Visual analysis of patterns and trends in UK unemployment

Dmitry Borisovskiy

Abstract — It is argued that due to the structural changes in UK economy the patterns and trends in unemployment are changing. Using a time series visualization approach this paper investigates the changes in UK unemployment with particular attention to variation in unemployment rate between different industries and gender factor in unemployment. Taking into consideration economic crisis 2007-2010 the analysis concentrates on two components of unemployment time series: seasonality and trends. It was found that some industries are very similar in their seasonal unemployment which was then supported by time series correlation analysis. Even though there was no difference found between the industries in their long-term trends to decrease in unemployment the results cannot be conclusive due to the dramatic effect of economic crisis on unemployment. Also, it was found that male and female unemployment rates have comparable patterns and in the recent years levels of unemployment rate for both genders became especially similar.

1. Problem Statement

Technology change, production relocation and rise of the service economy in the last 25 years have led to a significant sectoral shift and change in structure of employment of most of the developed western countries (Basile et al, 2012). According to Robson (2006) a good example of the sectoral changes is the rise of unemployment in traditional manufacturing industry during a record low unemployment is the service industry.

At the same time Kooman, Ooms and Hindrayanto (2009) argue that unemployment is highly affected by the seasonal factor and rise or drop in unemployment rate might be just seasonally of the specific industry and not a long-term trend. Also, it is argued that unemployment rate significantly differs by gender and that in many countries male unemployment is lower that female unemployment (Azmat, Guell and Manning, 2006).

Using unemployment rates by industry data this paper aims to run a longitudinal analysis of seasonality and trends in UK unemployment between years 1995-2018 and answer the following questions:

- What is happening with UK overall unemployment rate and how does it differ between the various industries?
- Can the seasonal factor be removed to see what are the long-term trends in unemployment for each industry? Is it true that some of the industries are dying in UK?
- Are there any similarities between the industries in terms of their seasonal unemployment?
- To what extent male unemployment differs from female unemployment in UK?

2. State of the Art

Time series approach is widely used to analyse, explore and predict unemployment rate. The are many studies and research papers that are concentrating on time series as a method to forecast an unemployment rate or to establish the relationship of unemployment with other factors such as crime rate or health (Joyce, 1989).

Kooman, Ooms and Hindrayanto (2009) concentrate on unobserved components (UC) time series models which are able to split time series in the three components: trend, season and cycle. The scientists are offering a range of techniques that help to decompose time series in to seasonal and long-term trends. It is argued that those models can be easily implemented to time series analysis and forecasting with seasonal adjustment.

According to Kooman, Ooms and Hindrayanto (2009) unemployment rate is a highly seasonal and periodic indicator which is also depend on the business cycle and therefore traditional time series might not be always effective in examining changes in unemployment rate. They adopted UC models to examine a post war US unemployment data with attention to seasonality and cycles. This study gives a good starting point of how to approach unemployment time series. In order to achieve a deeper level of analysis the error component of time series needs to be removed so seasonal and trend components can be investigated separately.

However the study by Kooman, Ooms and Hindrayanto (2009) concentrates more on the technical side of time series decomposition rather than on visual exploration and comparison of the decomposed time series. They are using two-dimensional graphs to visualize the aggregate unemployment level from all the industries which is a simple but a very effective way to visualise time series. In our study we have unemployment rates for 14 different industries and since we aim to investigate the interconnection of the industries with focus on comparison of how change in unemployment rate in one industry is correlated with unemployment rate in the other industry we need a more powerful visual approach.

Kothur et al (2015) argue that a correlation-based comparison approach is very useful for comparing time series. The scientists are offering the approach that is based on windowed cross-correlation (WCC) model strengthened by visual analytics techniques. The computational part is done by WCC which calculates the correlation values of a different combination of multiple time series. The output is then passed through an interactive visualisation which helps the users to explore and interpret the results so they can modify the original input parameters if necessary (Kothur et al, 2015). The offered model has been applied to real world time series data and has been found very effective in calculating and demonstrating the interdependence of the complex time series which is hard to be observed otherwise. Since we aim to compare industries by their unemployment, the elements of the offered approach can be used in our study. If we manage to decompose our time series then we can compare time series components using correlation method and visualise the results for the further analysis.

3. Properties of the Data

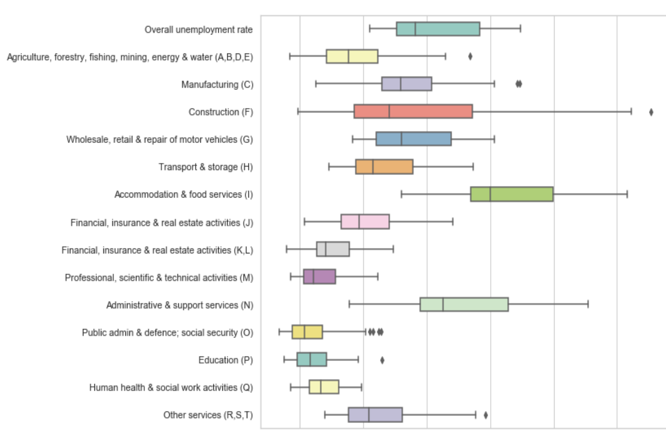
The dataset that is used in this paper is taken from Office For National Statistics (https://www.ons.gov.uk/) and called 'Unemployment by previous industrial sector'. The estimates used in the dataset are sourced from the Labour Force Survey.

The unemployment rates are calculated based on people who are being unemployed by industry of the last job as a percentage of those who are currently in employment in the same industry. The period covered is from year 1995 until year 2018 and include figures for February, May, August and November of each year. The dataset provides figures for overall unemployment rate and unemployment rate for 14 different industries which are classified according to Standard Industrial Classification SIC 2007. The industries are: Agriculture, forestry, fishing, mining, energy & water (A,B,D,E), Manufacturing (C), Construction (F), Wholesale, retail and repair of motor vehicles (G), Transport & storage (H), Accommodation & food services (I), Financial, insurance & real estate activities (J), Financial, insurance & real estate activities (K,L), Professional, scientific & technical activities (M), Administrative & support services (N), Public admin & defence, social security (O), Education (P), Human health & social work activities (Q), Other services (R,S,T). Later in the analysis we might use the names of the categories (capital letter) when discussing specific industries. Also the dataset provide a separate unemployment figures (by industry) for men and women.

All the data cleaning and data wrangling processes have been done in python notebook. Overall the original dataset was clean without any missing values, a part of those that were result of loading dataset in to python notebook but which were removed at the data cleaning stage. One of the initial problems with data was that the date figures were in the format not suitable for time series analysis so the python function 'pd.date_range' has been used to convert the values in the appropriate date format.

To check the properties of the data it was decided to use boxplot (figure 1) as one of the best methods to explore and visualize the distributional characteristics of time series data. Figure 1 shows that construction industry has the biggest spread out in values, which means a substantial variation in unemployment rate over the years. This is an interesting observation that would need to be further explored in our analysis. We also can see that the highest median unemployment is in accommodation and food services industry and lowest median unemployment is in public admin & defence, social security industry. Generally, we can observe that majority of the industries are slightly right-skewed which might mean that most of the time unemployment rate was low but there were times of a rapid increase in unemployment. This anomaly also needs to be further investigated. There are a few possible outliers shown by the boxplot but taking in to account the big size of the data, they should not affect our analysis.

Figure 1: Distributional characteristics of the data



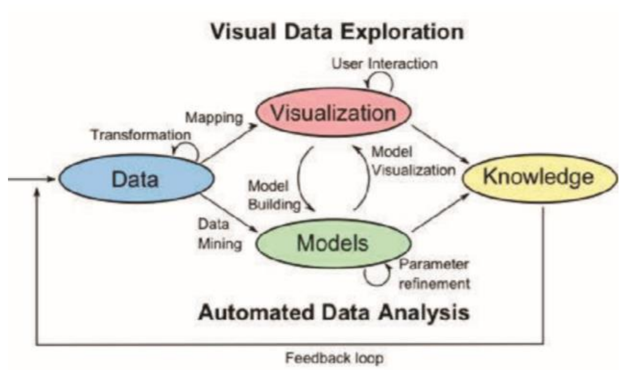
4. Analysis

4.1. Approach

Not long time ago it was widely accepted that due to the limited ability of humans to deal with uncertainty and complexity of a modern data, computer technologies would be doing majority of tasks or even replace humans in making decisions. However soon it became clear that to deal with an exponentially growing amount of complex data a far more effective way would be to use computers to do complex calculations and leave decision making to human reasoning. (Reani, Peek and Jay, 2019). Visual Analytics is known as well recognised approach that combines machine computational abilities and power of human thinking, creativity and decision making (Keim and Zhang, 2011).

According to Gogolou et al (2019) finding patterns and comparing complex time series are extremely challenging tasks if no proper visualization approach is used. In order to be able to achieve the goals of our analysis we are going to employ a famous visual analytics approach (figure 2) described in (Keim and Zhang, 2011).

Figure 2: Visual Analytics Process (Keim and Zhang, 2011).



With reference to the diagram above the following steps will be taken:

- Loading the data and transforming it to a format appropriate for our study. Since we are working with a longitudinal data, we will use a time series format for our analysis.
- We will plot the data to have some visual representation of it to get some sense of the

data we are working with and possibly identify some issues or other areas of interest that we were not aware of at the planning stage.

- If the initial visualisation shows us that something is wrong with original data or if there is an important additional point that needs to be investigated we can go back to the data to do a further transformations, explorations and computations. The results will be then visualised so we can try to extract a possible new knowledge that can add to our understanding of unemployment rate in UK.
- We are planning to decompose the time series to show any seasonal or long-term trends to understand what is going on with unemployment in UK in the short and long runs. If the visualization demonstrates anything valuable, then we will try to go deeper in our analysis and model the interconnection in unemployment rates between different industries.

Please note that the above steps do not have any specific order and can be performed at different points of the study depending on the outcome of each part of the analysis.

4.2. Process

Once data cleaning and data transformation activities are done we are going to plot the overall unemployment rate and unemployment rate by industry (figure 3) to get some understanding of the time series we are working with. Figure 3 shows that unemployment time series are complex with a lot of ups and downs in unemployment rate, so it is difficult to spot any trends or patterns just by looking at this time series graph. We can observe that for majority of industries there was a rapid increase in unemployment approximately between years 2008 and 2010/2011. This specific period has been annotated to bring attention to this event so we can have a closer look at it at the later stage of analysis.

Figure 3: Overall Unemployment Rate and Unemployment Rate by Industry

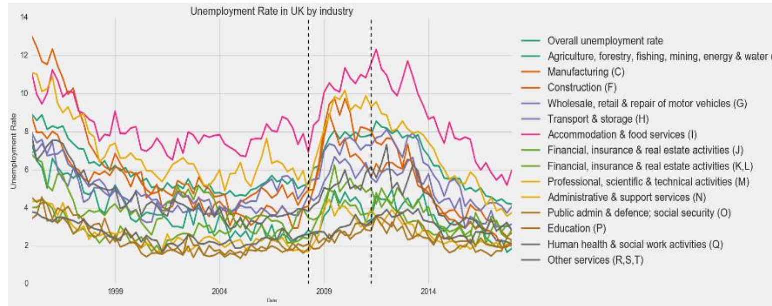
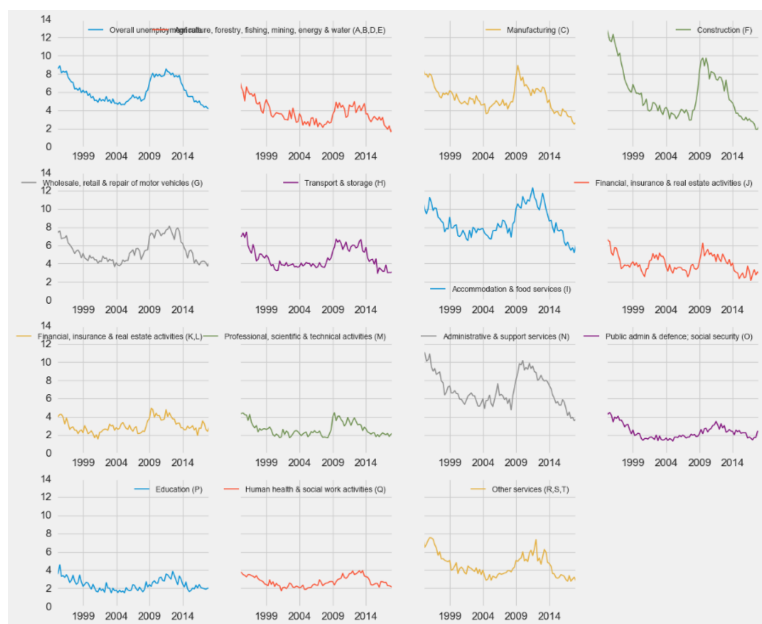


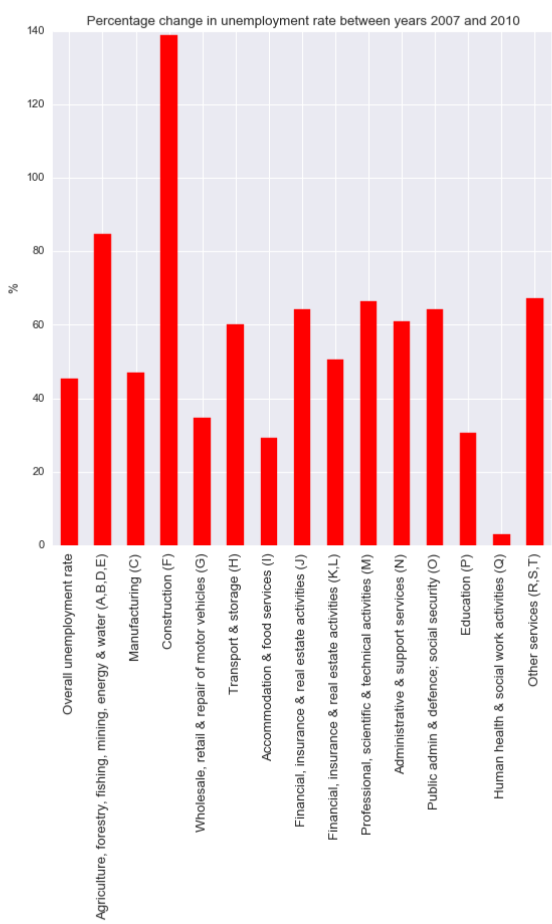
Figure 3 looks very busy and complex, so it is hard to see the changes in unemployment by industry. To resolve this issue, we are going to separate the above time series and build simultaneously unemployment rate by industry on the separate subplots. Figure 4 shows a much clearer picture of changes in unemployment rate for each industry. Figure 4 has confirmed our previous observation that there was a significant increase in unemployment around year 2008. Having done some research it was discovered that between 2007 and 2010 there was a worst economic crisis in 50 years, which caused a massive increase in unemployment rate all over the world (Mertens and Beblo, 2016).

Figure 4: Unemployment Rate by Industry



Even though it was not part of our initial plan, now once we can see that economic crisis had a dramatic effect on unemployment in UK, we should investigate to what extent each industry was affected by it. Let's find out what was the percentage change in unemployment between February 2007 and February 2010 – the period of time when unemployment rate has dramatically risen as per our previous figures. Figure 5 demonstrates that construction industry has been worst affected by crisis, the unemployment rate has risen by 140 %. This supports a well known fact that construction is a very volatile industry which has a significant dependence on economic conditions. Categories A,B,D,E were the other industries that were significantly affected by crisis. It is seen that historically education sector and human health & social activities sector have the lowest unemployment rate and also these industries were the least affected by crisis.

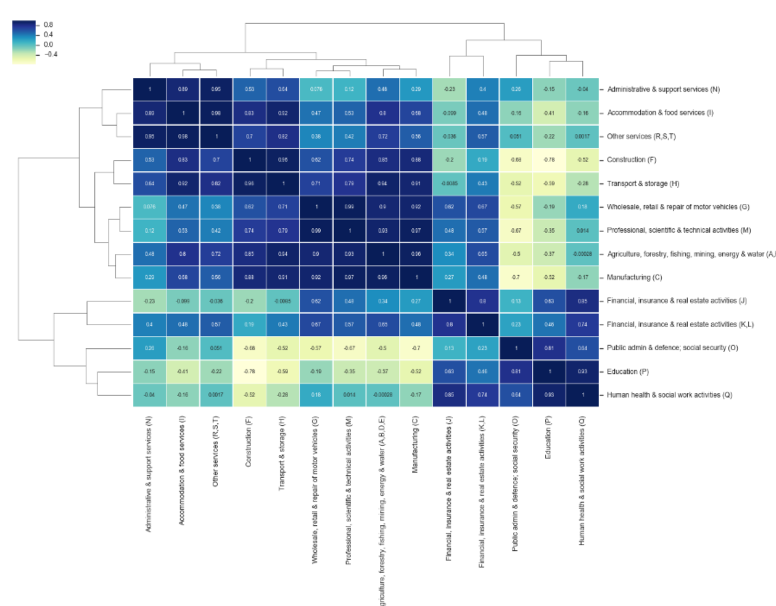
Figure 5: Percentage Change in Unemployment Rate between Years 2007 and 2010



As discussed in the introduction part due to sectoral shifts and movement of production to eastern countries we were expecting to see a long-term rise in unemployment in industries such as manufacturing or agriculture but figure 4 showed us that the unemployment rate for those industries is not rising but actually dropping. However, our initial observation might be misleading as according to Theodosiou (2011) time series consist of the three main components: seasonality, trend and noise (error) so what we observe on figure 4 might be just an element of seasonality. As previously mentioned unemployment is a highly seasonal indicator and therefore to do a more detailed analysis, we need to decompose time series in to separate components and analyse each component on its own (Kooman, Ooms and Hindrayanto, 2009).

We will use a decomposition function in python and extract seasonal component of unemployment time series first. Visualising seasonality does not give any benefits in terms of achieving goals of this paper as it looks very messy and difficult for human eye to extract any valuable information out of it. However to compare the industries in terms of seasonality, with reference to the approach offered by Kothur et al (2015) described earlier, we can find Pearson correlation coefficient between seasonal unemployment components of each industry over a whole period of time and visualise it using cluster map (figure 6).

Figure 6: Seasonal Correlation Matrix (Cluster Map)

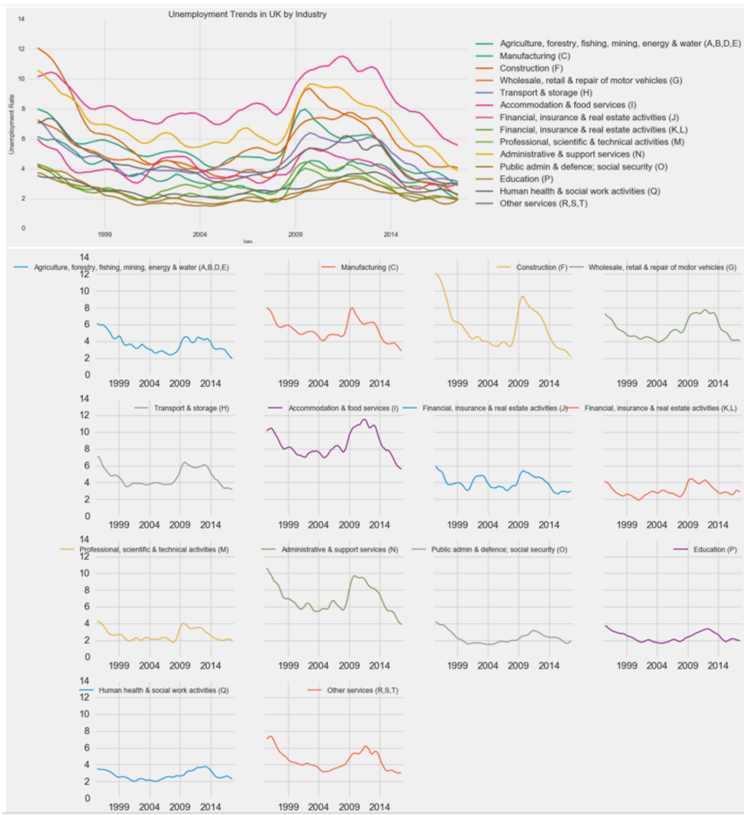


As can be seen on the cluster map the industries are grouped together by the strength of the correlation, the darker the colour the stronger the correlation between the industries seasonal unemployment, which helps us to judge about seasonality of each industry. It is interesting that seasonal unemployment rate in manufacturing industry is strongly correlated with seasonal unemployment in transport and storage industry. One possible interpretation of that is that during high production season, more transport and storage facilities are required and as the result unemployment rate drops in both industries. The same explanation can be given to a strong correlation between group of industries with agriculture and group of industries with wholesale and retail – during summer season more workers are required in agriculture sector and at the same time more employees are needed in wholesale and retail for agriculture goods to be sold. Also we can see that industries such as education or health have very weak correlation in seasonality with industries such as manufacturing or food services. This makes sense as those sectors in real life are very different with very little interconnections.

Now we are going to extract the trend component of time series. Figure 7 demonstrates the unemployment trends all together and at the same the separate subplots for each industry, which makes it easier to compare and analyse long-term unemployment trends. We can observe that for almost whole period of time group of categories A,B,D and E (Agriculture etc) have the highest level of unemployment, even during the crisis time. However, after year 2010 we can see a strong trend to decrease in unemployment. Surprisingly we can observe a similar trend to decrease in unemployment in manufacturing industry. To check this observation, we build a similar correlation cluster map as we did for seasonal components. We found a strong correlation between the two sectors' trends, which supports our visual understanding of graph 7. This goes against our original expectation that unemployment rate in manufacturing will be rising.

We were also expecting to see a long-term trend in decrease in unemployment rate for a different kinds of service industry such as categories K,L and N, however this is not observed. Correlation cluster map also did not show that trends in service sectors are much different to trends in any other sectors. Generally, we can see that among all industries there is a steady trend to decrease in unemployment rate. This explains such a strong decrease in overall unemployment rate that we could observe on figures 3 and 4.

Figure 7: Unemployment Trends in UK by Industry



As mentioned in the introduction historically women's unemployment rate was much higher than men's unemployment rate. (Azmat, Guell and Manning, 2006). Also according to Komerade and Richardson (2018) the difference in male/female unemployment can become especially obvious during the economic downturn as women's jobs are more vulnerable to crisis. However recently there were a number of studies that showed that in western countries there is no difference in unemployment rate between men and women and both indicators are interlinked with each other. So it would be interesting to compare male/female unemployment rates in UK, taking into account that in previous part of analysis we identified a big increase in unemployment due to economic crisis. Let's plot overall unemployment rate for men and women to find out if we can observe any obvious differences (figure 8).

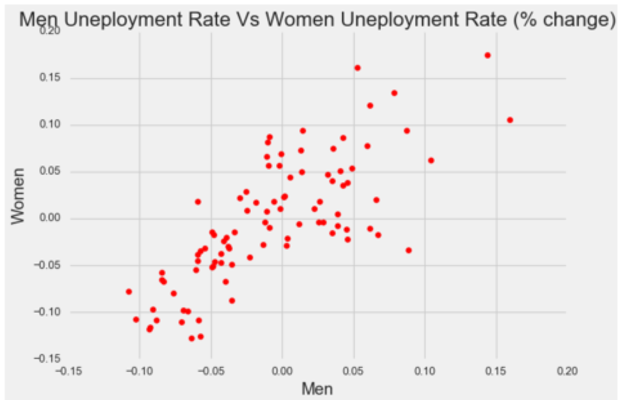
Figure 8: Overall Unemployment Rate for Men and Women

Surprisingly, we can see that over the whole period of observation the unemployment rate for men was higher than for women. Furthermore, during the crisis the rise in unemployment rate was more significant for men than for women as can be clearly seen on figure 8. Another interesting observation is that after recession both levels of unemployment became very similar to each other which possibly can be explained by the fact that in recent years gender become a less important factor on the labour market.

Figure 8 shows that unemployment rate moves very similar for both genders so it might be interesting to investigate the relationship between men and women unemployment rates and test to what extent it is correlated with each other. Instead of using nominal values we will find how percentage change in men's unemployment rate is correlated with percentage change in women's unemployment rate, which might be a more realistic representation of the true picture. Also we will build a simple linear regression model to test the relationship between unemployment of different genders. To build a regression model we will use percentage change in female unemployment as independent variable and percentage change in male unemployment as dependent variable.

Figure 9 shows a positive correlation between the two unemployment rates, which is supported by the regression model results. The R-squared value (0.588) tells us that almost 59 % of variation in male unemployment rate can be explained by female unemployment. This is a fascinating outcome that supports our initial impression that male and female unemployment rates behave in a similar way. It might be interpreted as indicator of gender equality on UK labour market.

Figure 9: Men Unemployment Rate Vs Women Unemployment Rate (% change)




```
                            OLS Regression Results
==============================================================================
Dep. Variable:     men_unemp_pct_change   R-squared:                       0.588
Model:                              OLS   Adj. R-squared:                  0.584
Method:                   Least Squares   F-statistic:                     128.5
Date:                  Thu, 19 Dec 2019   Prob (F-statistic):           4.98e-19
Time:                          10:38:28   Log-Likelihood:                 177.10
No. Observations:                    92   AIC:                            -350.2
Df Residuals:                        90   BIC:                            -345.2
Df Model:                             1
Covariance Type:              nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                   -0.0059      0.004     -1.576      0.118      -0.013       0.002
women_unemp_pct_change   0.6407      0.057     11.338      0.000       0.528       0.753
==============================================================================
Omnibus:                       11.946   Durbin-Watson:                   1.930
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               12.393
Skew:                           0.831   Prob(JB):                      0.00204
Kurtosis:                       3.686   Cond. No.                         15.2
==============================================================================
```

4.3. Results

This paper has investigated the different aspects of unemployment rate in UK between years 1995 and 2018. It was found that after a dramatic increase in overall unemployment rate due to economic crisis in 2007 there was a steady trend to decrease in unemployment. It was also found that even though some industries (e.g. construction) were hit by economic crisis more than other sectors, we can observe a clear tendency to decrease in unemployment rate among the all industries. There was not much difference found in the long-term unemployment trends between the industries and therefore basing on that indicator we cannot conclude that some industries (e.g. manufacturing) are dying in UK as initially expected.

We managed to extract the seasonal factor from unemployment rate of different industries and found that some sectors are very similar (e.g. manufacturing and transport) and some sectors are very different (e.g. construction and education) in terms of their seasonality. Also, it was discovered that levels of unemployment rate for men and women became more similar closer to the end of the observed time series and both unemployment rates are interconnected with each other. This shows that UK labour market is moving towards gender equality.

5. Critical Reflection

We used a time series visualisation approach to do our analysis. Time series is a powerful approach that enables users to investigate a complex longitudinal data. However, without appropriate visualisation techniques time series can be confusing and misleading. We could see that just putting all time series on one plot can make extremely difficult for human reasoning to extract the required information.

So it is crucially important to visualise time series in a correct way using techniques such as annotation of the important events, simultaneous subplots and time series decomposition. The use of those techniques has helped us to understand the important issues associated with unemployment time series and navigated us throughout the study. For example, we were able to identify that there was economic crisis during the period we were observing, which had a massive effect on unemployment rate. Therefore, we had to amend our plan of the study and give attention to the effect of crisis on unemployment. Also, separation and then visualisation of time series in the form of seasonal and trend components have triggered the idea that some sectors might be similar in their seasonality and long-term trends, which was then proved by computing of a cluster correlation map.

One of the issues with visualising time series is that in order to extract some valuable information and important trends you need to have a long observation period. The observation time used in our study was 23 years (1995-2018), which originally seemed to be long enough to analyse long-term changes in UK unemployment. However, since right in the middle of observation period there was a long and deep economic crisis, the observed time series were split in to two parts: before economic crisis and after economic crisis. This made it extremely difficult to identify long-term trends in unemployment rate by industry. It would be useful to analyse a much longer period to see how unemployment rate behaves over 50 years period.

The other issue that became obvious during the analysis part of the study is the way the industries were categorised in the used dataset. For example the unemployment rate for industries agriculture, forestry, fishing, mining, energy & water (A,B,D,E) was given in the dataset as one joined unemployment rate whereas seasonality for agriculture might be different from seasonality in mining. So we couldn’t fully investigate a seasonal correlation of just agriculture sector with some other sectors. In the future studies in order to have more representative findings it is importation to make sure that dataset provides separate unemployment rate figures for each specific industry.

Table of word counts

Problem statement	239
State of the art	501
Properties of the data	478
Analysis: Approach	415
Analysis: Process	1500
Analysis: Results	199
Critical reflection	428

References:

- Azmat, G., Guell, M. and Manning, A. (2006) 'Gender Gaps in Unemployment Rates in OECD Countries', Journal of Labor Economics, 2006, vol. 24, no. 1.
- Basile, R., Girardi, A., Mantuano, M. and Pastore, F. (2012) 'Sectoral shifts, diversification and regional unemployment: evidence from local labour systems in Italy', Science and Business Media, New York.
- Datacamp – Learn Data Science Online, Available at https://www.datacamp.com/
- Gogolou, A., Tsandilas, T., Palpanas, T. and Bezerianos, A. (2019) 'Comparing Similarity Perception in Time Series Visualizations', ieee transactions on visualization and computer graphics, vol. 25, no. 1.
- Joyce, T. (1989) 'A Time-Series Analysis Of Unemployment And Health', journal of Health Economics 8 (1989) 419-436.
- Kamerāde, D. and Richardson, D. (2018) 'Gender segregation, underemployment and subjective well-being in the UK labour market', human relations 2018, Vol. 71(2) 285 –309.
- Keim, D. and Zhang, L. (2011) ' Solving Problems with Visual Analytics: Challenges and Applications', i-KNOW '11 Proceedings of the 11th International Conference on Knowledg
- e Management and Knowledge Technologies.
- Koopman, S.J., Ooms, M. and Hindrayanto, I. (2009) ' Periodic Unobserved Cycles in Seasonal Time Series with an Application to US Unemployment', Oxford bullet in of economics and statistics, 71, 5.
- Mertens, A. and Beblo, M. (2016) 'Self-Reported Satisfaction and the Economic Crisis of 2007–2010: Or How People in the UK and Germany Perceive a Severe Cyclical Downturn', Soc Indic Res (2016) 125:537–565.
- Office For National Statistics, Available at https://www.ons.gov.uk/.
- Reani, M., Peek, N. and Jay. (2019) 'How different visualizations affect human reasoning about uncertainty: An analysis of visual behaviour', Computers in Human Behavior 92 (2019) 55–64.
- Robson, M. (2006) 'Sectoral Shifts, Employment Specialization and the Efficiency of Matching: An Analysis Using UK Regional Data' , Regional Studies, Vol. 40.7, pp. 743–754.