Definitions/notspecified-logo-eps-converted-to.pdf

Article

Weak Supervision in Analysis of News: Application to Economic Policy Uncertainty

Paul Trust [1,], Ahmed Zahran [1] and Rosane Minghim [1,]

[1] School of Computer Science and Information Technology [2] School of Computer Science and Information Technology [3] School of Computer Science and Information Technology

‡ These authors contributed equally to this work.

Abstract: The need for timely data for economic decisions has prompted most economists and policy makers to search for supplementary sources of data. In that context, text data is being explored to enrich traditional economic data sources due to its abundance and ease to collect. Our work focuses on studying the capability of textual data, in particular news pieces, for detecting and measuring economic policy uncertainty. Understanding economic policy uncertainty is of great importance to policy makers, economists and investors since it influences their expectations about the future economic fundamentals with impact on their policy, investment and saving decisions. This research tackles the data bottleneck challenge that has hindered the adoption of machine learning in measuring economic policy uncertainty from text data. We test various approaches of classifying news pieces in regards to presenting economic uncertainty content. We propose a solution involving a weak supervision approach, which expresses domain knowledge and heuristics through labeling functions. These labeling functions are used to generate probabilistic labels that can be used for training an end model without need for human annotated data, after we generated a weak supervision based economic policy uncertainty index that we used to conduct extensive econometric analysis along with the Irish macroeconomic indicators to validate whether our generated index foreshadows weaker macroeconomic performance.

Keywords: weak supervision, neural models, economic uncertainty

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Modern machine learning (ML) approaches especially neural network approaches have achieved state of the art performance close to human performance in many applications. Their success is largely attributed to the availability of large annotated data sets, which are necessary for training these models. However, annotating data is not a trivial task and usually requires expensive human effort. The data bottleneck problem is significantly worse in low resource fields like economics, where there are few to no large annotated public data sets with most of the labeling tasks requiring economic knowledge. As a consequence, ML approaches have not achieved full potential in various fields. In order to tackle this problem, several learning methodologies, such as few shot learning [1], distant supervision [2], weak supervision [3], transfer learning [4], have been proposed to apply ML models in low resource settings requiring less training data sets. (see Section 2.1 for details).

We focus on the recently proposed data programming methodology by Ratner et.al(2016) [3] that aims at creating training data for ML models in less time through labeling functions(LFs); labeling functions are user defined heuristics that can be as simple as just keywords and patterns, or may include knowledge bases, distant supervision or existing models. The labeling functions provided are used to train a generative model that produces probabilistic noisy labels which are used to train end models such as BERT [5]. The approach aims at generalizing beyond the information provided in the labeling functions.

We specifically apply weak supervision in macroeconomics by supporting automatic detection of economic policy uncertainty(EPU) from news articles. Economic Policy Uncertainty(EPU) can be defined as the public's inability to predict the outcomes of their decisions under new policies and future economic fundamentals [6]. Measuring policy uncertainty from news articles was first proposed by Baker et.al(2016) [7], who constructed an EPU index by retrieving news articles that satisfied a keyword occurrence related to economy, policy and uncertainty as a ratio of the number of published news pieces in that period of time. The constructed index spiked during tight presidential elections, Gulf wars, the 9/11 attack, Brexit vote and the failure of Lehman Brothers. The index also showed higher correlation with macroeconomic indicators like gross domestic product(GDP),investment and market volatility. Periods of high policy uncertainty were associated with low levels of investment and employment; this is not surprising since investors have an incentive to wait rather than exposing themselves to a high risk [8].

Even though this simple keyword search yielded a good EPU index, almost half of the news articles were found to be false positives according to human auditors with economic level knowledge[7]. To mitigate the false positives and false negatives, Baker et.al(2016) [7] employed several research assistants to manually audit the retrieved articles; this manual process, used in [7] to validate the retrieved articles, is laborious, time consuming and not sustainable due to the need of specialized human annotators each time a new index is to be constructed. A more sustainable approach to coping up with the time consuming labeling process is to employ ML algorithms to automatically detect policy uncertainty from these news articles. Our work proposes an approach to use weak supervision combined with neural language models to automatically detect news articles describing policy uncertainty from news articles (see Figure 1). The success of this approach may impact the progress of locating sources of economic uncertainty in news more effectively by shortening

the laborious labeling step.
The key contributions of this work is as follows;

- Proposing a weak supervision approach to detecting economic policy uncertainty from news pieces.
- Generating an Irish weak supervision based economic policy uncertainty index and conducting extensive econometric analysis with Irish macroeconomic indicators to understand whether the generated index foreshadows weak macroeconomic fundamentals.

2. Background and Related Work

.

2.1. Machine learning Models

ML involves using available data to learn patterns to make predictions for unseen data. Learning can fall in different categories depending on the nature of data. The major learning paradigms include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning aims to learn a predictive model by taking an input data point $x \in X$ and predicting its corresponding output $y \in Y$ by learning on a training dataset comprised of input-output pairs $\{x_n, y_n\}_{n=1}^{N}$. Unsupervised learning on the other hand, has no access to output variables and aims at extracting the underlying structure, patterns or characteristics directly from the input data $\{x_n\}_{n=1}^{N}$. Semi-supervised learning leverages both supervised and unsupervised learning and is often used when there is only a small portion of labeled data and a large amount of unlabeled data. Reinforcement learning involves the task of learning how agents should take sequences of actions in a dynamic environment in order to maximize the cumulative reward. With this learning paradigm, the goal of the agent is to learn good behavior by incrementally acquiring and updating its skills by interacting with the environment in a trial and error fashion, the agent then receives a reward or penalty for the actions taken[9].

Traditional ML algorithms, such as support vector machines [10], random forests [11], are well studied and are known to perform well for small data sets. The key limitations of such models is that they are heavily reliant on feature engineering to achieve better performance. Additionally, these traditional algorithms are much more prone to the curse of dimensionality [12]. Recently, deep neural networks have become preferred alternatives to traditional models. This preference is partly due to the exponential growth of data and improvement in computing software since deep neural networks scale well with the large data size and are better suited for learning complex non-linear patterns from data

Recurrent Neural Networks(RNN) [13,14] are among the most prominent neural models that are customized for sequence modeling tasks, text analysis, due to their ability to incorporate information from previous time steps. However, one of the key limitations of RNN models is their inability to handle very long sequences because of the problem of vanishing gradients [15] and performing computations in a sequential fashion which leads to inefficient use of parallel hardware[16]. Attention [17] is one of those mechanisms that were proposed to reduce the effects of the vanishing gradients problem. Transformer [16] leveraged attention to get rid of the recurrence used in RNNs to build a neural model that is solely based on attention mechanism. Transformer achieves superior performance to the state-of-the-art.

Bi-directional Encoder from Transformers(BERT) [5] is one of the transformer models that achieved state of art performance across most natural language processing tasks. BERT [5] was trained in a bi-directional context with two objectives: masked language modeling and next sentence prediction on a very large generic dataset. Additionally, the trained model can be fine tuned for downstream tasks. Robustly Optimized BERT Pretraining Approach (RoBERTa) [18] is another neural model based on the transformer model. RoBERTa replicates BERT pre-training architecture with the following modifications: it trains the model for longer period of time with larger batches, uses more data, and removes the next sentence prediction objective. There are other several variants of BERT including but not limited to the following [19–22].

The incredible performance of these models is largely enabled by the presence of high quality annotated data, which may not be readily available and can be expensive to create. Hence, the application of these models is challenging, especially in low resource settings where there are few to no annotated data sets. Hence, there is an increasing interest in ML techniques that can achieve high performance with limited data. Transfer learning [4] is one approach where models are pretrained on generic tasks and fine-tuned on domain tasks. Data augmentation [23] creates more training data points by artificially modifying the existing data set. Active learning [24] trains models with few labeled data selected from a large pool of unlabeled data using acquisition functions that select the most informative data points. Meta learning or few shot learning [25] aims to learn from very few examples by learning from other learning algorithms to be able to discover structure among tasks enabling them to learn fast on new tasks.

Weak Supervision is an evolving ML paradigm for programmatic creation and modeling of training data sets, often known as data programming [3]. Data programming [3] involves users or domain experts providing simple rules or heuristics, known as *labeling functions*, which can be patterns, keywords, pretrained models, existing knowledge bases or similar domain heuristic instead of the laborious manual labeling. Unlabeled data records are processed using expert-defined labeling functions to assign a number of labels to each record. These labels can generate potentially conflicting and correlated set of labels. Data programming addresses the noisy and conflicting nature of labeling functions outcome by modeling them as a *generative process*, which denoises the resulting training set by learning the accuracy and correlation structure of the labeling functions. The generative model produces probabilistic labels which are then used to train an end classifier that generalizes beyond labeling functions. Data programming has been employed in various real world applications, such as medical applications [26] [27] [26], social media analysis [28] [29], and autonomous driving [30]. Our work explores the application of weak supervision to measuring economic policy uncertainty, aiming to replace current query based approaches.

2.2. Economic Policy Uncertainty

The impact of policy uncertainty on macroeconomic variables like employment, Gross Domestic Product(GDP) and investment has for long been of great interest to economists and policy makers. Frank Knight [31] defines *uncertainty* as the inability of people to predict the likelihood of happening of future events. According to his definition of uncertainty, people can not construct a probability distribution for their beliefs on what will happen in future. He further defined *risk* as the people's known probability distribution over the occurrence of future events. Economists have however used these terms (risk and uncertainty) interchangeably.The subjective belief expressed by people about the future economic conditions influences their current actions which subsequently affects their consumption, investment and savings. The assumed casual relationship between uncertainty and other macroeconomic indicators has resulted into several attempts to quantify the uncertainty concept. There is however no direct way of measuring uncertainty and all the current solutions including ours are rather proxy measures. The difficulty in measuring uncertainty arises from the fact it an obstruct term that is not directly observed, but is about beliefs in the minds of consumers, firms and policy makers.

We highlight a few prominent proxies used to understand uncertainty which include; text based approaches, forecaster disagreements, Chicago Board Options Exchange Volatility Index (VIX), business surveys of subjective uncertainty, stock market and GDP volatility. VIX measures the market expectations of the relative strength of near-term price changes of the S&P 500 index over the next 30-days backed out from option prices, it is often used for understanding the market sentiment and the degree of fear among market participants [32]. Another often rather controversial measure of uncertainty is disagreements among forecasters [33], which measures the deviation of projections of macroeconomic indicators among several professional forecasters, it is plausible to think that when many experts disagree about what the future GDP is, then we could say that the future GDP is highly uncertain. Businesses can also give their subjective assessment of their own uncertainty providing an elicit point probability distribution of their expected sales growth for one year ahead.

The next set of uncertainty are text based measures which counts the mentions of terms related uncertainty in text based resources like newspapers, twitter, Biege books, annual reports from companies and public releases from financial institutions like central banks or International Monetary Fund(IMF).The most significant work in the text-based approaches to measuring EPU is that of Baker et.al(2016) [7] on which our work builds. Their work measures economic policy uncertainty from news articles by counting the number of news articles containing keywords related to EPU, as a proportion of the total number of articles in the same newspaper and month. The monthly newspaper-level series is standardized to a unit standard deviation and then averaged across the number of newspapers by month of publication. Our work differs from this work in that we use ML approach to detect EPU related text from newspapers.

Azqueta et.al(2017) [34] handled economic policy uncertainty detection by using latent Dirichlet allocation(LDA) [35], a topic modeling technique that outputs topics from documents, and then creates a time series of the counts of the topics over the years. The topics generated by the algorithms are often not intuitive and may not capture the true classification of economic policy uncertainty. Our work differs from this work in that it employs proper classification rather than a topic modeling or counting. Keith et.al(2020) [36] did propose supervised ML methods for EPU classification. They further studied annotator uncertainty and casual assumptions for measuring EPU from newspapers. In that work, the need for supervision limits the reach of the approach, we propose to add a weak supervision strategy to classification as a candidate solution.

3. Proposed Weak Supervision Framework

3.1. Overview

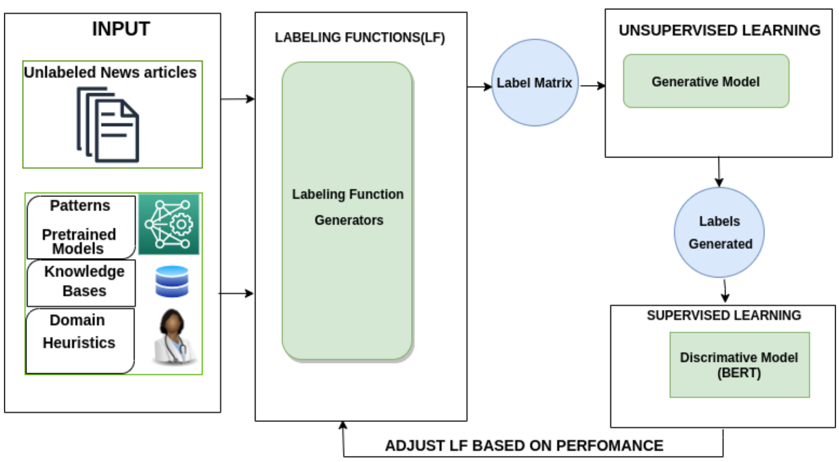

Figure 1. The Proposed Weak Supervision Framework

Figure 1 illustrates our proposed framework that integrates three key stages. The first stage leverages expert-defined labeling functions to automatically generate a label matrix, in which each article is assigned a number of labels. The second stage includes an unsupervised generative model that assigns every article an auto-generated noisy label by only observing the conflicts and correlations in the label matrix. In the third stage, a discriminative model is trained, in a supervised fashion using the generated labels to provide the final label. The following subsections present these stages in more details.

3.2. Automated labeling Model

3.2.1. Labeling Functions

The input of labeling functions is a data set, denoted as $X^{n\times 1}$, consisting of independent and identically distributed(i.i.d) news articles. The dataset is processed by m labeling functions (LFs), denoted by $\lambda = \{\lambda_1, ..., \lambda_m\}$, that are derived from domain heuristics provided by the domain experts or existing prior knowledge about the task. Each labeling function λ_j noisily labels the news article with $\lambda_j(x) \in \{-1, 0, 1\}$, In our case 1 indicates that the news article describes policy uncertainty, 0 indicates that the news article does not and -1 means that the labeling function abstained from labeling the news article.

Data programming applies these m labeling functions on n unlabeled news articles to produce a label matrix of LFs outputs $\Lambda \in \{-1, 0, 1\}^{n\times m}$, the label matrix is then processed by the generative model to produce a vector of noisy labels $\bar{Y} = \{\bar{y}_1, ..\bar{y}_k\}$, with $k \leq n$ for only those articles with an associated noisy label, these are then be used to train a discriminative model.

Table 1 highlights our proposed labeling functions used for economic policy uncertainty detection from news articles. The design of these labeling functions are inspired by the coding guide that domain experts employed to label news articles [7].

- **Keywords**: The first labeling function attempts to register keywords believed by economists to describe policy uncertainty when present in a news article. These keywords were generated by using those used in [7] and also Irish specific keywords as defined in [37].
- **Patterns**: We also searched in news articles for the occurrence of key words related to EPU and one of the words known to be associated with uncertainty. For example articles that describe uncertain events like Brexit or financial crisis are more likely to be describing policy uncertainty.
- **Sentiment Polarity**: We also hypothesized that articles describing policy uncertainty are more likely to have a negative sentiment polarity. This is our hypothesis and is not necessarily supported by the literature in economics.
- **Semantic Similarity**: We computed the semantic similarity between the keywords and news articles. This was achieved by first representing both the keywords and news articles with continuous representations using sentence embeddings provided by SBERT [38] and then calculating the cosine similarity between the embeddings of the keywords and news articles. From that perspective, a news article was considered to be describing policy uncertainty if the cosine similarity was greater than a certain threshold, which is considered a hyper parameter. We used a threshold of 0.1 in our experiments.

Table 1: Some domain heuristics written as labeling functions to express domain knowledge to be fed into a generative model for creation of probabilistic labels that will be used for training an end model

uncertainty ={'uncertain','uncertainty'}
economy={'economic','economy'}
policy={'regulation','legislation','regulation','deficit','federal','spending','budget','tax','congress',regulatory'}
negative terms={'fall','pessimistic','bothered','recession','unemployment'}
known uncertain events ={ '9/11','brexit','covid','Greece crisis','gulf wars','elections',Lehman brothers','great depression'}

Labeling Functions	Description
keywords1	if article contains **uncertainty** AND **economy** AND **policy**, return 1 else return 0
pattern1	if an article contains **known uncertain events** AND **keywords1**, return 1, else return 0
pattern2	if an article contains **known uncertain** AND **negative term** AND **keyword1**, return 1, else return 0
sentiment	if the **sentiment polarity** of news article is less than threshold, return 1 else return 0
Semantic Similarity	if the **cosine similarity** of news article and the keywords is greater than threshold, return 1 else return 0

The labeling function can be adjusted based on the performance from a validation set.

3.2.2. Generative Model

The objective of this stage is to automatically assign each news article a label in an unsupervised fashion. We achieve this objective by modeling the joint probability of the outputs of the labeling functions Λ and the true class label of the news article Y, In our case we don't have access to the true labels Y, and these are considered as latent variables, by this we mean that we don't observe these variables prior to modeling, this is possible because these variables will be marginalized out when computing the parameters of the model w.

This generative model, denoted by $P_w(\Lambda, Y)$, encodes the conflicts, correlations, and propensity(where LFs did not abstain) in the label matrix as factor graphs. Assuming that LFs are independent given the true latent class label Y, we can obtain the label model as follows:

$$P_w(\Lambda, Y) = Z_w^{-1} exp(\sum_{i=1}^{m} w^T \phi_i(\lambda_i, y_i)) \tag{1}$$

where Z_w is the normalizing constant and $\phi_i(\Lambda_i, y_i)$ is a concatenation of factors for all labeling functions for a sample news article $x \in X$. We learn parameters of the model w using maximum likelihood by only observing the agreements and disagreements in the label matrix Λ since we don't have access to the ground truth:

$$w = \underset{w}{\operatorname{argmax}}\, log \sum_{Y} P(\Lambda, Y) \tag{2}$$

3.2.3. Discriminative Model

This stage involves training a standard supervised learning algorithm using the noisy labels $\bar{Y} = P_w(Y|\Lambda)$ from the generative model with a goal of generalizing beyond information obtained in the labeling functions.

Given that now we have a labeled data set of $(X, \bar{Y})$, we used this noisy data set to train a discriminative transformer model(BERT[5]) , This was achieved by fine-tuning a pre-trained BERT model by adding a classifier on top of the pretrained model to enable learning on the noisy data set, the model is trained by minimizing expected loss using a noisy aware objective function.

$$\theta = \arg\min_{\theta} \frac{1}{k} \sum_{i=1}^{k} E_{\bar{y}\sim\bar{Y}}[L(x_i, \bar{y}_i)] \tag{3}$$

where E denotes expectation, L denotes the loss function and k denotes the number of training examples.

The objective function used is the same as a standard supervised learning loss except that for this case we are minimizing the expected value with respect to the noisy probabilistic labels $\bar{Y}$ generated by the label model.
Theoretical analysis guarantees that the generalization error of the discriminative model decreases at the same asymptotic rate as with traditional hand-labeled data [3].

3.2.4. Construction of the EPU index

The predictions from our model are used to construct the EPU index for Ireland from 1992 to August 2021. This was achieved by first aggregating the number of articles published in a given month that were classified as describing policy uncertainty by our model and then generated a monthly index X_t for counts of articles describing policy uncertainty, the standard deviation σ_t^2 and the mean M are of the monthly series X_t. The EPU index is generated using the same procedures as in Baker et.al(2016) [7] by multiplying $(100/M)$ with monthly series X_t

4. Results and Discussion

This section describes the elements of implementation of the weak Supervision framework for Economic Policy Uncertainty classification, as well as its evaluation.

4.1. Data Sets

Irish Newspapers: We employed a data set with the news contents of two Irish newspapers: Irish Times and Irish Independent in the range from January 1992 to August 2021. These newspapers were preferred because they had the highest coverage in the country and had been in publication for a long time compared to other newspapers. The articles that were retrieved are those that passed the key word query; {('uncertain' OR uncertainty') AND ('economy' OR 'economic') AND ('regulation' OR 'legislation' OR 'dail' OR 'deficit' OR 'Taoisearch')}. In total, 10700 articles were retrieved according to the previous keyword queries, the contents of the articles were extracted using Lexis-Nexis tools [39], an R package that extracts contents of newspapers into an MS Excel file to be used for analysis. We decided to manually label 10% of the retrieved articles for our experiments and these were randomly selected.

The annotation process followed the coding guide provided by [7], where the newspaper article is classified as to be describing economic policy uncertainty if: the article talks about uncertainty over who makes or will make policy decisions that have economic consequences; the article talks about current and past uncertainty over what economic policy actions will be undertaken; and the news article talks about uncertainty regarding the economic effects of policy actions.The labeled data set(1070 news pieces) was split into training, validation and testing in the ratio of 8:1:1 respectively.

USA Newspapers: We also used a sample of 10200 from the human annotated news articles provided by Baker et.al(2016) [7]. This data set was created by searching 10 leading USA newspapers; USA Today, Miami Herald, Chicago Tribune, Washington Post, Los Angeles Times, Boston Globe, San Francisco Chronicle, Dallas Morning News, New York Times, Los Angeles Times, Boston Globe from 1900 that contained the following triple: uncertainty' or uncertain; 'economic' or economy and one of the policy terms: 'congress','federal reserve',regulatory','legislation','regulation'.The data set was split into training, validation and testing in the ratio of 8:1:1 respectively. Full details of the data set can be found in [7].

Volatility of Stock price for Ireland: We used the volatility index for the stock price of Ireland, the index measures the 360-day ahead standard deviation of the return on the national stock market index. This was obtained from Organisation for Economic Co-operation and Development(OECD) [40] on 1st November, 2021 at 12:41 pm.

Irish Economic Policy Uncertainty Index: We used the Irish economic policy uncertainty index which was generated by a keyword search from the Irish times provided by Baker et.al(2016) [37].

Consumer Price Index(CPI): We used the Irish monthly consumer price index(CPI) from January 1992 to August 2021, CPI is a measure of the change in the prices of a basket of goods and services that are purchased by specific groups of households and is commonly used by economists as a proxy for inflation. The data was obtained from OECD [40] on 1st November, 2021 at 12:41 pm.

Unemployment rate: Monthly Irish unemployment rate from January 1992 to August 2021 was used, the indicator measures the proportion of unemployed people of working age who dont have work, who are available for work and have also taken a step to obtain work. This was obtained from OECD [40] on 1st November, 2021 at 12:41 pm.

Unemployment rate: Monthly Irish unemployment rate from January 1992 to August 2021 was used, the indicator measures the proportion of unemployed people of working age who dont have work, who are available for work and have also taken a step to obtain work. This was obtained from OECD [40] on 1st November, 2021 at 12:41 pm.

Industrial Production: Irish industrial production from January 1992 to August 2021 was used, this indicator measures the output of industrial establishments in a given period of time.This was obtained from OECD [40] on 1st November, 2021 at 12:41 pm.

Short Term Interest Rates: The Irish short term interest rates from January 1992 to August 2021 obtained from OECD [40] on 1st November, 2021 at 12:41 pm was used, this measures the rate at which short term government paper is issued or traded in the market.

Business Confidence Index: The Irish monthly business index(BCI) from January 1992 to August 2021 was used, the index uses opinion surveys on developments in productions, orders and stocks to provide information on future developments. We obtained the indicator from OECD [40] on 1st November, 2021 at 12:41 pm.

4.2. Evaluation Setup and Metrics

We implemented transformers; BERT [5], RoBERTa [18] using simple transformers Library [41], weak supervision approaches were implemented using snorkel [42]. We ran experiments with Adam optimizer [43], an initial learning rate of $2e-5$, a batch size of 8 for 10 iterations and sequence length of 200 tokens. The training and validation loss for each training epoch were monitored and the models with the best accuracy on the validation set were saved before testing on the test set. LSTM was implemented using Keras library [44] and we used scikit-learn library [45] for implementing Support Vector Machine (SVM) classifier.

We compare our proposed solution with the baselines and other models using precision, accuracy and f1-score metrics as follows;

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{5}$$

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{6}$$

where TN are the true positives,FP are the false positives,FN are the false negatives and TN are the true negatives

4.3. Experiments

To evaluate the performance of weak supervision on automatic economic policy uncertainty detection from news articles, we performed several experiments and compared the results of weak supervision(model trained with noisy labels produced by labeling functions) with keyword search (currently being employed by economists), and also with state of art of methods that were trained on human annotated labels requiring many hours of expert labeling. Table 2 summarizes the performance of state of the art methods and the proposed weak supervision for Irish news articles.

Table 2: Precision, Accuracy and F1 scores scores(all in percentages) for the Irish news articles test set

Test Results on Irish dataset			
Model	**Precision**	**Accuracy**	**f1-score**
BERT	60.9	64.9	57.8
RoBERTa	64.28	66.34	60.67
LSTM	56.36	57.69	50.76
SVM	48.9	53.62	50.07
Weak Supervision	60.1	62.4	55.6
Keyword Search	39.4	-	-

The results on the Irish news articles data set show that weak supervision provides significant improvement in precision(+20%) from just using keyword occurrences (the current solution [7]) to construct an EPU index. The comparison with other state of art ML methods trained on human annotated data show that weak supervision (61.9%) still outperforms SVM and LSTM , while it is outperformed by RoBERTa(66.34%) the best performing model on this data set. It should be noted that with weak supervision we only used noisy labels generated by labeling functions and no human labels were needed while state of art methods were heavily reliant on human annotated labels. To explain the difference in human effort involved in this case, for example, it took us just a week to write labeling functions while it took a team of human annotators 6 months to label 12000 USA news articles [7]. The difference in performance(3%) can be traded off in most applications of economics as it is, though labeling functions can evolve and be improved.

Table 3: Precision, Accuracy and F1 scores scores(all in percentages) for the USA news articles test set

Test Results on USA News articles			
Model	**Precision**	**Accuracy**	**f1-score**
BERT	61.11	68.66	60.59
RoBERTa	60.90	70.71	63.98
LSTM	60.97	67.70	51.52
SVM	51.52	67.6	59.14
Weak Supervision	58.7	69.4	63.1
keyword Search	40.12	-	-

Table 3 shows the results on the test set consisting of news articles from USA news papers. The USA news articles data set is 10 times larger compared to the Irish dataset. Weak supervision presents a significant improvement from just using keyword search(+19%) that is currently being used in constructing EPU indices. State of the art models still outperform weak supervision. We reinforce the observation that this better performance of competitive models (about 5% in precision and accuracy, and comparable in f1-score) comes at the cost of a painstaking labeling process.

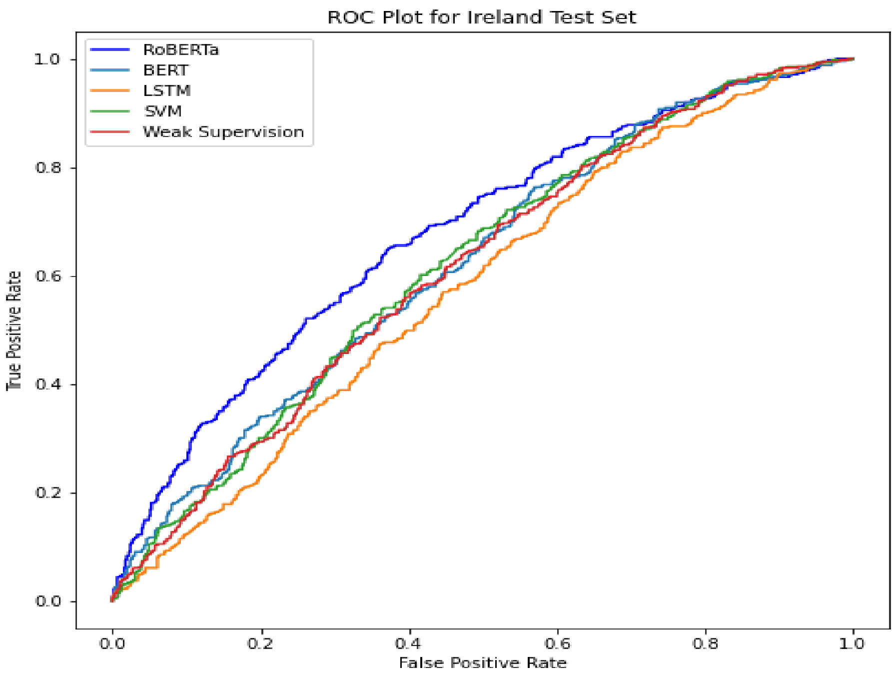

Figure 2. ROC Curve on the Irish Test Dataset

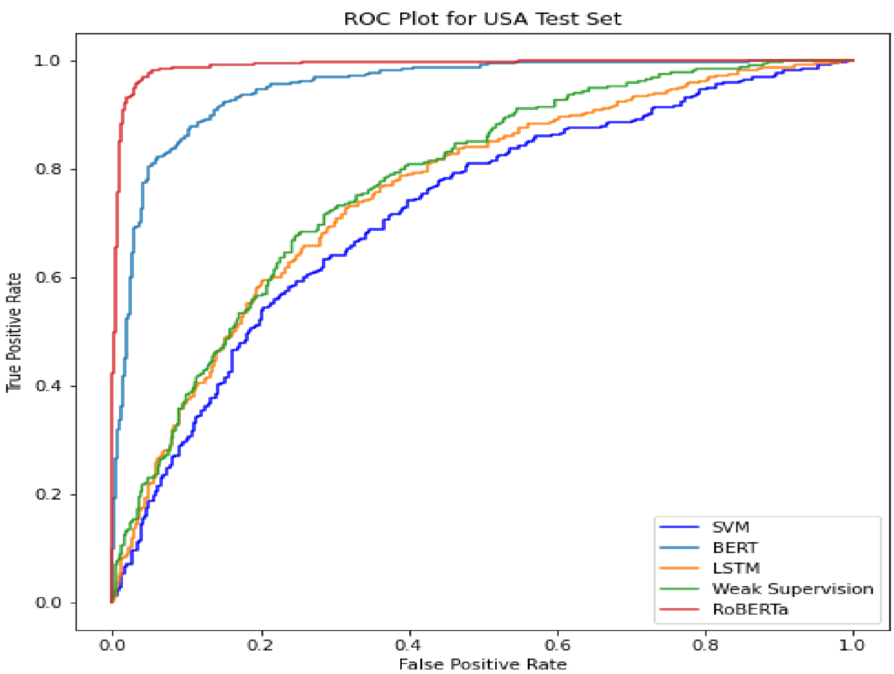

Figure 3. ROC Curve on the USA Test Data Set

Figure2 and figure3 show the Receiver Operating Characteristic curves(ROC) of models on both the Irish and USA test data sets,the ROC curves demonstrate that all models are more confident in discriminating between the two classes on USA data sets than Ireland data set with the area under the curve for models being higher on USA test set compared to Irish test, in particular transformer based models become more confident in discrimination on USA test sets compared to other models.

The general observations from both the ROC curves and other evaluation metrics results show that results of most models are better on the USA data sets compared to the Irish news data sets. We hypothesize that this is explained by the data size since the USA data set used for training is much larger than that used for training models on the Irish data set.

Another observation showed that state of the art models are outperforming weak supervision for large data sets compared to small data sets, meaning that when a significant labeling budget is available, economists are better off employing neural models trained with human labels against the particular labeling functions we have employed. In most applications, however, that trade off is not attractive and designing labeling functions that accommodate new 'views' of the data set is a straightforward activity.

4.4. Economic Policy Uncertainty Index

4.4.1. Weak Supervision Generated EPU index for Ireland

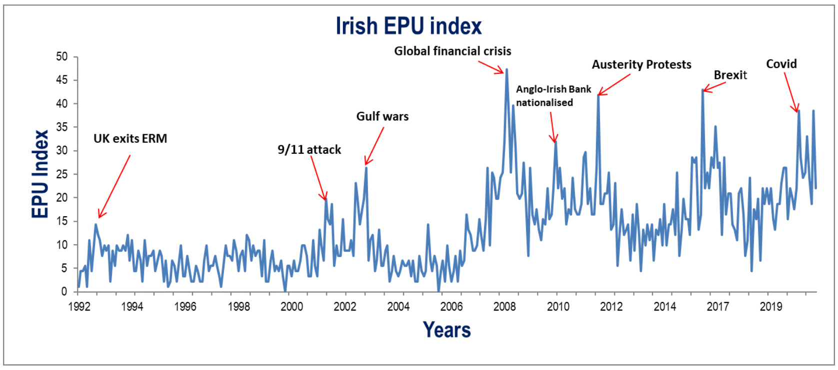

Figure 4. Weak supervision based EPU Index for Ireland

We used the predictions from the weak supervision model to construct a monthly EPU index from January 1992 to July 2021 for Ireland.Figure4 shows that our index captures both local and global events that are known to have caused immerse uncertainty not only to Ireland but also to the entire world. The index spiked highest in 2008 which we suspect is the uncertainty due to the global financial crisis following the failure of Lehman and brothers [46,47], followed by a spike in 2016 which may be attributed to the uncertainty due to brexit effects and local uncertain events like austerity protests.The index demonstrates that the Irish economy is more prone to policy uncertainty from abroad, this is partly explained by the open nature of the Irish economy [48].

4.4.2. Comparison of Irish EPU index with other indices

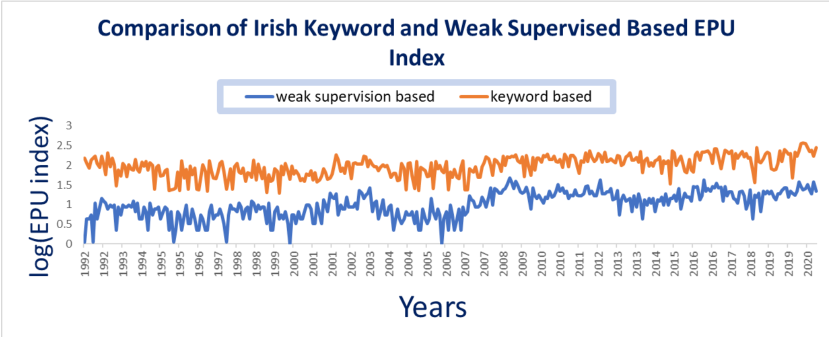

Figure 5. Irish Keyword based and weak supervision based EPU index

Figure 5 shows a comparison of the Irish EPU index generated by keyword search with minimal human audit and the index generated by weak supervision, the EPU index generated by weak supervision has a significant positive relationship$(r = 0.61, p - value < 0.05)$ at 5% level of significance where r is the Pearson correlation coefficient, figure5 also demonstrates that the keyword generated index is much higher in magnitude compared to the weak supervision based EPU index, this is due to the fact keyword based indices are more prone to including many false positives [7].

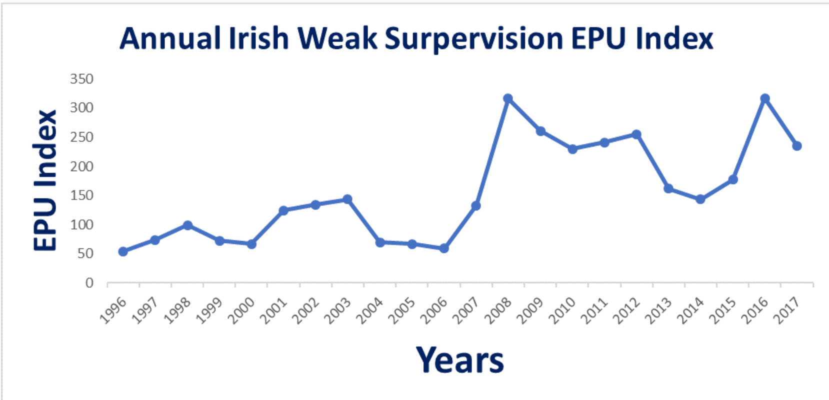

Figure 6. Irish weak supervision EPU Index

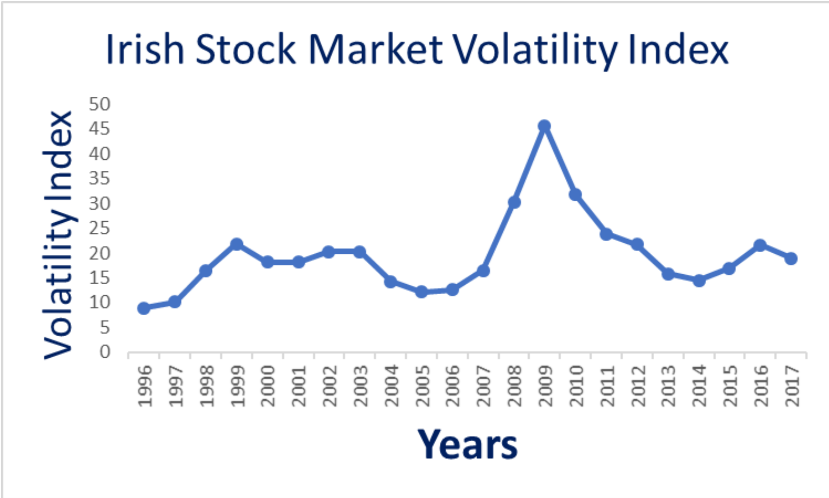

Figure 7. Irish Annual stock market volatility index

We also compared the generated index with the volatility for stock price index for Ireland, our annual EPU index shows a positive significant relationship at 5% level of significance $(r = 0.68, p - value < 0.05)$, the index generated by just keyword currently employed by economists has no significant relationship with the volatility index for Ireland at 5% level of significance.

4.4.3. Economic Signal from Weak Supervision EPU Index

We conducted an econometric analysis using vector auto-regressive models(VAR) [49] to exploit time series variations within the Irish macroeconomic indicators.
More formally, consider $n \times 1$ macroeconomic time dependent variables $Y_t = (y_{1t}, ..y_{nt})^T$, we define a $p - lag$ vector auto-regressive$(VAR(p))$ as:

$$Y_t = c + \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} + \varepsilon_t, t = 1, ..T \quad (7)$$

where Φ_i are $n \times n$ coefficient matrices and ε_t is an $(n \times 1)$ unobserved zero mean white noise vector process with time invariant co-variance matrix Σ [50].

Our goal is to use the estimated model to understand whether our generated EPU index foreshadows weaker macroeconomic performance conditioned on standard macro and policy variables, this was achieved by generating impulse response functions from the estimated VAR model, Impulse response functions are well established tools in econometric analysis that are used to investigate how changes in a policy variable at time t causes changes in another variable after time period t with consideration of the interaction among the variables.

We fit a VAR model to monthly Ireland data from January 1992 to August 2021 using cholesky decomposition [51] in order to recover orthogonal shocks using the following macroeconomic variables; Weak supervision EPU index, Unemployment rate, logarithm of industrial production, logarithm of consumer confidence index, short term interest rates and logarithm of of consumer price index(CPI). To ensure that the series are stationary(a pre-condition for VAR analysis), we tested for unit root in macro economic variables using augumented dicker fuller test(ADF)test [52] with results shown in Table4.

Table 4: Shows the results of the Stationarity tests carried out using augmented Dickey Fuller Test(ADF) for VAR model fitting

Stationarity tests using Augmented Dickey Fuller Test(ADF)		
Macroeconomic Indicator	**Test-statistic**	**P-value**
Economic Policy Uncertainty	0.7	0.01933
Consumer Price Index	3.5957	0.0043
Industrial Production	2.7372	0.00016
Interest Rates	-0.9647	0.0007001
Unemployment Rate	-1.4947	0.000

Observations from Table4 reveal that our macroeconomic variables of interest are stationary since their $p - values$ are less than the critical value 0.05, we therefore rejected the null hypothesis of presence of a unit root and concluded that our variables are stationary at 5% level of significance. Akaike's Information Criterion(AIC) [53] was then used to find an optimal time lag to be used to fit a VAR model, this was found to be 4 months based on this criterion.

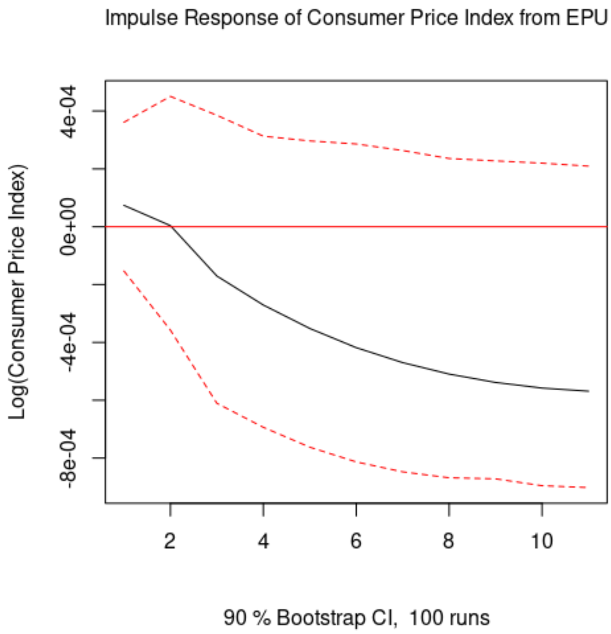

Figure 8. Impulse Response Function of the Response of Consumer Price Index to policy uncertainty shocks for 10 months ahead

Figure 8 shows the impulse response function of consumer price index to economic policy uncertainty, consumer price index responds slowly to policy uncertainty shocks but the effects of policy uncertainty continue to reduce consumer price index for the next 10 months after the shock. This negative relationship is further supported by our causality analysis with granger causality test [54] which is statistically significant at 5% level of significance $(f - value = 3.42014, p = 0.009255)$ which means that policy uncertainty negatively impacts consumer price index.

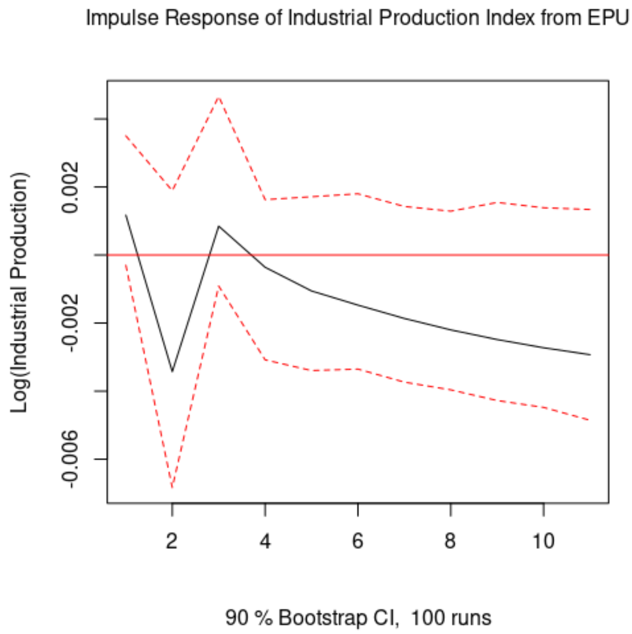

Figure 9. Impulse Response Function of the Response of Industrial Production to policy uncertainty shocks for 10 months ahead

Industrial production responds sharply to policy uncertainty shock within the first two months of the shock and then rises back to normal and starts to decline gradually after the 4th month as shown by figure 9.Our granger casuality tests however shows that policy uncertainty is not predictive of industrial production at 5% level of significance but the test is significant at 10% level of significance $(f - value = 2.3126, p = 0.08783)$

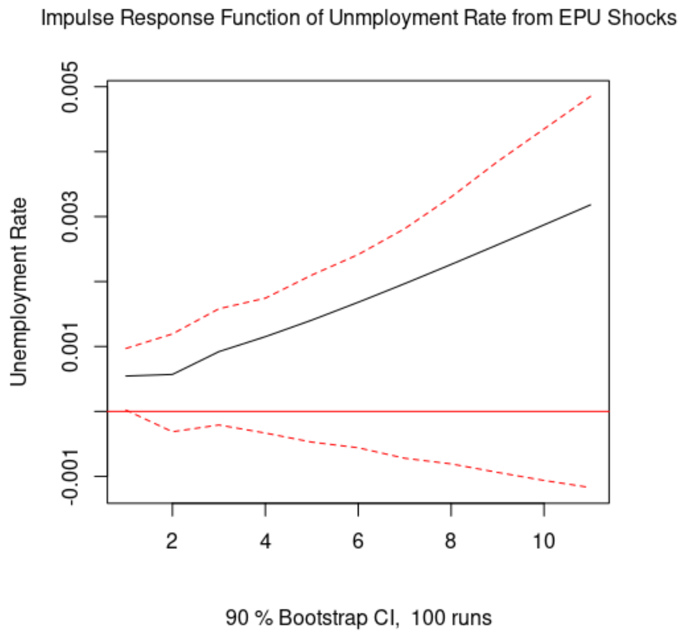

Figure 10. Impulse Response Function of the Response of Unemployment rates to policy uncertainty shocks to EPU for 10 months ahead

Unemployment rate responds gradually to a policy uncertainty shock, with the least effects of the shock experienced within a month after the shock but the effects of the shock continues to be experienced even after 10 months. We should note that such a relationship is not being supported by the granger causality test at 5% level of significance.

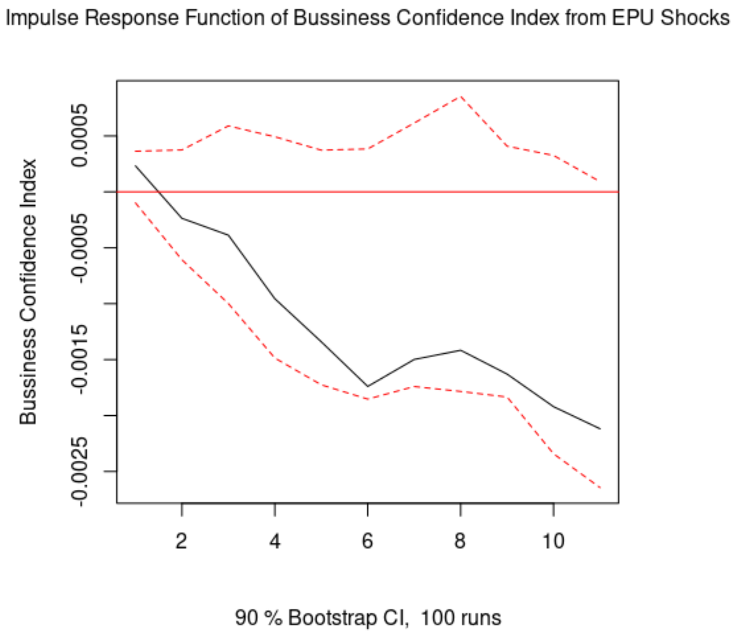

Figure 11. Impulse Response Function of the Response of Business Confidence Index to policy uncertainty shocks to EPU for 10 months ahead

Figure 9 demonstrates that policy uncertainty reduces business confidence with immediate effects and the effect of the policy uncertainty shock continues to be felt until the 6th month. Statistical analysis at 5% level of significance shows that policy uncertainty can also be predictive of business confidence$(f - value = 2.4835, p = 0.04412)$.

5. Conclusions

In this paper, we presented and evaluated the results of applying neural models on automatic economic policy uncertainty detection from news articles. Both USA and Ireland news articles were employed using weak supervision and extensive labels. We find that even though state of art methods trained with many labels outperform weak supervision in some cases, the gap in performance is small and the trade off can be accommodated in most economic applications. With the weak supervision set up presented here, we aim at timely results for policy decisions, compared to spending hundreds of hours on data annotation. Our results show that weak supervision can play a significant role in applying ML methods in measuring policy uncertainty from text, with much higher precision of current policy, which is based in counting words from a query to construct EPU indices. For future work, we intend to explore complementing weak supervision with a small set of carefully selected annotated examples through active learning or data subset selection as well as working on strategies for labeling functions and multi-label classification in regards to different types of policy uncertainty.

6. References

1. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)* **2020**, *53*, 1–34.
2. Smirnova, A.; Cudré-Mauroux, P. Relation extraction using distant supervision: A survey. *ACM Computing Surveys (CSUR)* **2018**, *51*, 1–35.
3. Ratner, A.J.; De Sa, C.M.; Wu, S.; Selsam, D.; Ré, C. Data programming: Creating large training sets, quickly. *Advances in neural information processing systems* **2016**, *29*, 3567–3575.
4. Ruder, S.; Peters, M.E.; Swayamdipta, S.; Wolf, T. Transfer learning in natural language processing. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials, 2019, pp. 15–18.
5. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* **2018**.
6. Liu, B. Uncertainty theory. In *Uncertainty theory*; Springer, 2007; pp. 205–234.
7. Baker, S.R.; Bloom, N.; Davis, S.J. Measuring economic policy uncertainty. *The quarterly journal of economics* **2016**, *131*, 1593–1636.
8. Azqueta-Gavaldon, A.; Hirschbühl, D.; Onorante, L.; Saiz, L. Economic policy uncertainty in the euro area: an unsupervised machine learning approach **2020**.
9. François-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J. An introduction to deep reinforcement learning. *arXiv preprint arXiv:1811.12560* **2018**.
10. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intelligent Systems and their applications* **1998**, *13*, 18–28.
11. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
12. Valentini, G.; Muselli, M.; Ruffino, F. Cancer recognition with bagged ensembles of support vector machines. *Neurocomputing* **2004**, *56*, 461–466.
13. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* **1997**, *45*, 2673–2681.
14. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
15. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT press, 2016.
16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. Advances in neural information processing systems, 2017, pp. 5998–6008.
17. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* **2014**.
18. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* **2019**.
19. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* **2019**.
20. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150* **2020**.

21. Zaheer, M.; Guruganesh, G.; Dubey, K.A.; Ainslie, J.; Alberti, C.; Ontanon, S.; Pham, P.; Ravula, A.; Wang, Q.; Yang, L.; others. Big Bird: Transformers for Longer Sequences. NeurIPS, 2020.
22. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems* **2019**, *32*.
23. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Journal of Big Data* **2019**, *6*, 1–48.
24. Settles, B. Active learning literature survey **2009**.
25. Vanschoren, J. Meta-learning: A survey. *arXiv preprint arXiv:1810.03548* **2018**.
26. Kuleshov, V.; Ding, J.; Vo, C.; Hancock, B.; Ratner, A.; Li, Y.; Ré, C.; Batzoglou, S.; Snyder, M. A machine-compiled database of genome-wide association studies. *Nature communications* **2019**, *10*, 1–8.
27. Callahan, A.; Fries, J.A.; Ré, C.; Huddleston, J.I.; Giori, N.J.; Delp, S.; Shah, N.H. Medical device surveillance with electronic health records. npj Digital Medicine, 2 (1), 94, 2019.
28. Bach, S.H.; Rodriguez, D.; Liu, Y.; Luo, C.; Shao, H.; Xia, C.; Sen, S.; Ratner, A.; Hancock, B.; Alborzi, H.; others. Snorkel drybell: A case study in deploying weak supervision at industrial scale. Proceedings of the 2019 International Conference on Management of Data, 2019, pp. 362–375.
29. Hammar, K.; Jaradat, S.; Dokoohaki, N.; Matskin, M. Deep text mining of instagram data without strong supervision. 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE, 2018, pp. 158–165.
30. Weng, Z.; Varma, P.; Masalov, A.; Ota, J.; Ré, C. Utilizing Weak Supervision to Infer Complex Objects and Situations in Autonomous Driving Data. 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019, pp. 119–125.
31. Langlois, R.N.; Cosgel, M.M. Frank Knight on risk, uncertainty, and the firm: a new interpretation. *Economic inquiry* **1993**, *31*, 456–465.
32. Bekaert, G.; Hoerova, M. The VIX, the variance premium and stock market volatility. *Journal of econometrics* **2014**, *183*, 181–192.
33. Barrero, J.M.; Bloom, N. Economic uncertainty and the recovery. *Unpublished manuscript. https://www. kansascityfed. org/˜/media/files/publicat/sympos/2020/20200806 bloom. pdf* **2020**.
34. Azqueta-Gavaldón, A. Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters* **2017**, *158*, 47–50.
35. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *the Journal of machine Learning research* **2003**, *3*, 993–1022.
36. Keith, K.A.; Teichmann, C.; O'Connor, B.; Meij, E. Uncertainty over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty. *arXiv preprint arXiv:2010.04706* **2020**.
37. Zalla, R. Economic policy uncertainty in Ireland. *Atlantic Economic Journal* **2017**, *45*, 269–271.
38. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* **2019**.
39. Gruber, J. *LexisNexisTools. An R package for working with newspaper data from 'LexisNexis'*, 2021. R package version 0.3.4.
40. OECD data.
41. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; others. Huggingface's transformers:

State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771* **2019**.
42. Ratner, A.; Bach, S.H.; Ehrenberg, H.; Fries, J.; Wu, S.; Ré, C. Snorkel: Rapid training data creation with weak supervision. *The VLDB Journal* **2020**, *29*, 709–730.
43. Zhang, Z. Improved adam optimizer for deep neural networks. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, 2018, pp. 1–2.
44. Chollet, F.; others. Keras: The python deep learning library. *Astrophysics Source Code Library* **2018**, pp. ascl–1806.
45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
46. Reeves, E.; Palcic, D. Getting back on track: the expanded use of PPPs in Ireland since the global financial crisis. *Policy Studies* **2017**, *38*, 339–355.
47. Rudd, K. The global financial crisis. *Monthly, The* **2009**, pp. 20–29.
48. Rice, J.; others. Economic policy uncertainty in small open economies: a case study in ireland. Technical report, Central Bank of Ireland, 2020.
49. Mertens, K.; Ravn, M.O. Measuring the impact of fiscal policy in the face of anticipation: a structural VAR approach. *The Economic Journal* **2010**, *120*, 393–413.
50. Zivot, E.; Wang, J. Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-Plus®* **2006**, pp. 385–429.
51. Lin, J.L. Teaching notes on impulse response function and structural VAR. *Institute of Economics, Academia Sinica, Department of Economics, National Chengchi University* **2006**, pp. 1–9.
52. Mushtaq, R. Augmented dickey fuller test **2011**.
53. Hu, S. Akaike information criterion. *Center for Research in Scientific Computation* **2007**, *93*.
54. Seth, A. Granger causality. *Scholarpedia* **2007**, *2*, 1667.