The Data Open 2019

Citadel

Economic Impact of Brexit on UK Cities and Allocation of Government Resources

Yao Lei Xu, Alexandros Haliassos, Hal Frigaard, Kriton Konstantinidis

1 EXECUTIVE SUMMARY

1.1 Motivation

Brexit has been the center of modern political and economic landscape since its vote in 2016, causing strong disruptions in the status-quo with profound consequences at an international and domestic level for the UK. Amongst the chaos and the uncertainty surrounding the topic, we tried to find clarity on how and where government should dedicate its resources from an economic point of view. Specifically, we aim to address the following question:

Based on UK cities' industry sector composition profiles, what has been the economic impact of Brexit on these cities and which high-risk areas should the government target?

A main reason for the choice of this specific question is its utility in guiding the UK government's decision-making. Specifically, insights gained through this analysis can allow the government to pursue region-specific spending and support based on the relative extent of negative impact on these regions. The ample data provided that links job listings to various regions ("job_listings" dataset) as well as to each sector and the corresponding market capitalisation ("lse_historical_data dataset") make this analysis possible. Finally, our intuitions were that (1) differences between sector market capitalisation before and after Brexit provide concise information on the effects of Brexit on each sector, and (2) regions can be clustered well based on sector composition. These intuitions further suggested that this would be a fruitful topic to explore.

1.2 Main Findings

Through our analysis, we discovered the following insights:

1. The economic impact of Brexit can be measured through our proposed metric (see section 2), which is in line with current Brexit passporting consequences.

2. Overall, the impact of Brexit has been negative across UK cities. However, large cities with diversified sector presence are less affected by Brexit than smaller ones, which tend to rely on fewer sectors.

3. UK cities with similar economic structure can be identified through clustering, which helps in identifying groups of cities where similar policies can be applied.

2 TECHNICAL EXPOSITION

In this section we discuss the technical details of our approach, including data wrangling and analysis (unsupervised learning), assumptions, decision-making process based on results at various stages, and, where appropriate, the conclusions that we drew as well as possible shortcomings.

2.1 Quantifying the Economic Impact

Firstly, we wanted to quantify the economic impact of Brexit broken down to a sector level. To this end, assuming that the financial markets have efficiently priced the Brexit event on the stock market, we decided to compute the percentage difference in market capitalization per company, using the median value of 1.5 years before and after the Brexit vote (equation 1). The median is a more robust metric of central tendency than the arithmetic average, which we deemed more appropriate given the highly volatile time series at hand. The equation is

$$s_{co} = \frac{V_{T^+} - V_{T^-}}{V_{T^-}}, \tag{1}$$

where V_{T^+} and V_{T^-} denote the market capitalization 1.5 years after and before the Brexit vote respectively.

Secondly, the impact scores computed per company were aggregated and averaged on a sector level basis. The averaging operation should cancel out the risks that are company specific, which should give a more reasonable view of the sector as a whole. Finally, to make the scores comparable, the sector specific impacts were scaled as z-scores, and we denote these as s_i for sector i; these are shown in Figure 1.

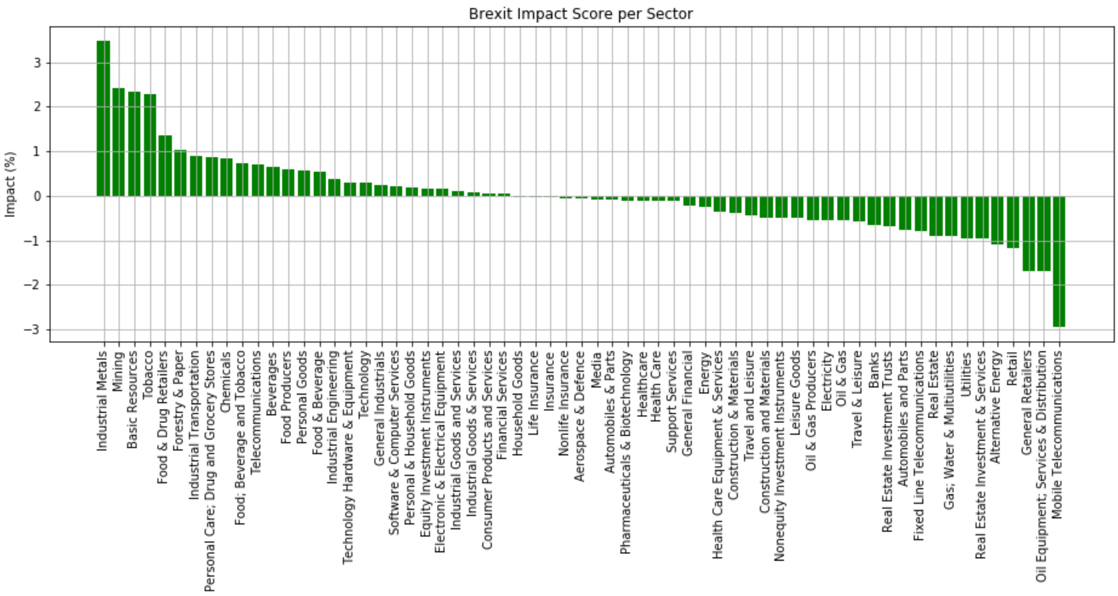

Figure 1: *Brexit economic impact z-score per sector*

The figure shows that the most negatively affected companies tend to be in the professional service sector (e.g., Travel & Leisure, Banking, etc). This is expected since the Brexit passporting issues create a strong barrier for these particular industries.

2.2 Infer Regional Economic Impact

We wanted to see the implications of our findings on a regional level. As such, we wanted to profile each UK city based on its sector-wise profile. This was done by looking at the job-listing dataset, which showed the employment opportunity per city offered by different companies. Using the LSE dataset, it is possible to breakdown the job opportunities in different cities on a per sector level by combining the ticker of the company offering a given job from the two datasets. We calculated the percentage of job opportunities from each sector on a per city level. This is shown in Figure 2.

city	Travel and Leisure	Personal Care; Drug and Grocery Stores	General Financial	General Retailers	Consumer Products and Services	Banks	Industrial Goods and Services	Financial Services	Food; Beverage and Tobacco	Construction and Materials	...	Technology Hardware & Equipment	Industrial Metals	Industrial Goods & Services
London	0.001056	0.203369	0.002160	0.0	0.000192	0.005280	0.042575	0.00048	0.000096	0.016463	...	0.011040	0.000048	0.057502
Manchester	0.002939	0.134039	0.000882	0.0	0.000000	0.009700	0.047619	0.00000	0.002939	0.000000	...	0.000000	0.000294	0.050852
Leeds	0.000000	0.105644	0.000000	0.0	0.000000	0.002894	0.069465	0.00000	0.000000	0.000000	...	0.000482	0.000482	0.060781

Figure 2: *Relative frequency of sector presence per city for three sample cities.*

We then computed the weighted average of Brexit impact scores per city. To calculate this, we used a linear combination of the sector percentages (weights) and the sector-specific economic impact scores for each city:

$$I_c = \sum_{i=0}^{N-1} w_{i,c} s_i$$

where I_c denotes the (weighted) averaged impact score for each city, $w_{i,c}$ the weights for each city and each sector, and s_i the impact score per sector.

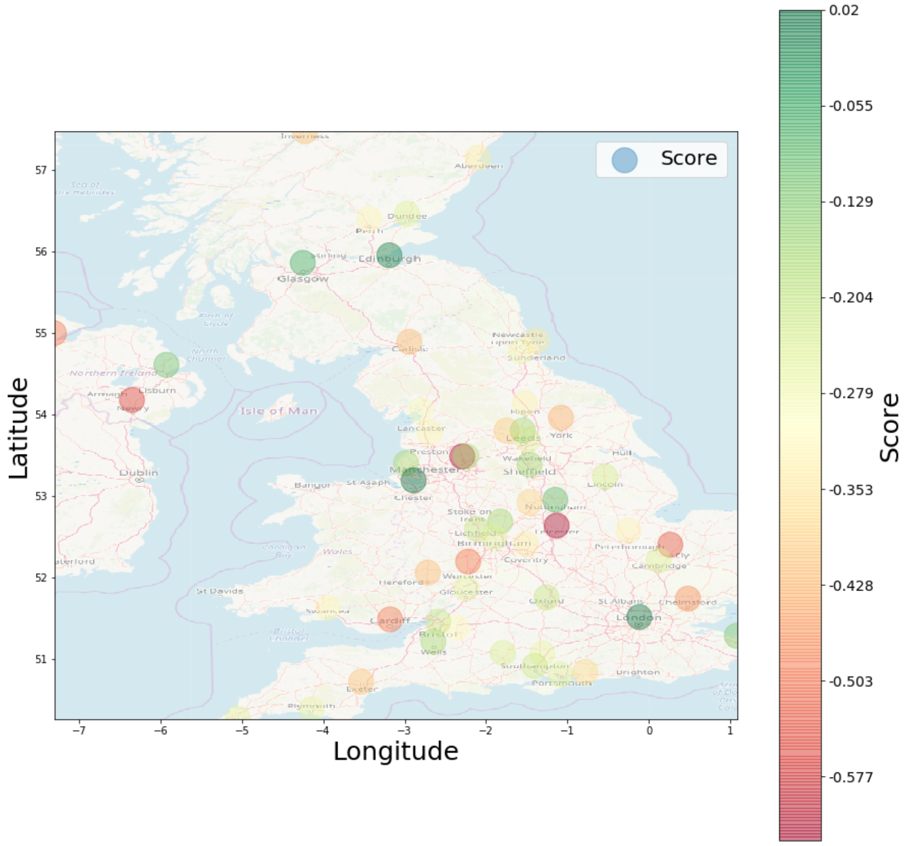

Figure 3: *Impact scores for major cities.*

As a result, we were able to identify which cities are most impacted by the Brexit event. This is shown in Figure 3 for some UK cities, where the colors are proportional to the score. Overall, the scores are negative across most of the cities, with the worst affected ones being Letterkenny, Finchampstead, Coleman Street, Hove and Chilwell, which are small cities that tend to be concentrated on few sectors. Larger cities such as London and Edinburgh are less affected since they are more diversified in terms of sector presences.

2.3 Clustering

Due to the large number of cities and the associated difficulty for the government to provide tailored funding to each of these cities, we decided to cluster them based on their feature vectors. The vectors consisted of city-specific sector percentages of job listings. The idea is that the government could instead draw up plans that would correspond to every city in each cluster and therefore would significantly simplify their task, as the number of clusters was expected to be much smaller than the number of cities. Of course, the assumption embedded in this argument is that cities can be effectively grouped in clusters for which the intra-variance is small and inter-variance is large.

2.3.1 Initial Approach: K-Means

For clustering we first considered the K-Means algorithm and found the silhouette scores[1] for each number of clusters $k \in [3, 20]$. To visualise the clusters, we projected the data points to 2-D using the t-SNE algorithm, and found that the points were not optimally clustered. We suspected that this is due to the fact that K-means assumes clusters of similar shapes and densities, as well as with spherical shapes.

2.3.2 Better Approach: Gaussian Mixtures

In order to generate better clusters, we applied a Gaussian mixture model, which is a probabilistic model that assumes that the data points come from a mixture of a given number of Gaussian distributions. Despite the somewhat unfavourable computational complexity of the algorithm, this model was preferred over K-Means due to the non-prohibitive size of the dataset as well as the model's ability to model non-spherical clusters.

Since silhouette scores are only informative when the clusters are spherical, we used two theoretical information criteria to choose the optimal number of clusters: Bayesian information criterion (BIC) and Akaike information criterion (AIC). The corresponding equations are as follows:

$$\begin{aligned} \text{BIC} &= p \log(N) - 2 \log(l_{\max}) \\ \text{AIC} &= 2p - 2 \log(l_{\max}), \end{aligned}$$

where p is the number of parameters in the model, N is the number of instances (cities), and $l_{\max}$ is the maximum value of the likelihood function of the model (note that the model finds the parameters that maximise its log-likelihood function). These information criteria give low (favourable) scores to models with few clusters and which fit the data well. Figure 4 shows the BIC and AIC values as functions of the number of clusters.

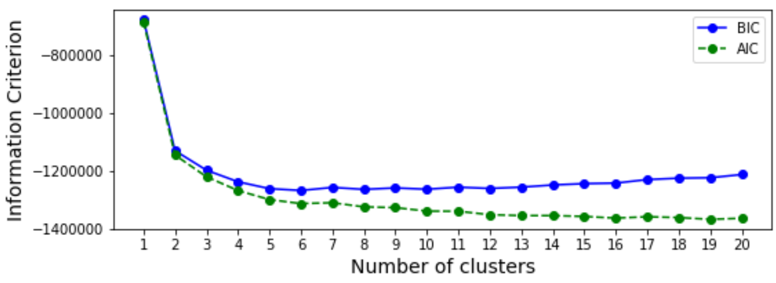

Figure 4: *BIC and AIC information criteria for optimal cluster number selection*

[1]The silhouette score is the average silhouette coefficient over the instances. The silhouette coefficient is a measure of the distance of a given instance to its neighbouring clusters.

We see that the number of clusters that minimises BIC is six, whereas AIC seems to keep decreasing beyond six clusters. The reason for this discrepancy is that BIC penalises more heavily the number of clusters than AIC. Since the goal was to keep the number of clusters low to allow tailored targeting of a manageable number of different groups of cities, we decided to set the number of clusters to six.

Given the high dimensionality of the feature vectors used (larger than 2 or 3 dimensions making it hard to visualise), we applied t-SNE (a state-of-the-art dimensionality reduction algorithm) for visualisation purposes (see Figure 5). We see that clusters 2 and 4 are for the most part separated from the rest of the clusters, which suggests that the corresponding cities would benefit from being targeted as part of their respective clusters rather than separately. The fact that there seems to be overlap between the rest of the clusters can be interpreted in at least two ways: (1) the overlap exists because of the dimensionality reduction and in fact the points are well clustered in the original high dimensional space; or (2) the feature vector is not "expressive" enough to cluster the cities well because there are other factors that distinguish the regions.

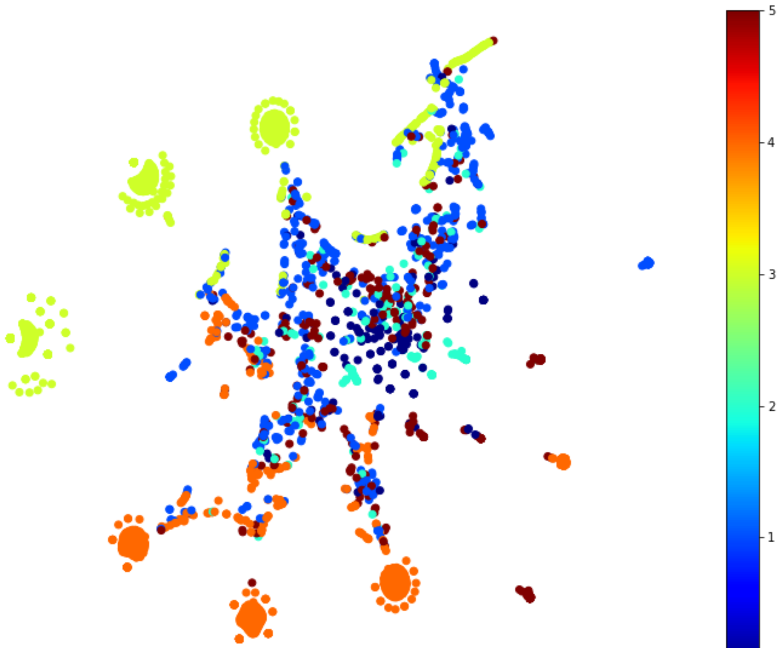

Figure 5: *t-SNE on the instances (cities) where the colour of each point corresponds to the cluster it was assigned to for the Gaussian Mixtures model.*

2.3.3. Anomaly Detection

The rationale for collectively targeting cities in each cluster becomes tenuous for cities that deviate substantially from the centroids of the clusters. To address this issue, anomaly detection was performed, where the instances residing in low-density regions were taken to be the outliers. Outlier cities (such as Basingstoke, Bradford, Cork) should be targeted separately based on further analysis.

2.3.4. Clustering: Conclusions

Based on this analysis we suggest that the safest option for the government would be to target the cities in clusters 2 and 4, and perform further analysis to better distinguish the cities in the rest of the clusters as well as the outlier cities. Figure 6 shows a sample of the regions (downsampled from 2764 towns to 85 for clarity of presentation), superimposed on UK's map.

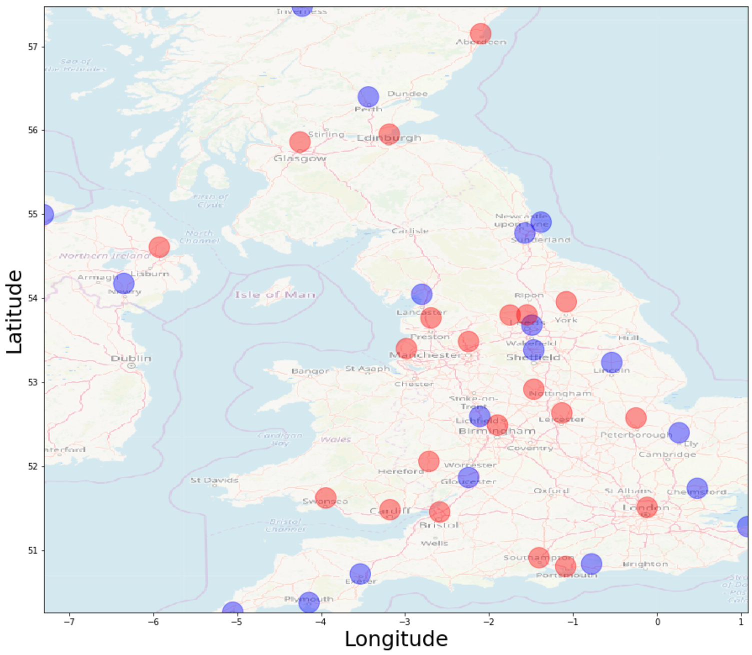

Figure 6: *Cities (sampled for clarity of presentation) with colour corresponding to cluster 2 (blue) and cluster 4 (red).*

3. CONCLUSION & FURTHER INVESTIGATION

In conclusion, we used the percentage of each sector in a large set of UK cities and the post-Brexit vote fluctuations in the market capitalization across sectors to construct city-specific Brexit impact scores.

We found that cities that rely on many different industrial sectors are characterized by a low Brexit impact score and thus their economy is less prone to be heavily affected by Brexit. On the other hand, areas heavily dependent on a small number of sectors are at higher risk. Clustering of cities based on their sector percentage profiles can help the government prepare collective plans for areas with similar Brexit impact scores. We believe that the practical aspect of being able to apply financial strategy plans is crucial for the feasibility of the proposed approach.

Finally, we believe it would be very insightful to perform time-series analysis on the temporal evolution of available job listings in order to predict how the sector profiles are expected to evolve. Combining that with predictions on the market capitalisation across sector could be useful to calculate future city impact scores. We expect that such a procedure can provide the government with further insights on the local sensitivity of different regions and help navigate its financial strategy.