Summary:

This report presents a detailed analysis of Tuberculosis (TB) infection, aiming to identify trends and patterns associated with this disease. It employs a comprehensive methodological approach, ranging from an ARIMA model to a POMP model - SEIRS, to study and simulate the time series of both incidence rate and incidence number. Following the methodology taught in class, the authors meticulously describe the model, estimate parameters, conduct model assessment and diagnosis, and compare the performance of ARIMA and POMP models using this dataset. The report provides valuable insights derived from this analysis. Furthermore, the authors outline plans for further investigation and improvement of the existing model. These plans include conducting a global search and upgrading to a more advanced SEIQRV model.

Major Comments

exploratory data analysis

The report includes a detailed Exploratory Data Analysis (EDA) section, providing clear descriptions of trends, data range, and periodicity. It appropriately utilizes analyses in both the time and frequency domains, aligning with the methods covered in class. To improve the report's flow, it would be beneficial to establish a closer connection between the EDA findings and the subsequent model selection process. By linking these sections more closely, the report can demonstrate how the EDA insights informed the choice of modeling approach and parameter selection.

model description and assumptions

In the ARIMA section of the report, while a thorough estimation and model assessment are provided, it would be beneficial to include a brief description of the ARIMA model, including formulas and clarification of variables. Additionally, it is worth mentioning the rationale for incorporating integration in the ARMA framework. This decision is informed by the exploratory data analysis, which revealed a clear decreasing trend in both the incidence number and rate, while the percentage change appeared stationary. Therefore, a more detailed explanation of the ARIMA model setting would enhance the clarity and justification of the chosen approach.

The POMP section of the report presents a thorough and comprehensive model description. The authors demonstrate their effort to upgrade the SEIR model to SEIRS, providing clear explanations of the model's states and processes. This section is supported by sufficient formulas and definitions of variables, which greatly enhances the clarity and depth of the analysis.

model diagosis

In the ARIMA section of the report, the residual plot exhibits clear heteroscedasticity, and the QQ-plot indicates heavier tails than a normal distribution, both of which are comprehensively analyzed. However, in the POMP section, the model diagnostics are insufficient as some parameters fail to converge adequately according to the MIF2 convergence diagnostics plots. And I think it better to do some explanation and analysis on it. Additionally, the simulation plots may not adequately represent the overall trend of the data due to an insufficient number of simulations and the presence of higher peaks in the later part of the timeline, contrary to the original data trend. Consequently, it is not prudent to conclude that 'The simulations seem to capture the overall trend of the data well.

data selection

Your decision to analyze TB data instead of the more commonly studied COVID-19 data demonstrates innovation in your approach. However, there are valid concerns about the extensive time range (1960 to 2020) covered in the dataset. Such a wide timeframe might lead to certain epidemic disease features becoming less significant over time and could potentially limit the effectiveness of modeling techniques like the SEIRS model. Furthermore, the application of the SEIR model, typically used to capture outbreaks of epidemic diseases, may not yield desired results in this context. This is due to the observed general decreasing trend in the TB incidence data, which differs from the acute, episodic outbreaks that the SEIR model is designed to model.

logic continuity in POMP model

In the main part of the POMP model, there appears to be a lack of clarity in how the model initialization and parameter optimization were conducted. The report does not clearly explain the source of initial parameter values used in the model, which can be confusing for readers. Additionally, the interpretation of the results from local and global search methods is missing, which can be disappointing for those seeking a deeper understanding of the modeling process. To enhance the logical progression of the analysis, it would be more effective to first adjust the initial parameters based on empirical insights to improve simulation results. Subsequently, local and global search methods can be applied around these adjusted parameters. Providing a clearer explanation of these processes and their outcomes would greatly enhance the comprehensibility and impact of the report.

project submission

It seems that the zip you submitted does not include the html file or any rds file, and it takes lots of time to knit the Rmd and reproduce the results.