Summary:

The report presents a comprehensive analysis of the influenza outbreak in the Netherlands, employing both ARIMA and SEIR models for epidemiological investigation. The study rigorously explores parameter estimation methods, including local and global search techniques, to effectively calibrate the models using reported case data.

An innovative aspect of the study involves integrating vaccination status into the SEIR model, which is crucial for understanding the impact of vaccination on disease transmission dynamics. By examining parameter ratios between vaccinated and unvaccinated populations, the analysis provides nuanced insights into the influence of vaccination on outbreak patterns.

While acknowledging areas for potential improvement, the report's well-structured format and clarity contribute to its success in elucidating the dynamics of the influenza outbreak. This detailed analysis serves as a valuable resource for informing public health strategies and further research in infectious disease modeling.

Main comments

Introduction

This study's introduction is well-structured and sets a clear foundation for the research objectives. You effectively highlight the significance of exploring vaccination's impact on influenza transmission, particularly in The Netherlands, given its compact geography and dense population. The decision to move beyond a traditional ARIMA model to develop a custom SEIR model tailored to separate vaccinated and unvaccinated populations reflects a sophisticated approach to understanding flu dynamics. One suggestion is that it might be beneficial to briefly address the limitations of existing approaches (like ARIMA) in capturing the nuances of influenza transmission, further emphasizing the necessity of your innovative SEIR model.

Data

The description of the data source and selection process is thorough, outlining the rationale behind choosing The Netherlands as the study area and focusing on sentinel data. The decision to narrow the dataset to a specific influenza season aligns well with the research objectives, ensuring a targeted analysis of flu transmission patterns during peak periods. To enhance this section, consider briefly introducing the length of the time series, which might be helpful for some diagnostics and help us judge if it could affect model performance.

EDA

The EDA section effectively summarizes key observations from the influenza data, emphasizing the seasonal nature of flu transmission and the strong autocorrelation among adjacent observations. The use of time series plots and autocorrelation analysis demonstrates a methodical approach to understanding the underlying data structure. And it's nice to see specific challenges (such as whether to difference) encountered during the EDA phase and how these challenges inform subsequent modeling decisions.

ARIMA

You offered a thorough ARIMA part to study on the time series from choosing appropriate models, estimating parameters to model diagnostics and conclusions following the steps of what we learned in class. Specifically, you not only use Auto Arima but also mannually choose models and then compare their AIC value to decide on the final model. The diagnostics and conclusions summarize the shortcoming of ARIMA and naturally introduce POMP model.

POMP

In the POMP model, a novel approach was introduced using the SEIR model to analyze the impact of vaccination. The analysis involved multiple iterations of parameter estimation using both local and global search methods to fit the model to reported case data. A notable finding was the substantial improvement achieved with global search methods compared to local search. Furthermore, the profile likelihood analysis, which studied the ratios of estimated parameters, provided insightful implications for real-world applications, offering a detailed and comprehensive discussion of the results. There are only two concerns for me. One is that there missed some interpretation of the initial guess of μ and β. Are they from some resources or just some random guesses. The other is that the local search plot seems a little strange. In the latter part of the iteration, all parameter search just stopped and remains constant. Is it because the codes running is stopped mannually? Or the number of particles are just too large for the dataset? Maybe it is worth to have some discussion on the result. Besides, in the profile likelihood part, it would be better if you can calculate the rough range of the confidence interval and make some comments on it.

small tips on the code

Use `warning = FALSE` in the R chunk option can help reduce warning message in the html output and makes the report looks nicer.