categories over time. Our uncertainty measure is broadly similar to the economic policy uncertainty measure of Baker et al. (2016) during the 1985–2018 period (0.68 correlation). Moreover, *P&G Uncertainty* correlates highly (0.55–0.66) with similar measures, such as tax expiration and categorical uncertainty on taxes, fiscal policy, government spending, health care, and entitlement programs.

4. Assessing the impact of different types of uncertainty on the economy

In order to explore the relationship between uncertainty and the real economy, we specify a quarterly vector auto-regressive (VAR) framework, similar to Baker et al. (2016), from 1970-Q2 to 2017-Q4.[3] We let uncertainty be followed by the unemployment rate and (log) GDP, as well as the federal funds rate and the (log) S&P 500 Index, in order to control for business-cycle and general financial conditions.

The first column in Fig. 4 shows the response to a shock to *Uncertainty*, while the second and third columns, show the responses to a shock to *B&E* and *P&G Uncertainty*, respectively. A one standard deviation increase in *Uncertainty* is about 2.8 points. To provide some context, this magnitude is roughly equal to the increase from the average level of uncertainty in 2005–2006 to 2007–2008. Notoriously, the uncertainty index has muted effects; but when divided into categories, it is clear that *B&E Uncertainty* is associated with economically and statistically significant changes in the economy; whereas *P&G Uncertainty* is not.

A one standard deviation shock to *B&E Uncertainty* relates to an increase in the unemployment rate of 0.3%.[4] This effect is statistically significant at the 90 percent confidence level and quite persistent, lasting for over 3 years. A shock to *B&E Uncertainty* is also associated with a decline in economic activity of about 1.8%.[5]

The impact of *B&E Uncertainty* shocks is in line with the effects of other publicly available measures of uncertainty, such as the financial and macroeconomic uncertainty indices from Jurado et al. (2015); the historical economic policy uncertainty index from Baker et al. (2016); and the financial uncertainty index from Chuliá et al. (2017).

The main results are robust to altering the order of the variables so that uncertainty is first, in the middle, or last; allowing for different lags as specified by several information criteria rules; including different variables to control for general optimism/pessimism in financial markets; incorporating the shadow rate to account for the stance of monetary policy during the zero lower bound period; first-differencing variables; excluding the unemployment rate; considering non-farm payroll employment; and removing the zero lower bound period.

In all cases, the same results hold: *B&E Uncertainty* is associated with stronger, more persistent, and statistically significant changes in the real economy, whereas *P&G Uncertainty* is not.[6]

---

[2] This result is consistent with work by Larsen (2017), who finds that various types of uncertainty from news paper articles in Norway have different effects on the economy.

[3] It is worth noting that a structural model is necessary to identify the causal effects of uncertainty and deal with potential endogeneity as in Richter and Throckmorton (2017).

[4] This magnitude is consistent with the response of the unemployment rate to different uncertainty shocks previously found in the literature: 0.05 p.p. (Bundick et al., 2017); 0.08% (Creal and Wu, 2017); 0.15% (Castelnuovo and Tran, 2017); 0.2 p.p. (Leduc and Liu, 2012); 0.25%-0.4% (Mumtaz, 2018).

[5] In their quarterly VAR specification, Baker et al. (2016) found uncertainty to lower GDP by 1.2%. Others have identified a decline in industrial production with similar magnitudes: 0.25% (Husted et al.); 0.4% (Bundick et al., 2017); 1% (Choi, 2013); 1.2% (Baker et al., 2016); 1% during expansions and 2% during recessions (Caggiano et al., 2017).

[6] A potential explanation is that *P&G Uncertainty* captured in the Beige Books is not associated with changes in the real economy because it reflects uncertainty that is perhaps already priced in the economy or agents do not believe these documents provide special insight into fiscal policy.