***Abstract*—An explanation behind the initial treatment of the data is discussed before the (S)ARIMA models are iteratively fitted to the treated data, and a short introduction to the apparent seasonality of the data is also included. Models obtained from separate data sets are then compared by analysing their log-likelihoods and Akaike Information Criterion (AIC) values. Then, 3 types of forecasting for the data are explored and analysed by the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The models are then transferred across data sets, and the sufficiency of the fits of the models are tested to find whether the data sets can be simulated by one shared model. Finally, a summary of the findings are given, as well as a brief outline of the benefits and drawbacks of the methods used, as well as further methods that could have been used on the same data set.**

I. Introduction

The use of a time series analysis approach to analyse and model data arising from the COVID-19 pandemic is motivated by the behaviour of the data points around the general trend, whereby fluctuations show temporal dependency. We also see periodic peaks in the data, which could potentially be modelled using a Seasonal approach. The use of ARIMA to model data related to COVID-19 has previously been attempted in [1], and we wish to expand on this by considering seasonality, different forecasting approaches, and comparing the models that arise form different countries. If the models we arrive at are accurate, they could be used to predict future behaviour of the pandemic. For example, if a continued increase in cases was predicted, this could be used by governments or health authorities to prepare for the increase. On the other hand, if a prolonged decrease was predicted, governments could prepare to ease restrictions.

II. Methods

A. Preliminary Analysis

In this study we focused on daily cases and deaths of three European countries: France, Germany and the UK. We used data from the date of the first case onwards in each of these countries to fit our models. Our models (after differencing if needed) will take the general form

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} \\ + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} + e_t \end{aligned}$$

where we seek to estimate p, q and the parameters ϕ_i and θ_j for $1 \leq i \leq p$ and $1 \leq j \leq q$, and e_t is white noise. The analysis and model fitting will all be done using the statistical package R. In the rest of this section, we will focus on UK's daily death data, so as to demonstrate our methodology for a given data set. Firstly, we check the stationarity of the time series by plotting data as shown on the left of Figure 1. There is clear non-stationarity, so we proceed to difference the data which gives us the stationary series in Figure 1 on the right. Stationarity of this data is confirmed by conducting a Augmented Dicky Fuller Test (ADF). Studying the figures we notice heteroscedasticity, which is a feature that cannot be captured by an ARIMA model, and therefore we need to try a different approach.

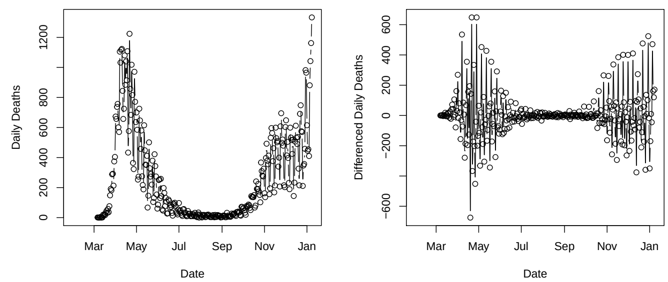

Fig. 1.

We now try logging the data, since it is a common transformation in time series analysis and may help to stabilise the data.

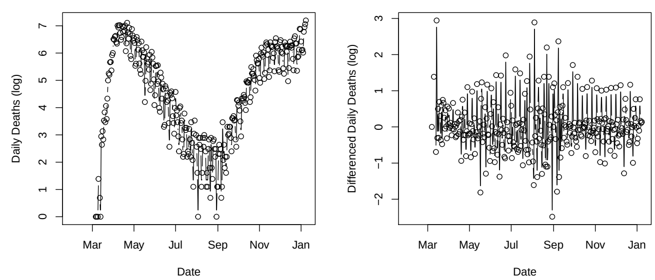

Fig. 2.

In addition to this the spread of disease often follows an exponential growth, motivating a log transformation to manipulate the data into something better suited to an ARIMA approach. The logged data can be seen in Figure 2. Now we must difference the log data to achieve stationarity, seen in Figure 2 on the right. As we can see this data produces a model much better suited to an ARIMA due to its seemingly constant variance.

B. Fitting the Model

We will try two different approaches, a seasonal and non-seasonal model, to achieve the best possible model to help us forecast COVID-19 . In order to avoid repetition, we will focus our demonstration in this paper on the Seasonal ARIMA model. We will use the Box-Jenkins method [2], whereby we add complexity to the model one step at a time with an iterative approach. Initial parameters of the SARIMA model are estimated by autocorrelation function (ACF) graph and partial autocorrelation (PACF).

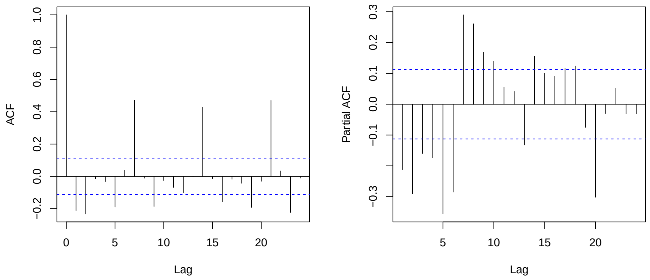

Fig. 3.

We will start with an SARIMA$(0,0,0)(1,0,0)(7)$ model. This is motivated by the large repeating spikes in the ACF at multiples of 7, with far smaller measurements in the PACF. Now, by using the tsdiag function we can analyse the performance of this model by looking at the ACF of its residuals and by running the Ljung-Box test on the residuals.

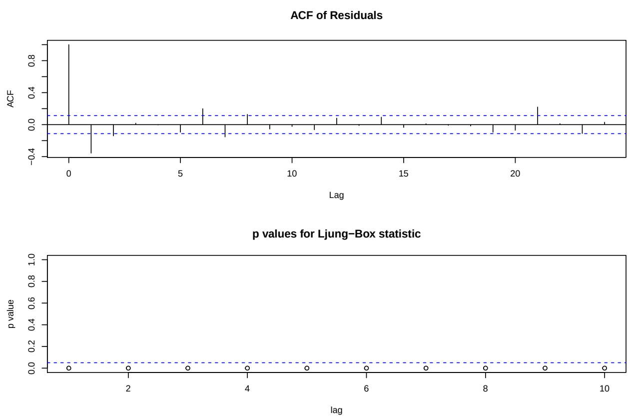

Fig. 4.

As we can see in Figure 4, we reject this model by the Ljung-Box test, and next try the SARIMA$(0,0,1)(1,0,0)(7)$ model. We try this model because there was a large non-repeating spike in the ACF at lag 1. We continue this iterative process of choosing a model, analysing the residuals, and altering the model several times, eventually arriving at a final model when a model passes the Ljung-Box test. We repeat this process for Daily Cases data, and repeat the whole process for Germany and France's data also.

C. Model Comparison

We will demonstrate two main approaches to model comparison in this paper. The first is to compare the log-likelihood (LL) and AIC of models. It is clear to choose one model over the other if one model has both a higher log-likelihood and a lower AIC. The other approach is to compare the forecasting abilities of the two models. Within this approach, we can test both fixed origin and rolling origin forecasting, and compare the results using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Whilst we will demonstrate the latter method for models fitted to the UK Death Data, we must note that this approach can be very computationally intensive, and we will therefore use the former approach for the mass comparison of models used when we discuss multiple countries and daily cases data. When we later test whether or not models can be transferred between countries, we will also first use the Ljung-Box test before either of these methods. If a model does not pass this test for a given country's data set, we reject it outright.

III. Results

A. UK Models

For the UK's daily death data, we arrive at an SARIMA$(0,1,6)(1,0,1)(7)$ model for the undifferenced log data. We chose this model despite an SARIMA$(0,1,1)(1,0,1)(7)$ model technically passing the Ljung-Box test, since it only barely did so at lag 6, motivating a further search for a stronger model. The SARIMA$(0,1,6)(1,0,1)(7)$ model has a higher log-likelihood and lower AIC (see Table I), so we select this model over SARIMA$(0,1,1)(1,0,1)(7)$. Using a similar iterative method for a non-seasonal approach, we arrived at an ARIMA$(7,1,1)$ model. Models for daily cases data and the other countries shall be investigated later.

B. Investigating Forecasts for UK Daily Deaths

Model	Log-Likelihood	AIC
ARIMA(7,1,1)	-242.37	502.73
SARIMA(0,1,1)(1,0,1)(7)	-211.5	433.01
SARIMA(0,1,6)(1,0,1)(7)	-202.33	424.67

TABLE I

Now comparing these models as you can see from the data above that the SARIMA performs better than the ARIMA as it produces a lower AIC and a greater LL. We now turn our attention to the analysis of forecasting ability.

1) Fixed Origin Forecast

Model	MAE	RMSE
ARIMA(7,1,1)	246.17	356.13
SARIMA(0,1,6)(1,0,1)(7)	157.55	238.94

TABLE II

We will start by discussing our results for the fixed origin forecast. As we can see above they are both very high in terms of MAE and RMSE; however, the non-seasonal graph on the left of Figure 5 shows the forecast tapering off the further we try to predict into the future, and doesn't follow the same upwards trend and weekly fluctuations that the seasonal model does. So by a heuristic observation here we can say that the seasonal model is better suited for forecasting longer periods ahead.

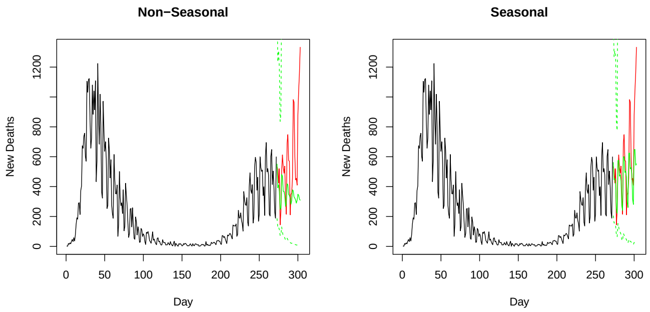

Fig. 5.

2) Rolling Day-Ahead Forecast

Model	MAE	RMSE
ARIMA(7,1,1)	50.96	91.98
SARIMA(0,1,6)(1,0,1)(7)	42.02	79.24

TABLE III

In Figure 6, the green line represents the prediction of a day's death rate given all of the data from up to the day before, with confidence intervals excluded to avoid cluttering. We forecast from day 50 onwards, to give a burn-in period. In this case, the SARIMA model performs better in terms of day-ahead forecasting, as shown by the lower MAE and RMSE.

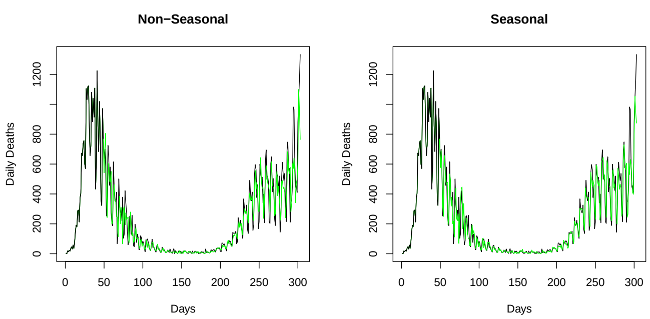

Fig. 6.

3) Rolling Week-Ahead Forecast

Model	MAE	RMSE
ARIMA(7 ,1,1)	52.82	99.58
SARIMA(0,1,6)(1,0,1)(7)	53.67	96.10

TABLE IV

Similarly to before, the green line represents a model's prediction for a given day given all the data from up to a week before. We see that the ARIMA model is better performing in terms of MAE and the SARIMA is better performing in terms of RMSE. As the square difference in RMSE "amplifies" larger error terms, this indicates that while on average the ARIMA model is more accurate than SARIMA by a narrow margin, the ARIMA model does feature some larger individual errors.

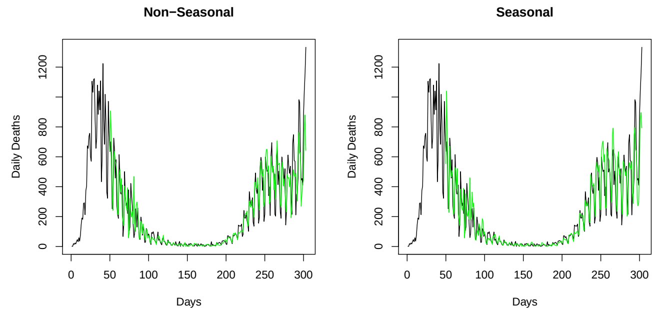

Fig. 7.

C. *Other countries and Model Transference*

We will now shift our attention over to the Case and Death data of the UK, France and Germany, in order to determine whether it is the case that the models obtained from the data are transferable across different data-sets. This would be useful in the sense that well performing transferable models could be implemented as a starting point or preliminary estimation of how new Covid outbreaks may behave - to base new containment strategies/policies from etc - therefore resulting in a more effective initial controlling of the spread. In turn, more lives would be saved, and the affected area would be economically hit less hard. Also, comparing how our models perform when pitted against other data-derived models could potentially highlight any errors in the generation of our models, ensuring that we obtain an even more accurate model - or at least minimise the error in the selection of our model.

We begin with determining whether, generally speaking, the models we focus on are seasonal as opposed to non-seasonal. To do this, we compare LL and AIC scores. Using the previously mentioned methodology (Box Jenkins etc), we fit both ARIMA and SARIMA models for the three countries' Case Data - with corresponding AIC scores and LLs:

Country	Model	Log-Likelihood	AIC
UK	ARIMA(0,1,2)	-147.04	300.07
UK	SARIMA(0,1,2)(0,0,1)(7)	-144.37	296.74
France	ARIMA(2,1,4)	-687.38	1390.77
France	SARIMA(2,1,4)(0,0,4)(7)	-677.16	1378.31
Germany	ARIMA(1,1,1)	-336.12	678.24
Germany	SARIMA(1,1,1)(1,0,0)(7)	-287.26	582.53

On each model for the Case Data, it is clear to see that the SARIMA models meet our criteria for being the favourable model. To test whether seasonality represents a more accurate base model further, it is important to also check this using the Death Data:

Country	Model	Log-Likelihood	AIC
UK	ARIMA(7,1,1)	-242.37	502.73
UK	SARIMA(0,1,6)(0,0,1)(7)	-222.47	462.94
France	ARIMA(5,1,3)	-541.88	1103.75
France	SARIMA(5,1,3)(1,0,1)(7)	-521.37	1066.74
Germany	ARIMA(3,1,2)	-378.52	771.03
Germany	SARIMA(3,1,2)(1,0,0)(7)	-368.18	752.37

Similarly, the Death Data's SARIMA models for all countries outperformed the ARIMA models in terms of LL and AIC, confirming that we should move forward only considering the seasonal models.

With our UK SARIMA Case and Death models now obtained, we will see how these perform when fitted to France and Germany's Case and Death Data, carefully ensuring that any AIC scores and LLs are only recorded following the model fit passing the Ljung Box test.

Our UK seasonal Case and Death Model passed the Ljung Box test for each fit to each of our different country's two (Case and Death) data-sets, resulting in the following LLs/AIC scores:

Data-set	Log-Likelihood	AIC
Germany (Cases)	-690.23	1388.47
Germany (Deaths)	-348.21	714.43
France (Cases)	-296.51	601.01
France (Deaths)	-479.08	976.15

France's seasonal Case and Death Model passed the Ljung Box test for each fit to our different country's data-sets, giving us the LLs/AIC scores being:

Data-set	Log-Likelihood	AIC
Germany (Cases)	-274.55	573.1
Germany (Deaths)	-351.3	726.61
UK (Cases)	-169.17	362.35
UK (Deaths)	-221.60	467.20

Germany's seasonal Case Model passed the Ljung Box test for only France's Case Data - with a failure when tested on the UK Case Data. The seasonal Death Model passed the Ljung Box test when fitted to both France and the UK's Death Data, with the following LLs/AIC scores:

Data-set	Log-Likelihood	AIC
France (Cases)	-532.06	1072.12
France (Deaths)	-545.27	1104.54
UK (Deaths)	-270.85	555.71

Based on the success rate of each of the models passing the Ljung Box test when applied to new data, it is evident that generally, the Case and Death Models produced seem to be transferable across data-sets.

Through cross-fitting our obtained SARIMA Case and Death Models across data-sets, we notice that Germany's Models on it's own data are actually being outperformed (defined through our criteria of a favourable model) by our other observed countries' models:

France's Case Model on Germany's Case Data gives a LL/AIC score of -274.55/573.1 compared to Germany's Case model on the Germany Case Data giving a LL/AIC score of -278.26/582.53. These results are relatively close in terms of performance, so we conclude that Germany's Case Model is well-functioning and requires no further attention. However, both the UK and France's Death Model's on Germany's Death Data - giving LLs/AIC scores of -348.21/714.43 and -351.3/726.61 respectively - heavily outperformed Germany's own Death Model on it's own Death Data, which has a LL/AIC score of -368.18/752.37.

We hence revisit Germany's Death Model to see if any improvements could potentially be made. Iterating through the Box-Jenkins approach resulted in a new SARIMA(0,1,1)(1,0,1)(7) Death model being accepted as a more optimal model fit for the Germany Death Data, with LL/AIC scores of -353.42/716.83, which is not outperformed by the UK and France's Death Model on Germany's Death Data, with LLs/AIC scores being -348.21/714.43, -351.30/726.61 respectively. Therefore, we conclude that our new Death Model for Germany is more favourable, and an improvement to our old Germany Death Model.

IV. Discussion

We see that, in general, our seasonal models outperform our regular models when comparing LLs and AIC scores, hence the preference and focus on seasonal models rather than non-seasonal. Furthermore, we see that the models gained from individual data sets seem to be generally transferable between each other, as all the seasonal models perform sufficiently for all the data sets.

When looking at our forecasts, we see also that we tend to prefer the season models over the non-seasonal models. When looking at the fixed origin forecast, we see that the seasonal model better maintained the upwards and weekly trends, wheres the non-seasonal model did not. The rolling day-ahead forecast shows that the seasonal model is better performing from our error metrics. However, when looking at the rolling week-ahead forecast, we do see more nuance in terms of performance, with SARIMA better in terms of RMSE, but ARIMA better in terms of MAE.

A. Different time series models

As mentioned previously, when treating our data, we notice it displays heteroscedasticity, which is not ideal as we want a (somewhat) static variance to analyse our models. We continued by logging our data, however, we could have instead used a different model to observe the evolution of our variance over time, such as an Autoregressive Conditionally Heteroscedastic (ARCH) model. The model is described using the following equations:

$$a_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2$$

where $\alpha_0, \alpha_m > 0$, $\alpha_1, ..., \alpha_{m-1} \geq 0$, and $\{\epsilon_t\}$ is a white noise process of variance 1. The second equation is known as the volatility equation, and behaves analogously to an Autrogressive model, which of course makes up a component of the ARIMA models we used for our analysis.

B. SIR Model

We could also use the epidemiological SIR model, which uses a system of differential equations to model the number of Susceptible individuals, Infectious individuals, and Removed (dead or fully immune) individuals. This could be used to fit a general trend, while ARIMA could be used to model the fluctuations or residuals around this trend. [3]

REFERENCES

[1] N. Mashinchi, "Predicting number of covid19 deaths using time series analysis (arima model)," *towards data science*, 2020.

[2] A. K. Sahai, N. Rath, V. Sood, and M. P. Singh, "ARIMA modelling forecasting of COVID-19 in top five affected countries," 2020.

[3] P. Nadler, S. Wang, R. Arcucci, X. Yang, and Y. Guo, "An epidemiological modelling approach for COVID-19 via data assimilation," *European Journal of Epidemiology (2020)*, 2020.