VARYING-COEFFICIENT MODEL FOR INTERPRETING MONTHLY HOSPITALIZATION RATES OF COVID-19

by

BENJAMIN GARCIA
B.S. University of South Florida, 2023

A research paper submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the Department of Statistics and Data Science
in the College of Science
at the University of Central Florida

Spring Term
2025

ABSTRACT

This research paper focuses on using a varying-coefficient model to explain monthly hospitalization rates of COVID-19 within the states in the U.S. that are a part of COVID-NET. The monthly hospitalization rates of other communicable diseases such as the flu and RSV (respiratory syncytial virus) and seasonal terms were used as predictors. The motivation behind using a varying-coefficient model is that month by month and year by year the COVID-19 hospitalization rate fluctuates in both cyclical patterns and in a gradual overall decrease as vaccines, antiviral therapies, and non-pharmaceutical interventions (NPIs) become more available. The monthly rates are dependent on the variable time in such a way that we need a model that is flexible enough to account for cyclical and nonlinear patterns. The performance of the varying-coefficient model was compared to a linear model and a random forest model using the same data. Leave-one-out cross-validation (LOOCV) was conducted to compute the accuracy of the models in predicting the true COVID-19 hospitalization rates. The varying-coefficient model more closely approximated the actual hospitalization rates and showed to have the smallest mean squared error. This analysis highlights the necessity of varying coefficients in modeling data that evolves over time in a nonlinear manner.

ACKNOWLEDGMENTS

I would like to thank Dr. Suyeon Kang, the chair of my committee, for her assistance in this project. Her insights and guidance helped keep me focus and understand how varying coefficent models could be implemented with this specific data. I appreciate the work we were able to do this semester and how it expanded my knowledge of statistical models. I am additionally grateful for committee members Dr. Hsin-Hsiung Huang and Dr. Jongik Chung, who provided useful commentary for this project and ideas going forward. I appreciate how helpful and encouraging they have been as both professors and mentors since I started my academic journey at the University of Central Florida. I am very thankful for all of the faculty and staff in the Department of Statistics and Data Science at the University of Central Florida, as their teachings and guidance have been valuable to my personal and academic growth. Finally, I would like to thank all of my family and friends who have supported me in becoming the person I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

COVID-19 is a communicable disease that has spread across the world and garnered a lot of global attention. The initial response to COVID-19 included social isolation or distancing, personal protective equipment (PPE) such as masks, and eventually vaccines [13]. The implementation of these things mitigated the effects and transmission of COVID-19 according to [13]. Distancing and PPE not only mitigated COVID-19, but other transmissible diseases such as the flu and respiratory syncytial virus (RSV) [1]. As a result, the seasonal patterns of the flu and RSV were greatly impacted. [13] states that with lower immunity to the flu and RSV along with stronger variants of COVID-19, a wave of all three diseases ravaged public health resources during the winter of 2022. If we look at figures 1.2 and 1.3 we see the spike in hospitalization rate for the flu and RSV corresponding to what [1] referred to as the Tripledemic.

Being able to model and predict these spikes is very important in the field of epidemiology. In order to be sufficiently prepared for a public health emergency or a seasonal illness, it is important to study and understand the patterns and relationships communicable diseases have with each other and with the population. The goal of this paper is to use a varying coefficient model to better understand COVID-19 hospitalization rates using flu and RSV hospitalization rates along with seasonal terms as predictors. The adequacy of a varying coefficient model with this type of data will be tested against other predictive models such as a linear model and random forest.

Figure 1.4 highlights the trend of the seasonal terms s1, s2, s3, s4, which are to be used with the monthly hospitalization rate of the flu and RSV for predicting COVID-19 monthly hospitalization rates. The terms are set up such that s1 and s2 represent annual cycles, while s3 and s4 represent semiannual cycles. The idea is that communicable diseases like the flu have yearly peaks in the winter and lapses in the summer months. Based on the COVID-19 data collected so far, we can see

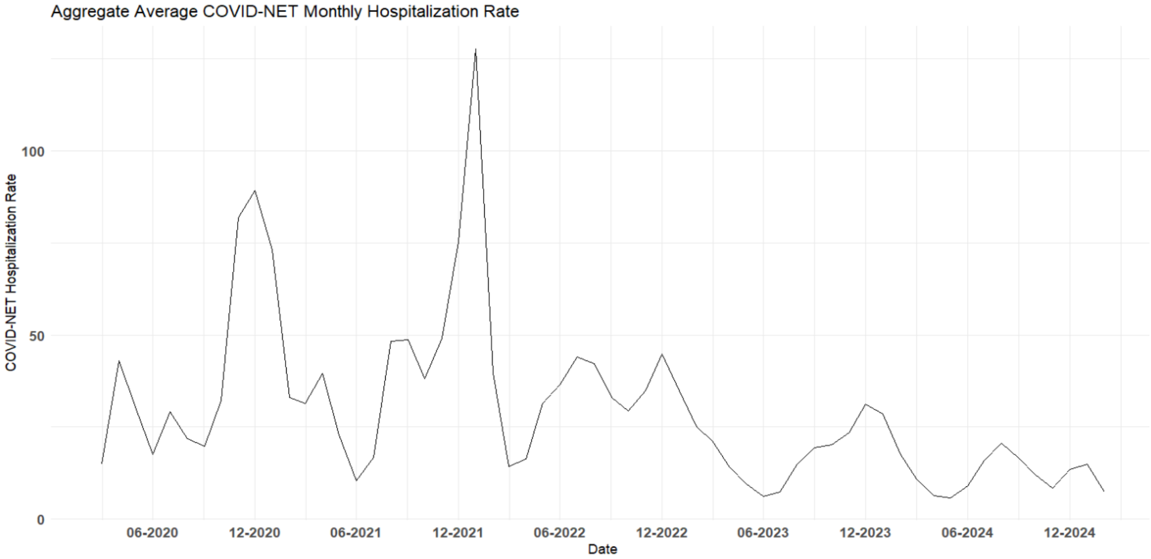


Figure 1.1: COVID-NET Monthly Hospitalization Rate

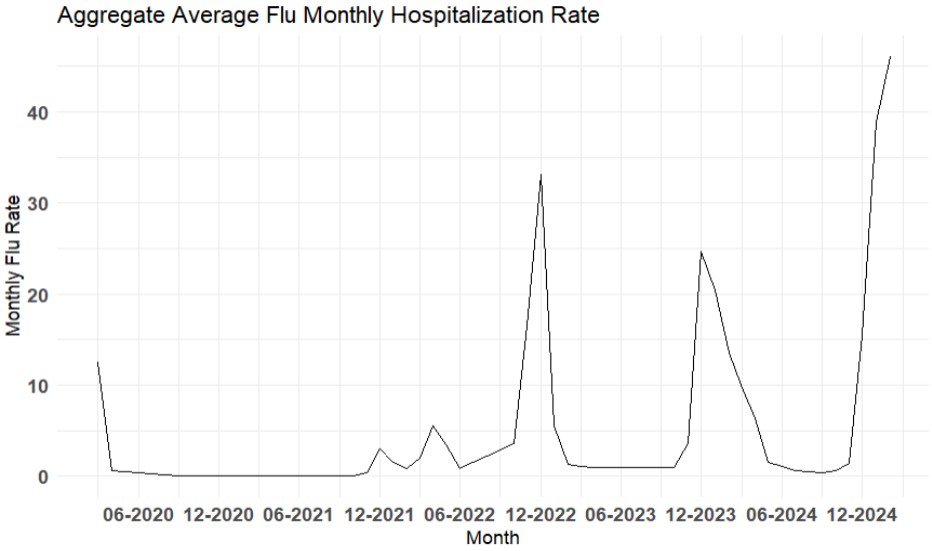


Figure 1.2: FluSurv-NET Monthly Hospitalization Rate

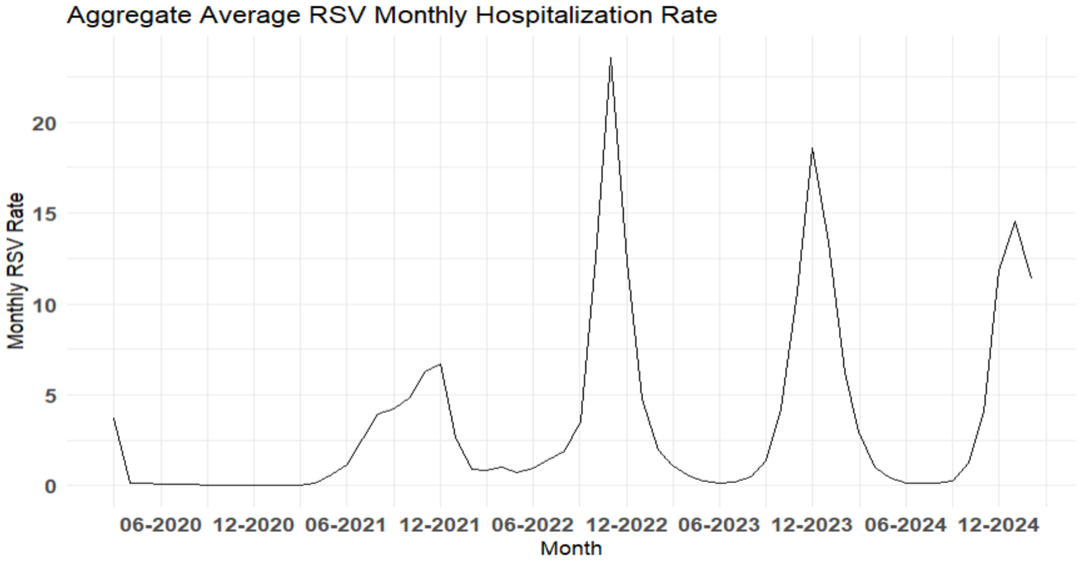


Figure 1.3: RSV-NET Monthly Hospitalization Rate

a similar overall pattern. Where COVID-19 differentiates from diseases like the flu is that there has been a change in the immunity and rate of transmission since the beginning of the pandemic. Seasonal terms that fluctuate consistently and cyclically take into account the trends of all communicable diseases in different seasons, yet are not in tune with yearly differences. These terms give us a way to predict seasonal COVID-19 patterns, however we need more information (potentially flu and RSV rates) to understand how COVID-19 has varied in more severe years compared to milder years.

```
 s1 = cos(2 * pi * Month / 12),
s2 = sin(2 * pi * Month / 12),
s3 = cos(2 * pi * Month / 6),
s4 = sin(2 * pi * Month / 6).
```

Our motivation for using a varying coefficient model comes from the pattern exibited by the data.

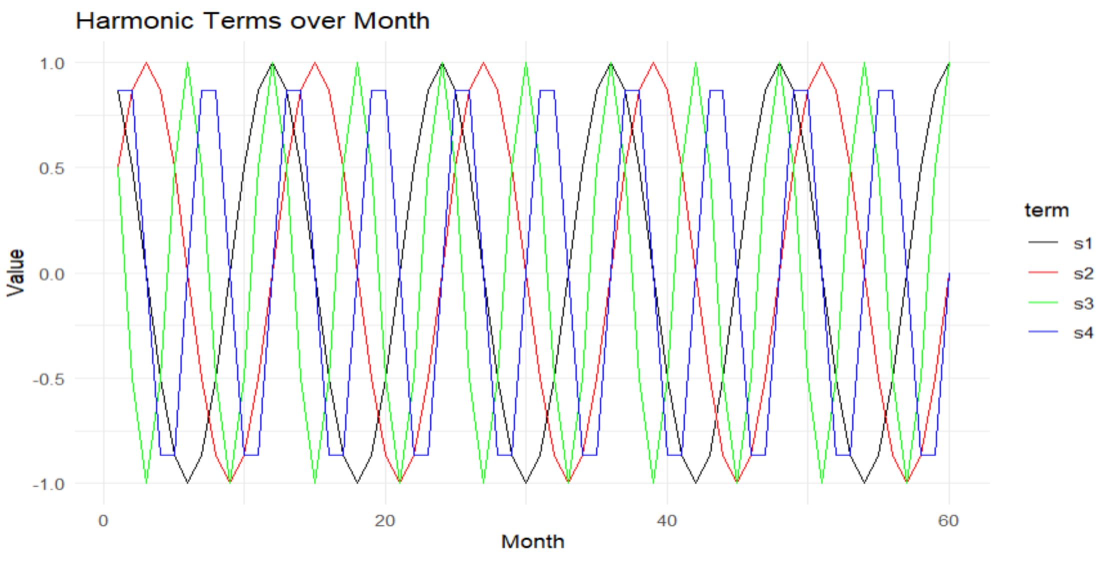


Figure 1.4: Seasonal Terms Plot

The data for this paper was retrieved from the CDC which uses certain surveillance networks to monitor the prevalence of various communicable diseases. These networks include COVID-NET, FluSurv-NET, and RSV-NET which are all apart of Respiratory Virus Hospitalization Surveillance Network (RESP-NET) [4]. As we can see in Figure 1.5 and Figure 1.6, the slope of the scatterplot or the relationship between the flu and COVID-19 and RSV and COVID-19 changes as the months increase. In earlier months, the monthly rate of hospitalization of COVID-19 increased greatly as flu and RSV changed very little. During the later months, COVID-19 seemed to not reach the peaks that it did previously, but the flu and RSV both varied and increased more steeply. This change in the relationship between the diseases over time indicates that a constant slope or coefficient term would be unwise. Rather, we may need a coefficient that varies to account for the trend of the data.

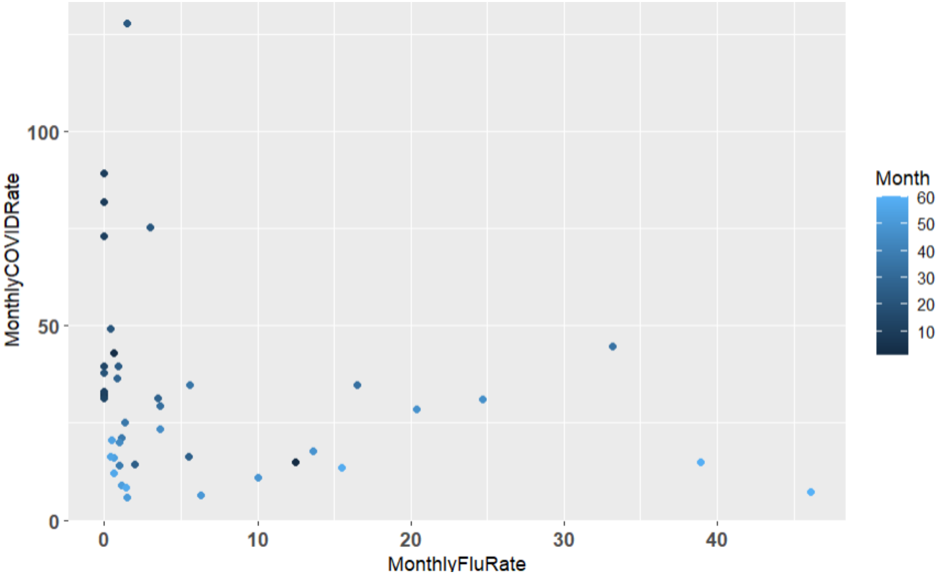


Figure 1.5: Scatterplot of COVID-19 vs Flu Hospitalization Rates Colored by Month

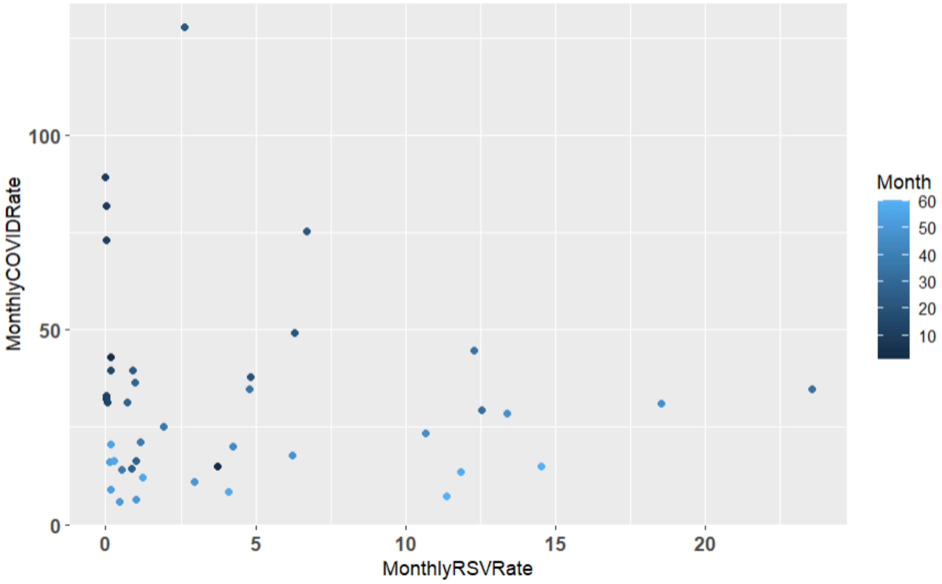


Figure 1.6: Scatterplot of COVID-19 vs RSV Hospitalization Rates Colored by Month

CHAPTER 2: LITERATURE REVIEW

The usage of varying coefficient models for modeling the trend of a communicable disease, such as COVID-19, have been applied previously. In [12], the authors decided to implement a time-varying coefficient regression model to understand the relationship between cumulative death count and cumulative case count. More specifically, the authors used local polynomial regression and examined the lag between the case and death count. [12] found that local polynomial regression could be used as a complementary tool for predicting future death count for COVID-19 and potentially other diseases. Additionally, the local polynomial regression could work better than other methods such as piecewise linear regression models, especially when the pattern of the data gets more complicated [12]. Concessions made about the method were that local polynomial regression was sensitive to outliers and that no missing data was present.

What differentiates the analysis conducted in [12] from this research paper is the consideration of other diseases as predictors. The usage of case and death counts as predictor and response will naturally have a positive relationship. While the relationship can vary over time, especially as global health policies have changed through the years, the two counts will most likely reflect the same pattern. Additionally, the source uses cumulative daily counts. [12] justifies that daily fluctuations create additional noise that makes fitting the model more complicated and not as representative of the overall pattern.

In [8] we see how the varying coefficient model can be used to examine the relationship between hospital admission for circulatory and respiratory problems and the level of air pollutants in Hong Kong. The model proposed in the source includes an intercept term and the coefficients as a function of time. The time in this instance was measured in days as was also in [12]. The overall conclusion in [8] is that the impacts of the pollutants vary with time as predictors for circulatory and respiratory hospitalization. The varying coefficient model with the coefficients as functions of time, is able to model the relationship between various predictors and hospitalization which is similar to the target variable of this analysis.

While the coefficients of a varying coefficient model don't necessarily have to be functions of time as seen in the introductory example of [10], for the purposes of measuring communicable diseases, time seems to be essential. According to the literature, multiple predictors can be used in the analysis. [8] shows that examining the relationship of air pollutants in totality as opposed to individually is valid and can be used to draw conclusions about the general impact of air pollutions over time. This helps solidify the use of both the flu and RSV as predictors of COVID-19 instead of separately. Additionally, the change in [12] from daily to cumulative daily counts represents the desire to avoid excess noise and deviations that are not representative of the overall trend. In a similar way using COVID-19 monthly hospitalization rates which according to [4], are the total number of events divided by the total population at risk (per 100,000), may avoid the excess noise present in daily responses and provide a better idea of the trend over several years. We continue with our analysis better understanding the potential of a varying coefficient model for COVID-19 data and how similar studies have been effective in their efforts.

Another method used for modeling nonlinear data is a Bayesian zero-inflated negative binomial (ZINB) using nearest-neighbor Gaussian process from [11]. Zero-inflated models have consistently been used to handle data that contains zeros which derive from individuals not at risk or individuals at risk, but still observed as zero. The usage of Bayesian methods over asymptotically approximate distributions for estimating of parameters has worked well. This model was used to model COVID-19 death counts in the state of Florida during the beginning months of the COVID-19 pandemic to examine relationships between social vulnerability and COVID-19 deaths [11]. The motivation for using this model with this data is that the death count in the beginning of the COVID-19 pandemic in some counties of Florida were near or essentially zero. Counties like

Miami-Dade county varied in death count compared to the 74% of other Florida counties that had zero reported deaths [11]. If we were to accurately predict death counts going forward, we would need to account for the zero-inflated death count present early in the data. The social vulnerability index (SVI) was analyzed to determine its relationship with COVID-19 death count. In Florida, moderate SVI risk was associated with the highest COVID-19 death count. The conclusion is that the Bayesian framework for zero-inflated negative binomial regression models with spatiotemporal effects is comparable to existing methods like a B-spline spatiotemporal model for modeling COVID-19 death counts. While this project looks to model hospitalization rates, there is no doubt that death and hospitalization have a close relationship and that the time and space in which the data is being reported from will influence these values greatly.

CHAPTER 3: METHODOLOGY

Varying Coefficient Model

Given the data present for this project, we wish to use a varying coefficient model to understand how the monthly hospitalization rate of COVID-19 can be predicted by the monthly hospitalization rate of the flu and RSV and other seasonal effects. Using the kernel-local polynomial smoothing method from [8] we get the following equation.

$$y = X^T a(U) + \varepsilon$$

Which for our model, is equivalent to

$$y = a_1(U) + a_2(U)x_2 + a_3(U)x_3 + a_4(U)x_4 + a_5(U)x_5 + a_6(U)x_6 + a_7(U)x_7 + \varepsilon$$

where $a_1(U)$ is for our intercept term, $a_2(U)x_2, a_3(U)x_3, a_4(U)x_4, a_5(U)x_5$ are the four terms for our seasonal effects, $a_6(U)x_6, a_7(U)x_7$ correspond to the flu and RSV, U is time or month, and ε is random error with $E(\varepsilon) = 0$ and $var(\varepsilon) = \sigma^2(U)$.

In order to solve for the functional coefficients from [8], we use the following equation

$$\hat{a}(u) = (I_p, 0_p)(\Gamma_u^T W_u \Gamma_u)^{-1} \Gamma_u^T W_u Y$$

where,

$$Y = (y_1, ..., y_n)^T, \mathbf{X} = (X_1, ..., X_n)^T, U_u = diag(U_1 - u, ..., U_n - u),$$

$$W_u = diag(K_h(U_1 - u), ..., K_h(U_n - u), \Gamma = (\mathbf{X}, U_u, \mathbf{X}).$$

We have then that Y is a vector of the observed values. Likewise **X** is a vector of the predictors including the intercept, $x_2, x_3, x_4, x_5, x_6,$ and x_7.

U_u is a matrix that only has diagonal entries. along these diagonal entries is the difference between U_1, which for our data is 1 for month one, and u which corresponds to the month number we are trying to estimate the coefficient for. This continues along the diagonal until we get to $U_{60} - u$ where U_{60} equals 60 for the 60th and last month in the data set. For each $\hat{a}(u)$ where u = 1,...,60 we have a corresponding U_u from u = 1,...,60.

Similar to U_u, W_u is a diagonal matrix except $U_1 - u$ through $U_{60} - u$ is applied to the kernel K_h(t). The kernel K_h(t) is equal to $K(t/h)/h$, where h is equal to the bandwidth. Γ is a matrix which takes the product of **X**, U_u, and **X**. For each u (from 1 to 60) there is a unique U_u, W_u, and Γ matrix.

We then have the following minimizer, where $\hat{a}(u)$ corresponds to **a** and the minimizer is solved for each given u as,

$$L(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} = (y_i - X_i^T\mathbf{a} - X_i^T\mathbf{b}(U_i - u))^2 K_h(U_i - u).$$

We can implement this code in programming languages like R, such that for each u or in this case, each month from 1 to 60, we have coefficients for the intercept and all of the predictors.

Prediction Methods

Multiple packages were used to conduct the prediction of COVID-19 monthly hospitalization rates using the predictor variables from our data. The lm function or linear model located in the stats package in base R from [5] was used. The linear model provides an estimation of the COVID-19 rates using the predictor variables, without consideration for the month the observation occurred in. While the linear model is simple in nature, it allows great interpretability over the coefficients and their significance. The random forest function in the random forest package from [2] was also used to model COVID-19 rates. The random forest method generates multiple decision trees which improves the accuracy of prediction by considering them together. Both of these methods are included in this analysis to interpret the ability of methods that do not consider time or allow the coefficients to vary to predict this data. While these methods may not be the most ideal with this data, it is important to consider models that could question the validity of the data being dependent on time if their accuracy is similar or better than methods that do interpret time or allow the coefficients to vary over time. Also, if parametric assumptions are violated by the VCM, these models could be good substitutes. Random forest specifically is good at being flexible in modeling nonlinear relationships and interactions between predictors.

The other two methods include a time-varying linear model or TVLM from the tvReg package from [3] and a generalized additive model from the mgcv package from [15]. Both of these models consider the effect that the month number has on the rate of COVID-19 hospitalization. The implementation of tvlm on the data follows the methodology presented in the varying coefficient model section of this chapter. It uses local polynomial kernel smoothing to estimate the coefficients as functions of time that best fit the target variable. According to [10], the generalized additive model is a special case of the varying coefficient model. Where the generalized additive model differs is that the parametric functions of the regressors are replaced by smooth nonparametric functions [10]. Such is the case that the coefficients do vary, however not as a coefficient for the predictor, but as unspecified or smoothing function byitself [10]. Typically, smoothing splines are used to estimate the function needed to model the predictor for the target [15].

Parameter Tuning Using Leave-One-Out Cross Validation

Parameter tuning was conducted using leave-one-out cross-validation or LOOCV. LOOCV can be implemented in R such that for each observation, a training data set with every observation except the observation to be tested on is included and a test data set with the one observation not included in the training. The model is built on the defined training data and then used to predict the target variable that was not included in training. LOOCV can be conducted to test several bandwidth values. The bandwidth value that results in the best accuracy will be the one used when the model is implemented on the full data set. Given that our target variable is COVID-19 monthly hospitalization rate and our models are being used for regression, our predicted value should be a continuous, numeric value. Thus, we shall use mean squared error or MSE as the metric to indicate how close the predicted value was to the observed value.

Missing Value Imputation

In our data, there were some missing flu and RSV monthly hospitalization rates. The data, retrieved from the CDC, had months where the rates were not reported due to the diseases being in their off-season. In order to make sure that each COVID-19 rate had a corresponding flu and RSV rate, missing values were imputed using the na_seadec function from the imputeTS package from

[6]. The function performs seasonal decomposition which removes any seasonal effects before imputation occurs. This is done with loess regression which smoothes the variable of interest using the tricube weight function. This helps assign a neighborhood weight from [6], where values closer to the missing value being imputed have more weight than values further away from the missing value. After seasonality was removed, imputation using an interpolation algorithm was used. Interpolation considers the values in the data and chooses a value that best fits in the range between the last and next complete values. In the case of our data, the rates that were filled in followed either a decreasing, increasing, or zero rate trend between the two closest non-missing observations. Ultimately, since the data that was missing was during the off-season, the values being imputed were quite small. Multiple values were either zero or near zero since the trend leading up to the off-season was a constant decrease from the winter peak months. The nonzero values that were imputed made sense contextually, do not create significant changes to the analysis, and seem fit as approximate estimates for the true values. As such, we can continue with the analysis and results knowing that our data is representative of the overall trend of the flu and RSV.

CHAPTER 4: FINDINGS

Coefficients of the Model

We begin this section by interpreting the coefficients for the varying coefficient model. In Figure 4.1, the plots of the values of the coefficients for the predictors and the changing intercept term are shown from 1 to 60 months. It can be seen that the intercept, s4, and the RSV rate seem to fluctuate more notably than the other predictors as time increases. This helps support the idea that varying coefficients are necessary to model the changing COVID-19 rates. We also have 95% confidence intervals motivated by the fact that the coefficient estimates are asymptotically normally distributed [8]. If we were to consider whether a coefficient does significantly vary over time, the lack of linear confidence intervals could serve as evidence.

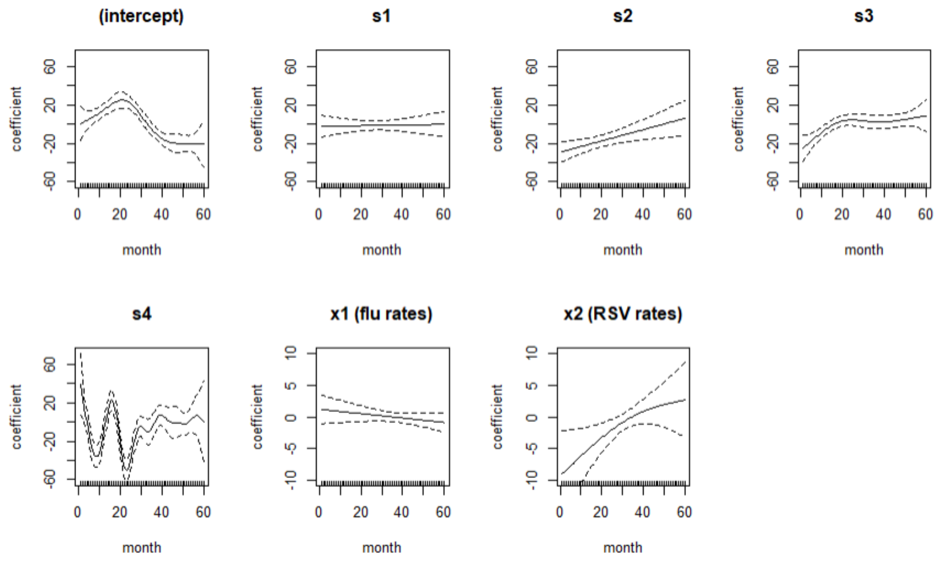
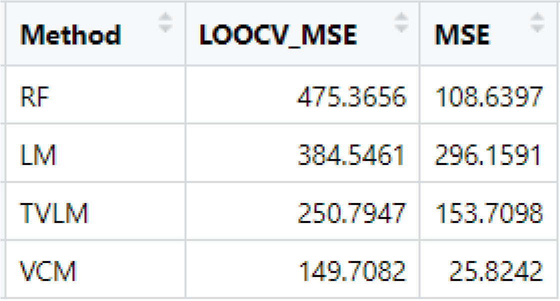


Figure 4.1: Coefficient Confidence Interval Plots

Model Comparison

We now consider the accuracy of the models in predicting the monthly hospitalization rate of COVID-19 using LOOCV mentioned in the previous section to compute the MSE, along with the MSE for all observations in the data when the entire data set is used to build the models. We see in Table 4.1 that the random forest technique had the highest MSE when LOOCV was applied followed by the linear model, TVLM, and then the varying coefficient model or VCM. What this may indicate is that the random forest method had a hard time predicting the COVID-19 rates when the data to be predicted was not already seen by the model. On the other hand, the VCM could make better predictions for new data especially if the month helped indicate where the observation resided in the seasonal cycle. The third column of of Table 4.1 shows the MSE when predictions are made using a model that has seen every observation. Naturally, we would expect the MSE to decrease since the models were constructed with the observation. For the MSE using the entire data set to build the model, the linear model had the highest MSE followed by the TVLM, random forest, and then VCM. Random forest makes a jump from having the highest LOOCV MSE to the second lowest MSE. Random forest may then work well when the predicted data has already been seen, but its performance trying to predict new values hinders its effectiveness with the data. Ultimately, the VCM is consistently the best model because it generates the lowest MSE. VCM may be best suited to explain and predict the target value.

We can also visualize how well the models were fitted to the observed values. Figure 4.2, considers each models ability to predict the target when trained on the entire data set. We can see that the linear model struggled the most, as a majority of the points deviate from the line and over-approximate and under-approximate the true values. The random forest has a tighter fit, with only a few points that deviate below the line. The random forest seems to make closer approximations than the TVLM technique. The method that makes the closest approximations is the VCM. Aside



Method	LOOCV_MSE	MSE
RF	475.3656	108.6397
LM	384.5461	296.1591
TVLM	250.7947	153.7098
VCM	149.7082	25.8242

Table 4.1: Model MSE Table

from a few points that deviate, nearly all the points are on the line and take on the form of a straight line.

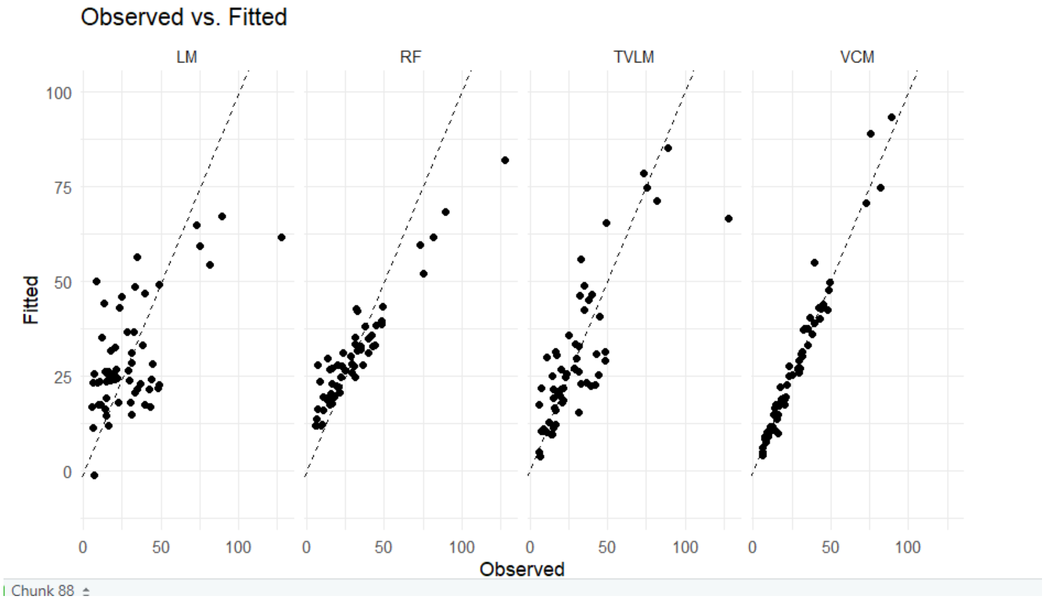


Figure 4.2: Observed vs. Fitted Plot

With Figure 4.3, we see the actual trend of the COVID-19 hospitalization rates compared to the predicted trend for each of the models from month 1 to 60. Like with Figure 4.2, the linear model deviates the most, especially past 35 months as it goes from under-estimating to over-estimating the true values. Because a linear model is attempting to minimize the sum of squared errors using a straight line, the method is unable to approximate any of the observed values closely. The data has multiple upward and downward curves that contribute to the nonlinear pattern of diseases with seasonal changes. The green and pink fitted lines which correspond to the random forest and the TVLM methods vary in terms of which provides the closer estimate. Between months 25 and 40, the random forest gets closer to the observed COVID-19 rate, while at month 10 and from month 43 to 60, TVLM seems to provide a closer estimate. Figure 4.3 provides less distinction between TVLM and random forest compared to Table 4.1 and Figure 4.2 where random forest appeared more accurate. It is unclear why at different time points one model does better than the other, however both represent good options when using all observation to predict the target values. The VCM model really fits to the observed values well. The only point where it deviates from the actual value is around month 22 where a large, sharply increasing spike made it difficult for all of the models to accurately estimate the true value. The VCM accounts for all of the data's peaks and valleys may indicate overfitting, however the LOOCV MSE for VCM assures that the model can predict new data as well.

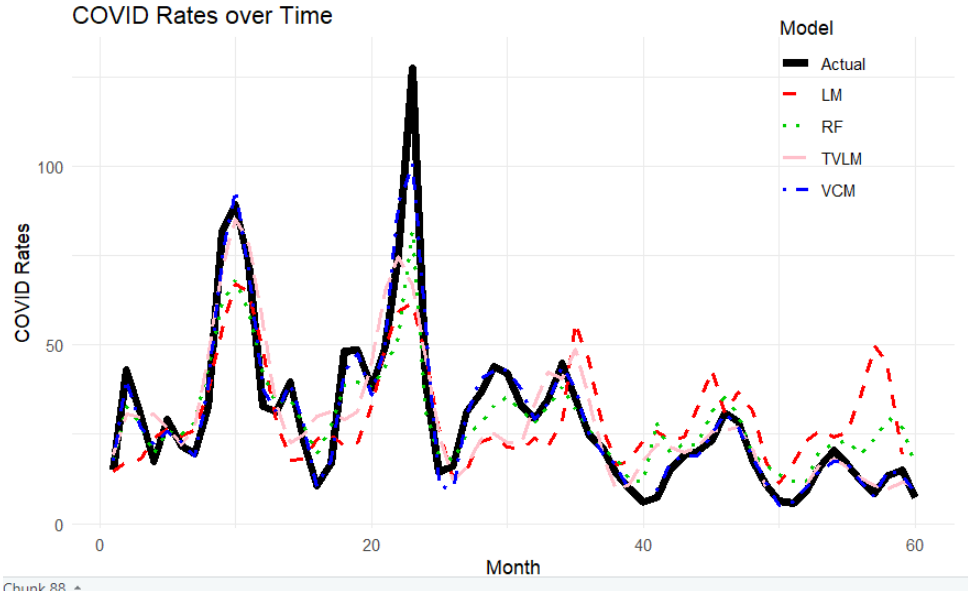


Figure 4.3: Actual COVID-19 Rates vs Model Predicted Rates

CHAPTER 5: CONCLUSION

Based on our results, we conclude that for the population present in our data, the monthly hospitalization rates of COVID-19 can be modeled by seasonal terms and the monthly hospitalization rates of the flu and RSV. In conjunction, the predictors are able to account for the yearly winter peaks and the summer lows, which are common in many communicable diseases. The predictors also account for the change in the magnitude of the rates of peak months. As COVID-19 immunity increased and mortality decreased, the hospitalization rate was sure to decrease as well. While the rate of COVID-19 cases may tell a different story, the distinction between the severity and transmissibility of a disease indicates that a population can be more immune to the severe effects of a disease while its prevalence is still high. Unlike the flu and RSV, COVID-19 is a disease that has more recently spread across the globe. The beginning period of the data represents strict public health safety measures due to the lack of immunity to the virus's severe effects. Even after the introduction of vaccines around month 14, hospitalizations would still be high, as the increase in transmission due to decreased adherence to COVID-19 safety protocols kept the disease in circulation. It was not until the virus had been present in the population for multiple years could true immunity be acquired which has now resulted in COVID-19 no longer being considered a pandemic and for current and future peaks of the disease to never be quite as severe as they once were. Thus, in regards to the relationship COVID-19 has with other communicable diseases, the social response to its global dominance resulted in the suppression of other diseases like the flu and RSV, which led to the previously mentioned tripledemic, which was followed by the flu and RSV returning to their normal severity and transmissibility as COVID-19 rates continues to decrease. This trend is present in the data and the one that can be modelled by the predictors using the previously mentioned methods.

The coefficients, set as a function of the month number since the beginning of COVID-19 prevalence in the surveyed area, help predict the target when they vary. The ability for the coefficients to vary is important for the COVID-19 hospitalization rates, as they seemed to display nonlinear patterns. Regardless of the varying coefficient method chosen, there is serious usage and need for flexible models that account for time and are proven not to overfit.

Additionally, based on the comparison of models, VCM seems to best explain and predict the COVID-19 hospitalization rates in the data. The VCM with parameter tuning greatly increased its overall ability to predict the observed value and smoothness in its account for so many small and large peaks and valleys. One worry is if the VCM is too overfit to the data which would result in it not being able to predict new values well. It seems though that for this project that the model fits new and trained data well enough. The random forest varies greatly between its ability to predict already trained data and new data. These inconsistent results make it hard to justify the implementation of this method with this kind of data. Future research understanding the difficulty random forest had at predicting new data whether it be because of the specific data presented or insufficient parameter tuning, should be considered. Without consideration for time or the ability to account for nonlinear patterns, the linear model's lack of fit indicates that the data needs a more flexible model that can adjust to multiple peaks and valleys. Meanwhile, the TVLM provides more consistent results and takes into account the month the observation was recorded in. TVLM is another method that could be useful for tracking communicable diseases, especially since rates can greatly vary over time. The tvReg package has other methods that consider autoregressive and seemingly unrelated regression equation (SURE) models. Overall, our data suggest allowing time-varying effects of coefficients in order to understand the relationship between predictors and targets. Additional time-varying, coefficient varying, time series, or nonlinear models could also be tested with this kind of data to see how it compares to the tested methods in this project. More research into generalized additive models could improve the results on this type analysis

Things to consider regarding this analysis include the surveyed population. While the surveillance network consisted of millions of people in the United States, they only originated from 13 or 14 states (depending on the network). The rates and trends present in this analysis may not generalize to the entire United States as a whole. It would be interesting to see how the analysis would be changed if the surveyed population included individuals from every state. Another consideration is that while the choice of imputed values for the flu and RSV rate are justified, they could still differ from the actual values. While this would not take away from the relationship observed during the peak months since flu and RSV were always reported between November and February, the low months may not be as low as estimated. The CDC's lack of reporting during nonseasonal months made the analysis more challenging, but still provided enough data to derive notable results. While our results make sense given the context and pattern of our data, it remains to be seen how variable selection would plays a role in model fit and accuracy. In [13] only the case count and death count for COVID-19 were used. With further investigation, a simpler model or a model that accounts for another disease or seasonal may provide better results for one or multiple methods. Future work in the area of applying varying coefficient models to predict communicable disease could involve a comparison based on gender, race, age, or another demographic status. As in [11], usage of socioeconomic factors can better help us understand the how rates and counts of COVID-19 differ amonst our population. The COVID-19 pandemic has impacted different socioeconomic groups in the United States differently. Knowing the extent of these differences could help future public health leaders protect the various people in their communities.

LIST OF REFERENCES

[1] Kaiming Bi, Shraddha Ramdas Bandekar, Anass Bouchnita, Spencer J. Fox, and Lauren Ancel Meyers. Annual Hospitalizations for COVID-19, Influenza, and Respiratory Syncytial Virus, United States, 2023–2024. *Emerging Infectious Diseases*, 31(3), pages 636–638, 2025.

[2] Leo Breiman. Random Forests. *Machine Learning*, 45:5-32, 2001.

[3] Isabel Casas and Rubén Fernández-Casal. tvReg: Time-varying Coefficients in Multi-Equation Regression in R. *The R Journal*, 14(1), pages 79-100, 2022.

[4] Centers for Disease Control and Prevention. Covid-Net. *Centers for Disease Control and Prevention*, https://www.cdc.gov/covid/php/covid-net/index.html. .

[5] John M. Chambers and Trevor J. Hastie. Linear models. *Statistical Models in S edited by John M. Chambers and Trevor J. Hastie*, pages 95-144, 1992.

[6] Robert B. Cleveland, William S. Cleveland, Jean E. McRae, and Irma Terpenning. STL: A seasonal-trend decomposition. *Journal of Official Statistics*, 6(1), pages 3-73, 1990.

[7] Gerco den Hartog, Puck B van Kasteren, Rutger M Schepp, Anne C Teirlinck, Fiona R M van der Klis, and Robert S van Binnendijk. Decline of RSV-specific antibodies during the COVID-19 pandemic. *The Lancet Infectious Diseases*, 23(1), pages 23–25, 2023.

[8] Jianqing Fan and Wenyang Zhang. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1(1), pages 179-195, 2008.

[9] Trevor J. Hastie, Robert Tibshirani, and Jerome Friedman. Kernel Smoothing Methods. *In The Elements of Statistical Learning Data Mining, Inference, and Prediction*, 2nd edition, pages 191-218, 2009.

[10] Trevor J. Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55(4), pages 757-779, 1992.

[11] Qing He and Hsin-Hsiung Huang. A Framework of Zero-Inflated Bayesian Negative Binomial Regression Models For Spatiotemporal Data. *Journal of Statistical Planning and Inference*, 229, 2024.

[12] Juxin Liu , Brandon Bellows, X. Joan Hu , Jianhong Wu, Zhou Zhou, Chris Soteros and Lin Wang. A new time-varying coefficient regression approach for analyzing infectious disease data. *Scientific Reports*, 13(1), 2023.

[13] Wei Luo, Qianhuang Liu1, Yuxuan Zhou, Yiding Ran, Zhaoyin Liu, Weitao Hou, Sen Pei, and Shengjie Lai. Spatiotemporal variations of "triple-demic" outbreaks of respiratory infections in the United States in the post-COVID-19 era. *BMC Public Health*, 23(1), 2023.

[14] Byeong U. Park, Enno Mammen, Young K. Lee and Eun Ryung Lee. Varying coefficient regression models: A review and New Developments. *International Statistical Review,* 83(1), pages 36–64, 2013.

[15] Simon N. Wood. GAMs in Practice: mgcv. *In Generalized Additive Models An Introduction with R Second Edition* , pages 325-397, 2017.