Traditional or social media: which capture employment better?

MARIJA HRUŠKA, Ph.D.*
MIRJANA ČIŽMEŠIJA, Ph.D., Full Professor*

Article**
JEL: E69, C19, C49, D81
https://doi.org/10.3326/pse.48.4.2

* The authors would like to thank two reviewers for their helpful comments and advice.

** Received: April 17, 2024
Accepted: August 27, 2024

Marija HRUŠKA
University of Zagreb, Faculty of Economics and Business, Department of Statistics, Trg J. F. Kennedy 6, 10000 Zagreb, Croatia
e-mail: mlogarusic@efzg.hr
ORCiD: 0000-0001-6608-1067

Mirjana ČIŽMEŠIJA
University of Zagreb, Faculty of Economics and Business, Department of Statistics, Trg J. F. Kennedy 6, 10000 Zagreb, Croatia
e-mail: mcizmesija@efzg.hr
ORCiD: 0000-0002-9873-7291

Abstract

Political discourse has the ability to spread either uncertainty or calm among the population. Economic upheavals of considerable magnitude can also spread ambiguity. Both newspaper articles and Twitter posts reflect important events that have the potential to increase or decrease uncertainty from a citizen's perspective. We employ two measures of media uncertainty, one reflecting the uncertainty perceived by journalists and the other characterizing the uncertainty associated with Twitter users. More specifically, we use the Twitter Economic Uncertainty and the Economic Policy Uncertainty Index. To investigate which uncertainty source better captures employment variations, we apply a regression decision tree and linear regression. Our results speak in favour of the more traditional media uncertainty source. Linear regression outperforms the decision tree in both models. Namely, we find a statistically significant negative relationship between both uncertainty measures and employment, while controlling for other macroeconomic aspects.

Keywords: economic policy uncertainty, employment, machine learning, decision tree, Twitter economic uncertainty

1 INTRODUCTION

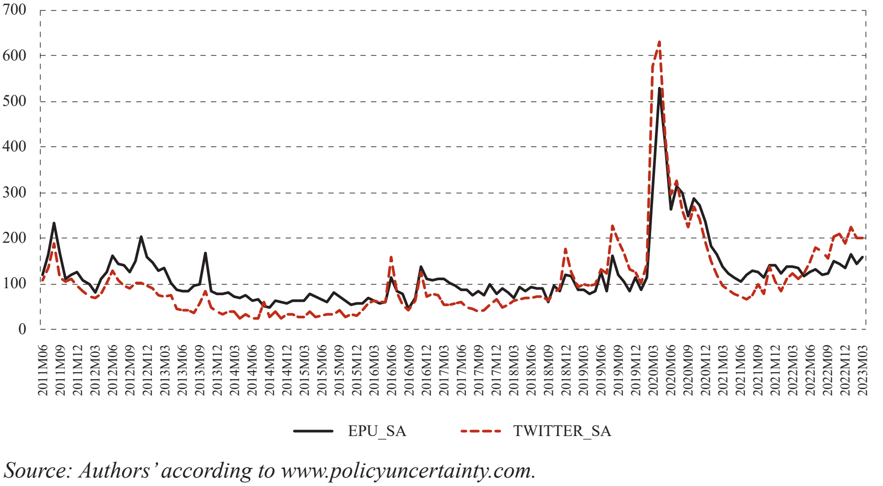
The media is believed to be the main source of information for economic agents seeking knowledge about various economic indicators and trends in economic policy. As a result, economic agents rely on newspaper articles to form their perceptions of the uncertainty inherent in economic policy. One method used to gauge uncertainty involves analysing the frequency with which predefined keywords related to uncertainty are found in media reports. Baker, Bloom, and Davis (2016) have done notable work in this area, seeking to measure uncertainty in economic policy by examining the content of widely accessible media reports. Baker, Bloom, and Davis (2016) use newspaper articles to quantify an index of economic policy uncertainty in an aggregate manner, and specific policy categories. To determine if an article reflects uncertainty in economic policy, they look for at least one keyword from each of three selected word groups: economy, policy, and uncertainty. For example, an article that contains the following paragraph: "When people argue that uncertainty about taxation and regulation is freezing corporate decision-making, they are generally arguing that more certainty would be a good thing for the economy" (Appelbaum, 2013) would be classified as an article that depicts economic policy uncertainty: it contains the following three words: "economy" [from the word group economy], "uncertainty" [from the word group uncertainty], and "regulation" [from the word group policy]. A negative relationship is found between media uncertainty measured via the Economic Policy Uncertainty (EPU) Index and macroeconomic variables such as output, investment, and employment by various authors such as Baker, Bloom and Davis (2016), Nilavongse, Michal and Uddin (2020), Colombo (2013), and Alam and Istiak (2020). There is growing literature that uses newspaper articles and media in general to measure economic uncertainty.

In July 2023, Twitter rebranded to X. However, below the name Twitter will be used as it coincides with the analysed data frame. Twitter users in the US account for about 22% of the total adult population (Baker et al., 2021), or about 76 million users (World Population Review, 2023). The US has the highest number of Twitter users in the world, followed by Japan and India. Therefore, the US Twitter posts could adequately capture the prevailing uncertainty in the US population. Baker et al. (2021) quantify economic uncertainty through Twitter posts. This indicator reflects the perceptions and attitudes of Twitter users. The limitation of the Twitter economic uncertainty (TEU) is the availability of Twitter posts only from June 2011 until mid-April 2023, while the newspaper uncertainty indicator is available from January 1985 and is updated on a regular basis. The limited time frame of the TEU index comes from the removal of academic research access to the Twitter application programming interface. On the one hand, the TEU index is a relatively new measure of uncertainty, and its impact on macroeconomic variables is still insufficiently explored. On the other hand, correlation, and causality between TEU and cryptocurrency markets has received considerable scientific attention (Gok, Bouri and Gemici, 2022), especially during the COVID-19 pandemic (Aharon et al., 2022). A common conclusion is that the TEU index has as negative effect on cryptocurrency returns (Bashir and Kumar, 2022).

Both the EPU and the TEU show similar movements and depict important events such as the US debt ceiling crisis, US-China trade conflicts, and the COVID-19 crisis. Therefore, Twitter users and journalists have similar perceptions of uncertainty in the economy. EPU and TEU are shown in graph 1.

Graph 1

Economic policy uncertainty and Twitter economic uncertainty in the US, June 2011 – March 2023



Source: Authors' according to www.policyuncertainty.com.

Newspapers are available to a wide variety of people, while Twitter posts are read by only a part of the population. Therefore, we assume that there is a more intense transmission channel between the journalist's impression of uncertainty and the transfer of uncertainty perceptions to the readers of the newspapers than the transmission from the authors of Twitter posts to Twitter users. Due to consumer expectations, which are subjective and adaptive, higher media-related uncertainty would lead to a lower willingness to spend and decrease the estimate of individual's future income and mood or attitudes towards spending. In other words, higher media uncertainty will lead to lower consumption, employment, and output. The aim of this paper is to investigate the impact of uncertainty (measured through newspaper media and social media) to employment. Also, we test the effectiveness of parametric vs. nonparametric methodologies in capturing the uncertainty-employment relationship, and at the same time econometric and machine learning methods.

The remainder of the paper is organized as follows. First, a brief literature review is presented. Afterwards, the data and methodological approach are described. Finally, the results of our econometric and machine learning analysis are provided, followed by a concise conclusion.

2 LITERATURE REVIEW

This literature review is divided into three parts. First, we show the theoretical background of the relationship between uncertainty and employment. Second, we focus on the quantification procedure of EPU and the literature depicting the relationship between EPU and a set of macroeconomic variables. Finally, the TEU quantification procedure is presented, followed by a discussion of the few papers that tackle uncertainty measured through Twitter posts (hereafter tweets).

Rising economic uncertainty can slow down employment if employers decide to postpone job creation. The effects of economic uncertainty on employment are empirically investigated by Baker, Bloom and Davis (2016), and Jurado, Ludvigson and Ng (2015). The authors show a negative relationship between uncertainty and employment. The theoretical background for such conclusions lies in the theory of irreversible decisions via the real-options transmission channel. The real-options transmission mechanism is explained by Bloom (2014) as two solutions that can be applied to businesses or consumers depending on economic circumstances. If the economy is in stable conditions, i.e. there is no pronounced uncertainty, businesses will act, which means they invest and create new jobs. On the other hand, rising uncertainty in the economy leads to postponed actions on both sides, businesses and consumers. Businesses delay their activities as they decide to wait for better economic circumstances to employ and invest. Consumers can postpone their decisions regarding expenditures and increase their savings. Such mechanisms are present only in situations where the decisions are irreversible.

Baker, Bloom and Davis (2016) construct the newspaper media indicator of uncertainty, namely the EPU index. It should be noted that the newspapers used in uncertainty measuring include the 10 most popular US newspapers: USA Today, Miami Herald, Chicago Tribune, Washington Post, Los Angeles Times, Boston Globe, San Francisco Chronicle, Dallas Morning News, the New York Times, and the Wall Street Journal. The EPU indicator is constructed through a series of steps. Firstly, media articles must contain at least one keyword from each of the following three groups: economy, policy, and uncertainty. To qualify for inclusion, media articles must contain at least one word from all three word groups: "economic" or "economy"; "uncertain" or "uncertainty"; "Congress", "deficit", "Federal Reserve", "legislation", "regulation", "White House". After the frequency count, such time series are divided by the total number of all articles in each newspaper, and afterwards divided by their standard deviation. Further, the average across all ten newspapers is calculated. Finally, the series are normalized. Baker, Bloom and Davis (2016) have shown that the EPU indicator registers important events such as presidential elections, Gulf Wars, the 9/11 attacks, the Lehman Brothers collapse, and the debt-ceiling dispute. Also, EPU is closely related to other important uncertainty measures, such as the implied stock market volatility, and uncertainty measured via the Federal Reserve System's Beige Books.

Baker, Bloom, and Davis (2016) apply a VAR model to 12 countries and show that EPU innovations anticipate declines in investment, output, and employment. Numerous authors have also demonstrated a negative relationship between EPU and a set of macroeconomic variables.

Empirical research shows that European economies, as well as the economies of other countries around the world, react to changes in the US EPU index. Thus, industrial production (Nilavongse, Michal and Uddin, 2020), aggregate price indices (Colombo, 2013), and interest rates (Alam and Istiak, 2020) usually decline after a shock to the EPU index in the US. Stockhammar and Osterholm (2016, 2017) show significant negative EPU effects on GDP growth, especially on investment growth and export growth in small open economies. Using monthly micro-panel data for urban households in China, Aaberge, Liu, and Zhu (2017) find a negative correlation between aggregate consumption and EPU. With the emergence of Bitcoin and other cryptocurrencies, many researchers began to investigate the impact of the EPU index on Bitcoin returns (Demir et al., 2018; Wu et al., 2021). The authors find predictive characteristics of the EPU index in terms of Bitcoin returns and show that the relationship between those two variables is negative.

To sum up, previous research shows that macroeconomic variables react more strongly to changes in the US EPU index than to changes in domestic uncertainty indices (Colombo, 2013; Alam and Istiak, 2020).

Uncertainty shocks create short strong recessions and recoveries (Bloom, 2009). Namely, major crises and shocks increase uncertainty while production and investments decrease, and unemployment rises. After the shock, the increased dynamics of change cause an excess of production, employment, and productivity. This leads to output, employment, and productivity overshoot in the medium term. The impact of the EPU index on selected macroeconomic variables increases after the 2008 financial crisis (Coronado, Martinez and Venegas-Martinez, 2020; Kido, 2016). A general conclusion is reached that recession and crisis periods increase the impact of the EPU index on macroeconomic variables. Čižmešija, Lolić and Sorić (2017) have presented interesting results using the Toda-Yamamoto causality test between the EPU index and economic activity in the US and several European countries. Their conclusion is that causality exists in both directions only for the US, while only in one direction for France and Germany. The next conclusion is that the main result did not change during and after the 2008 recession. Further, Karnizova and Li (2014) apply a probit recession forecasting model and confirm that the EPU index can predict a recession up to five quarters in advance.

Researchers use different methodological approaches in investigating the effects of economic policy uncertainty shock on macroeconomic variables. Most authors apply the Structural VAR model (Nilavongse, Michal and Uddin, 2020; Colombo, 2013, Alam and Istiak, 2020; Coronado, Martinez and Venegas-Martinez, 2020). Lolić, Sorić and Logarušić (2022) made an additional contribution to the methodological improvement of the analysis of the relationship between EPU and macro-economic variables. These authors apply ensemble learning techniques (ensemble linear regression, and random forest) and gradient boosting techniques (Gradient Boosting Decision Tree and Extreme Gradient Boosting). Their main conclusion is that EPU is more strongly correlated to financial volatility measures than to consumers' assessments of uncertainty.

On the other hand, Baker et al. (2021) quantify economic uncertainty using tweets. Namely, the authors count the occurrence of the terms related to "economy" and "uncertainty" in tweets. Keywords related to uncertainty are: "uncertain", "uncertainly", "uncertainties", "uncertainty". The second set of terms is connected to economics: "economic", "economical", "economically", "economics", "economies", "economist", "economists", "economy". The data spans from June 2011 and is updated every day. Several measures of Twitter uncertainty are derived. First, an English-language version of the TEU indicator which captures only tweets written on English. The second indicator includes only posts from Twitter users located in the US. Nevertheless, many tweets are not specifically related to the area of the US. Baker et al. (2021) apply a random forest model for classifying tweets written in the US. The third variation of the Twitter uncertainty indicator uses weights for each tweet regarding the number of times it is reposted. Lastly, to control for changes in the intensity of Twitter usage over time, the fourth indicator scales the number of tweets each day by the total number of tweets. Baker et al. (2021) compare their four TEU indicators with the EPU index from Baker, Bloom and Davis (2016). The highest correlation is present in the case of TEU connected to Twitter users located in the US. In a monthly specification, the correlation coefficient is 0.90. Like the EPU indicator, the TEU reflects important economic disturbances such as the US debt ceiling crisis, US-China trade conflicts, and the COVID-19 crisis.

PUBLIC SECTOR ECONOMICS
48 (4) 399-419 (2024)

The TEU index was developed in the time of the COVID-19 pandemic and studied primarily in the context of predicting changes in financial markets, but it also has forecasting power in terms of firms' bankruptcy. Fedorova et al. (2022) employ machine learning methods in the sample of French, Italian, Russian, and Spanish firms to show that the inclusion of the TEU index into bankruptcy prediction models significantly increases their accuracy. Bashir and Kumar (2022) use a simple linear regression, quantile regression (QR), exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model, and sentiment analysis with the aim of investigating the impact of the TEU index on the performance of cryptocurrencies. Aharon et al. (2021) find a strong causal link between the TEU index and cryptocurrency returns. In their research, they use a battery of methods: quantile regressions, Granger-causality in distributions using copula functions, and directional predictability tests. The effects of TEU on stock and energy market are also investigated. Lee, Choi and Kim (2023) show that shocks in the TEU index are significantly related to future returns in the Chinese stock market, investment, consumption, unemployment, and output. The standard uncertainty index of the Chinese economy is less efficient than the TEU index. Further, the recent energy crisis has brought the need for better forecasting of energy prices. Uncertainty indices are also used for this purpose. The TEU index and the corresponding methodological modifications may significantly improve prediction accuracy for oil price future volatility (Lang et al., 2022).

The calculation of TEU is being methodologically improved, even though it is a relatively new indicator. Lang et al. (2022) developed the Twitter-based Market Uncertainty (TMU) index using a novel Markov-regime GARCH-MIDAS model, which showed excellent prognostic properties in predicting oil prices during COVID-19 pandemic. Yesiltas et al. (2022) formulated a Twitter-based high-frequency Economic Policy Uncertainty (TEPU) index based on tweets of experts opinions on the topic. Comparing changes in the TEPU index and in financial indicators (exchange rate and stock market index), the authors find that they are correlated. In addition, it is observed that fluctuations in the TEPU index can be a key indicator for predicting the country risk premium in emerging market economies.

The literature discussed above can be summed up as follows. There is a negative relationship between EPU and a set of macroeconomic indicators as investment, output, employment, interest rates, export, and consumption. There is also a negative relationship between TEU, and firm performance indicators and cryptocurrency returns. To the best of our knowledge, the previous literature has not tackled the question of which media uncertainty measure captures the macroeconomy better. To be specific, in this paper we investigate whether EPU or TEU has better explanatory power regarding employment in the US. To do so, we will apply two methodological approaches that are explained in continuation of this paper.

PUBLIC SECTOR ECONOMICS
48 (4) 399-419 (2024)

3 DATA AND METHODOLOGY

In this section the analysed data and methodological approach are briefly explained. Our dataset includes the following variables for the US: employment, industrial production, S&P 500, interest rate, EPU, and TEU. Employment represents the employment-population ratio expressed as percentage in monthly frequency, and seasonally adjusted. The industrial production is in monthly frequency and seasonally adjusted. The Federal Funds Effective Rate (interest rate) is in percentage, aggregated from daily to monthly frequency using simple averaging, and seasonally adjusted using the ARIMA X-12 method. Also, the TEU index is in daily frequency, aggregated to monthly frequency using averaging. EPU index is in monthly frequency, while both media uncertainty indices are seasonally adjusted with ARIMA X-12. Employment, industrial production, and interest rate are from Federal Reserve Bank of St. Louis (FRED), while the S&P 500 is from finance.yahoo.com, and the EPU and TEU are from policyuncertainty.com. The data span is from June 2011 to February 2023. The augmented Dickey-Fuller test results are in the appendix shown in table A1. Due to the unit root test results, variables are included in their levels, except the S&P 500 and interest rate which are in first and second differences, respectively.

The analysis consists of two methodological approaches. The main idea is to use simple but powerful techniques to investigate whether parametric or nonparametric methods are more suitable for explaining employment via media uncertainty. Therefore, we use a linear regression model and decision tree model to determine the relationship between our variables of interest. The regression model is very common in economic analysis and a simple technique. The applied model is shown in equations 1 and 2.

$$emp_t = c + indp_t + kta_t + sp500_t + epu_t + \varepsilon_t \tag{1}$$

$$emp_t = c + indp_t + kta_t + sp500_t + teu_t + \varepsilon_t \tag{2}$$

The notation *emp* represents employment, *indp* industrial production, *kta* is interest rate, *sp500* S&P 500, *epu* and *teu* are the media uncertainty indices EPU and TEU. The model with a better fit indicates which uncertainty type is more successful in explaining employment in the US.

The variable selection is based on the economic theory and is driven by the idea of parsimony. If such a parametric approach meets high dimensional data, some of the variable selection procedures are usually applied (for example stepwise regression). A competitor to the linear regression, including stepwise regression, is a regression decision tree (Breiman, 2017).

Therefore, the second methodological approach applied in this paper is a machine learning technique. The decision tree model is developed by Breiman et al. (1984). It is a nonparametric approach that can be used both in classification and regression tasks. Our variables of interest are continuous numerical values, which implies the application of a decision tree regression model. The main idea behind this machine learning model is to partition the input space by using a variable that provides the best split of the input data (regressors). The regression decision tree algorithm (Breiman, 2017) is shown below.

1) The mean squared error can be formulated as follows:

$$R(d) = N^{-1}\sum_{n}\left(y_n - d\left(x_n\right)\right)^2 \tag{3}$$

where y is the dependent variable, and $d(x)$ is the estimate of dependent variable.

2) We search for the value of $y(t)$ that minimizes $R(d)$. This is the average of y_n for all pairs (x_n, y_n) which minimize $y(t)$. This notation can be shown as follows:

$$\bar{y}(t) = N(t)^{-1}\sum_{x_n \in t} y_n \tag{4}$$

where $N(t)$ is the total number of pairs in t.

3) Therefore, the predicted value in any node t is $\bar{y}(t)$.
4) We replace the notation $R(d)$ with $R(T)$.

$$R(T) = N^{-1}\sum_{t \in \tilde{T}}\sum_{x_n \in t}\left(y_n - \bar{y}(t)\right)^2 \tag{5}$$

$$R(t) = N^{-1}\sum_{x_n \in t}\left(y_n - \bar{y}(t)\right)^2 \tag{6}$$

$$R(t) = \sum_{t \in \tilde{T}} R(t) \tag{7}$$

For every node t, the notation $\sum_{x_n \in t}\left(y_n - \bar{y}(t)\right)^2$ is the within node sum of squares, i.e. the total of squared deviations of y_n from the mean. The sum over all $t \in \tilde{T}$ represents the total within node sum of squares. Multiplying this notation with N^{-1} gives the average within node sum of squares.

5) The best split from a set of splits S for a terminal node t in $\tilde{T}$ is the one which the most decreases $R(T)$. Any split s of t that forms t_L and t_R can be written as:

$$\Delta R(s,t) = R(t) - R(t_L) - R(t_R) \tag{8}$$

6) The best split s^* can be defined as:

$$\Delta R(s^*,t) = \max_{s \in S} \Delta R(s,t) \tag{9}$$

PUBLIC SECTOR ECONOMICS
48 (4) 399-419 (2024)

This procedure includes iterative splitting nodes to maximize the decrease in the mean squared error ($R(T)$). We grow a tree starting from the root node, splitting the data into two branches that maximize the decrease in the $R(T)$. The estimated two models use the same variables as the regression models, i.e. the endogenous variable is employment, while candidates for exogenous variables are industrial production, interest rate, S&P 500, and EPU for the first model, while EPU is replaced with TEU in the second model. The decision tree model is estimated in the programming language R within the package *rpart*.

4 EMPIRICAL RESULTS

The previous literature (see, e.g. Baker, Bloom and Davis, 2016) shows negative connectedness between uncertainty measures and employment. A negative relationship is expected because higher uncertainty leads to more careful decision-making regarding different economic activities. This means that economic agents, i.e. consumers and firms, are more likely to spend less during uncertain economic times, and wait for better economic circumstances to invest, employ, and spend. The recent COVID-19 pandemic is an extreme example of an economic activity slowdown when most economic activities literally stopped. High uncertainty regarding health slowed down consumers spending, firms' investment and employment, export, and other economic activities.

The results of the econometric regression analysis in levels are shown in tables 1 and 3. The first estimated model, which is the newspaper media model that includes EPU as an exogenous variable, is shown in table 1. The first model depicts a statistically significant relationship between industrial production, interest rates, and EPU with employment. The social media model shows that the media variable is also statistically significant in the model shown in table 3. The relationship between TEU and employment is negative, as expected.

Table 1

Newspaper media regression analysis results in levels

Variable	Estimate
Intercept	36.195***
Industrial production	0.239***
Interest rate	-0.837**
S&P500	-0.001
EPU	-0.008***
R^2	0.735

*Note: ***, **, * depict 1%, 5%, and 10% significance levels, respectively.*

Source: Authors' calculation.

PUBLIC SECTOR ECONOMICS 48 (4) 399-419 (2024)

Table 2
Newspaper media regression analysis results with uncertainty in first lag

Variable	Estimate
Intercept	35.114***
Industrial production	0.250***
Interest rate	-0.373
S&P500	-0.001
EPU (lag 1)	-0.007***
R^2	0.731

*Note: ***, **, * depict 1%, 5%, and 10% significance levels, respectively.*

Source: Authors' calculation.

Table 3
Social media regression analysis results in levels

Variable	Estimate
Intercept	29.139***
Industrial production	0.304***
Interest rate	-0.955**
S&P500	-0.001*
TEU	-0.003***
R^2	0.685

*Note: ***, **, * depict 1%, 5%, and 10% significance levels, respectively.*

Source: Authors' calculation.

Table 4
Social media regression analysis results with uncertainty in first lag

Variable	Estimate
Intercept	30.821***
Industrial production	0.288***
Interest rate	-0.464
S&P500	-0.001
TEU (lag 1)	-0.004***
R^2	0.699

*Note: ***, **, * depict 1%, 5%, and 10% significance levels, respectively.*

Source: Authors' calculation.

Both analysed models show a statistically significant relationship between uncertainty measures and employment, as expected and as shown in previous studies (see, e.g. Baker, Bloom and Davis, 2016; Baker et al., 2021). The newspaper media regression model shows a better model fit. The coefficient of determination for the first model is 0.735, while the coefficient of the social media regression model is 0.685. Therefore, we could note that EPU is better in explaining employment variations than TEU. Although the US is the country with the largest proportion of population that uses Twitter, tweets still do not capture the prevalence of

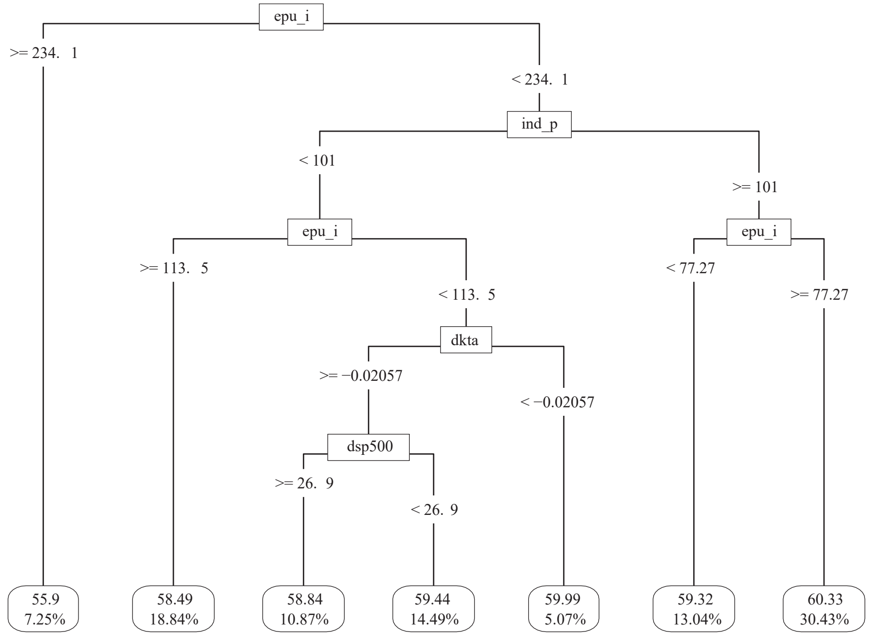
PUBLIC SECTOR
ECONOMICS
48 (4) 399-419 (2024)

uncertainty in the population better than traditional media. Despite the high popularity of Twitter in US, the Twitter uncertainty indicator has not managed to outperform the uncertainty captured through classic newspaper articles. There are two possible reasons for such results. The first is that Twitter posts are not written and read by the total US population, rather one fifth of the US population. On the other hand, newspapers are widely available. In the time of globalisation and easy access to the internet, newspaper articles are more than ever before available to the public. The second reason for the better performance of EPU in explaining employment could be that there is additional space for improvement in the construction of the TEU indicator. Some methodological alternatives to Twitter uncertainty construction have already been suggested in the literature (for example Lang et al., 2022; Yesiltas et al., 2022).

The previous analysis shows the contemporaneous relationship between uncertainty measures and employment. Additionally, we investigate this relationship including one lag in both uncertainty measures, EPU and TEU. The inclusion of one lag in EPU and TEU in our models could show the predictive properties of EPU and TEU. The regression results with one lag in uncertainty measures are shown in tables 2 and 4. We can conclude that in both models, the relationship remains negative and statistically significant.

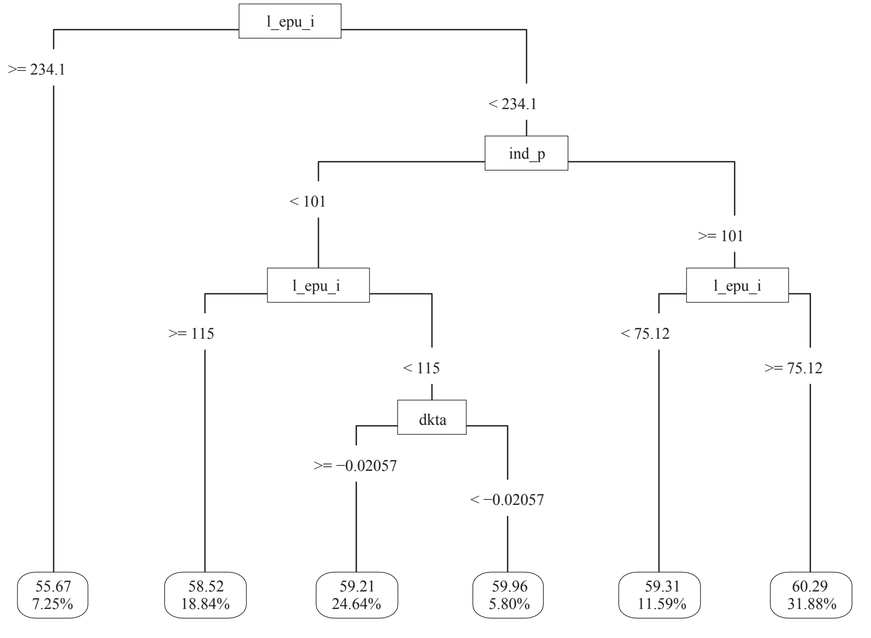
As an alternative approach to model the impact of uncertainty to employment, we estimate a regression decision tree model. Precisely, four regression decision tree models as comparison to the shown linear regression models. The estimated decision tree models are shown in graphs 2 – 5. Each leaf represents the average value of the dependent variable employment to total population ratio for the observations that are included in it.

Graph 2
Newspaper media decision tree results in levels



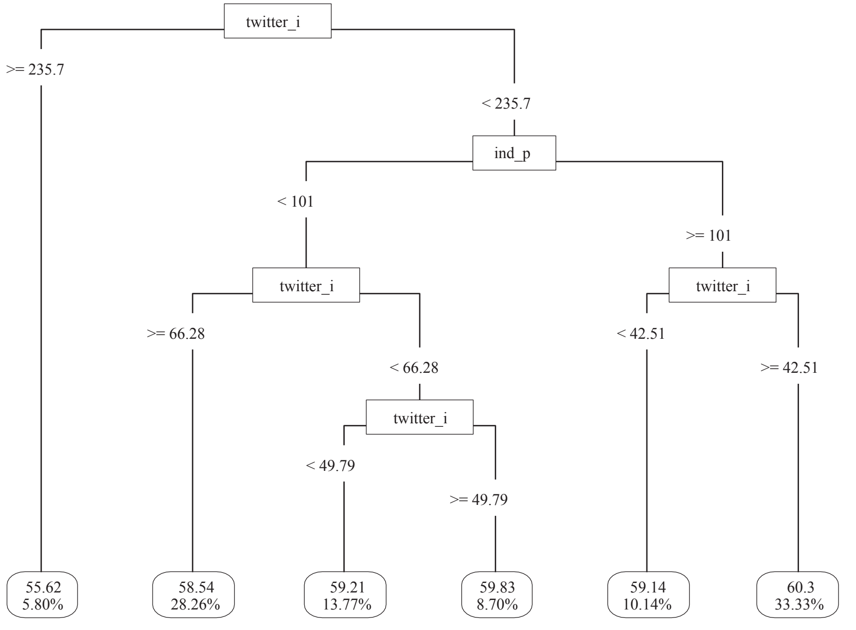
Source: Authors' calculation.

Graph 3
Newspaper media decision tree results with uncertainty in first lag



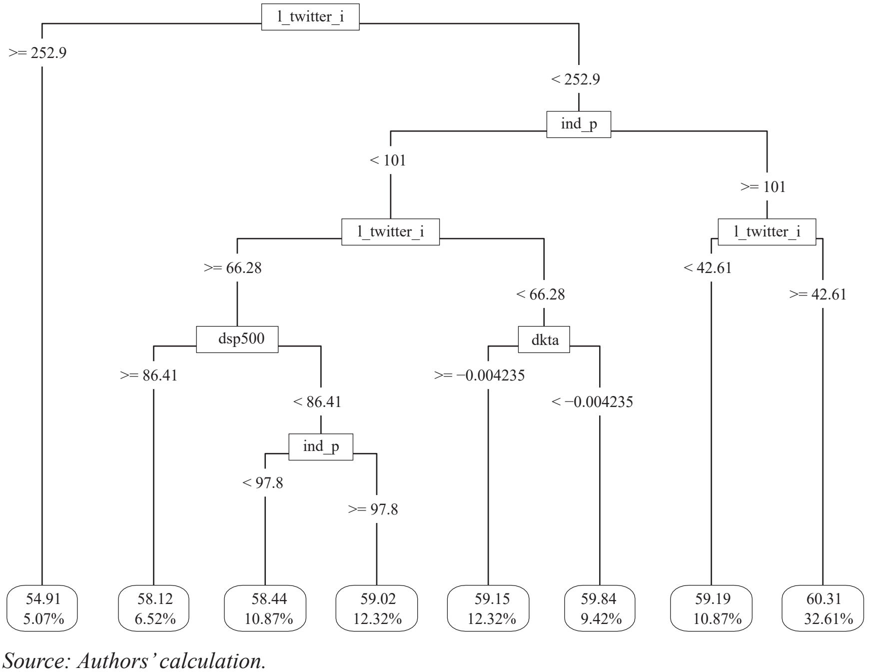
Source: Authors' calculation.

Graph 4
Social media decision tree results in levels



Source: Authors' calculation.

Graph 5
Social media decision tree results with uncertainty in first lag



Source: Authors' calculation.

The newspaper media model (graph 2) shows that the variable EPU (in levels) is chosen at the root node. The left branch shows the scenario when EPU is greater than or equal to 234.1, while the right branch presents an EPU lower than 234.1. At the left branch, variable EPU maximizes the decrease in the mean squared error in this node. For EPU greater than or equal to 234.1 the employment to total population ratio is 55.9%. For lower uncertainty levels than 234.1 employment is between 58.49% and 60.33% depending on the levels of industrial production, interest rates, and S&P 500. As with the linear regression models, we can conclude that higher uncertainty levels lead to lower employment, and vice versa. Further, the combination of lower EPU levels and higher industrial production leads to higher employment. The second model shown in graph 3 uses the EPU in first lags. The conclusion remains the same, i.e. higher uncertainty levels lead to lower employment, and vice versa. The coefficients of determination for our regression decision tree model in levels and including one lag in EPU are 0.572, and 0.540, respectively.

Our second set of models, shown in graphs 4 and 5, considers TEU as an alternative uncertainty variable to EPU. The regression decision tree model is similar to the newspaper decision tree model, and the main conclusion stays the same. That is, higher uncertainty levels (higher TEU) are connected to lower employment levels in the US. The coefficients of determination for our regression decision tree model in levels, and including one lag in TEU are 0.610, and 0.638, respectively.

Comparison of our linear regression models with the regression decision tree models shows that linear regression provides a better fit. Therefore, the linear regression would be preferred for modelling employment variations via uncertainty and macroeconomic variables. The second conclusion is that EPU has better explanatory power than TEU. Therefore, newspaper articles are a better source for measuring uncertainty in the economy than tweets. Nevertheless, this might not always be true. There is additional space for improvement of the Twitter uncertainty indicator.

Media uncertainty indicators have also experienced negative criticism. Some of the negative connotations refer to media that favour negative over positive news. The media might be prone to publishing more negative news with catchy titles to attract more readers than positive news. But does this affect the media uncertainty indicators? The quantification procedure for both media indicators, namely economic policy uncertainty and Twitter economic uncertainty, considers the total number of articles or Tweets in each day/month so that the rising number of published articles or tweets through time has no effect on uncertainty levels. Another critique regarding media uncertainty indicates that media could be politically aligned more to the left or to the right. Hemphill, Culotta and Heston (2016) show that Democrats and Republicans use different Twitter hashtags to discuss overlapping issues. Authors Hemphill, Culotta and Heston (2016) calculate Twitter polarization scores based on the connectedness between hashtags and political parties.

Twitter hashtags are effective measures for the estimation of political polarization. In contrast, Niven (2001) finds no media bias towards political orientation but shows the dominance of negative news over positive ones. Partisanship bias is a strong loyalty to a political party or ideology that can be more left or right aligned. Azzimonti (2021) shows that higher partisan conflict leads to higher uncertainty and consequently can cause economic crises. Further, this can pause reforms and disrupt economic activities. Political polarization is important not only for political scientist, but also for economists as it strongly impacts economic policies. Azzimonti-Renzo (2023) emphasizes that higher levels of partisanship can lead to more pronounced policy uncertainty, which delays consumer spending, employment, investment, and aggregate economic growth. Shultziner and Stukalin (2021) discuss partisan bias in the US media. Authors find that partisan bias can be easily observed through the types of newspaper articles that the media highlights on their cover page and in the sizing and emphasizing of articles.

As already mentioned, uncertainty is a latent variable and there is no such thing as a one fits all measure of uncertainty that is suitable for all macroeconomic and/or financial problems. This paper has shown that EPU is better at explaining employment in the US than TEU. Those two uncertainty measures are chosen as they depict the picture of macroeconomic uncertainty better than other known uncertainty measures. However, media uncertainty indicators can be constructed using different sets of keywords to depict specific macroeconomic topics. Therefore, one suggestion for future research could be to construct specific media uncertainty indices connected to employment. Constraints of this research are the application of two methodological aspects and limited data availability. This leads to another potential direction of future research which would include other machine learning and econometric methods.

Baker et al. (2021) point out that using Twitter as a database source for quantifying economic uncertainty has a few limitations. First, as already mentioned, the database has a limited time span, starting only in June 2011. Second, Twitter users are younger population, so they do not represent the whole US population. Therefore, we already point out that only about 22% of the total adult population uses Twitter, while newspapers are available to the total population. Finally, bots are a real problem in the online world. Bots can generate automated tweets and disseminate false information. Recently, bot detection has improved (see for example Antenore, Camacho Rodriguez and Panizzi, 2023).

5 CONCLUSION

Uncertainty can be measured through different approaches, for example through media articles, macroeconomic volatility or professional forecaster disagreement. This paper focuses on measuring uncertainty through newspaper articles and Twitter posts. We use the two constructed uncertainty measures from Baker, Bloom and Davis (2016), and Baker et al. (2021) to investigate which one has better explanatory power for employment in the US. We apply two methodological approaches, linear regression and a regression decision tree. The estimated models show statistically significant negative relationship between both uncertainty measures and employment while controlling for other macroeconomic variables like the industrial production, interest rates, and S&P 500. Our results speak in favour of linear regression and uncertainty measured through newspaper articles. Although Twitter is popular in the US population, the availability of newspaper articles to almost every citizen in the US could be one of the reasons why modelling employment with EPU provides a better fit. Our recommendation for future research is that it should focus on additional improvements in the Twitter uncertainty indicator so that it can better capture the uncertainty in the total US population.

Disclosure statement

The authors have no conflicts of interest to declare.

REFERENCES

1. Aaberge, R., Liu, K. and Zhu, Y., 2017. Political uncertainty and household savings. *Journal of Comparative Economics*, 45(1), pp. 154-170. https://doi.org/10.1016/j.jce.2015.12.011
2. Aharon, D. Y. [et al.], 2021. Twitter-Based uncertainty and cryptocurrency returns. *Research in International Business and Finance*, 59, 101546. https://doi.org/10.1016/j.ribaf.2021.101546
3. Alam, M. R. and Istiak, K., 2020. Impact of US policy uncertainty on Mexico: Evidence from linear and nonlinear tests. *Quarterly Review of Economics and Finance*, 77, pp. 355-366. https://doi.org/10.1016/j.qref.2019.12.002
4. Antenore, M., Camacho Rodriguez, J. M. and Panizzi, E., 2023. A Comparative Study of Bot Detection Techniques with an Application in Twitter Covid-19 Discourse. *Social Science Computer Review*, 41(5), pp. 1520-1545. https://doi.org/10.1177/08944393211073733
5. Appelbaum, B., 2013. The Benefits of Uncertainty. *The New York Times*, January 7.
6. Azzimonti, M., 2021. Partisan Conflict, News, and Investors' Expectations. *Journal of Money, Credit and Banking*, 53, pp. 971-1003. https://doi.org/10.1111/jmcb.12810
7. Azzimonti-Renzo, M., 2023. Partisan Conflict in the US and Potential Impacts on the Economy. *Richmond Fed Economic Brief*, 23(20), pp. 1-5.
8. Baker, S. R. [et al.], 2021. *Twitter-derived measures of economic uncertainty*.
9. Baker, S. R., Bloom, N., and Davis, S. J., 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), pp. 1593-1636. https://doi.org/10.1093/qje/qjw024
10. Bashir, H. A. and Kumar, D., 2022. Investor attention, Twitter uncertainty and cryptocurrency market amid the COVID-19 pandemic. *Managerial Finance*, 49(4), pp. 620-642. https://doi.org/10.1108/MF-09-2021-0414
11. Bloom, N., 2009. The Impact of Uncertainty Shocks. *Econometrica*, 77(3), pp. 623-685. https://doi.org/10.3982/ECTA6248
12. Bloom, N., 2014. Fluctuations in uncertainty. *Journal of Economic Perspectives*, 28(2), pp. 153-176. https://doi.org/10.1257/jep.28.2.153
13. Breiman, L. [et al.], 1984. *Classification and Regression Trees*. Wadsworth International Group.
14. Breiman, L., 2017. *Classification and regression trees*. Routledge. https://doi.org/10.1201/9781315139470
15. Colombo, V., 2013. Economic policy uncertainty in the US: Does it matter for the Euro area? *Economics Letters*, 121(1), pp. 39-42. https://doi.org/10.1016/j.econlet.2013.06.024
16. Coronado, S., Martinez, J. N. and Venegas-Martinez, F., 2020. Spillover effects of the US economic policy uncertainty in Latin America. *Estudios de Economia*, 47(2), pp. 273-293. https://doi.org/10.4067/S0718-52862020000200273
17. Čižmesija, M., Lolić, I. and Sorić, P., 2017. Economic policy uncertainty index and economic activity: what causes what? *Croatian Operational Research Review*, 8(2), pp. 563-575. https://doi.org/10.17535/crorr.2017.0036

18. Demir, E. [et al.], 2018. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters*, 26, pp. 145-149. https://doi.org/10.1016/j.frl.2018.01.005
19. Fedorova, E. [et al.], 2022. Economic policy uncertainty and bankruptcy filings. *International Review of Financial Analysis*, 82, 102174. https://doi.org/10.1016/j.irfa.2022.102174
20. Gok, R., Bouri, E. and Gemici, E., 2022. Can Twitter-based economic uncertainty predict safe-haven assets under all market conditions and investment horizons?. *Technological Forecasting and Social Change*, 185, 122091. https://doi.org/10.1016/j.techfore.2022.122091
21. Hemphill, L., Culotta, A. and Heston, M., 2016. Polar Scores: Measuring partisanship using social media content. *Journal of Information Technology & Politics*, 13, pp. 365-377. https://doi.org/10.1080/19331681.2016.1214093
22. Jurado, K., Ludvigson, S. C. and Ng, S., 2015. Measuring uncertainty. *American Economic Review*, 105(3), pp. 1177-1216. https://doi.org/10.1257/aer.20131193
23. Karnizova, L. and Li, J. X., 2014. Economic policy uncertainty, financial markets and probability of US recessions. *Economics Letters*, 125(2), pp. 261-265. https://doi.org/10.1016/j.econlet.2014.09.018
24. Kido, Y., 2016. On the link between the US economic policy uncertainty and exchange rates. *Economics Letters*, 144, pp. 49-52. https://doi.org/10.1016/j.econlet.2016.04.022
25. Lang, Q. Q. [et al.], 2022. Oil futures volatility predictability: Evidence based on Twitter-based uncertainty. *Finance Research Letters*, 47, 102536. https://doi.org/10.1016/j.frl.2021.102536
26. Lee, K., Choi, E. and Kim, M., 2023. Twitter-based Chinese economic policy uncertainty. *Finance Research Letters*, 53, 103627. https://doi.org/10.1016/j.frl.2023.103627
27. Lolic, I., Soric, P. and Logarusic, M., 2022. Economic Policy Uncertainty Index Meets Ensemble Learning. *Computational Economics*, 60(2), pp. 401-437. https://doi.org/10.1007/s10614-021-10153-2
28. Nilavongse, R., Michal, R. and Uddin, G. S., 2020. Economic policy uncertainty shocks, economic activity, and exchange rate adjustments. *Economics Letters*, 186, 108765. https://doi.org/10.1016/j.econlet.2019.108765
29. Niven, D., 2001. Bias in the news: Partisanship and negativity in media coverage of presidents George Bush and Bill Clinton. *Harvard International Journal of Press/Politics*, 6, pp. 31-46. https://doi.org/10.1177/108118001129172215
30. Shultziner, D. and Stukalin, Y., 2021. Distorting the News? The Mechanisms of Partisan Media Bias and Its Effects on News Production. *Polit Behav*, 43, pp. 201-222. https://doi.org/10.1007/s11109-019-09551-y
31. Stockhammar, P. and Osterholm, P., 2016. Effects of US policy uncertainty on Swedish GDP growth. *Empirical Economics*, 50(2), pp. 443-462. https://doi.org/10.1007/s00181-015-0934-y

32. Stockhammar, P. and Osterholm, P., 2017. The Impact of US Uncertainty Shocks on Small Open Economies. *Open Economies Review*, 28(2), pp. 347-368. https://doi.org/10.1007/s11079-016-9424-x
33. World Population Review, 2023. *Twitter Users by Country 2023*.
34. Wu, W. S. [et al.], 2021. Does economic policy uncertainty affect cryptocurrency markets? Evidence from Twitter-based uncertainty measures. *Research in International Business and Finance*, 58, 101478. https://doi.org/10.1016/j.ribaf.2021.101478
35. Yesiltas, S. [et al.], 2022. A Twitter-Based Economic Policy Uncertainty Index: Expert Opinion and Financial Market Dynamics in an Emerging Market Economy. *Frontiers in Physics*, 10, 864207. https://doi.org/10.3389/fphy.2022.864207

APPENDIX

Table A1
Unit root test results

	Included in test equation		
Variable	**Intercept**	**Trend and intercept**	**None**
Employment	-3.406**	-3.427*	0.161
Industrial production	-2.861*	-2.919*	0.430
d(S&P500)	-14.840***	-14.793***	-14.365***
d_2(interest rate)	-17.152***	-17.187***	-17.174***
EPU	-2.948**	-3.345*	-1.471
TEU	-2.780*	-3.542**	-1.619*

*Note: ***, **, * depict 1%, 5%, and 10% significance levels, respectively.*

Source: Authors' calculation.