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

Paul Trust, Ahmed Zahran and Rosane Minghim
School of Computer Science and Information Technology, University College Cork, Ireland.

Contributing authors: 120222601@umail.ucc.ie; ahmed.zahran@ucc.ie; r.minghim@cs.ucc.ie;

Abstract
The need for timely data analysis for economic decisions has prompted most economists and policy makers to search for non-traditional supplementary sources of data. In that context, text data is being explored to enrich traditional data sources because it is easy to collect and highly abundant. Our work focuses on studying the potential of textual data, in particular news pieces, for measuring economic policy uncertainty (EPU). Economic policy uncertainty is defined as the public’s inability to predict the outcomes of their decisions under new policies and future economic fundamentals. Quantifying EPU is of great importance to policy makers, economists, and investors since it influences their expectations about the future economic fundamentals with an impact on their policy, investment and saving decisions. Most of the previous work using news articles for measuring EPU are either manual or based on a simple keyword search. Our work proposes a machine learning based solution involving weak supervision to classify news articles with regards to economic policy uncertainty. Weak supervision is shown to be an efficient machine learning paradigm for applying machine learning models in low resource settings with no or scarce training sets, leveraging domain knowledge and heuristics. We further generated a weak supervision based EPU 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, Causality

1 Introduction
Modern machine learning (ML) approaches especially neural network approaches have achieved a state of the art performance close to human performance in many applications. Their success is largely attributed to the availability of high quality annotated data. However, annotating data is not a trivial task and usually requires expensive human effort with domain expertise making the process significantly expensive. This pain-stacking
task of annotating data has limited the potential of adoption of ML techniques in many low resource fields like economics where there are few to no large annotated public data. There has been several efforts by the ML community to enable ML models work efficiently with limited training data which include; few shot learning [1], distant supervision [2], weak supervision [3] and transfer learning [4] (see Section 2.1 for details).

We study the potential of automatically measuring economic policy uncertainty in a low resource setting using ML models. Economic policy uncertainty (EPU) is defined as the public’s inability to predict the outcomes of their decisions under new policies and future economic fundamentals [5]. The fact that people cannot correctly know what will happen in the future, makes them to rely on their subjective beliefs (uncertainty) to make current decisions and actions regarding their consumption, investment and savings with a significant impact on national economic aggregates.

There is thus an assumed casual relationship between EPU and other macroeconomic indicators making EPU an interesting concept for macroeconomic monitoring and planning. Several methodologies have been proposed for measuring EPU including but not limited to text based approaches, forecaster disagreements, Chicago Board Options Exchange Volatility Index (VIX), business surveys of subjective uncertainty and stock market volatility (see Section 2.2 for details). However, quantitatively measuring EPU is not a trivial task since it is an abstract concept that is not directly observable but rather within the minds of consumers and policy makers and all the described measures of EPU rely on proxies to quantify it.

Our case study involves using news articles as a proxy for measuring the level of EPU [6] in an economy. Our approach is inspired by Baker et.al (2016) who constructed an EPU index by retrieving news articles that satisfied a keyword occurrence related to economy, policy and uncertainty expressed 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 investment risk [7].

Even though this simple keyword search used by Baker et.al (2016) yielded a promising EPU index, almost half of the retrieved news articles were found to be false positives according to
human auditors with economic level knowledge[6]. To mitigate the false positives and false negatives, Baker et al. (2016) [6] employed several research assistants to manually audit the retrieved articles. The manual process used by Baker et al. (2016) [6] 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 classify these news articles with regards to EPU categories. The challenge with using the ML approach is that we needed a large amount of training data which was not readily available in our case. We thus opted for the weakly supervised learning (BERT + WS) by fine-tuning BERT (Bidirectional Encoder Representations from Transformers) one of the state of art deep learning models using noisy labels generated by cheap weak sources. Weak supervision (WS) is an evolving machine learning paradigm that generates training data in less time by aggregating multiple domain heuristics expressed in the so called labeling functions (LF) [3]. These 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 outputs from these labeling functions are combined using a generative model producing noisy labels which are used to train end models such as transformers [8]. In short, subject matter experts express their domain knowledge about the data through labeling functions instead of individually labeling each data point.

Our work proposes an approach that uses weak supervision combined with neural language models to automatically classify news articles describing economic policy uncertainty (see Figure 1). The success of this approach may impact the progress of locating sources of economic uncertainty in news articles more effectively by shortening the laborious labeling step.

*The key contributions of this work is as follows:

- Proposing a weak supervision approach (BERT + WS) for automatic classification of economic policy uncertainty from news pieces.
- Generating an Irish weak supervision based economic policy uncertainty index based weak supervision and neural language models.
- 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

Machine Learning (ML) involves using available data to learn patterns for making predictions
for unseen data. ML can fall in different categories depending on the nature of data. The dominant learning paradigms include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning aims to learn a predictive model using input data points \( x \in X \) and predicting its corresponding outputs \( y \in Y \) 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 (SVM)[10], random forests [11], are well studied and are known to perform well for small datasets. 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. The 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) and Long Short Term Memory (LSTM) [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 models [16] leverages attention mechanism to get rid of the recurrence used in RNNs and build a neural model that is solely based on attention mechanism. Transformer has achieved superior performance to the state-of-the-art before it and most of state of the art models are based on this architecture.
Bi-directional Encoder Representations from Transformers (BERT) [8] is one of the transformer models that achieved state of art performance across most natural language processing tasks. BERT [8] was pre-trained in a bi-directional context with two objectives: masked language modeling and next sentence prediction using the book-corpus (800 million words) and English wikpedia (2,500 million words). 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 (160GB of uncompressed text), and removes the next sentence prediction objective. The two models (BERT and RoBERTa) can be fine-tuned by adding domain task specific inputs and outputs into the pre-trained BERT and fine-tuning all the parameters end to end. There are other several variants of BERT including but not limited to the following [19–21].

The incredible performance of these models is largely enabled by the presence of high quality annotated data, which may not be readily available or 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 datasets. Transfer learning [4] is one approach where models are pretrained on generic tasks and fine-tuned on domain tasks. Data augmentation [22] creates more training data points by artificially modifying the existing dataset. Active learning [23] 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 [24] 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 datasets, 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 be potentially conflicting or correlated. Data programming addresses the noisy and conflicting nature of labeling functions outcome by modeling them as a generative process using a factor model. 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 [25] [26] [25], social media analysis [27] [28], and autonomous driving [29]. 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 [30] 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 with certainty. 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.

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 [31]. Another often rather controversial measure of uncertainty is disagreements among forecasters [32], 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 by providing an elicited 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) [6] 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 divided by the number of news articles within 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) [33] handled economic policy uncertainty detection by using latent Dirichlet allocation (LDA) [34], a topic modeling technique that outputs topics from documents, and then created 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. In our work, we employ a proper classification rather than a topic modeling or counting. Keith et.al (2020) [35] did propose supervised ML methods for EPU classification, and 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 Methodology

3.1 Overview

Our proposed framework as shown in Figure 1 involves three key stages; The first stage uses expert-defined labeling functions (LFs) to automatically generate a label matrix, in which each news article is assigned a number of noisy labels for each LF. The second stage uses unsupervised generative factor model to combine the outputs of multiple LFs into a single auto-generated noisy label for each news article denoted as \( \bar{Y} \). In the third stage, we fine-tuned BERT to perform EPU classification using the auto-generated noisy labels. The following subsections present these stages in more details.

3.2 Automated labeling Model

3.2.1 Labeling Functions

Labeling functions are a means of expressing domain knowledge to label a subset of data points without individually labeling each data point. The input of labeling functions is a dataset, denoted as \( X^{n \times 1} \), consisting of independent and identically distributed (i.i.d) news articles and \( m \) labeling functions (LFs), denoted by \( \lambda = \{ \lambda_1, ..., \lambda_m \} \), that are derived from heuristics provided by the domain experts or existing prior knowledge about the task. Each labeling function \( \lambda_j \) noisily labels each news article with \( \lambda_j(x) \in \{-1, 0, 1\} \), where 1 indicates that the news article is about EPU, 0 indicates that the news article is not about EPU 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 \( \Lambda \in \{-1, 0, 1\}^{n \times m} \).

The label matrix is then processed by the generative factor model to produce a vector of noisy labels \( \bar{Y} = \{\bar{y}_1, \ldots, \bar{y}_k\} \) where \( k \) denotes only those articles with an associated noisy label with \( k \leq n \) (only those where the LFs did not abstain), these are then used to fine-tune BERT.

We describe some of the details of the labeling functions used in our case as follows:

- **Keywords**: The experts are only required to provide a few keywords associated with each class. This is plausible because experts can not always give an exhaustive list of keywords associated with the classes. We expand the expert provided keywords by mining the top nearest words to the provided words in the semantic space. This was achieved by first obtaining the embedding of these expert provided keywords and also an embedding for each word in the corpus, the embeddings of both are normalized to reside on a unit sphere representing a joint semantic space from where we can retrieve top \( k \)-nearest neighbors words. We choose to focus on the spherical space as opposed to the euclidean space due to its superiority in exploiting structure and geometry of the manifold[36–38].

More formally, let \( U \) be the embeddings of the few user provided seed words for each
class, the semantics of each class is modeled as a Von-Mises Fisher (vMF) distribution. A \(d\)-dimensional unit random vector \(u\) is said to have \(d\)-variate Von Mises-Fisher (vMF) distributions if its probability density function (pdf) is given by:

\[
f(U \mid \mu, k) = c_d(k) \exp k \mu^T u
\]  

(1)

where \(\mu, k \geq 0\), The normalizing constant \(c_d(k)\) is given by:

\[
c_d(k) = \frac{k^{d/2 - 1}}{(2\pi)^{d/2} I_{d/2}(k)}
\]  

(2)

where \(I_t(.)\) is a modified Bessel function of order \(t\).

We used an Expectation Maximization (EM) algorithm to find the parameters of the Von Mises-Fisher (vMF) distributions. Given vMF distributions \(f_v(u \mid \theta_v)\) with parameters \(\theta_v = (\mu_v, k_v)\) for \(1 \leq v \leq k\), we can modify the initial pdf to include a mixture of the \(k\) vMF distributions:

\[
f(U \mid \Theta) = \sum_{v=1}^{k} \alpha_v f_v(u \mid \theta_v)
\]  

(3)

where \(\Theta = \{\alpha_1, \ldots, \alpha_k, \theta_1, \ldots, \theta_k\}\)

Let \(H = \{h_1, \ldots, h_n\}\) be a set of hidden variables for vMF distributions where the points are sampled.

\[
\ln(P(U, H \mid \Theta)) = \sum_{i=1}^{n} \ln(\alpha_{h_i} f_{h_i}(u_i \mid \theta_{h_i}))
\]  

(4)

Assuming the values of \(H\) were observed, we could have obtained the values of the parameters using complete log-likelihood of the observed data, but since these are not observed we shall optimize the incomplete likelihood using E-step and M-step;

– E-step:

In the Expectation step, we used the observed data to estimate or guess the values of the missing data:

\[
p(h \mid u_i, \Theta) = \frac{\alpha_h f_h(u_i \mid \Theta)}{\sum_{l=1}^{k} \alpha_l f_l(u_i \mid \Theta)}
\]  

(5)

– M-step:

In the Maximization step, we used the complete data from the E-step to update the parameters of the model:

\[
\alpha_h = \frac{1}{n} \sum_{i=1}^{n} p(h \mid u_i, \Theta),
\]  

(6)

\[
r_h = \sum_{i=1}^{n} u_i p(h \mid x_i, \Theta)
\]  

(7)

\[
\mu_h = \frac{r_h}{r_h}
\]  

(8)

\[
I_{d/2}(k_h) = \frac{r_h}{I_{d/2-1}(k_h)} = \sum_{i=1}^{n} p(h \mid u_i, \Theta)
\]  

(9)

The expanded keyword list is used as keyword lookup for labeling functions, as well for semantic similarity tasks.
• **Semantic Similarity**: We found contextualized vector representations (embeddings) of our expanded keywords and for each of the news articles $x$ within our corpus using siamese sentence transformer[^39]. These embeddings are used to generate soft labeling functions using the semantic similarity of the expanded user provided keywords and for each news article in our corpus using cosine similarity. The hypothesis behind this labeling function is that the news article with a higher cosine similarity with the embedding of keywords for an EPU class is most likely to belong to that class. We assign a new article $x$ to a pseudo EPU class $Y$ if:

$$\text{COSINE}(x, k) < \phi$$

where $\phi \in [0, 1]$ is a hyper parameter.

• **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.

• **Zero shot classifier**: We used BART (Bidirectional Auto encoder Regressive Transformers)[^40] to perform zero shot inference on our news articles producing noisy labels for each EPU class.

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

### 3.2.2 Generative Model

The objective of this stage was to automatically assign each news article a *noisy* label in an unsupervised fashion. The outputs of the labeling functions described above can be conceptualized as multiple annotators labeling the same news article just like in the crowd-sourcing setting[^41]. The produced labels will have conflicts and potentially correlations, this will happen even if these news articles were labeled by domain experts[^35]. Instead of simply taking a majority vote to obtain a final label, we adopt a probabilistic framework to exploit the structure and correlations within the label matrix. The probabilistic model $P_w(\Lambda, Y)$ is formulated as a joint probability of the outputs of the labeling functions (label matrix) $\Lambda$ and the latent (unobserved) true class labels of the news articles $Y$. In particular, we encoded the label matrix with a factor model using three factor types representing the conflicts, correlations, and propensity (where LFs did not abstain) of the labeling functions.
The generative model can be defined as follows:

\[
P_w(\Lambda, Y) = Z_w^{-1} \exp \left( \sum_{i=1}^{m} w^T \phi_i(\lambda_i, y_i) \right),
\]

where \(Z_w\) is the normalizing constant and \(\phi_i(\lambda_i, y_i)\) represents an aggregation representing factors for all labeling functions given a sample news article \(x \in X\) and \(y_i\) is the latent class label.

In order to learn parameters \(w\) of the model \(P_w(\Lambda, Y)\), we minimized the negative marginal likelihood given the observed label matrix \(\Lambda\), by only observing the agreements and disagreements in the label matrix \(\Lambda\) since we do not have access to the ground truth labels with the formulation below:

\[
w = \arg \max_w \log \sum_{Y} P_w(\Lambda, Y)
\]

The learned parameters are used to generate the noisy labels \(\bar{Y} = P_w(Y | \Lambda)\) which can be used in fine-tuning the BERT classifier.

### 3.2.3 Discriminative Model

In this stage, we fine-tuned BERT (BERT + WS) using the noisy labels generated by weak supervision sources instead of the human annotated labels. This was achieved by adding a feed forward neural network on the last layers and leaving the other layers frozen to facilitate adaptation on our downstream EPU classification task. The model was trained by minimizing expected loss using a noise-aware objective function:

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

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 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].

### 4 Results and Discussion

This section describes the implementation of our weak supervision framework (BERT + WS) and its comparison to other available models for economic policy uncertainty (EPU) classification.

#### 4.1 Datasets

We employed news articles published in Ireland and USA newspapers, as well as data on economic indicators to understand the usefulness of our approach in predicting economic fundamentals.
4.1.1 News Datasets

Irish Newspapers: We searched for news articles from the Irish Times ¹ and Irish Independent ² newspapers in the timeframe of January 1992 to August 2021. These newspapers were selected because they were among those that 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 contained the following keyword combinations; {('uncertain' OR uncertainty') AND ('economy' OR 'economic') AND ('regulation' OR 'legislation' OR 'dail' OR 'deficit' OR 'Taoiseach')}. In total, 10070 articles were retrieved according to the keyword query. 10% of the retrieved articles were randomly selected and manually labeled for our experiments. The annotation process followed the coding guide provided by Baker et. al (2016) [6]. In this guide, the newspaper article is labeled as describing economic policy uncertainty if: 1. The article talks about uncertainty over who makes or will make policy decisions that have economic consequences. 2. The article talks about current and past uncertainty over what economic policy actions will be undertaken. and 3. The news articles talk about uncertainty regarding the economic effects of policy actions. The labeled dataset (1070 news pieces) was split into training, validation and testing in the ratio of 8:1:1 respectively.

USA Newspapers: We used 12000 news articles that were selected as an audit sample from the retrieved USA articles containing the terms ‘economy’ and ‘uncertainty’. The domain experts labeled these selected news articles into binary categories of presence or absence of EPU and into further EPU categories like taxes, fiscal policy, monetary policy and others. For our case we only used the binary EPU categories for the experiments. Further details about the dataset can be obtained from³ [6]. The dataset was split into training, validation and testing in the ratio of 8:1:1 respectively.

4.1.2 Economic Indicators

In this section, we describe in details of the economic indicators that were used in econometric analysis. The source of the datasets was Organization for Economic Co-operation and Development (OECD)[42] obtained on 1st November, 2021 at 12:41 pm.

Volatility Index: We used the Volatility Index for the stock price of Ireland. The index measures the 360-day ahead standard deviation on the return on the national stock market index. The deviation is a representation of the market’s expectations of the future changes on the stock returns.

---

¹https://www.irishtimes.com/
²https://www.independent.ie/
³https://www.policyuncertainty.com
**Consumer Price Index (CPI):** We employed Irish monthly consumer price index (CPI) from January 1992 to August 2021 to perform our econometric analysis. CPI is a measure of the changes in the prices of a basket of goods and services that are purchased by specific groups of households. The index is commonly used by economists as a proxy for inflation.

**Unemployment Rate:** Monthly Irish unemployment rate from January 1992 to August 2021 was also utilized for econometric analysis. The indicator measures the proportion of unemployed people of working age who do not have work, who are available for work, and who have also taken a step to obtain work.

**Industrial Production:** We employed Irish industrial production from January 1992 to August 2021 to assess how our generated index is predictive of outputs in the economy. The indicator measures the output of industrial establishments in a given period of time.

**Short Term Interest Rates:** The Irish short term interest rates from January 1992 to August 2021 measures the rate at which short term government paper is issued or traded in the market.

**Business Confidence Index:** The Irish monthly business consumer 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.

**Irish Economic Policy Uncertainty Index:** We used the Irish economic policy uncertainty Index. The index was generated by a keyword search from the Irish Times newspaper provided by Baker et. al (2016) [43].

### 4.2 Evaluation Setup and Metrics

We implemented BERT [8] and RoBERTa [18] models alongside their embeddings using huggingface 4 and simple transformers libraries [44]. Weak supervision approaches were implemented using Snorkel 5 [45] and Rubrix. Long short term memory (LSTM) [14] was implemented using Keras library 7 [46] and support vector machine (SVM) classifier using Scikit-Learn library 8 [47].

The experiments for neural models (BERT, RoBERTa, BERT + WS) were conducted with an Adam optimizer[48], an initial learning rate of $2e^{-5}$, a batch size of 8 and a maximum length of 512 tokens. BERT and RoBERTa generated their own embeddings to be used for classification, BERT + WS used the embeddings generated by BERT and the standard bag of words models using term frequency-inverse document frequency (TF-IDF) were fed into the classification pipeline for LSTM and SVM. The training and validation

---

4[https://huggingface.co/](https://huggingface.co/)
5[https://www.snorkel.org/](https://www.snorkel.org/)
4[https://rubrix.readthedocs.io/en/stable/guides/weak-supervision.html](https://rubrix.readthedocs.io/en/stable/guides/weak-supervision.html)
7[https://keras.io/](https://keras.io/)
8[https://scikit-learn.org/](https://scikit-learn.org/)
9Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions based on adaptive estimates of lower-order moments[48]
losses for each training epoch were monitored and the models with the best accuracy on the validation set saved before comparison with the test set.

The comparison of our proposed solution with the other models was done using F1-score precision and accuracy (eq. 13 to 15).

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (13)
\]

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)
\]

\[
F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (15)
\]

where \(TP\) are the True Positives, \(FP\) are the False Positives, \(FN\) are the False Negatives and \(TN\) are the True Negatives.

### 4.3 Experiments

This section describes the experimental results and evaluation of our weak supervision framework (\(BERT + WS\)) involving BERT model fine-tuned with noisy labels generated by weak sources. We compared this framework with the models described in section 4.2. Our baseline comparison is with the keyword search, the current solution employed by economists to generate the EPU index from news articles.

The results on the Irish news articles dataset (see Table 1) demonstrated that (\(BERT + WS\)) presents a significant improvement in precision (+20%) from just using keyword occurrences.

| Model            | Precision | Accuracy | F1-score |
|------------------|-----------|----------|----------|
| BERT             | 60.90     | 64.90    | 57.80    |
| RoBERTa          | 64.28     | 66.34    | 60.67    |
| LSTM             | 56.36     | 57.69    | 50.76    |
| SVM              | 48.90     | 53.62    | 50.07    |
| **BERT + WS**    | 60.10     | 62.40    | 55.60    |
| **Keyword Search** | 39.40     | 39.00    | 28.00    |

Table 1: Evaluation metrics; Precision, Accuracy and F1 scores scores (all in percentages) for (\(BERT + WS\)) and other comparison models on the Irish news articles test set

\(BERT + WS\) also outperforms SVM and LSTM, but it is outperformed by RoBERTa (62.40%' against 66.34% ) and BERT fine-tuned using human annotated data (62.40%'against 64.90% ).

Table 2 illustrates the results on the USA test dataset which is 10 times larger compared to the Irish dataset. BERT fine-tuned with weak supervision presents a significant improvement from just using keyword search (+19%) the current state of the art solution for EPU indices. State of the art neural models trained on human annotated data
Fig. 2: t-sne projections of the embedding space of Irish news articles segregated according to economic policy uncertainty (EPU) categories. Point colors represent EPU classes and the values shown in the parenthesis represent the silhouette coefficient in the projected space. Projections labeled human annotation are the model’s embeddings space segregated by human provided labels for the articles, BERT and (BERT + WS) projections were generated using BERT embeddings, LSTM and SVM were generated using TF-IDF embeddings and RoBERTa projections were generated on top of RoBERTa model embeddings.
Fig. 3: t-sne projections of the embedding space of USA news articles segregated according to economic policy uncertainty (EPU) classes. Point colors represent EPU classes and the values shown in the parenthesis represent the silhouette coefficient in the projected space. Projections labeled human annotation are the model’s embeddings space segregated by human provided labels, BERT and (BERT + WS) projections were generated using BERT embeddings, LSTM and SVM were generated using TF-IDF embeddings and RoBERTa projections were generated on top of RoBERTa model embeddings.
still outperform models trained with weak supervision based noisy labels, at about 5% in precision and accuracy, but with comparable in F1-score.

It should be stressed that while the other models reported in this work require large amounts of expert annotated labels, that effort is replaced by designing labeling functions in (BERT + WS). To illustrate the difference in human effort involved in this case, it took one of the authors of this work (who is trained in economics) just a week to write proper labeling functions for either dataset with inspiration from the coding guide provided by [6], while it took a team of 14 expert human annotators 6 months to label 12000 USA news articles [6]. The difference in performance (3%) can be traded off in most applications of economics, also making (BERT + WS) largely reproducible in new contexts.

Figure 4 shows the performance of the models using the Receiver Operating Characteristic (ROC) curves of both the Irish and USA test datasets. The area under the curve for most models is larger on the USA dataset compared to the Irish dataset; this indicates that models have better discriminative power with respect to the EPU classes on USA datasets than Irish dataset. There is also much more area covered by (BERT + WS) on a large dataset (USA news articles) implying that fine-tuning with weak supervision also benefits from more training examples. Similar results were earlier observed with evaluation metrics in Tables 1 and 2.
Figures 2 and 3 are t-distributed stochastic neighbor embedding (t-sne) [49] multidimensional projections of the neural embedding space of the news articles by different models. We evaluated the models both visually and analytically using the silhouette coefficient [50]. In both datasets, BERT had the best silhouette coefficients. The visual analysis and silhouettes coefficients of all models is not good, this is explained by the difficulty that comes with EPU classification.

Experimental results suggest that when a significant labeling budget is available, policy makers and 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 (3% to 5%) in performance is not attractive. Additionally, designing labeling functions that accommodate new 'views' of the dataset is a straightforward activity allowing flexibility in the process of adaptation to new tasks.

4.4 Economic Policy Uncertainty Index

We used the predictions from \((BERT + WS)\) to construct an Irish monthly EPU index from January 1992 to July 2021 using the following steps as proposed by Baker et.al(2016) [6]:

**Step 1:** Collected news pieces from relevant newspapers in a time window of interest.

**Step 2:** We used \((BERT + WS)\) to classify news articles describing policy uncertainty.

**Step 3:** Counted the number of articles that have a positive label for economic policy uncertainty level for each newspaper across months.

**Step 4:** Computed the time-series variance \(\sigma_i^2\) in the selected time interval for each newspaper, normalized the time series of counts of new articles with positive EPU label using the standard deviation of the time series.

**Step 5:** Computed the mean \(M\) of the normalized time series of counts of news articles

**Step 6:** Generated an EPU index by multiplying \((100/M)\) with the normalized time series of counts.

4.4.1 Economic Signal from Weak Supervision EPU Index

The generated Irish EPU index captures both local and global events that are known to have caused immense uncertainty not only to Ireland but also to the entire world as shown by Figure 5. The index spiked highest in 2008 which we suspect is the uncertainty that was caused by the global financial crisis following the failure of Lehman and brothers [51, 52], the spike in 2016 may be attributed to the uncertainty due to brexit effects and local uncertain events like austerity protests. We can learn from the index that the Irish economy is more prone to policy uncertainty.
Fig. 5: Economic Policy Uncertainty (EPU) Index for Ireland plotted from 1992 to 2021 generated using weak supervision based methodology. Higher index counts means that there was much policy uncertainty perceived at that time. We also annotate the graph with the potential events responsible for the extreme high values from abroad, this is partly explained by the open nature of the Irish economy [53].

we defined a $p-\text{lag}$ vector auto-regressive (VAR($p$)) as:

$$Y_t = c + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t, \ t = 1, \ldots, T \quad (16)$$

where $\Phi_i$ are $n \times n$ coefficient matrices and $\varepsilon_t$ is an $(n \times 1)$ unobserved zero mean white noise vector process with time invariant co-variance matrix $\Sigma$ [55].

Our goal was to use the estimated model to understand whether our generated EPU index can be predictive of standard macroeconomic variables by using this impulse response functions generated from the estimated VAR model.

Impulse response function [56] 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

5 Economic Signal from Weak Supervision EPU Index

5.1 Econometric Model

5.1.1 Vector Auto-Regression Models (VAR)

We conducted an econometric analysis using vector Auto-Regression models (VAR) [54] to exploit time series variations within the Irish macroeconomic indicators.

More formally, consider $n \times 1$ macroeconomic time dependent variables: $Y_t = (y_{1t}, \ldots, y_{nt})^T$
time period $t$ with consideration of the interaction among the variables.

### 5.1.2 Granger Causality

In econometric analysis, we are often interested in determining how variables affect or influence one another. Even though this is a well studied problem in the statistics literature, determining the actual cause of a certain phenomena is a non-trivial task. The complexity of the problem arises due to the existence of confounders, confusion from spurious correlations and determining the direction of the relationship. For example, consider two variables $X$ and $Y$, there are 4 possible relationships; $X$ causes $Y$ ($X \rightarrow Y$), $Y$ causes $X$ ($X \leftarrow Y$), Both $X$ and $Y$ cause each other ($X \leftrightarrow Y$) or the two variables are independent ($X \mid Y$).

Granger causality assumes that if $X$ is a cause of $Y$, then $X$ (Cause) must occur before $Y$ (effect), and for that matter can be used to predict it. We can thus deduce that forecasting future values of $Y_t$ with both past target and past source time series $E(Y_t \mid Y_{<t}, X_{<t})$ is significantly powerful than only using past time series $E(Y_t \mid Y_{<t})$.

To infer Granger causality of two variables, we consider the following two equations:

$$Y_t = \sum_{i=1}^{n} \alpha_i Y_{t-1} + N_t$$  \hspace{1cm} (17)

$$Y_t = \sum_{i=1}^{n} \alpha_i Y_{t-1} + \sum_{i=1}^{n} \beta_i Y_{t-1} + \tilde{N}_t$$  \hspace{1cm} (18)

where $N_t$ and $\tilde{N}_t$ are assumed to be i.i.d time series. $X$ is said to be a Granger-cause of $Y$ whenever the noise term $\tilde{N}_t$ with predictions of $X$ included has a significantly smaller variance than the noise term $N_t$ obtained with $X$ [57].

### 5.2 Econometric Analysis

We fitted a VAR model to monthly Ireland data from January 1992 to August 2021 using Cholesky decomposition [58], this was done in order to recover the orthogonal shocks using the following macroeconomic variables; EPU index, unemployment rate, logarithm of industrial production, logarithm of consumer confidence index, short term interest rates and logarithm of consumer price index (CPI).

To ensure that the series are stationary (a precondition for VAR analysis), we conducted unit root tests of the variables using augmented dicker fuller test (ADF) test [59], the results are shown in Table 3.

| Variable              | Statistic | P-value  |
|-----------------------|-----------|----------|
| EPU                   | 0.700     | 0.01933  |
| CPI                   | 3.5957    | 0.0043   |
| Industrial Production | 2.7372    | 0.00016  |
| Interest Rates        | -0.9647   | 0.0007001|
| Unemployment Rate     | -1.4947   | 0.000    |

**Table 3**: Shows the results of the Stationarity tests carried out using augmented Dickey Fuller Test (ADF) for VAR model fitting.
Observations from Table 3 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) [60] was then used to find an optimal time lag to be used to fit a VAR model and 4 months were found to optimal.

The impulse function shown in Figure 6 indicates that 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 [61] 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.

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 7. Our Granger causality 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 ($f-value = 2.3126, p = 0.08783$) Unemployment rate responds gradually to a policy uncertainty shock, with the least effects of the shock experienced within a month after the shock.
Figure 8 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).

6 Conclusion

In this paper, we presented and evaluated the results of fine-tuning BERT with noisy labels generated by weak supervision ($BERT + WS$). We find that even though state of art methods trained with human annotated labels outperform $BERT + WS$ in many cases. The gap in performance is small and the trade off can be accommodated in most economic applications. The weak supervision framework presented here ($BERT + WS$) aims 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 compared to current keyword based approaches used by economists in constructing EPU indices. For future work, we intend to explore complementing weak supervision with a small set of carefully selected human 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.

**Supplementary information.**

**Acknowledgments.** This work is supported by the Science Foundation Ireland (SFI) Center for Research Training in Advanced Networks and Future communications at University College Cork, project code, R19784

**Declarations**

- **Funding**
  The work was funded by Science Foundation Ireland (SFI) Center for Research Training in Advanced Networks and Future communications

- **Conflict of interest/Competing interests** The authors declare no Conflict of interest

- **Ethics approval**
  Not applicable

- **Consent to participate**
  Not applicable

- **Consent for publication**
  Yes

- **Availability of data and materials**
  The used in the experiments can be found on [https://www.policyuncertainty.com/](https://www.policyuncertainty.com/)

- **Code availability**
  The paper code is found on with sample data [https://github.com/TrustPaul/Weak-supervision-in-Economic-Policy-Uncertainty](https://github.com/TrustPaul/Weak-supervision-in-Economic-Policy-Uncertainty)

- **Authors’ contributions**
  All authors contributed equally to the work

**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) **53**(3), 1–34 (2020)

[2] Smirnova, A., Cudré-Mauroux, P.: Relation extraction using distant supervision: A survey. ACM Computing Surveys (CSUR) **51**(5), 1–35 (2018)

[3] Ratner, A.J., De Sa, C.M., Wu, S., Sel-sam, D., Ré, C.: Data programming: Creating large training sets, quickly. Advances in neural information processing systems **29**, 3567–3575 (2016)

[4] Ruder, S., Peters, M.E., Swayamdipta, S., Wolf, T.: Transfer learning in natural language processing. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials, pp. 15–18 (2019)

[5] Liu, B.: Uncertainty theory. In: Uncertainty
Theory, pp. 205–234 (2007)

[6] Baker, S.R., Bloom, N., Davis, S.J.: Measuring economic policy uncertainty. The quarterly journal of economics 131(4), 1593–1636 (2016)

[7] Azqueta-Gavalдон, A., Hirschbühl, D., Onorante, L., Saiz, L.: Economic policy uncertainty in the euro area: an unsupervised machine learning approach (2020)

[8] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (2019). https://doi.org/10.18653/v1/N19-1423. https://aclanthology.org/N19-1423

[9] François-Lavet, V., Henderson, P., Islam, R., Bellemare, M.G., Pineau, J.: An introduction to deep reinforcement learning. Found. Trends Mach. Learn. 11(3–4), 219–354 (2018). https://doi.org/10.1561/2200000071

[10] Hearst, M., Dumais, S., Osman, E., Platt, J., Schölkopf, B.: Support vector machines. IEEE Intelligent Systems and their Applications 13(4), 18–28 (1998). https://doi.org/10.1109/5254.708428. Max-Planck-Gesellschaft

[11] Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)

[12] Valentini, G., Muselli, M., Ruffino, F.: Cancer recognition with bagged ensembles of support vector machines. Neurocomputing 56, 461–466 (2004)

[13] Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing 45(11), 2673–2681 (1997)

[14] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735–1780 (1997)

[15] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning, (2016)

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 30 (2017). https:
[17] Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015). http://arxiv.org/abs/1409.0473

[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:1907.11692 (2019). http://arxiv.org/abs/1907.11692

[19] Beltagy, I., Peters, M.E., Cohan, A.: Longformer: The long-document transformer. CoRR abs/2004.05150 (2020)

[20] Zaheer, M., Guruganesh, G., Dubey, K.A., Ainslie, J., Alberti, C., Ontanon, S., Pham, P., Ravula, A., Wang, Q., Yang, L., et al.: Big bird: Transformers for longer sequences. In: NeurIPS (2020)

[21] 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 32 (2019)

[22] Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data 6(1), 1–48 (2019)

[23] Settles, B.: Active learning literature survey (2009)

[24] Vanschoren, J.: Meta-learning: A survey. arXiv preprint arXiv:1810.03548 (2019). https://doi.org/10.1007/978-3-030-05318-5_2

[25] 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 10(1), 1–8 (2019)

[26] 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)

[27] Bach, S.H., Rodriguez, D., Liu, Y., Luo, C., Shao, H., Xia, C., Sen, S., Ratner, A., Hancock, B., Alborzi, H., et al.: Snorkel drybell: A case study in deploying weak supervision at industrial scale. In: Proceedings of the 2019 International Conference on Management of
[28] Hammar, K., Jaradat, S., Dokoohaki, N., Matskin, M.: Deep text mining of insta-
gram data without strong supervision. In: 2018 IEEE/WIC/ACM International Confer-
ence on Web Intelligence (WI), pp. 158–165 (2018). IEEE

[29] Weng, Z., Varma, P., Masalov, A., Ota, J., Ré, C.: Utilizing weak supervision to
infer complex objects and situations in autonomous driving data. In: 2019 IEEE
Intelligent Vehicles Symposium (IV), pp. 119–125 (2019). IEEE

[30] Langlois, R.N., Cosgel, M.M.: Frank knight on risk, uncertainty, and the firm: a new inter-
pretation. Economic inquiry 31(3), 456–465 (1993)

[31] Bekaert, G., Hoerova, M.: The vix, the variance premium and stock market volatil-
ity. Journal of econometrics 183(2), 181–192 (2014)

[32] Barrero, J.M., Bloom, N.: Economic uncertainty and the recovery. https://www.
kansascityfed.org/~/media/files/publicat/sympos/2020/20200806
bloom. pdf (2020)

[33] Azqueta-Gavaldón, A.: Developing news-
based economic policy uncertainty index with
unsupervised machine learning. Economics
Letters 158, 47–50 (2017)

[34] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent
dirichlet allocation. Journal of machine
Learning research 3, 993–1022 (2003)

[35] Keith, K.A., Teichmann, C., O’Connor,
B., Meij, E.: Uncertainty over uncertainty:
Investigating the assumptions, annotations,
and text measurements of economic policy
uncertainty. In: Proceedings of the
Fourth Workshop on Natural Language
Processing and Computational Social Sci-
ence, pp. 116–131 (2020). https://doi.org/
10.18653/v1/P17. https://aclanthology.org/
2020.nlpcss-1.13.pdf

[36] Meng, Y., Huang, J., Wang, G., Zhang, C.,
Zhuang, H., Kaplan, L., Han, J.: Spheri-
cal text embedding. In: Advances in Neural
Information Processing Systems (2019)

[37] Sra, S.: Directional statistics in machine
learning: a brief review. Applied Directional
Statistics: Modern Methods and Case Studies
225, 6 (2018)

[38] Pewsey, A., García-Portugués, E.: Recent
advances in directional statistics. Test 30(1),
1–58 (2021)

[39] Reimers, N., Gurevych, I.: Sentence-bert:
Sentence embeddings using siamese bert-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (2019). https://aclanthology.org/D19-1410.pdf

[40] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.acl-main.703. https://aclanthology.org/2020.acl-main.703

[41] Callison-Burch, C., Ungar, L., Pavlick, E.: Crowdsourcing for nlp. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts, pp. 2–3 (2015)

[42] OECD: Organisation for Economic Co-operation and Development (OECD). http://data.oecd.org (2021). Accessed 2021-11-05

[43] Zalla, R.: Economic policy uncertainty in ireland. Atlantic Economic Journal 45(2), 269–271 (2017)

[44] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45. Association for Computational Linguistics, Online (2020). https://www.aclweb.org/anthology/2020.emnlp-demos.6

[45] Ratner, A., Bach, S.H., Ehrenberg, H., Fries, J., Wu, S., Ré, C.: Snorkel: Rapid training data creation with weak supervision. The VLDB Journal 29(2), 709–730 (2020)

[46] Chollet, F., et al.: Keras: The python deep learning library. Astrophysics Source Code Library, 1806 (2018)

[47] 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 12,
[48] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2015)

[49] van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research 9(86), 2579–2605 (2008)

[50] Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65 (1987). https://doi.org/10.1016/0377-0427(87)90125-7

[51] Reeves, E., Palcic, D.: Getting back on track: the expanded use of ppps in ireland since the global financial crisis. Policy Studies 38(4), 339–355 (2017)

[52] Rudd, K.: The global financial crisis. Monthly, The (Feb 2009), 20–29 (2009)

[53] Rice, J., et al.: Economic policy uncertainty in small open economies: a case study in ireland. Technical report, Central Bank of Ireland (2020). https://www.centralbank.ie/docs/default-source/publications/research-technical-papers/01rt20-economic-policy-uncertainty-in-small-open-econ-a-case-study-of-ireland-(rice).pdf?sfvrsn=4

[54] Mertens, K., Ravn, M.O.: Measuring the impact of fiscal policy in the face of anticipation: a structural var approach. The Economic Journal 120(544), 393–413 (2010)

[55] Zivot, E., Wang, J.: Vector autoregressive models for multivariate time series. Modeling Financial Time Series with S-Plus®, 385–429 (2006)

[56] Inoue, A., Kilian, L.: Inference on impulse response functions in structural var models. Journal of Econometrics 177(1), 1–13 (2013)

[57] Peters, J., Janzing, D., Schlkopf, B.: Elements of Causal Inference: Foundations and Learning Algorithms. The MIT Press, ?? (2017)

[58] Lin, J.-L.: Teaching notes on impulse response function and structural var. Institute of Economics, Academia Sinica, Department of Economics, National Chengchi University, 1–9 (2006)

[59] Mushtaq, R.: Augmented dickey fuller test. Econometrics: Mathematical Methods & Programming eJournal (2011)

[60] Hu, S.: Akaike information criterion. Center for Research in Scientific Computation 93

[61] Seth, A.K., Barrett, A.B., Barnett, L.:
Granger causality analysis in neuroscience and neuroimaging. Journal of Neuroscience 35(8), 3293–3297 (2015)