2019

Forecasting Job Market in UK

GROUP A3:

Abhimanyu	61910668
Aseem Garg	61910250
Jagannath Panigrahi	61910981
Nagesh Singh	61910932
Sai Nikhit Grandhi	61910282
Srishti Saxena	61910904

EXECUTIVE SUMMARY

Problem Statement

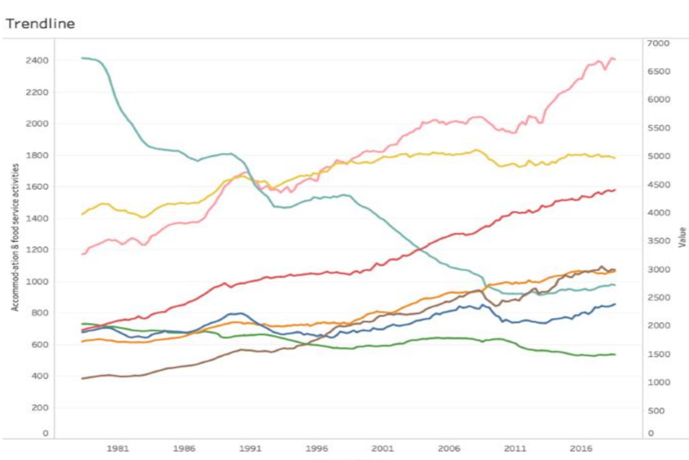
To determine and forecast the job growth and workforce in various industry in the United Kingdom. This will help the United Kingdom Government and the Education industry in the following ways:

- It will help in determining the training and course programs in the various industry sector based on the job growth
- Education sector in UK is a source of investment with investments of £33 billion annually through tuitions fees and grants and the government spends £87 billion. The forecast of labor and workforce in various industry will help the government in better allocation of investments, resource allocation.
- It will help the Education ministry to update, introduce new programs based on the prediction of job growth and workforce requirements

Data Collection

To forecast job in different industries in UK we took data from - **Office of National Statistics (GB)**, which contained time-series data of workforce employed in each sector.

The data source has quarterly data from 1978 of jobs in sectors such as construction, healthcare, manufacturing, education, retail and public administration.



If we visually analyze the data, we can see that there is some sectors where job has grown and others where there is a decline in the number of jobs. Further, analysis also shows that there is some seasonality in the quarterly data. Based on the analysis in visualization software we can say there is both **level and additive seasonality** in the data.

Final Forecasting Method:

From the exploratory analysis we determined that there is trend and seasonality in the data. But when looked upon closely, we can determine that the seasonality component is not present to the same extents across all the sectors. While Education sector shows the maximum seasonality, Public Admin and Defense sector shows the least seasonality present.

Therefore, since trend is present across all sectors and seasonality is present in varying degree across all sectors, we have mostly employed Double Exponential and Holt Winter models to perform the final forecasting. The double exponential model captures the trend (but not seasonality) present in the data. And the Holt Winter model captures both the trend and seasonality for a given time series.

After applying Double Exponential model and Holt Winter model to all the sectors, we realized that Holt Winter gave a more accurate prediction compared to that of Double Exponential.

Since Holt Winter model gave the most accurate results, we can observe in the plot that the forecasted data fit very well with the actual data.

Thus, this model was reliably used to predict the numbers of jobs expected to be created across all the sectors for the next 5 years.

Error Metrics

To measure the performance of the 2 employed models, we primarily relied on the MAPE metric. It is also to be noted that for all the sectors there is no stark difference in the MAPE value of training data and that of validation data. This confirms that the forecasting model is not subject to the issue of overfitting.

Table 1

Sector	MAPE (Naive model)	MAPE (Training Data)	MAPE (Validation Data)
Human Health	3.80	0.58	0.62

Education	2.86	0.61	0.99
Manufacturing	2.35	0.54	2.06
Construction	2.70	0.92	5.15
Wholesale & Retail	2.50	0.53	3.62
Public Administration & Defense	1.36	0.44	1.19

Conclusion & Recommendation

- There is an increasing trend in job growth in the construction and human health sector
- Public administration and manufacturing sector has a stagnant growth in the workforce.
- There is a steady decline in the number of jobs being generated in whole sale sector for the next 5 years.

Based on these findings we recommend the UK government to increase their spending in construction sector and human healthcare sector. Also, the education ministry should increase seats in human health & social work-related activities and also increase intake in construction related courses based on the projection of the forecast for the next 5 years in these sectors.

TECHNICAL SUMMARY

Data Preparation

The data had no missing values or outliers, and hence did not require any specific treatment. Since we required workforce numbers directly, no additional calculated fields were made.

The original dataset had the data for almost 20 industries. For the same of the project, we picked up 6 out of them. The industries which offered a higher number of jobs and/or growing at a faster rate were picked up.

Ideally we should have converted the quarterly data into annual data and then do an annual forecast as the current business goal only needs an annual forecast.

Forecasting Methods

Partitioning: Training Period: 142 quarters Validation Period= Forecast Horizon: 5 years (20 quarters)

Table 2

Industry/ Parameters	Human Health & Social Work	Construction	Education	Wholesale & Retail Trade	Public Admin	Manufacturing
Trend	Upward Linear Additive Trend	Linear Additive Trend	Upward Linear Additive Trend	Linear Additive Trend	Downward Linear Additive Trend	Linear additive trend
Seasonality	No Seasonality	No Seasonality	Quarterly Seasonality	No Seasonality	No Seasonality	Quarterly seasonality
Possible Methods	Moving Average (2), Double Exponential Smoothing, Regression	Moving Average (2), Double Exponential Smoothing, Regression	Holt-Winters Smoothing, Regression	Moving Average (2), Double Exponential Smoothing, Regression	Moving Average (2), Double Exponential Smoothing, Regression	Holt-Winters Smoothing, Regression
Best Forecasting Method	Double Exponential Smoothing	Double Exponential Smoothing	Holt- Winters Smoothing	Double Exponential Smoothing	Double Exponential Smoothing	Holts Winter smoothing

*More details on Comparison with Other Models in Appendix

*Performance metrics & benchmark (Naive) for each industry is given in Table 1 above

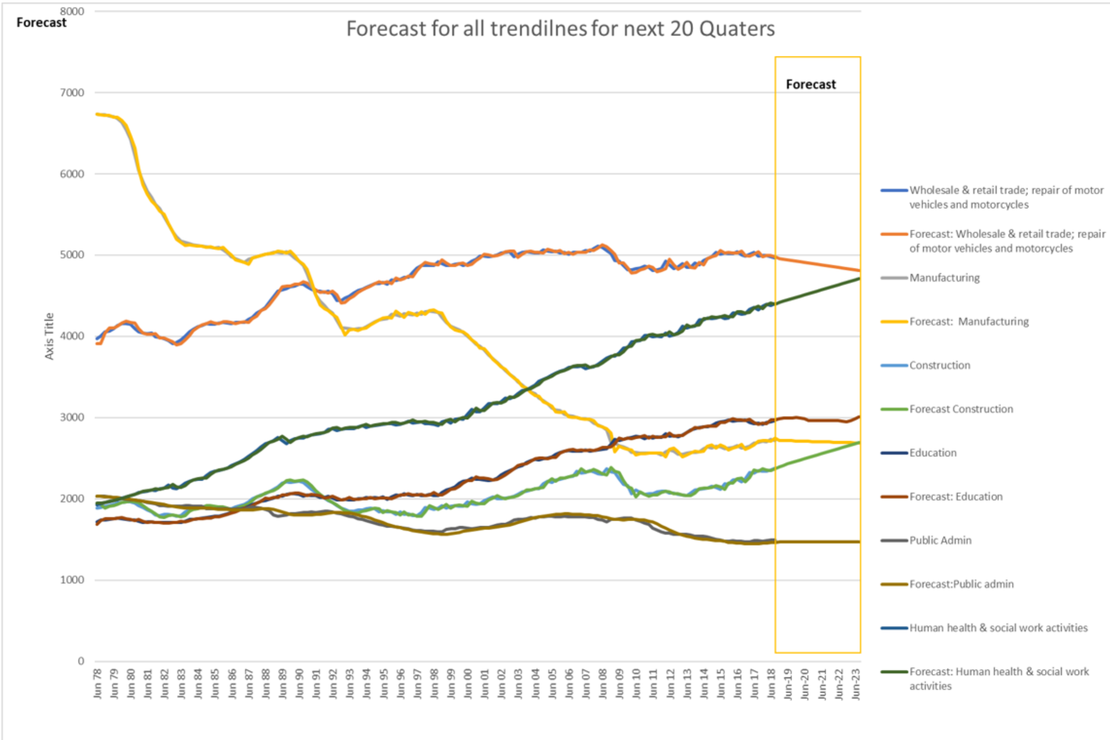
Since the selected methods produced low errors and good results, indirect methods which would employ de-trending or de-seasonalizing the data were not tested separately

Performance Evaluation:

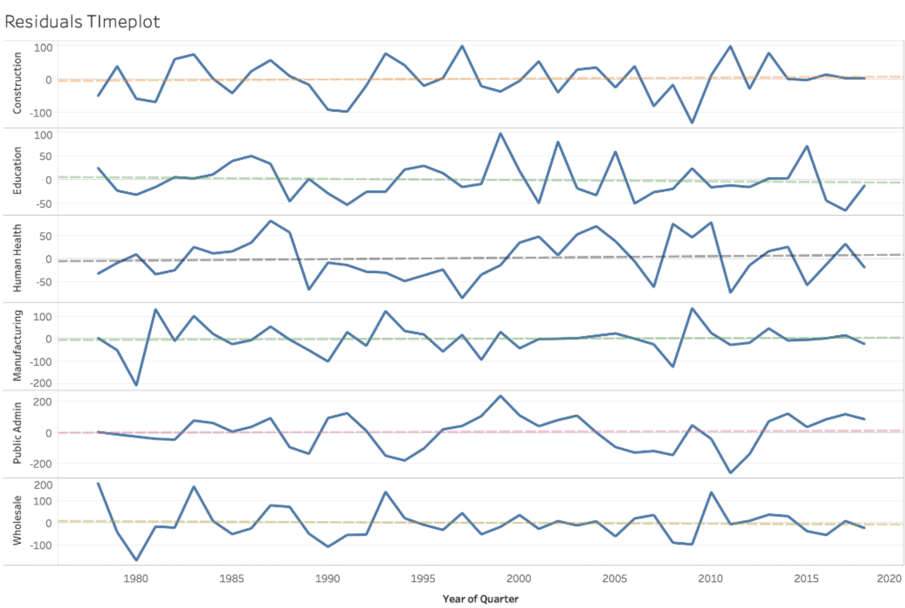
- Actual v/s Forecasted Plots for checking fit (over-forecasting or under-forecasting)
- Residual Plot with no trend (parallel to X axis), no seasonality, no autocorrelation (ACF plot), control plot
- Error Metrics: MAPE & RMSE. Comparison across Training & Validation period to check for overfitting

Actual vs Forecast for all the industries for past data as well as for Forecast Horizon

The plots of actual v/s fitted/forecasted values for the training and validation period of all 6 time series show minimum deviations, i.e. no significant under-forecasting or over-forecasting



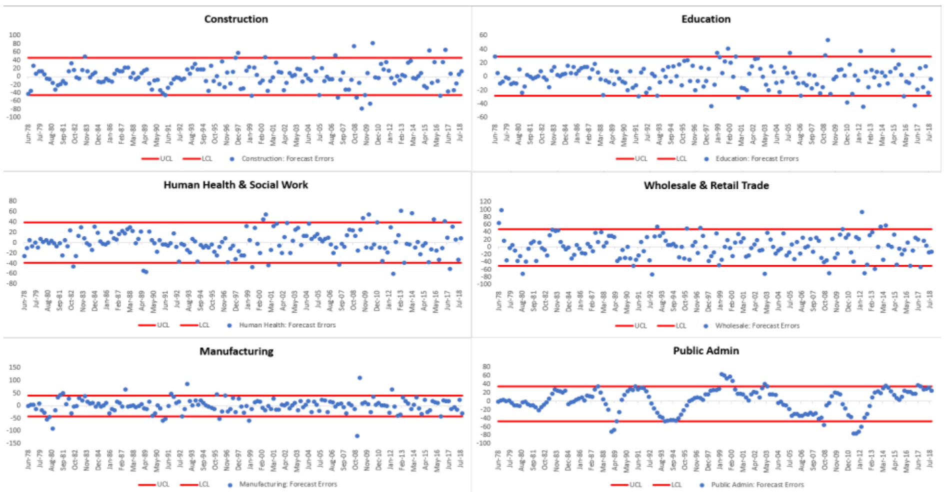
Residual Plot for all Industries



The trendline for the residual plots of all industries is mostly a straight line parallel to the X axis which shows that the trend has been captured well. Further, the plots show no seasonality, and hence seasonality has also been captured well by the models

APPENDIX

1. Control Charts for Residuals:



The forecast errors are mostly within the control limits

2. Comparison of Different Methods:

We compared different forecasting methods for all the time series and based on the performance metrics, picked up the best performing method. A comparison of different analyses done is as follows:

Time Series	Naïve Forecast	Best Method		Alternate Method 1		Alternate Method 2	
	MAPE Validation	MAPE Training	MAPE Validation	MAPE Training	MAPE Validation	MAPE Training	MAPE Validation
Construction	Naïve Forecast	Double Exponential Smoothing		Moving Average (2)			
	6.95	0.92	5.15	1.26	7.37		
Human Health & Social Work	Naïve Forecast	Double Exponential Smoothing		Moving Average (2)			
	3.80	0.58	0.62	0.98	3.92		
Education	Naïve Forecast	Multiple Holt Winters		Moving Average (2)		Double Exponential Smoothing	
	2.86	0.61	0.99	4.18	5.34	0.6	1.6
Wholesale & Retail Trade	Naïve Forecast	Moving Average (2)		Double Exponential Smoothing			
	2.5	0.68	2.75	0.53	3.62		
Public Admin	Naïve Forecast	Double Exponential Smoothing		Moving Average (2)			
	1.36	0.44	1.19	0.71	4		
Manufacturing	Naïve	Multiple Holt Winters					
	2.35	0.52	2.06				

A detailed example of how this was done for one of the series is as follows:

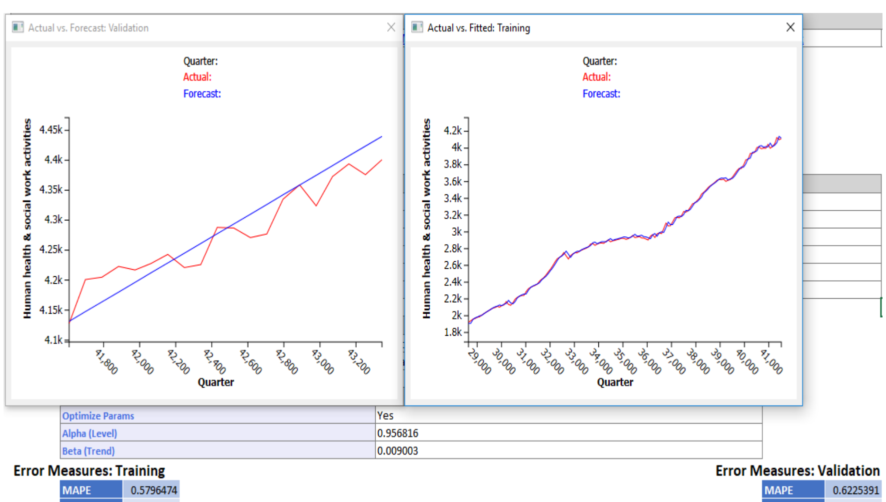
Human Health & Social Work

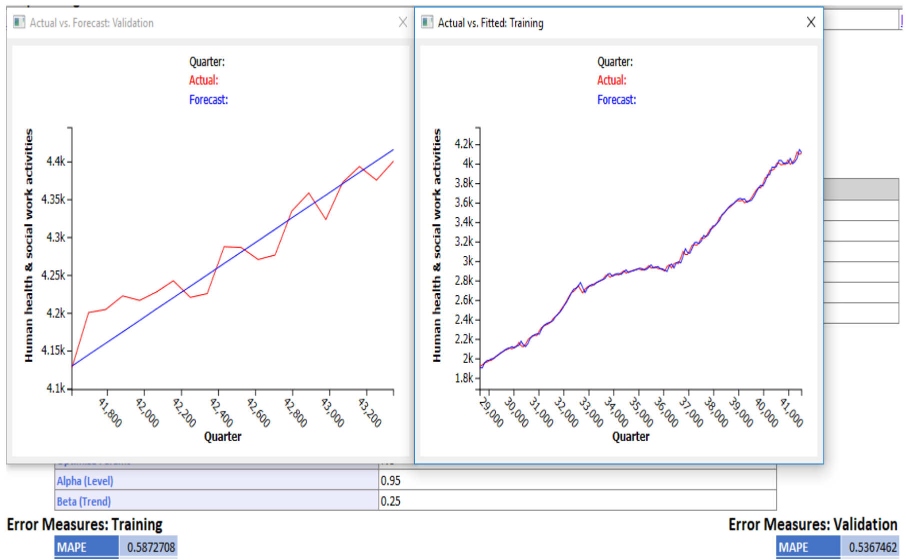
- Linear & Exponential Regression produced very high RMSE compared to the other 2 methods
- MA(2) was used as there was no seasonality and performed better than Naive on validation period
- Double Exponential Smoothing performed the best amongst the 3, in terms of MAPE and the difference in MAPE on training & Validation period is also the least, which eliminates the issue of overfitting

Training			
	Double Exponential Smoothing	Moving Average	Naïve Forecast
MSE	515.50	1,194.00	746.63
RMSE	22.70	34.55	27.32
MAPE	0.58	0.98	0.76

Validation			
	Double Exponential Smoothing	Moving Average	Naïve Forecast
MSE	1,005.25	33,984.31	32,323.75
RMSE	31.71	184.35	179.79
MAPE	0.62	3.92	3.80

- Different smoothing constants were tested for Holts and though another alternative gave better MAPE, it was rejected against the optimized one for fear of overfitting





3. Seasonality