Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Workforce Planning using Time Series

Rohan Girotra

August 19, 2022

Supervisor: Dr Bahman Honari

This dissertation is submitted to the University of Dublin, Trinity College Dublin for the degree of Master of Science in Computer Science - Data Science

Declaration

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at `http://www.tcd.ie/calendar`.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at `http://tcd-ie.libguides.com/plagiarism/ready-steady-write`.

Signed: ____________________ Date: ____________________

Abstract

This research demonstrates the multiple approaches to workforce planning. It mainly focuses on the healthcare sector workforce planning. Many healthcare institutions are under pressure to increase their treatment quality while becoming more cost-effective. Hospitals struggle with the problematic issue of maintaining an adequate number of General Physicians, Nurses, and Midwives staff to satisfy patient needs due to the ongoing variation in healthcare demands. The need to meet the same level of demand puts pressure on the remaining staff. For this reason, the second aim of this study is to use time series forecasting methods to curate the need of the healthcare workers.

We have explored different time series models and finally used Box- Jenkins ARIMA extension, i.e. Seasonal ARIMA, for forecasting with consideration of seasonality. The Vector Auto Regression (VAR) model is used for healthcare worker forecasting by leveraging the population increase.

Different statistical tests were leveraged to find the best parameter and model, like the Eyeball test of Autocorrelation and Partial autocorrelation to see the lag order, the Augmented Dickey-Fuller (ADF) for testing the stationarity of time series, Granger causality test to find the causal relationship between the two-time series.

The results were evaluated through Root Mean Square (RMSE),R-squared error (Multiple and Adjusted R square), Mean Absolute error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE).

Acknowledgements

I want to express my profound gratitude to Dr Bahman Honari for his constant guidance, feedback and support at every step and for sharing his vast knowledge. Without him, I would not have been able to complete this dissertation.

I am grateful to all of Trinity College Dublin's School of Computer science and statistics department's professors and members for their impeccable support during my tenure in the master's program. I would also like to thank my colleagues at Trinity college and friends for their advice and moral support.

I want to thank my friends Saloni and Archit, who kept me motivated and ensured that I looked after my mental and physical well-being.

Last, I would like to give heartfelt gratitude to my parents and siblings for their constant emotional and economic support and belief in my abilities.

Contents

List of Figures

List of Tables

Nomenclature

AR	Auto Regressive
MA	Moving Average
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
SARIMA	Seasonal Auto Regressive Integrated Moving Average
VAR	Vector Auto Regressive
GP	General Physicians
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
ADF	Augmented Dickey–Fuller
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
GDP	Gross Domestic Product
UK	The United Kingdom
STL	Seasonal and Trend decomposition using Loess

1 Introduction

1.1 Overview

The technique employed by Human Resource (HR) professionals to evaluate, project, and schedule the supply and demand for workers is known as "Workforce Planning". This technique combines forecasting for all the activities performed by HR professionals, i.e., recruiting workers, retaining the talent, redeploying them as and when the demand rises as well as development for leaders and employees. Structured workforce planning may be used to find any talent gaps that an organization could have and to have enough employees with the correct abilities in the right area at the right time.

Businesses have faced high demands, or a complete lack of it, during the COVID-19 pandemic which provides the best and most recent example of the need for workforce planning. With mass hiring in some sectors and major layoffs in many others, HR Professionals have been struggling, depending on their organisation, to either meet the needs of skilled workers or to reduce the workforce while limiting the damage caused to the morale of the employees. Thus, it is crucial for HR departments to adopt a formal strategy for maintaining the inventory of skilled talent at the right level which can be achieved with workforce planning.

The study has employed Statistical analysis, which enables the collection of data to analyse trends and make more precise predictions which are critical to achieving high accuracy in forecasting the workforce requirements of an organisation in accordance with its objectives and commercial strategies.

The Time series' Box Jenkins model, along with another statistical approach are used hereafter, along with carefully analysing the merits and demerits of other methods. These methods are applied to the UK health care data of General Physicians and Nurses & Midwifery data to forecast the demand considering seasonality and other factors like country population and GDP.

1.2 Motivation

Planning ahead is not easy; therefore, workforce planning is not prevalent in many firms. It isn't easy mainly because it requires defining a company's staffing needs based on projections for its future requirements and the supply of skilled workers that are available or would be available.

There are always periods of growth followed by periods of recession in an economy, as well as any business. HR Professionals need to be able to deal with these periods of rapid development and budget cuts by having a customised strategy for each business cycle [1]. The most recent example that truly motivated me to work on this project was the crisis created by the COVID-19 pandemic, particularly in the travel and healthcare sectors. In 2020, the demand for travel, both international and domestic, reduced drastically due to the COVID-19 pandemic, with countries closing their borders for international travellers as well as imposing lockdowns locally. At the same time, with the rising number of cases caused by the SARS-CoV-2 virus, the demand for healthcare professionals increased steeply across the world as the burden on the healthcare industry saw a sharp increase, but the supply of these skilled workers was very low. I was personally impacted by the effect of this dearth of healthcare workers during the wave of the Delta variant, when I was unable to secure an appointment with even a General Physician as my case was not one of COVID-19 and hospitals were focussed on treating COVID-19 patients at the time. Also, non-critical surgeries were postponed as hospitals closed their operation theatres to relieve the existing workforce from some pressure which impacted many patients.

Two years later, the world has returned to normal. However, the vast difference created in the demand and supply of the workforce still exists and needs to be addressed by HR Professionals. The stark example before us is the numerous reports of delayed flights, cancelled flights and lost baggage from the summer of 2022 due to the combination of the shortage created in professionals working in the travel industry, resulting from layoffs following budget cuts in 2020, and an increase in the demand as people sought to travel for leisure after two years of staying put. I was caught amid this crisis when my flight was cancelled due to staff shortage, and I was offered a flight for the next day. The employees of the airlines were unable to provide accommodation due to the high demand from passengers suffering the same fate as me and I had to spend the night at the Terminal.
These experiences encouraged me to conduct this research to determine an efficient way to achieve workforce planning to be prepared for such adverse scenarios. With the efficiency that workforce planning brings, the need for layoffs or for mass hiring can be avoided. My belief is that the approach should be taken to prevent problems rather than having to deal with the expense of fixing them. All of this can be achieved with structured workforce planning.

Previously, there were multiple methods through which we could do workforce planning and specifically for healthcare workers. However, they were not giving accurate results or are still undergoing improvements like not being able to incorporate the seasonality and trend factor.

Gathering time series data or transforming normal data into time series is more manageable. Other forecasting models like Markov and stock and flow require complete detailed data which is challenging due to privacy issues, and these datasets are difficult to maintain. Considering the simplicity, effectiveness, high accuracy and faster in predicting the short-run forecast [2], this is the reason we have chosen The Box Jenkins method of the Time series over another method.

1.3 Objective

There are two objectives of the thesis; the first objective is to study the different models of workforce planning and models related to healthcare and what their benefits are.

The second one is to predict the demand for healthcare workers in the country using the different Time series techniques where multiple Box-Jenkins models are used, fitted, and evaluated based on their performance and incorporating the prediction into the workforce planning model where statistical analysis will help in planning the future workforce in healthcare.

A time series data is picked from the Office for National Statistics. The number includes the quarter-wise data of General physicians, nurses and midwifery, the population of the UK and GDP data, starting from 2007 Quarter 1 till 2022 Quarter 1.

This data will be pre-processed and converted into the respective Time series. The Box Jenkins method of Time series forecasting will be used here. The main models we have used to get the results are Seasonal ARIMA and Vector Auto Regression.

1.4 Thesis Structure

The thesis structure is stated below:

- **Chapter 1** states the Thesis overview, the motivation behind taking up this research and the objective.
- **Chapter 2** It will describe the Background and Literature review of Workforce planning and healthcare worker workforce planning, which includes the previous work done in this field of study.

- **Chapter 3** will introduce the methodology where it describes the History of Time series and Time series method, the type of Times series and how it is used and evaluated. It also covers the detailed description of the dataset we use in this study and how it is processed.
- **Chapter 4** covers the Implementation of how the study went about implementing the models for forecasting the healthcare workers and evaluating the final results and discussion around it.
- **Chapter 5** brings us to a conclusion from all the chapters and shares the potential future work and the limitation that can be corrected.

2 Background and Literature Review

2.1 Work Force Planning

Workforce planning is the process followed by Human Resource professionals to handle shortages of skilled professionals or a surplus of the same by attempting to predict the future demands for the services provided by an organisation. This can be summarised by saying that with the help of workforce planning, the right people possessing the right skills are available to fulfil the requirements of an organisation at the right time.[3]

The need for effective workforce planning arises from the fact that the number of skilled workers on the verge of retirement is growing rapidly, the competition that skilled workers face is increasing, the ethnic as well generational diversity of the workforce is on the rise, and the views on maintaining a work-life balance among the younger generation of workers is very different from those preceding them. As a result of these factors, the existing policies of the Human Resource department may not be able to adapt and respond to the ever-evolving workforce or to the changes in the environment within the organisation or outside it.[3]

For HR professionals, taking a proactive approach to workforce planning, rather than a reactive approach is essential in understanding the risks as well as recognising the opportunities for an organisation which in turn can help with planning the workforce needed to successfully tackle any situation that may unfold in the future.

Workforce planning is a multi-step process and broadly includes the following.[4] The figure 2.1 also show these steps:

Figure 2.1: Workforce Planning Process

Source: CIPD Ireland

1. Understand the organisation and its environment: The workforce needs to be aligned with the mission statement and the future goals of their organisation.
2. Analyse the current and potential workforce: It is important to identify any gaps that may exist between the human capital available currently and that available for the organisation's future requirements.
3. Identify workforce gaps against future needs: Recognise the gaps in the workforce requirements for the organisation based on its future objectives.
4. Actions to address shortages, surpluses, or skill mismatches: Once the gaps in workforce requirements are identified, steps must be taken to address them.
5. Monitor and evaluate actions: It is necessary to constantly evaluate the actions taken to address gaps in the workforce needs.

These steps for workforce planning require constant iteration and are not necessarily linear. It is essential for the Human Resource professionals to continually monitor the changing needs of the organisation and remain aligned with its current as well as future goals for effective workforce planning [4].

In practicality, workforce planning faces many challenges. To take into account the future priorities, except the immediate future, of the organisation and evaluating the skills that may be required to fulfil those priorities can be difficult. Another hurdle is obtaining exhaustive data about the workforce currently available as well as that which will be available in the future. And often, managers are also not able to effectively provide realistic workforce needs[5].

The benefits of effectively implemented workforce planning are numerous, some of which are:

1. A well implemented strategy for workforce planning can help an organisation allocate resources to achieve its goals in an efficient manner.
2. It can help prevent stressful situations that may be obstructing the path to accomplishing the company's targets.
3. Workforce planning can also enable finding the right path for progress for the organisation.[5]
4. If the means for continual improvement, such as learning resources, are provided it can help reduce the rate of turnover as employees can more efficiently contribute to changing requirements.
5. Maintaining the right strength of the workforce can also eliminate scenarios of shortages, and hence mass hiring, or of surpluses, and hence layoffs.[1]

2.1.1 Workforce Planning for healthcare workers

The objective of healthcare workforce planning is to have the appropriate skills at the correct time to deliver the required services to the population in need of it. [6] [7]. The prediction of future healthcare needs, the availability of resources to provide the necessary services, and the evaluation and testing of policy solutions to fill any anticipated gaps between supply and demand are all part of this process. The essential services may be guided by one of three notions. Utilization refers to the use of already available services, and demand refers to service usage if access could be improved, and need-based refers to service use if all patients eligible for treatment could access all forms of healthcare.[8]
There are three traditional approaches

- **The workforce-to-population ratio:** The approach is overly straightforward for determining how many healthcare professionals are needed to serve a particular population. The findings can then be compared to benchmarks or professional judgments. Demographic data, such as statistics on population growth and workforce data, are frequently taken into account in this technique. Numerous studies have also been modified to take into consideration variables like consumption rates by age or gender and attrition rates of the medical staff. [9] The benchmark is the best ratio from a reference nation or area with a somewhat better-developed healthcare sector than the one under investigation [10]. For instance, in 2012, France had 112, Germany had 81, and Switzerland had 99 ophthalmologists per million people. [11] Despite having an apparent benefit owing to its speed and convenience, this strategy frequently overlooks elements like productivity, use, and distribution of healthcare staff, making it challenging to evaluate the data. Therefore, even with the expected projections, the issue of uneven worker distribution in healthcare is likely to continue.

- **The needs-based approach:** According to this approach, a population's current estimated healthcare demands are used to determine the necessary health workforce. Healthcare demands are the quantity or number of healthcare experts needed to sustain a healthy population through optimal healthcare services. The foundation of this strategy is a population's demographic features, such as the incidence of diseases, age, gender, and educational attainment. [12] This strategy is predicated on the suppositions that all healthcare requirements will be satisfied, cost-effective means of meeting those requirements can be established, and healthcare resources are used under relative degrees of need.
  This strategy offers several benefits. It is unaffected by the existing use of health services and can meet the population's healthcare demands by utilizing various human resources for health. The strategy is logical, comprehensible, and in line with ethical standards for professionals. As a result, it can be used as a tool for advocacy. However, it needs significant epidemiological data, which is frequently lacking. Additionally, this method constantly updates factors, such as the levels of technology, and neglects the effectiveness of resource allocation. As a result, anticipated staffing and service goals might not be feasible. [10]

- **The Utilization based approach:** In this approach, the population's existing service usage levels were used as a proxy for satisfying demand for estimating the future healthcare workforce requirements. The degree of healthcare services that a population will need and be able to obtain within a specific time-frame at the present price is referred to as satisfying demand in this context. The utilization-based method, like the needs-based approach, is based on demographic data, including illness prevalence, age, gender, and educational attainment. Additionally, trends in healthcare service consumption and the market forces influencing these patterns are taken into account [12].

  The basic presumptions of this strategy are that age and gender specific requirements will remain constant into the future and that demographic changes over time may be projected based on present patterns [13]. This approach helps forecast economically achievable targets by the presumption that there would be slight or no change in the population-specific utilization method [10].

2.2 Previous Work on Workforce Planning

2.2.1 Previous Work in Workforce forecasting models

Multiple research happened in forecasting workers, and a few of them are discussed below. Wong, Chan, and Chiang (2012) [14] conducted a thorough analysis of current workforce planning techniques to identify improvements for upcoming workforce models created for the construction sector. They contrasted top-down forecasting with bottom-up approaches as the primary strategy for addressing manpower needs. Topdown models, often called stepwise models, are created based on an overarching understanding of the entire system without going into specifics about the subsystems. This strategy assumes that each job in a given industry would see proportionate growth. A bottom-up strategy, on the other hand, combines first-level subsystems to create more complex systems.

Roberfroid in 2009 [12] performed an assessment of the literature on physician supply models. They divided forecasting models into four categories: supply forecasts, demand-based methods, need-based approaches, and benchmarking against related healthcare systems. They determined that most models were static and lacked historical data to support a validity test.

O'Brien-Pallas et al. in (2001)[15] discussed the prediction approaches for nursing requirements from 1996 to 1999. They outlined the flaws in the current models, such as their disregard for important principles that affect human resources for health (HRH), such as how price competition affects workforce participation and how HRH decisions affect public health. They also discussed the need, demand, and utilization-based approaches. The predictions for future nurse requirements based on each of these three methodologies differ significantly.

2.2.2 Previous work using Time series for healthcare worker forecasting

There has been a significant amount of study done on healthcare workforce forecasting where the researchers have studied different models and approaches and studied varied factors that can be used in forecasting demand of the healthcare worker. Sen, Jaydip [16] discussed different time series approaches for the healthcare sector in India. They use India's healthcare sector's time series index values from January 2010 to December 2016. Then, they show that the time series decomposition technique gives them helpful insights into several of the time series' traits and behaviours. They comprehend the growth pattern, the seasonal traits, and the level of unpredictability displayed by the time series index values by using the trend, seasonal, and random component values of the time series. Additionally, they offer a comprehensive framework for time series forecasting that makes use of three separate models for forecasting the values of time series indexes. They provide critical analysis of the three models and an explanation of why some methodologies perform better and result in lower forecast error numbers than others. The approaches used were HoltWinters, ARMA, and ARIMA. They evaluated the result with the best combination of Min Error, Max Error, Mean Error, SD of Errors, RMSE and find out that ARIMA is performing best.

A comprehensive framework for theoretical analysis and health status forecasting techniques is provided by Soyiri and Reidpath (2012)[17]. The authors go through the main problems with health forecasting as well as the characteristics of health data that affect which forecasting method should be used.

A model that forecasts the evolution of the supply of medical specialties under three distinct demand scenarios is put out by Sense et al. (2015)[18] The authors' demand predictions consider various factors, including demographics, service consumption rates, and hospital beds. The methodology suggested by the authors employs a mixed integer programming model to find the best distribution of medical specialization awards for the various study years based on the model output.

Ostwald & Klingenberger(2016)[19] put out a novel method for calculating the economic importance of the German oral healthcare industry. Based on a variety of explanatory factors, including demographic change, take-up behavior, medical and technological advancements, oral morbidity, aggregated supply, and income levels, the authors established a model for predicting the expansion of the oral healthcare industry. The authors' estimate indicates that by 2030, the gross value added to the German healthcare industry will have increased by 19.2%.

A thorough, nine-step, quantitative demand forecasting model for healthcare services was put out by Finarelli and Johnson (2004)[20]: Creating a baseline forecast of future demand requires the following steps: (i) gathering historical data; (ii) analyzing historical trends; (iii) identifying key demand drivers; (iv) locating pertinent benchmarks; (v) modeling current conditions; (vi) developing core assumptions for population-based demand; (vii) developing core assumptions for provider-level demand; and (viii) testing the sensitivity of projections to changes in core assumptions.

Hospital emergency department bed occupancy may be forecasted in the short run, according to research by Schweigler et al. (2009)[21]. The authors found that a sinusoidal model with auto regression (AR) while a seasonal ARIMA generate reliable forecasts in a short-term time frame of 12 hours.

Beech (2001)[22] suggested an approach for obtaining demand estimates for healthcare services from a wide range of data sources. Data on primary and secondary service regions, population breakdowns by demographic subgroups, discharge usage rates, market size, and market share per service line were all included in the data sets the author used. The author noted that market changes can facilitate the creation of trend models that can accurately predict upcoming needs.

3 Methodology

3.1 History of Time series

The early natural sciences have already made significant use of time series. In order to forecast astronomical occurrences, Babylonian astronomy studied time series for the stars and planets relative position. The rules that JOHANNES KEPLER found were based on observations of the motions of the planets. By identifying patterns in a variable's observations and deriving "rules" from them, time series analysis can improve future prediction by utilizing all of the data included in the variable. These approaches are based on this fundamental [23] This analytical approach to astronomy was adopted in the middle of the 19th century by the economists CHARLES BABBAGE and WILLIAM STANLEY JEVONS. Traditional time series was developed by WARREN M. PERSONS in 1919[23]. He identified the following four elements, which is also known as time series decomposition:

- First element is long run development, which is trend.
- Second, is cyclic component which range up to more than 1 year, which is a business cycle.
- Third element is high and low within that year, which is season cycle.
- The fourth and the final is the movement contained which neither belonged to trend or business nor to the seasonal element, which is called as residual.

The traditional time series analysis takes into account that systematic elements, i.e. trend, business cycle, seasonality, are not determined by stochastic disturbances and they can be represented as pure function of time. The residuals, at the same time, do not include any systematic movements, therefore stochastic influence is limited to the residuals. As a result, it is described as a pure random process, which is a set of independent or uncorrelated random variables with expectation zero and constant variance.

3.2 Time Series Method

A time series is described as a collection of quantitative data organised in time sequence. The field of econometrics has traditionally utilised time series.

There are multiple methods for time series; it can be parametric or non-parametric, linear or non-linear. For decades, linear statistical approaches dominated time series forecasting [24]. We will talk about the Parametric time series.

3.2.1 Parametric Time series method

The most famous method of time series forecasting is Box-Jenkins, and it is used in different prediction applications spanning from socioeconomic to engineering and environmental challenges [25] Box and Jenkins popularised these methods in 1976; they are named traditional time series [26]. First, these basic methods are discussed like The Auto Regressive (AR), Moving Average (MA) and the combination of these are Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA).

Autoregressive (AR) This model forecasts the future value through a linear combination using its past observation. This model is the most basic model of the Time Series approach and the benchmark and foundation of other enhanced models.

$$Y_t = c_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t \tag{1}$$

This model is referred to as AR(p) model, it has order of p, Y_t is the target value, we use "t" as a subscript for time, y is measured on time period t.
$Y_{t-1}, Y_{t-2}..Y_{t-p}$ are the past values of series also known as Lag, c_0 is a constant e_t is white noise or error and $\emptyset$ is autoregressive parameter.

Moving Average (MA) This model uses anomalies which came in forecasting past values to use in forecasting future values.

$$Y_t = c_0 + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{1}$$

This model is referred to as MA(q) model, it has order of p, $e_{t-1}, e_{t-2}..e_{t-q}$ are the past error of series, c_0 is a constant and θ_1 is moving average parameter. [2].

Auto Regression Moving Average (ARMA) This model was introduced to accommodate the anomaly in AR model, instead of increasing AR order we can add MA(1) that will adjust the without affecting the rate of exponential decay. The output of this mix is the auto regressive moving average or ARMA(1,1) model, It is ALSO denoted as ARMA(p,q):

$$Y_t = c_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots . + \emptyset_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots . + \theta_q e_{t-q} \tag{3}$$

The ARMA model cannot be used directly. Using the differencing approach to convert a non-stationary series to a stationary series is one of the method [2]. Differencing may eliminate trend and seasonality in a time series by stabilizing the mean [2], although one or two orders of differencing are usually sufficient to make the data stable [27]. A differenced series is the difference between two consecutive timestamp observations in the original series and is expressed as:

$$Y'_t = Y_t - Y_{t-2}$$

Auto Regression Integrated Moving Average (ARIMA) This method is derived from the ARMA method, which can overcome the stationary limitation in the time series. It is one of the most popular methods for forecasting using time series.

It is denoted as ARIMA(p, d, q), where p is the order of the AR part, d is the degree of differencing, and q is the order of the MA part. "integration" is the reverse of differencing.

Seasonal ARIMA (SARIMA) It is an extended version of ARIMA. It is used when time series displays a seasonal variation, i.e. the mean and other statistical data for a specific season are not stationary throughout the years. It is denoted as SARIMA (p,d,q)(P,D,Q)S, The first p, d, q part is similar to ARIMA while the D is seasonal differencing order and cycle of seasonality is S. The P and Q represent the autoregressive and moving average components of the seasonal part of the data, respectively. For instance, above S for quarterly data will be 4, and for monthly data, it will be 12 [28]

ARIMA models, particularly in reliability forecasting, are preferable to Bayesian methods [29]; However, while ARIMA may be tailored to provide a very accurate linear forecasting model, particularly over a short time horizon, it cannot be used for forecasting non-linear time series [30]. The reason for this is that ARIMA models have two fundamental limitations: the first is being linear in nature; they assume that predicted values of a time series will have a linear connection with previous values and white noise. The second limitation Is data restrictions which means the ARIMA model requires a large amount of historical data (minimum of 50 observations) to achieve high accuracy in the prediction [31] [30].

George Box and Gwilym Jenkins investigated a streamlined method for obtaining extensive information about the ARIMA model and employing the univariate ARIMA and SARIMA models [26]. Box-Jenkins (BJ) approach has four iterative steps:

- **Step 1: Identification** At this step, we select the regular differencing (d) and seasonal differencing(D), then choose the order of Autoregressive order (p), (P) non-seasonal and seasonal, respectively. Similarly choose the non-seasonal and seasonal order of Moving Average (q),(Q) respectively. These orders are identified by the Autocorrelation function (ACF) and Partial autocorrelation function (PACF). The ACF and PACF are also discussed in detail ahead in this section.
- **Step 2: Estimation** At this step, the historical data is utilized to determine the parameters of the model we are going to finalize.
- **Step 3: Diagnostic checking** The diagnostic test is performed to validate the model to be selected.
- **Step 4: Forecasting** The model we have finalized in step 3 is used for forecasting the future time stamp values.

The Box-Jenkin methodology has several advantages, including the ability to find the adequate trend from the historical patterns. These small number of parameters help in extracting a large amount of information. It handles stationary and non-stationary time series with non-seasonal and seasonal components. [2].

Exponential smoothing Exponential smoothening uses a weighted moving average approach, in which the weights of previous observations are decreased exponentially over time, and weighted patterns are easily changed to accommodate particular requirements predicted by previous cycle forecasts. It is mainly interpreted as an extension of the ARIMA model, but they are different. ARIMA use autocorrelation in the data, while exponential smoothening is based on the trend and seasonality [2].

Vector Auto Regression (VAR) CHRISTOPHER A. SIMS (1980) developed the approach of Vector Autoregressive Systems (VAR) One of the multivariable techniques for forecasting is vector autoregression (VAR). The univariate Autoregressive (AR) model is a generalization of this model. Like the AR model, the VAR model represents a particular type of time-varying process. The sole distinction is that the VAR employs two or more time series in the form of a vector, whereas the AR model only uses one time-series data. VAR may be used to model many variables simultaneously [32].

Vector time series data of Yt may be expressed as a function using a VAR model with k variables and p lag.

$$Y_t = c_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \in_t \tag{1}$$

The c_0 denotes the constant $\in_t$ is the error variable of k $\times$ 1 column vector a, $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are k $\times$ k matrices of AR coefficient.

A three-variable VAR model with a lag value p=1 is represented as below

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \emptyset_{1,1} & \emptyset_{1,2} & \emptyset_{1,3} \\ \emptyset_{2,1} & \emptyset_{2,2} & \emptyset_{2,3} \\ \emptyset_{3,1} & \emptyset_{3,1} & \emptyset_{3,3} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \in_{1,t} \\ \in_{2,t} \\ \in_{3,t} \end{bmatrix} \tag{2}$$

Reflection

From the above subsection, most of the time series forecasting methods described generally have various traits and predictive power. Bespeak, Box-Jenkins techniques like ARIMA can accurately model data autocorrelations, whereas SARIMA and exponential smoothing can successfully predict any seasonality and trend. ARIMA can also model linear processes competitively.

3.2.2 Time Series Forecasting the Process

3.2.2.1 Data Partitioning for validation

A forecasting time series model has been trained on a portion of the dataset called the training set. It is assessed for predicted accuracy in the test set. The test set's size is determined by the overall dataset's size and the intended prediction horizon, or we can say future steps for which we are making predictions [2]. The general practice is that the test set should be at least as large as the prediction horizon. According to Hyndman and Athanasopoulos (2018), [2], the training and test sets typically account for an 80-20 ratio of the dataset.
The distribution of the Training and test set is shown in the below figure 3.1

Figure 3.1: Train test data distribution in Time Series

Source: Hyndman & Athanasopoulos, 2018

There are other ways to evaluate the time series models. Time series cross-validation is sometimes referred to as rolled forward validation [2]. For instance, divide training data into two folds. For instance, we have data from 2000 to 2020 we want to predict the values for the 2021 Timestamp. For the first fold, 2000 to 2017 as training data and 2016 as test data Second, fold 2000 to 2018 as training data and 2017 as test data. Till 2000 to 2020 as training and 2021 as test data. The final accuracy values are calculated by averaging the error found at each phase. The distribution of the Training and test set is shown in the below figure 3.2.

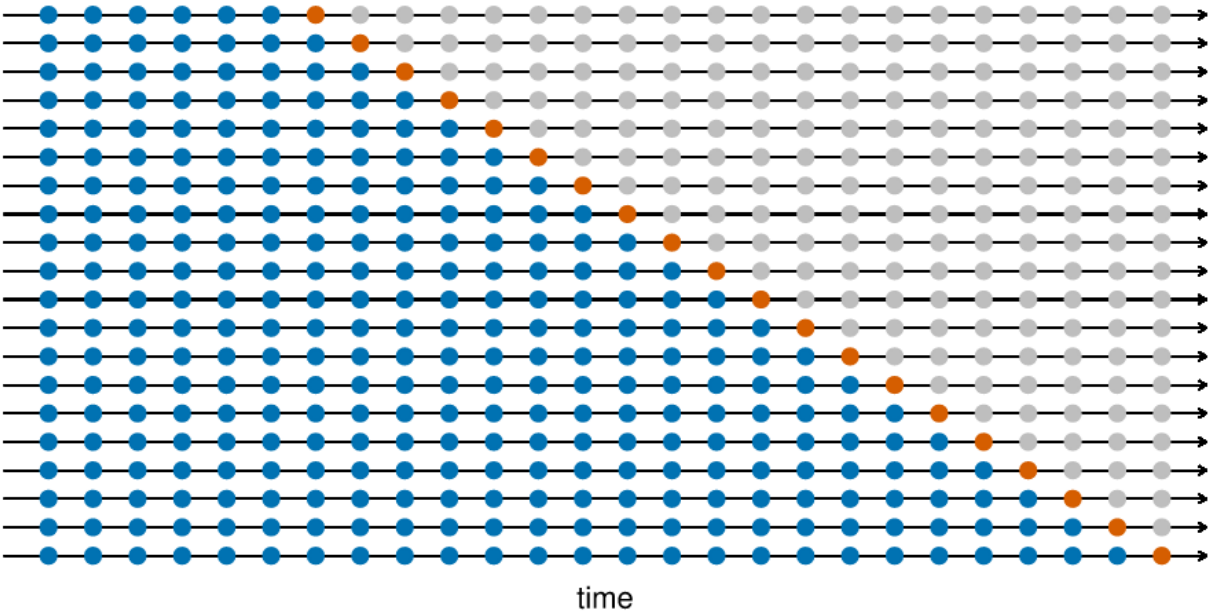


Figure 3.2: Time series cross-validation

Source: Hyndman & Athanasopoulos, 2018

3.2.2.2 Model Selection

In order for the model that is generated to have the appropriate qualities, a set of criteria is applied during the model creation process. These characteristics are precision (minimal errors) and parsimony (simple models with the fewest parameters), respectively [2].
The most common criteria are mostly used are: The Bayesian Information Criterion (BIC), the Corrected Akaike Information Criterion (AICc), and the Akaike Information Criterion (AIC).
They are represented as:

$$\text{AIC} = -2\log(L) + 2k,$$

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{T-k-1},$$

$$\text{BIC} = \text{AIC} + k[\log(T) - 2].$$

where k is the total number of estimated parameters and beginning states, and L is the model's log likelihood, T is the number of observation

The lowest the AIC, the best the forecasting model is. AIC uses the loglikelihood function to find the best fit. It also penalizes any increase in the number of parameters that help in preventing overfitting [33]. For the limited number of past data, AIC makes too many parameters therefore to overcome this limitation of AIC, we use the Corrected Akaike Information Criterion (AICc) [2]
Similar to AIC, the lower the BIC the good is model but BIC penalized the number of parameters more severely than AIC.

3.2.2.3 Residual Diagnostic

Residual is the difference between the predicted and actual values of $\hat{y}_i$ in the training and test sets, respectively, is what also known as a residual error.// This residual error Is represented as :

$$e_t = y_i - \hat{y}_i$$

The forecasting model should be evaluated to see how its residuals behave after the model-building phase. The appropriate tests on the residuals are run during this step of residual diagnostics. A great forecasting model's residuals often contain the following characteristics [2]:

1. Residuals are uncorrelated, if we see any relation between residuals then the information which is left in residuals should also be utilized for the forecasting.
2. The residual should have zero mean if it is not then the forecast is biased.

Eyeball test

ACF and PACF uses graph for residual diagnostics:

They're employed to determine if a time series has autocorrelation. ACF sometimes referred to as a correlogram, describes the auto-correlation versus the corresponding time lags. PACF assesses the degree of correlation between various lags after excluding the effects of other lags. Histograms and Quantile Plots of errors in the test set are used to determine if the residuals are normally distributed and, if so, whether 95% prediction intervals may be generated using the equation ŷ$\pm 1.96\sigma$ [34].

Process	ACF	PACF
AR(p)	Tails off towards zero (exponential decay or damped sine wave)	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off towards zero (exponential decay or damped sine wave)
ARMA(p,q)	Tails off after lag (q-p)	Tails off after lag (p-q)

Figure 3.3: Behaviour of ACF and PACF

Source: Elsevier : BuHamra

Ljung-Box

This test is named after statistician Greta M. Ljung and George E.P. Box. This test is used to check if a particular time series' null hypothesis of independence [2]. Typically, this test determines if the residuals are linked (stationary residual series). Additionally, the residuals' homoscedasticity (constant variance) examines if the Ljung-Box test was performed on squared residuals [35]. Regardless of sample size, the Ljung-Box test has the ability to be extremely effective [36].

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\rho_k^2}{n-k}, \tag{1}$$

where h is the number of lags, n is the sample size, and ρ_k is the sample auto correlation at lag k.

$$Q > \chi^2_{1-\alpha,h}. \tag{2}$$

Q's chi-squared distributed with h degrees of freedom can be proved. The crucial region where the null is rejected becomes the α -quantile if we assume a significance level α:

Granger Causality:

The universe is not made up of separate stochastic processes. Contrarily, general equilibrium theory states that economists typically believe everything is interdependent. The subsequent query thus concerns the (causal) links between different time series. In principle, there are two ways we may respond to this query. Following a bottom-up approach, one can first presume that the various time series' data-generation processes are independent of one another. One might check for relationships between various time series as a second step. This statistical method, which is commonly used nowadays when causality tests are conducted, is based on the ideas presented by CLIVE W.J. GRANGER (1969). A top-down approach is an alternative, which assumes that the producing processes are not independent and, in a second phase, determines if any particular time series are formed independently of the other considered time series. When employing vector autoregressive processes, this strategy is used. [37]

So, Granger causality test is a statistical hypothesis test to see if a time series helps forecast the other time series.

A vector autoregressive model (VAR) is usually fitted to a time series to perform multivariate Granger causality analysis. Let $X(t)$ be a d-dimensional multivariate time series. To perform Granger causality, fit a VAR model with L time lags as shown below.

$$X(t) = \sum_{t=1}^{L} A_\tau X(t-\tau) + \epsilon(t)$$

A time series X_i is called a Granger cause of another time series X_j if at least one of the elements $A_\tau(j, i)$ for $\tau = 1, ...L$, the absolute value of L is significantly larger than zero.

3.2.3 Evaluation of the Time series

Once we are done with the diagnostic test, the predictive model is assessed on their forecasting performance on the test data.
Specific accuracy metrics, which in fact is an error measure, are used to assess a model's prediction ability in the study.

$$e_t = Y_i - \hat{Y}_i$$

where e_t is error for given time, Y_i is Actual value $\hat{Y}_i$ is predicted value.

The most common error evaluation techniques are following:

- **Mean Absolute error (MAE)** It takes all the errors into account and calculates the mean of those errors; All the errors are given the same weight.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

- **Root Mean Square Error (RMSE)** In this method, Variance is penalised by RMSE because it pays greater weight to mistakes with more significant absolute values than errors with lower absolute values. When both measures are measured, the RMSE will always be greater than the MAE

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}$$

- **Mean Absolute Percentage Error (MAPE)** In this metric, the error is divided by the actual value, and This metric is used to compare the accuracy of several models on multiple datasets.

$$MAPE = mean\left(\left|\frac{100e_i}{y_i}\right|\right)$$

- **Mean Absolute Scaled Error (MASE)** This method is the MAE of forecasted value divided by the MAE of naïve forecast model on the training data. When this error metric in one-step forward forecasting is more than 1, it is implied that the naive technique will produce more accurate results [38]. If MASE id is greater than 1, then the data's forecasting should be reflected because data may contain white noise or the model is overfitting the data.

$$MASE = mean(\mid q_j \mid)$$

where q_j is represented as

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T} \mid y_t - y_{t-1} \mid}$$

- **R squared** It measures the quality of the regression line it is also represented as coefficient determination [39] . A line that fits poorly is represented by a value around 0, whereas a line that fits well is represented by a value near 1. In short to assess the accuracy of the fitting model and the impact of outliers on the regression analysis.

  It is represented as:

$$\begin{aligned} R^2 &= 1 - \frac{SSE}{TSS} \\ &= 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \end{aligned}$$

where SSE is the sum of squared errors, TSS is total sum of squares.

3.3 Dataset

3.3.1 Dataset Gathering

The dataset for this thesis has been procured from the Office of National Statistics for the UK [40]. The Office of National Statistics is responsible for producing the formal statistics for the UK, including employment statistics. They are also responsible for managing the census after every ten years in England and Wales. The data published by the Office of National Statistics are generated from a variety of sources that include social surveys like the Labour Force Survey. Unpublished data has not been used in this thesis as it may pose the threat of revealing confidential information.

On the Office of National Statistics website, within the section for "Employment and employee types", comprehensive figures about the employment rate have sourced the dataset for this thesis. This information provides an insight into the number of people who are salaried workers, categorised on the basis of their age and gender, as a proportion of the total population. Apart from details about the number of people who are employed, the statistics also present information about the number of open positions.

We have gathered four different datasets: the UK population,GDP, General Physicians, and Nurses & Midwiferies. It is a time series dataset in a quarter-wise format.

Time series representation of the dataset and the observation from the data

Gross Domestic Product (GDP) of the UK

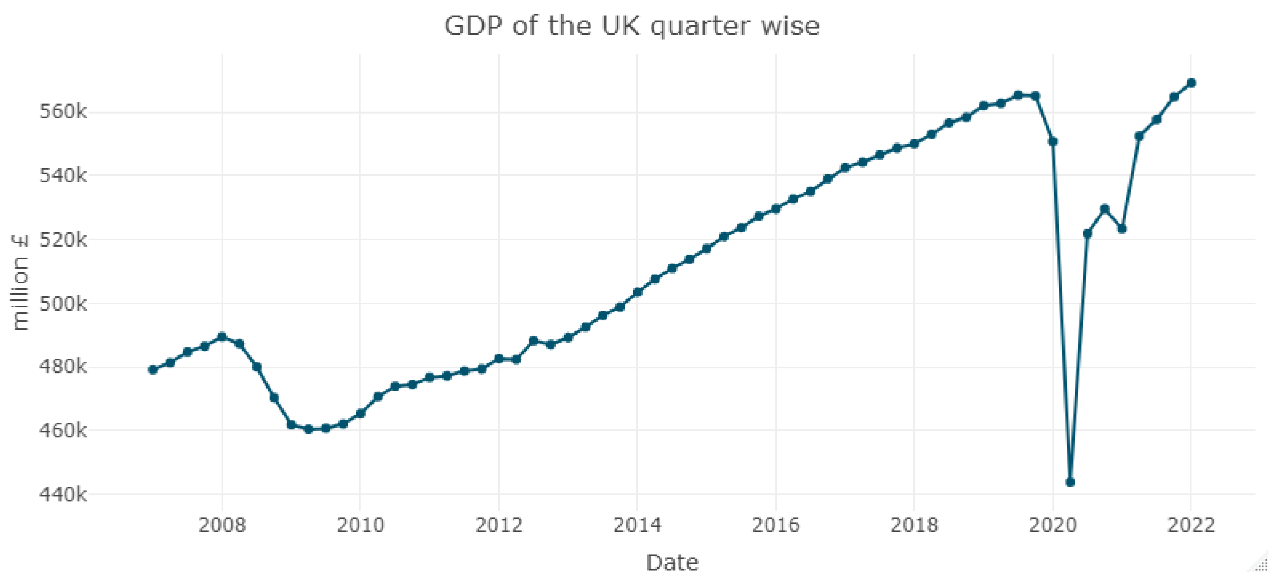


Figure 3.4: GDP of the UK

We have this GDP data from 2007 Q1 to 2022 Q1.
The UK GDP is expected to have climbed by an unrevised 0.8% in Quarter 1 (January to March) 2022.
At the end of the fourth quarter (October to December) of 2019, real GDP was 0.7% higher than before the Coronavirus (COVID-19).
There is the apparent effect of Coronavirus (COVID-19) on the UK GDP. [41]

Population of UK

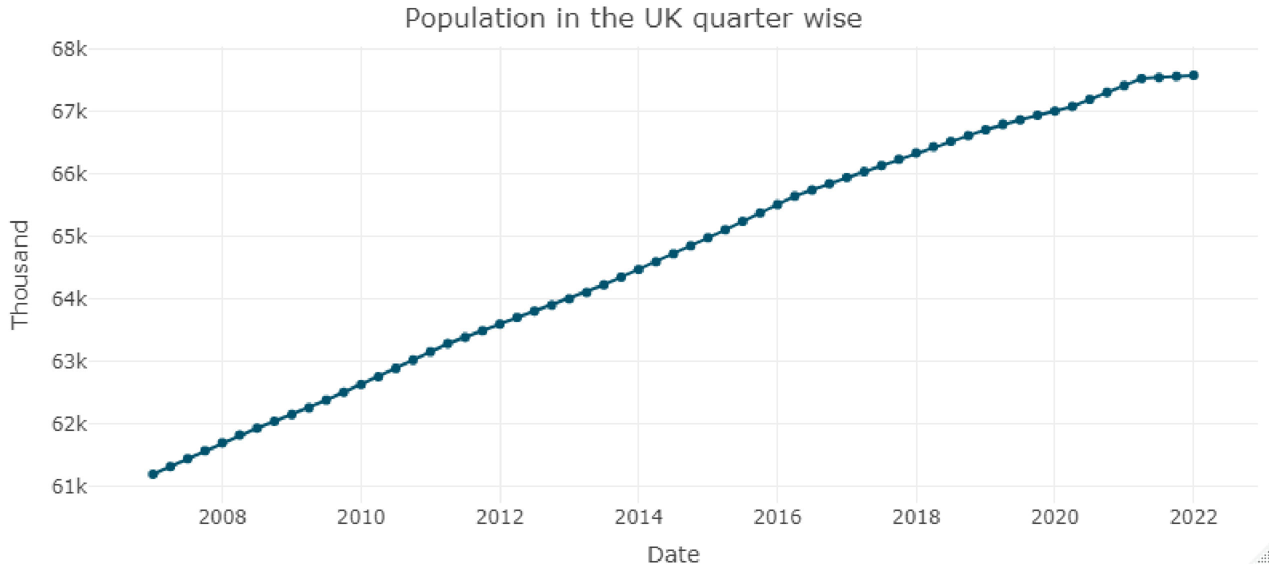


Figure 3.5: Population of the UK

The population of the United Kingdom was expected to be 66.8 million in mid-2019. The growth rate of the UK population from mid-2018 to mid-2019 was 0.5

General Physician

Figure 3.6: General Physician in the UK

The SOC stands for Standard Occupational Classification. In this dataset, jobs are categorised in terms of their skill level and skill content within the framework of the categorisation. It is used for providing career information to new entrants into the labour market, job matching by employment agencies, and the establishment of government labour market regulations.[42]

We have taken the data into two categories; one is General physicians, which consists of Partners, Salaried, Trainees, Retainers and Locums and the second is a combined dataset of Nurses & midwiferies. If we see the General physician dataset, we see a growing trend with a seasonality component.

Nurses & midwiferies

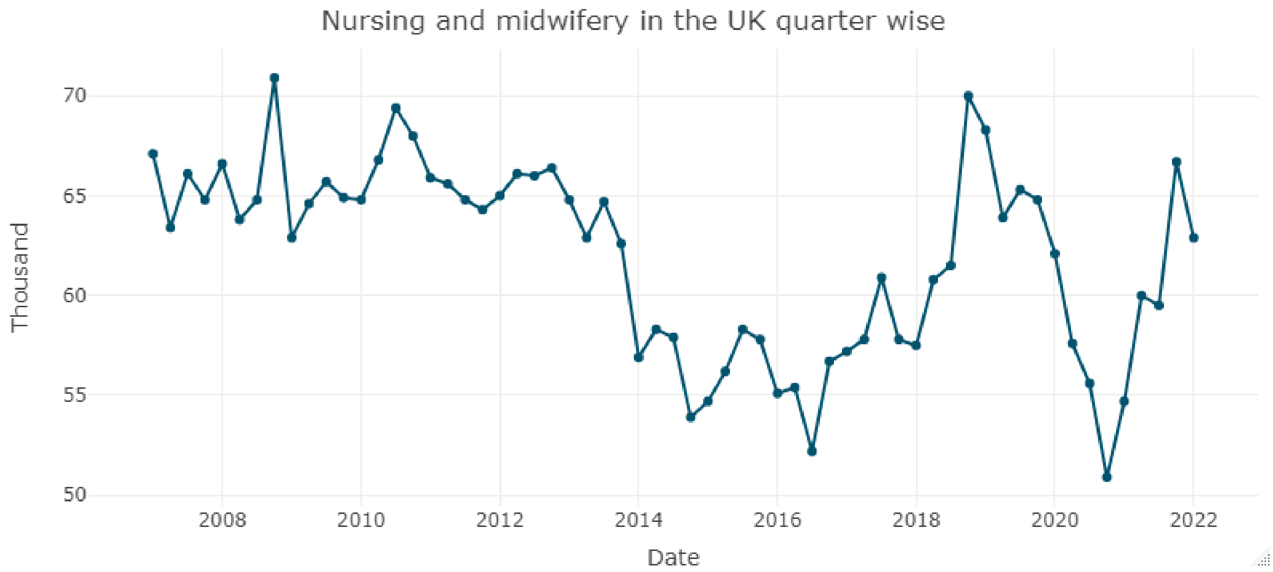


Figure 3.7: Nurses & midwiferies of the UK

From the Nurses and midwiferies dataset, we can observe that there is no constant growth and decreasing trend is seen in the data. But in 2016 Q3 and 2020 Q4, we have seen the number reach the lowest in the range of 2007 Q1 to 2022 Q1

3.3.2 Dataset Pre-processing

Three steps were involved in the data pre-processing: extracting, transforming, and loading. The extraction part was discussed above, i.e, how we found the data.

Transformation and cleansing: We had multiple datasets with multiple professions like construction, cleaner, administrator and healthcare workers, so the specific data was picked from it.
Population Data was interpolated quarter-wise from year data.

Interpolation: The technique of determining data between two points on a line or curve is known as interpolation. To recall what it implies, consider the initial half of the term, 'inter,' meaning 'enter,' which suggests we examine 'within' the data. It is crucial not just in statistics but also in science, commerce, and other situations where it is necessary to estimate values between two current data points. GDP data was from 1955, which was split from the original series. We are finally, converting the data into a unit of thousand so that our analysis will be more straightforward.

Load or merge the data: Then Finally, all the data from three different sources were cleaned and picked from their source and were kept in a time series format of ascending order starting from 2007 to 2022 quarter-wise in a CSV file so that it can be used directly in the analysis.

4 Implementation and Results

4.1 Implementation and Discussion

This subsection talks about the implementation where every step of implementation is discussed.

The process of implementing Time series forecasting can be divided into a few steps as shown in Figure 4.1

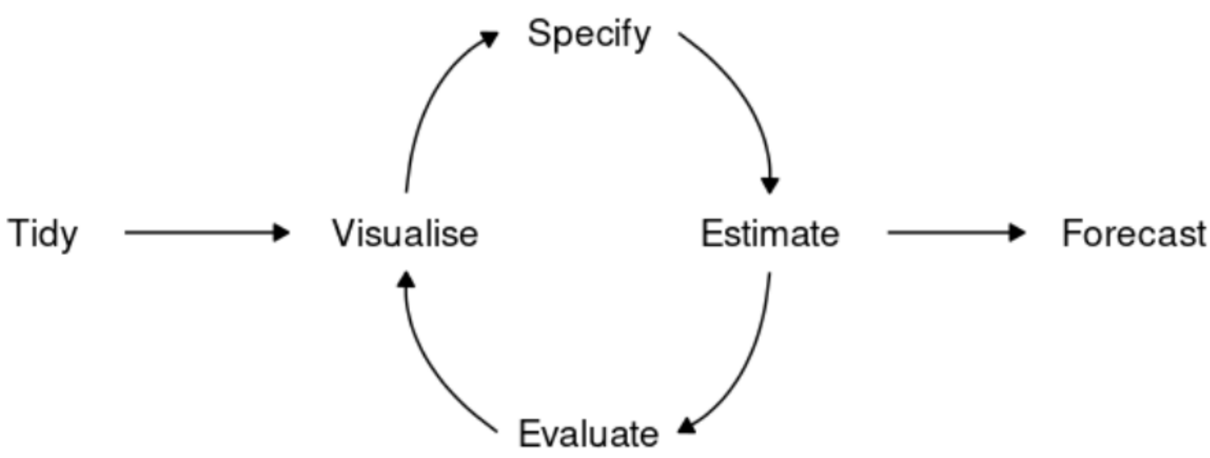


Figure 4.1: Implementation Time Series Forescasting

Source: Hyndman & Athanasopoulos, 2018

4.1.1 Data Preparation (Tidy)

The first step of our implementation was to look out for the time series data set and clean it according to the required time series model. Different models have different requirements, for instance, some models require data in a time order. The pre-processing and cleansing of the data and desired condition is discussed in the section 3.3.

We started with reading the processed data from the file we had created. Then we took the file and extract the column data of GP, Nurses & Midwifery, Population and GDP.

4.1.2 Plotting the data (Visualise)

Visualizing the data is a crucial step in having familiarity with the data. The visual of the data allows to identify the common pattern in the time series and helps identify the best model. The best way to visualise time series is to decompose it.

The actual time series is decomposed into three parts, and it is shown into four parts. The first is the original time series which is just for reference, the second is the Seasonality of the time series, and the third is Trend. And the fourth one is the residual, i.e., remaining, or we can also call it white noise.

Time series = Trend +Season + Remainder

This decomposition is done through the STL, the most renowned time series decomposition method, it uses the Loess [43], which is a method of predicting the non-linear relationship. Decomposition history is also discussed in the section 3.1.

Each dataset is decomposed below:

Nurses and Midwiferies

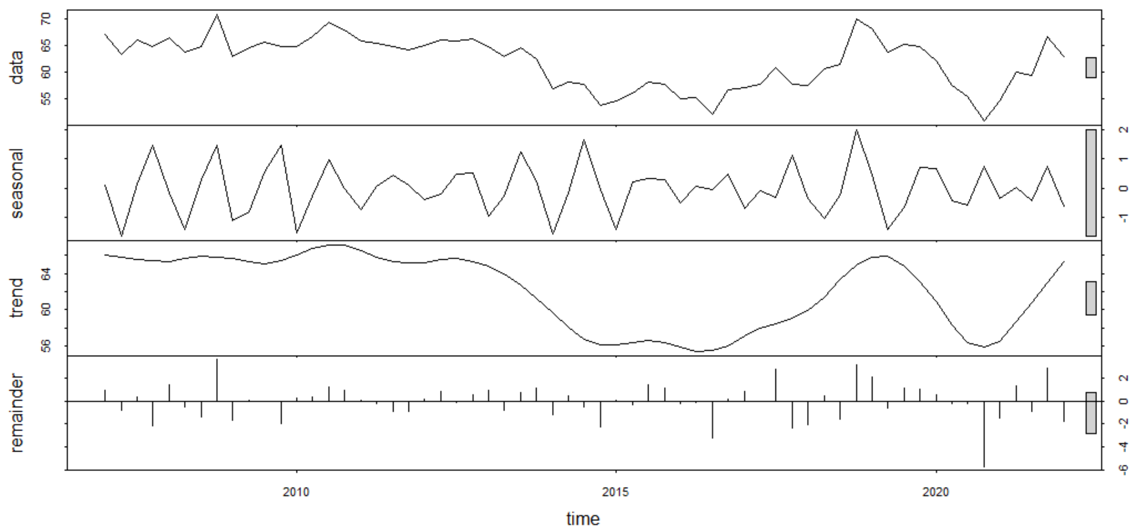


Figure 4.2: Decomposition of Nurses and Midwifery data

The time series of nurses and midwives are decomposed. From the above figure 4.2, we can see that most of the dip came in the 2016 Q3 and 2020 Q4. We can also see the impact of the great recession of 2008 to 2009, where the numbers are low, and after 2010 Q1, it started to grow. The grey bar on the right side of each graph shows the different scales.

Through the season component, we can see that seasonality exists in the data, while through the Trend, we can synthesise that no fixed trend is observed, but due to covid-19, a sudden fall and rise followed the Trend.

General physicians

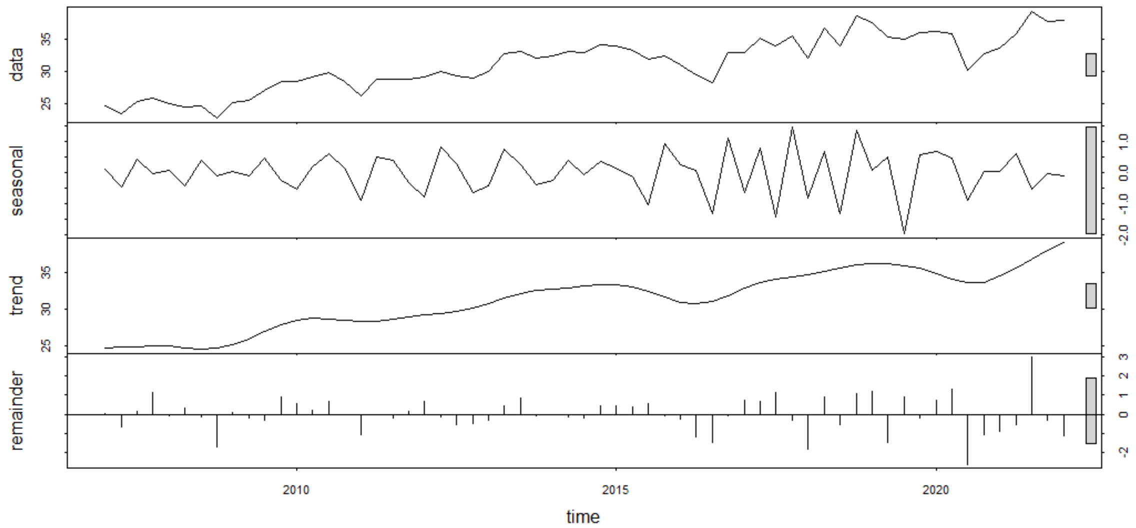


Figure 4.3: Decomposition of General Physician data

From the time series, we see the dip always comes mainly in the third quarter and sometimes in the fourth quarter, but there is growth in the number observed. Through the decomposition of the General physician time series data in the above figure 4.3 Their is a seasonal pattern in the data. For the Trend we see with the passing year trend is on the growing side, where we can see it grew to approx 25,000 in 2007 to almost 37,000 in the UK.

Population of the UK

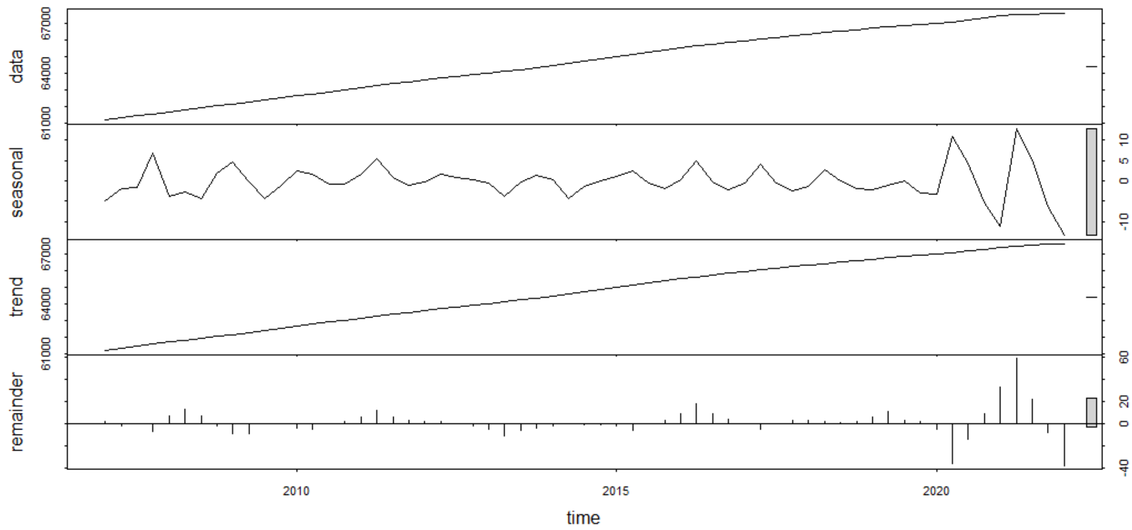


Figure 4.4: Decomposition of Population data

Through Time series data, it can be observed that the population is always on the growing side. We can see the population has risen from 61.32 million in 2007 Q1 to 67.44 million by 2022 Q1.

Through the decomposition of time series data in the above figure 4.4, the seasonality has significantly less impact on the series. Decomposition also suggests that there is no fixed trend in birth, death and immigration season.

The trend is similar to the original time series, and the residual has a slight impact.

GDP of the UK

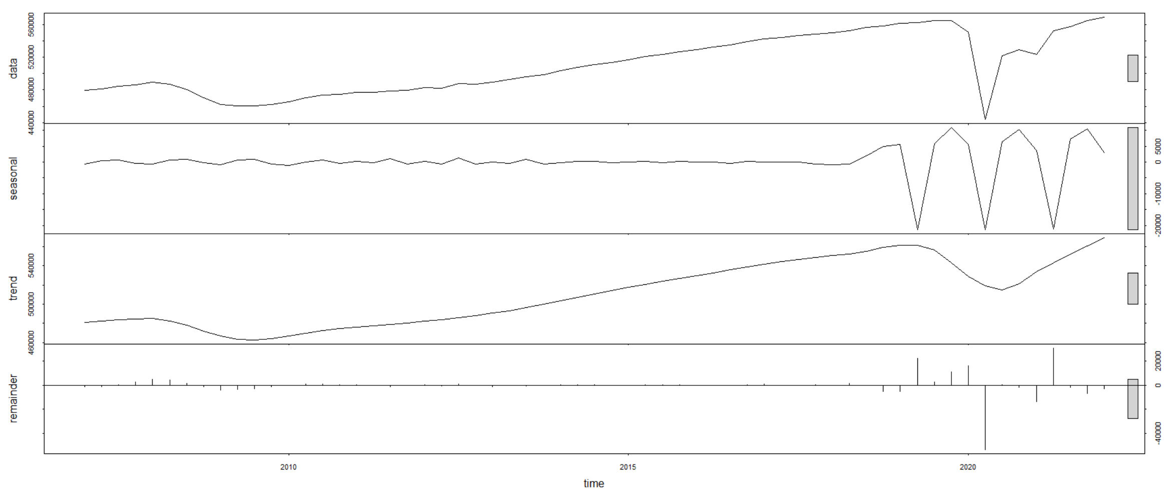


Figure 4.5: Decomposition of GDP data

Through the decomposition of GDP time series data in the above figure 4.5 we can see that the GDP has only observed seasonality in the Covid-19 time, where the decline and incline of the graph can be observed. Other times, there was very slight seasonality.

The Trend is growing except for two instances. The first is when it fell in 2008 -2009 during the great recession, and it started to increase again after 2010. The second is in the Covid-19 pandemic; it fell steeply and started recovering in 2021 and then again showing the growing trend. While the residual is negligible, there is less noise in the time series.

4.1.3 Specifying the model

Many different time series models can be used for forecasting, and multiple books, articles, journals and publications are devoted to discuss them. Selecting an appropriate model for the data is critical for creating accurate predictions. The initial model we choose for our analysis are AR, ARIMA, SARIMA and VAR. The description of these models is here in this subsection 3.2.1.

With the help of the analysis and iteration of the evaluation in this Chapter ahead, we finalised our model SARIMA and VAR because, as we can see from the decomposition, the General physician data and Nurse's data have a seasonality component in it, So the most appropriate and versatile first choice is using Seasonal ARIMA.

We have also chosen the Vector Auto regression model in this case because one Time series will help predict other time series.

4.1.4 Estimate and Evaluate

After specifying the model, the chosen dataset should be transformed into the time series format and, fed to the model and fitted model for the estimation.

After training the model, we will evaluate the model using a diagnostic tool that was discussed in the section 3.2.2

4.1.4.1 Eyeball test: Augmented Dickey-Fuller (ADF)

```
> ndiffs(population_UK, alpha = 0.05, test = c('adf'))
[1] 2
> ndiffs(gdp_UK, alpha = 0.05, test = c('adf'))
[1] 1
> ndiffs(GP, alpha = 0.05, test = c('adf'))
[1] 1
> ndiffs(Nursing_and_midwifery, alpha = 0.05, test = c('adf'))
[1] 1
> 
```

Figure 4.6: Augmented Dickey-Fuller Test

The above figure 4.6 shows that the ADF test is performed where the alpha value represents the p-value with a limit of 0.05. If the value is above 0.05 null hypothesis of non-stationarity cannot be rejected.

The ndiff function finds how much differencing is needed for time series to make the time series stationary, and the result is shown in the table 4.1 below.

Time Series	Differences Required
Population	2
GDP	1
GP	1
Nurses	1

Table 4.1: Augmented Dickey-Fuller Test Results

So all of the time series need single differencing except for population, which requires two differencing to be stationary.

4.1.4.2 ACF and PACF for checking the autocorrelation in a time series

We have plotted the graph of ACF and PACF below for all time series after making the time series stationary i.e. taking the differencing, as discussed in the above subsection 3.2.2.3.

ACF and PACF will help us identify the time series' p and q order.

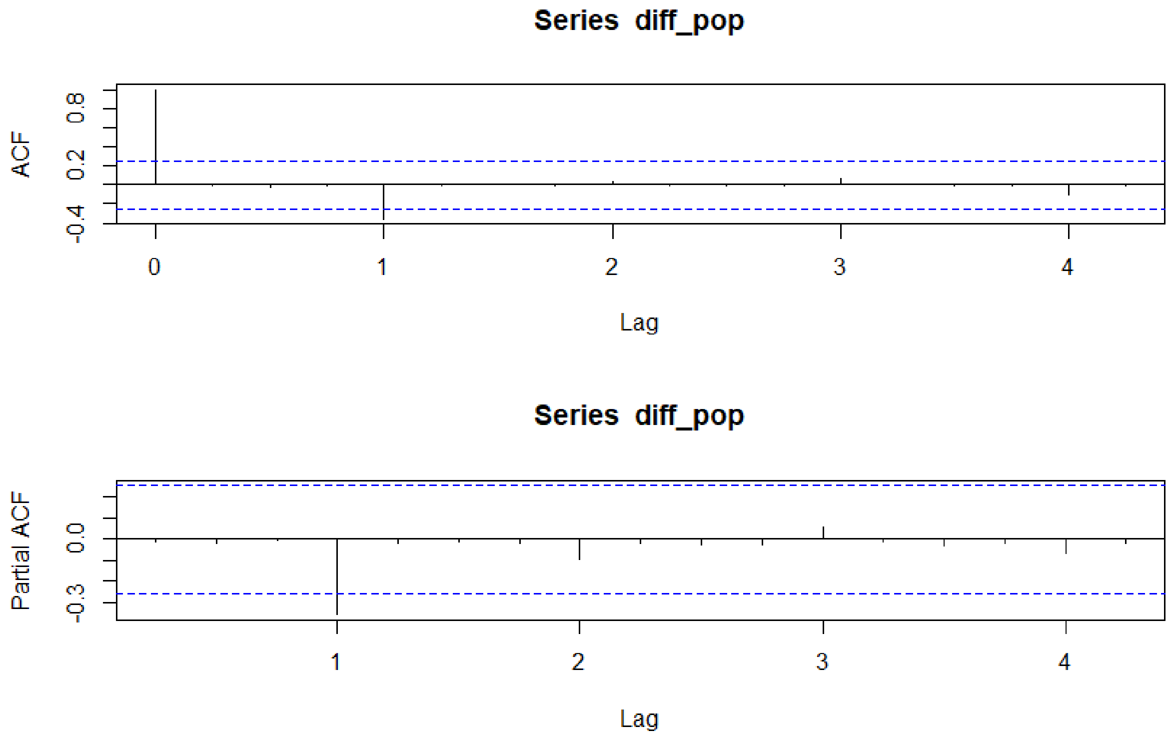


Figure 4.7: ACF and PACF plot of Population Time series

For the population times series the above figure 4.7 depicts the PACF value is 1 lag while the ACF is 1, so we can say AR(1) or MA(1) is suitable for predicting the population.

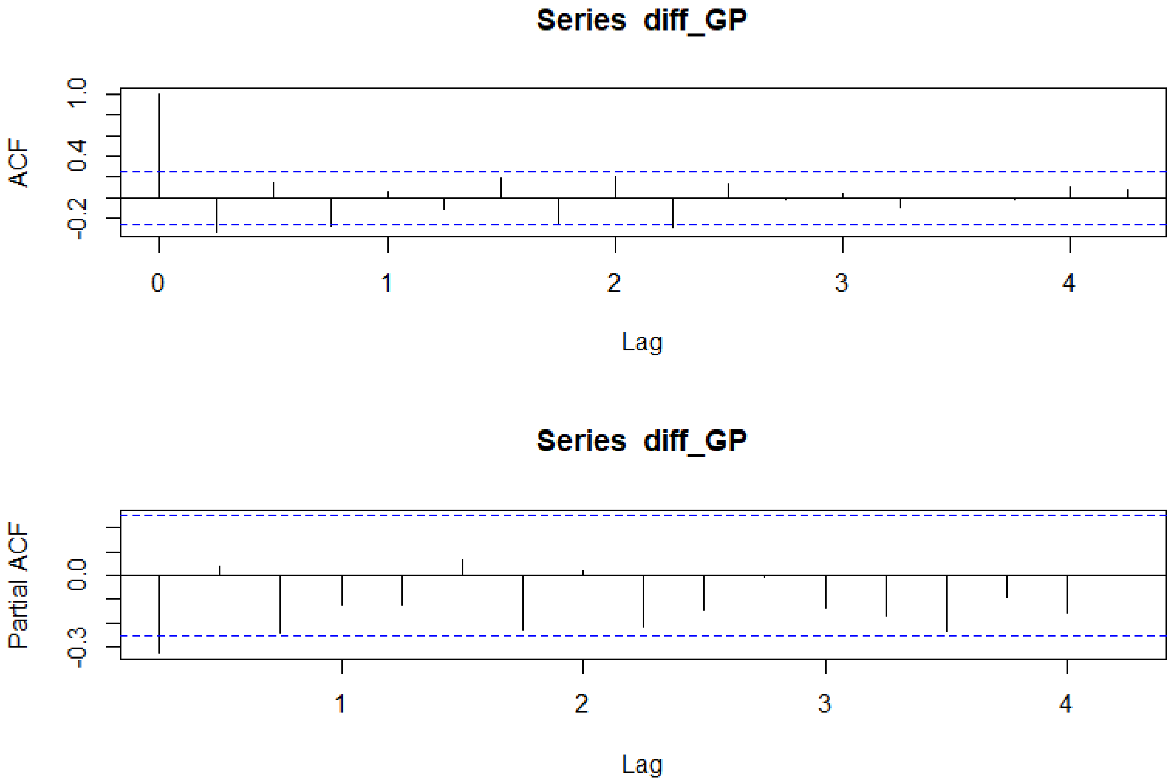


Figure 4.8: ACF and PACF plot of GP Time series

For General physicians the above figure 4.8 depicts that PACF has a threshold value between 3 and 4, and while the ACF is 2.

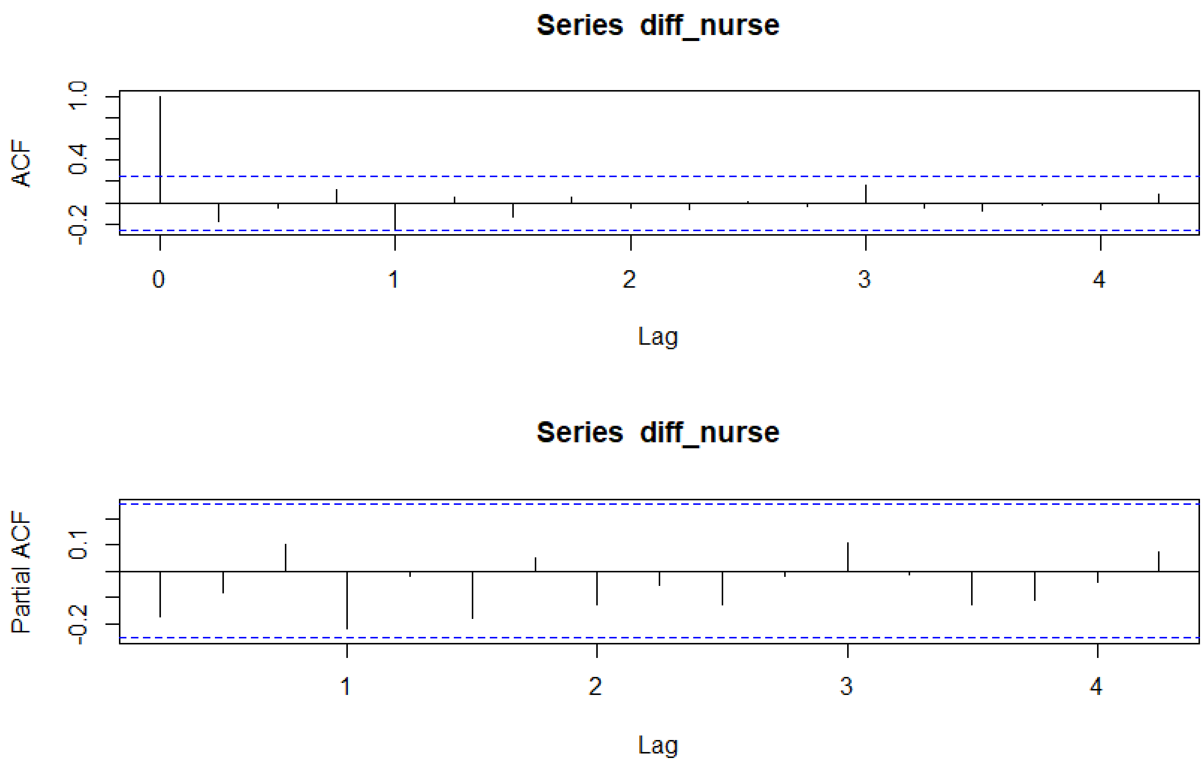


Figure 4.9: ACF and PACF plot of Nurses & Midwiferies Time series

For Nurses & Midwiferies as the above figure 4.9 depicts, we can see the order of ACF is 1 and the other lags are not significant. At the same time, PACF has the same lag of 1.

We have selected two models as specified in the subsection 4.1.3, and we will go through these diagnostics on both the model using the below statistics test.

4.1.4.3 Model parameter selection

These evaluations are performed on the training data as we had one limitation of cross-validation due to limited data and test data falling under COVID-19 times.

We have finalized the Seasonal Autoregressive integrated moving average (SARIMA), and Vector Auto Regression (VAR) model, and the best parameter selection is discussed below for the respective model.

Seasonal Autoregressive integrated moving average (SARIMA)

From the SARIMA, we are forecasting for the Nurses & midwiferies and for General Physician.

From the below table 4.2, we have taken the default model chosen by the R statistical tool AUTO ARIMA which is nothing but the form of SARIMA. The second function SARIMA (p,d,q) (P, D, Q) order is calculated by applying multiple statistical approaches like the

Eyeball test, ADF and PACF, residual diagnostic, and then taking the best possible outcome from the above tests.

Then a combination of the different sets is made to find the best (p,d,q) (P, D, Q) values using an iterative approach on these combinations. (2,1,1) (1,1,1), (2,1,1) (0,1,1), (2,1,1) (1,1,0), (2,1,2) (1,1,1), (2,1,2) (0,1,1), (2,1,2) (1,1,0), (3,1,1) (1,1,1), (3,1,1) (0,1,1), (3,1,1) (1,1,0),

So from above analysis (2,1,1) (1,1,1) order is chosen for Nurses & midwiferies and (3,1,1)(0,1,1) is chosen for General Physician.

Dataset	Model	AIC	AICc	BIC
Nurse	SARIMA(2,1,2)(1,1,1)[4]	295.48	297.81	309.65
	Auto ARIMA(0,1,0)(0,0,1)[4]	303.88	304.09	308.07
GP	SARIMA(3,1,1)(0,1,1)[4]	240.85	242.56	253
	Auto ARIMA (1,1,0)(0,0,1)[4]	245.31	245.52	249.5

Table 4.2: Criterion for model selection

Then these values are compared with the default value produced by Auto ARIMA with SARIMA best value in the table above 4.2.

If we compare SARIMA and Auto ARIMA for AIC, it is 8 and 5 lower, and for AICs, it is 7 and 3 lower for Nurses and midwiferies and General physician data(GP), respectively.

While BIC doesn't have a big difference, the lower the AIC, the better the forecasting model, as discussed above in subsection 3.2.2.2.

Predictive performance

The models are then evaluated on their forecasting performance.

The same (p,d,q) (P, D, Q) order in the above table 4.2 is used here to compare predictive performance.

Dataset	Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Nurse	SARIMA	0.130	2.391	1.807	0.135	2.974	0.773	-0.003
	ARIMA	-0.1310	2.9127	2.2809	-0.3374	3.7078	0.5740	-0.1412
GP	SARIMA	-0.1423	1.6495	1.1762	-0.5742	3.7607	0.8312	-0.0114
	ARIMA	0.2842	1.7912	1.3621	0.7272	4.3564	0.6157	-0.0097

Table 4.3: SARIMA Predictive Performance

From the analysis in the above table 4.3, the MAE is 0.48 and 0.19 less for SARIMA than Auto ARIMA for Nurses & midwiferies and General Physicians data respectively. Similarly, RMSE is 0.52 and 0.15 less for SARIMA than Auto ARIMA for Nurses & midwiferies and General Physicians data respectively.

The lower the Mean Absolute error (MAE) and Root Mean Square Error (RMSE), the better. Hence, the SARIMA model is better for the dataset.

Since Mean Absolute Percentage Error (MAPE) tells the accuracy of the model, so the lesser the error percentage, the better the model. SARIMA is performing better in terms of error as well here.

MASE is also under 1, a benchmark used in most forecasting techniques, so the overall performance of the SARIMA model is perfect for both datasets.

Finally, after all the evaluation (2,1,1) (1,1,1) order is chosen for Nurses & midwiferies and (3,1,1) (0,1,1) is chosen for General Physician.

Vector Autoregressive (VAR) model

One more test is also performed in the section below, which is specific to the VAR model.

Granger Causality Test

The Granger causality test was run on all-time series in the round robin. The observation we found out was that there is a robust causal relationship between the timeseries of General physicians and the population of the UK and the Time series of Nurses & midwiferies with the people of the UK and the results of the successful Granger test are shown in figures below, Where the F value is less than 0.05, as discussed in the subsection 3.2.2.3. GDP does not correlate with any other time series. So we are not further using GDP timeseries. So, the first combination of General physicians and the population of the UK is selected for the VAR model. Refer the test in Figure A1.1 The second combination of Nurses & midwiferies with the population of the UK is selected for the VAR model. Refer the test in Figure A1.2

Criterion for model selection using AIC(n)

We have extracted the AIC value for the fitted model for 30 lag values, which is shown in below Figure 4.10, if we see the AIC scores, for the lag 6 the value is the lowest compared to the other lags. Lags greater than 6 have less AIC score but if we choose a higher lag, we will come across the overfitting problem with the model. 10 also has very less AIC value. So, 6 and 10 will be the lags we evaluate going ahead.

```
> VARselect(yplot, lag.max=30)
$selection
AIC(n)  HQ(n)  SC(n) FPE(n) 
    15     15     15     16 

$criteria
                1           2           3           4           5           6           7           8
AIC(n)    8.629780    8.442759    8.091419    8.119176    7.964626    7.740121    7.987079    7.987926
HQ(n)     8.720253    8.593548    8.302523    8.390596    8.296361    8.132171    8.439444    8.500606
SC(n)     8.907326    8.905336    8.739026    8.951814    8.982295    8.942820    9.374809    9.560686
FPE(n) 5602.652186 4667.797534 3318.465187 3475.730905 3069.120013 2566.600142 3511.155482 3860.935608
                9          10          11          12          13          14   15   16   17   18   19   20
AIC(n)    8.061496    7.747800    7.693529    7.802287    6.365833    4.072012 -Inf -Inf -Inf -Inf -Inf -Inf
HQ(n)     8.634491    8.381112    8.387155    8.556229    7.180091    4.946585 -Inf -Inf -Inf -Inf -Inf -Inf
SC(n)     9.819286    9.690622    9.821381   10.115170    8.863746    6.754956 -Inf -Inf -Inf -Inf -Inf -Inf
FPE(n) 4741.530867 4169.035744 5140.325129 8464.746682 3752.807428 1251.967049  NaN    0    0    0    0    0
         21   22   23   24   25   26   27   28   29   30
AIC(n) -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
HQ(n)  -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
SC(n)  -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
FPE(n)    0    0    0    0    0    0    0    0    0    0
```

Figure 4.10: AIC value for Nurses & Midwiferies and Population Time series

Similarly, for the General Physicians and population time series, as can be seen in figure 4.11, we can tell that lag 6 and 10 have the lowest AIC values. Therefore, we will evaluate both these lag values as well.

```
> VARselect(yplot, lag.max=30)
$selection
AIC(n)  HQ(n)  SC(n) FPE(n) 
    15     15     15     16 

$criteria
                1           2           3           4           5           6           7           8
AIC(n)    7.599850    7.531900    7.603055    7.257418    7.103876    6.861518    6.943943    6.980696
HQ(n)     7.690323    7.682688    7.814159    7.528837    7.435610    7.253568    7.396308    7.493377
SC(n)     7.877396    7.994476    8.250662    8.090055    8.121544    8.064217    8.331672    8.553456
FPE(n) 2000.325298 1877.287269 2036.308585 1468.212854 1297.760635 1066.070042 1237.148447 1410.127022
                9          10          11          12          13          14   15   16   17   18   19   20
AIC(n)    7.018694    6.703621    6.555032    5.374418    4.476354 -0.1779492 -Inf -Inf -Inf -Inf -Inf -Inf
HQ(n)     7.591690    7.336932    7.248659    6.128361    5.290612  0.6966236 -Inf -Inf -Inf -Inf -Inf -Inf
SC(n)     8.776485    8.646442    8.682884    7.687301    6.974268  2.5049946 -Inf -Inf -Inf -Inf -Inf -Inf
FPE(n) 1671.227587 1467.419275 1646.447398 746.799301 567.238961 17.8590450  NaN    0    0    0    0    0
         21   22   23   24   25   26   27   28   29   30
AIC(n) -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
HQ(n)  -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
SC(n)  -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf -Inf
FPE(n)    0    0    0    0    0    0    0    0    0    0
```

Figure 4.11: AIC value for GP and Population Time series

VAR Evaluation

Dataset	Lag	Multiple R^2	Adjusted R^2
GP	6	0.785	0.7236
	10	0.8338	0.723
Nurse & Midwifery	6	0.8381	0.7919
	10	0.8721	0.791

Table 4.4: VAR model Predictive performance

VAR only uses regression technique i.e. Auto regressive in the form of vector; for regression, one of the best methods for regression analysis is the R squared value. Multiple R-squared is a simple R-squared measure for models with various predictor variables. When we add more predictors, the R-squared value will increase.

The model may have been overfitted if there is a significant disparity between the multiple and adjusted R-squared. As a result, it demonstrates a balance between the most parsimonious model and the best fitting model. In the above table, no vast difference between Multiple and Adjusted R-squared can be observed.

From the above table 4.4 For General Physicians, Lag 6 is more suitable because there is less difference between Multiple and Adjusted R squared, as discussed above.

For Nurses and Midwifery, lag 6 is more suitable with less gap between Multiple and Adjusted R squared, and 83% is a reasonable accuracy.

4.1.5 Forecast and Results

After multiple iterations for visualizing, changing models, and evaluating their performance, we have selected the best model based on various statistical tests and performance measuring techniques. Finally, we are plotting the results we extracted by fitting those models.

Results from SARIMA model

SARIMA can also be represented as ARIMA (p,d,q) (P, D, Q)[S]. Both of them represent the same thing.

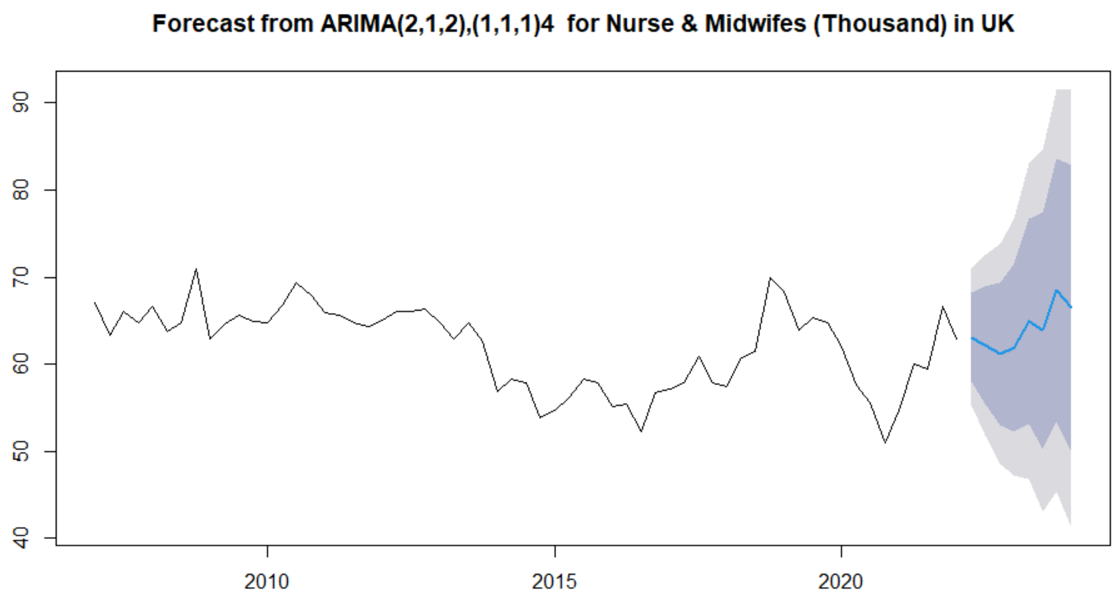


Figure 4.12: SARIMA forecast for Nurses & Midwiferies

From the above figure 4.12 for Nurses and midwives, we have forecasted the results for eight quarters ahead. The black line is the original data from 2007 Q1 to 2022 Q1, and the blue line is predicted results from 2022 Q2 to 2024 Q1. If we see 2021 Q4 had the number 66.7 thousand, then it fell to 62.9 and to follow this seasonality trend it fell again in next quarter, but it will rise again after 2022 Q3 and follow this growing trend.

The violet wave above and below forecasted line tells the possible maximum and minimum value of the forecast with 80% confidence, while the grey wave depicts 95% confidence interval. As we move further in the future, the confidence interval expands because when we predict close in the future, the Time series gives a perfect prediction. Results have more errors, and confidence decreases when we go further in the future. So, for the near future forecast, the Time series is a perfect model.

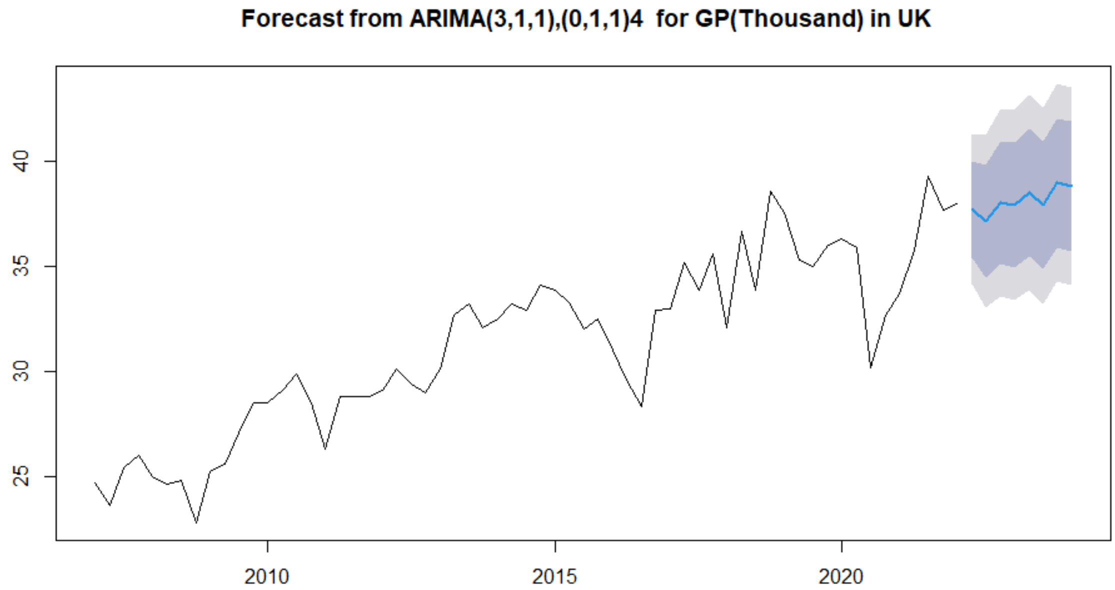
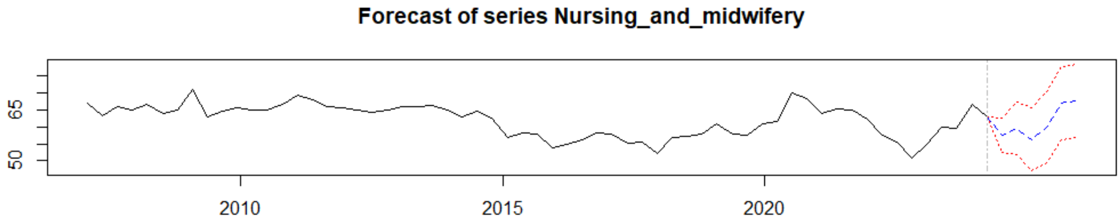


Figure 4.13: SARIMA forecast for GP

Similarly, we forecasted the results for eight quarters ahead for General Physicians which can be seen in figure 4.13. The black line is the original data from 2007 Q1 to 2022 Q1, and the blue line is predicted results from 2022 Q2 to 2024 Q1. In 2021 Q3, the number of doctors reached its peak, i.e. 39.5 thousand and after forecasting it showed slight seasonality, where in 2022 Q2 it fell and then in next quarter, i.e. Q3 it rose and till 2024 Q1 it is following this similar trend. But the overall trend is that the number of GP will be increasing.

Results from VAR model

For the VAR model, our main aim is to predict the workforce requirements for the healthcare workers, but the population is also predicted.



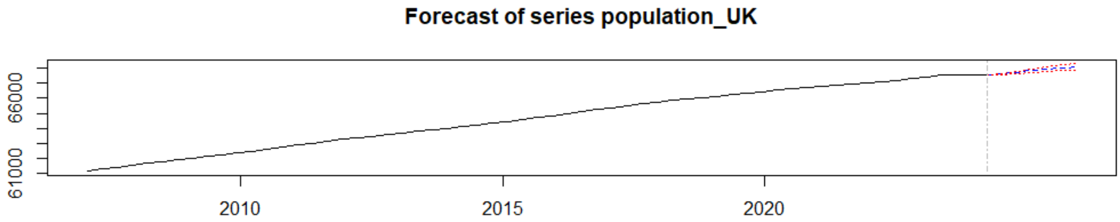


Figure 4.14: VAR forecast for Nurses & Midwiferies

While seeing the above results in figure 4.14 for Nursing and Midwifery, we have forecasted the results for six quarters ahead. The black line is the original data from 2007 Q1 to 2022 Q1, and the blue dotted line is predicted results from 2022 Q2 to 2023 Q3. A vertical line on the x-axis value 2022 Q1 separates original data from forecasted data.

The forecasted data shows similar trends to the SARIMA model, where the number falls for the first two predicted quarters and then starts growing.

The dotted red line around the forecasted line tells the possible maximum and minimum value of the prediction with 80% confidence.

While for population, there is a growing trend in the population of the UK, and it will continue to grow in the future, so it was very simple forecasting, and the red dotted line is very close as well.

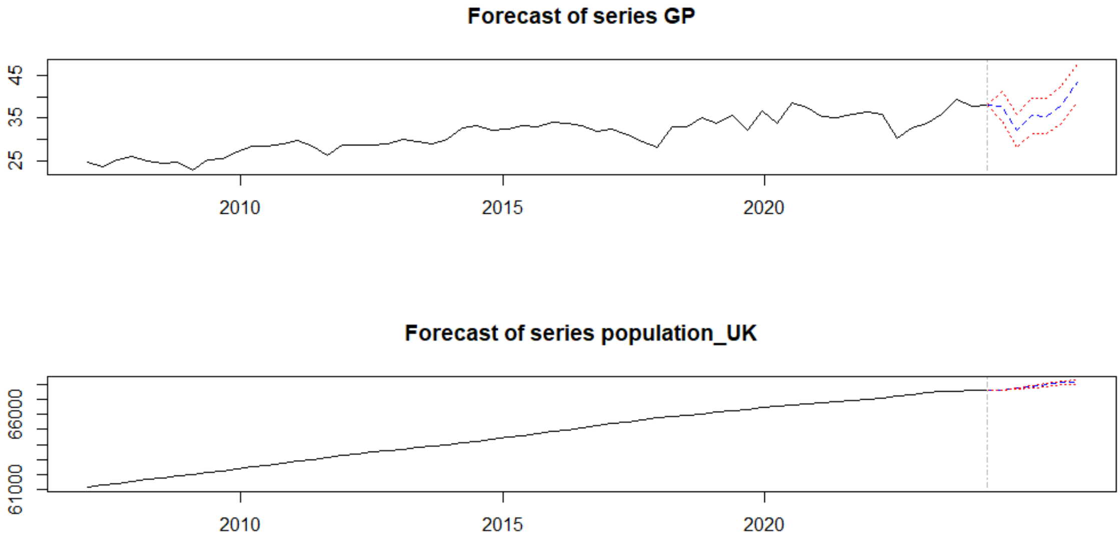


Figure 4.15: VAR forecast for GP

Similarly, we forecasted the results for General Physicians and Population six quarters ahead in figure 4.15. The black line is the original data from 2007 Q1 to 2022 Q1, and the blue line is predicted results from 2022 Q2 to 2023 Q3. The vertical line on 2022 Q1 separates the forecast and original data.

The forecasted data fell after the 2022 Q1, but in 2022 Q3, it rises again and see a constant growth trend.

5 Conclusion and Future Work

5.1 Conclusion

In this study, we have explored the different approaches to Workforce planning in the healthcare field and the other models suitable for planning the workforce. We took the dataset of the UK's General physicians and combined data of Nurses & Midwiferies. Then it is pre-processed into a Time series format by doing some transformation. We also extracted the data of the UK population and the UK GDP to check how it impacts the healthcare demand and how this can be leveraged in forecasting. Trends and some seasonality were present in healthcare workers' data, which was incorporated using the Seasonal Arima model by tuning the different hyperparameters. For model selection, we have used the Bayesian Information Criterion (BIC), the Corrected Akaike Information Criterion (AICc), and the Akaike Information Criterion (AIC) and compared the best value for the model. Also, different statistical tests were leveraged to find the best parameter and model, like the Eyeball test of Autocorrelation and Partial autocorrelation to see the lag order, and the Augmented Dickey-Fuller (ADF) for testing the stationarity of time series.

Similarly, to find the causal relationship between the two time series we used the Granger causality null hypothesis test through which we discovered the relation between population and healthcare worker. so we used the above tests to finalize our second model, i.e., the Vector Auto correlation model, which helped us in forecasting the future demand of healthcare workers.

The results were evaluated through Root Mean Square (RMSE), R-squared error (Multiple and Adjusted R squared), Mean Absolute error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE).

Overall, the General Physician prediction was more accurate with less error, and the SARIMA model was suitable for detecting seasonal patterns. At the same time, VAR was good with a forecast with other dependent time series. The time series method is the easiest and most efficient way of predicting linear data with high accuracy. The results suggest that from 2022 Q3, the demands of General Physicians and Nurses & midwives will increase.

5.2 Future Work

The study unavoidably has certain constraints because of the short time to complete the dissertation. The following enhancement can be added in the future:

1. Till now, we have worked on forecasting the demand using the supply, and the next step will look for the factor affecting supply and predict that supply so that we can complete the supply-demand data to make our workforce model more robust.

2. One more advanced predictive model can be used to predict time series. For instance, Prophet was developed by Facebook in 2017 [44], and it is open source. It is used for forecasting with consideration of seasonality along with the holiday component. It can be tuned according to the different domains and situations, and it is robust for shifting trends like the one we saw in the Nursing and Midwiferies data.

3. Sensitivity analysis is how the results will be impacted by changing the input variable like we have used the population to predict the output.

4. Cross-validation was unsuccessful due to the limited number of data points, i.e. 61. A good training data set for the time series model should have at least 50 data points. For cross-validation, time series-specific models' data can only be divided in a way where training data should be before the test data, unlike train test split, where data is split randomly. We tried to do that, but test data entries fell in the covid-19 times, which gave the wrong prediction of sudden shifts in the pattern. After 2021 Q4, things are returning to normal, which will provide us with more data in the future for cross-validation.

Bibliography

[1] John Sullivan. Why you need workforce planning. *Workforce*, 81(12):46–50, 2002.

[2] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[3] Ann Cotten. *Seven steps of effective workforce planning*. IBM Center for the Business of Government, 2007.

[4] Chartered Institute of Personnel and Development. Workforce planning, 2022. URL `https://www.cipd.ie/news-resources/practical-guidance/factsheets/workforce-planning#gref`.

[5] JOANNA O'Riordan. Workforce planning in the irish public service. *State of the Public Services Series. Research Paper*, (7), 2012.

[6] Karen Bloor, Alan Maynard, Jane Hall, P Ulmann, O Farhauer, and B Lindgren. *Planning human resources in health care: towards an economic approach: an international comparative review*. Canadian Health Services Research Foundation= Fondation canadienne de la ..., 2003.

[7] Anthony Scott, Peter Sivey, Catherine Joyce, Deborah Schofield, and Phillip Davies. Alternative approaches to health workforce planning. *Adelaide: National Health Workforce Planning and Research Collaboration*, 2011.

[8] John Ansah, Victoria Koh, Dirk De Korne, Steffen Bayer, Chong Pan, Thiyagarajan Jayabaskar, David Matchar, and Desmond Quek. Comparing health workforce forecasting approaches for healthcare planning: the case for ophthalmologists. *International Journal of Healthcare*, 3(1):84–96, 2017.

[9] Jerry Cromwell, Margo L Rosenbach, Gregory C Pope, Barbara Butrica, and Jane D Pitcher. Crna manpower forecasts: 1990-2010. *Medical Care*, pages 628–644, 1991.

[10] Norbert Dreesch, Carmen Dolea, Mario R Dal Poz, Alexandre Goubarev, Orvill Adams, Maru Aregawi, Karin Bergstrom, Helga Fogstad, Della Sheratt, Jennifer Linkins, et al.

An approach to estimating human resource requirements to achieve the millennium development goals. *Health policy and planning*, 20(5):267–276, 2005.

[11] Serge Resnikoff, William Felch, Tina-Marie Gauthier, and Bruce Spivey. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners. *British Journal of Ophthalmology*, 96(6):783–787, 2012.

[12] Dominique Roberfroid, Christian Leonard, and Sabine Stordeur. Physician supply forecast: better than peering in a crystal ball? *Human resources for health*, 7(1):1–13, 2009.

[13] Linda O'Brien-Pallas, Stephen Birch, Andrea Baumann, and Gail Tomblin Murphy. *Integrating workforce planning, human resources and service planning*. ITGPress, 2003.

[14] James Wong, Albert Chan, and Yat Hung Chiang. A critical review of forecasting models to predict manpower demand. *Construction Economics and Building*, 4(2): 43–56, 2004.

[15] Linda O'Brien-Pallas, Andrea Baumann, Gail Donner, Gail Tomblin Murphy, Jacquelyn Lochhaas-Gerlach, and Marcia Luba. Forecasting models for human resources in health care. *Journal of advanced nursing*, 33(1):120–129, 2001.

[16] Jaydip Sen and Tamal Datta Chaudhuri. A time series analysis-based forecasting framework for the indian healthcare sector. *arXiv preprint arXiv:1705.01144*, 2017.

[17] Ireneous N Soyiri and Daniel D Reidpath. An overview of health forecasting. *Environmental health and preventive medicine*, 18(1):1–9, 2013.

[18] Francesca Senese, Paolo Tubertini, Angelina Mazzocchetti, Andrea Lodi, Corrado Ruozi, and Roberto Grilli. Forecasting future needs and optimal allocation of medical residency positions: the emilia-romagna region case study. *Human Resources for Health*, 13(1):1–10, 2015.

[19] Dennis A Ostwald and David Klingenberger. Using morbidity and income data to forecast the variation of growth and employment in the oral healthcare sector. *Health economics review*, 6(1):1–11, 2016.

[20] Hugo J Finarelli Jr and Tracy Johnson. Effective demand forecasting in 9 steps: shifts in demand for a hospital's services can occur unexpectedly. demand forecasting can help you prepare for these shifts and avoid strategic missteps. *Healthcare Financial Management*, 58(11):52–58, 2004.

[21] Lisa M Schweigler, Jeffrey S Desmond, Melissa L McCarthy, Kyle J Bukowski, Edward L Ionides, and John G Younger. Forecasting models of emergency department crowding. *Academic Emergency Medicine*, 16(4):301–308, 2009.

[22] Alfred J Beech. Market-based demand forecasting promotes informed strategic financial planning. *Healthcare Financial Management*, 55(11):46–56, 2001.

[23] Marc Nerlove, David Grether, and José L. Carvalho. *Analysis of Economic Time Series*. Elsevier, 1 edition, 1979. URL https://EconPapers.repec.org/RePEc:eee:monogr:9780125157506.

[24] G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.

[25] Jie Zhao, Limei Xu, and Lin Liu. Equipment fault forecasting based on arma model. In *2007 International Conference on Mechatronics and Automation*, pages 3514–3518. IEEE, 2007.

[26] P GE. Box, gm jenkins, time series analysis: forecasting and control, revised ed, 1976.

[27] Saeid Dindarloo. Reliability forecasting of a load-haul-dump machine: a comparative study of arima and neural networks. *Quality and Reliability Engineering International*, 32(4):1545–1552, 2016.

[28] Adhistya Erna Permanasari, Indriana Hidayah, and Isna Alfi Bustoni. Sarima (seasonal arima) implementation on time series to forecast the number of malaria incidence. In *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 203–207, 2013. doi: 10.1109/ICITEED.2013.6676239.

[29] Julia A Beiser and Steven E Rigdon. Bayes prediction for the number of failures of a repairable system. *IEEE Transactions on Reliability*, 46(2):291–295, 1997.

[30] Mehdi Khashei, Mehdi Bijari, and Gholam Ali Raissi Ardali. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (anns). *Neurocomputing*, 72(4-6):956–967, 2009.

[31] Ting Ma, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan. Quantitative estimation of urbanization dynamics using time series of dmsp/ols nighttime light data: A comparative case study from china's cities. *Remote Sensing of Environment*, 124: 99–107, 2012.

[32] Firdos Khan, Alia Saeed, and Shaukat Ali. Modelling and forecasting of new cases, deaths and recover cases of covid-19 by using vector autoregressive model in pakistan. *Chaos, Solitons & Fractals*, 140:110189, 2020.

[33] Qing Fan and Hongqin Fan. Reliability analysis and failure prediction of construction equipment with time series models. *Journal of Advanced Management Science Vol*, 3 (3), 2015.

[34] Sana BuHamra, Nejib Smaoui, and Mahmoud Gabr. The box–jenkins analysis and neural networks: prediction and time series modelling. *Applied Mathematical Modelling*, 27(10):805–815, 2003.

[35] Lars Fiva Skarbøvik. Forecasting house prices in norway: A univariate time series approach. Master's thesis, Universitetet i Tromsø, 2013.

[36] Ramiro Serra and Andrés C. Rodríguez. The ljung-box test as a performance indicator for vircs. In *International Symposium on Electromagnetic Compatibility - EMC EUROPE*, pages 1–6, 2012. doi: 10.1109/EMCEurope.2012.6396834.

[37] Gebhard Kirchgässner, Jürgen Wolters, and Uwe Hassler. Introduction to modern time series analysis. 2007.

[38] Galit Shmueli and Kenneth C Lichtendahl Jr. *Practical time series forecasting with r: A hands-on guide*. Axelrod Schnall Publishers, 2016.

[39] Olivier Renaud and Maria-Pia Victoria-Feser. A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, 140(7):1852–1862, 2010.

[40] Office for National Statistics UK. Office for national statistics uk, 2022. URL https://www.ons.gov.uk/.

[41] Office for National Statistics UK. Gdp data of the uk, 2022. URL https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/abmi/ukea.

[42] Office for National Statistics UK. Soc data of nurses and midwifery and gp, 2022. URL https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010.

[43] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73, 1990.

[44] Facebook Open Source. Prophet model, 2022. URL https://facebook.github.io/prophet/#:~:text=Prophet%20is%20a%20procedure%20for,several%20seasons%20of%20historical%20data.

A1 Appendix

A1.0.1 Granger Casuality results

```
> grangertest(population_UK, GP, order = 4)
Granger causality test

Model 1: GP ~ Lags(GP, 1:4) + Lags(population_UK, 1:4)
Model 2: GP ~ Lags(GP, 1:4)
  Res.Df Df      F  Pr(>F)
1     48
2     52 -4 2.3486 0.06755 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> grangertest(GP, population_UK, order = 4)
Granger causality test

Model 1: population_UK ~ Lags(population_UK, 1:4) + Lags(GP, 1:4)
Model 2: population_UK ~ Lags(population_UK, 1:4)
  Res.Df Df      F  Pr(>F)
1     48
2     52 -4 3.0969 0.02397 *
---
```

Figure A1.1: Granger Causality Test on GP and Population Time series

```
> grangertest(population_UK, Nursing_and_midwifery, order = 3)
Granger causality test

Model 1: Nursing_and_midwifery ~ Lags(Nursing_and_midwifery, 1:3) + Lags(population_UK, 1:3)
Model 2: Nursing_and_midwifery ~ Lags(Nursing_and_midwifery, 1:3)
  Res.Df Df      F   Pr(>F)
1     51
2     54 -3 6.6374 0.000717 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> grangertest(Nursing_and_midwifery, population_UK, order = 3)
Granger causality test

Model 1: population_UK ~ Lags(population_UK, 1:3) + Lags(Nursing_and_midwifery, 1:3)
Model 2: population_UK ~ Lags(population_UK, 1:3)
  Res.Df Df      F  Pr(>F)
1     51
2     54 -3 3.2388 0.02957 *
---
```

Figure A1.2: Granger Causality Test on Nurses & Midwiferies and Population Time series

A1.0.2 Code link

The code is also deployed at github link