505777837_ECON-144_Project-3

Name: Andy Ho, UID: 505777837

2024-12-06

Project 3

Introduction

Education is a topic that many developing countries face problems with due to their unstable economies and social issues. The Gambia's education system faces multifaceted challenges that hinder equitable access to quality learning opportunities. Despite constitutional mandates for free and compulsory primary education, resource constraints and insufficient infrastructure impede full implementation. Although lower basic education is officially free, hidden costs such as uniforms, stationery, and books prevent an estimated 29% of children from attending school.

Gender disparities persist, with social norms often prioritizing marriage over girls' education, leading to lower enrollment rates for girls, especially in rural areas. Boys may face pressure to seek employment abroad, disrupting their educational paths. Additionally, approximately 20% of school-age children attend Quranic schools, which usually have a restricted curriculum, limiting their exposure to a comprehensive education.

Between the late 1990s and the 2010s, The Gambia's education system faced significant challenges despite efforts to improve access and quality. In 1998, the government abolished school fees for the initial six years of primary education, leading to increased enrollment rates. However, hidden costs such as uniforms, stationery, and books continued to prevent an estimated 29% of children from attending school. Resource limitations and inadequate infrastructure further hindered the effective implementation of free and compulsory primary education. The education sector grappled with a shortage of qualified teachers, limited teaching materials, and underdeveloped infrastructure, all of which adversely affected learning outcomes. Financial constraints also impacted educational quality. Although government expenditure on education increased from 15% to 21% of total government spending during the 1990s, the allocation was insufficient to address systemic issues.

The COVID-19 pandemic exacerbated existing issues, disrupting learning and widening educational inequalities. In response, initiatives such as community engagement and cash transfers have been implemented to re-enroll out-of-school children. In 2023, these efforts successfully brought over 30,000 children back to school, demonstrating the impact of targeted interventions.

Addressing these challenges requires coordinated efforts from the government, civil society, and international partners to ensure inclusive and quality education for all Gambian children. The data set that we will be looking at involves data from the World Bank organization which provides global education statistics across multiple countries and years. Key variables include government expenditure on education as a percentage of GDP (gov_exp_pct_gdp), adult literacy rates (lit_rate_adult_pct), primary completion rates (pri_comp_rate_pct), pupil-teacher ratios at the primary and secondary levels (pupil_teacher_primary and pupil_teacher_secondary), and school enrollment rates for primary, secondary, and tertiary education (school_enrol_primary_pct, school_enrol_secondary_pct, and school_enrol_tertiary_pct). However, the dataset contains many missing values, especially for earlier years and smaller countries, highlighting data availability issues. This dataset allows for comparative analyses of educational trends and challenges across different regions globally.

Since we are focusing on The Gambia, we will use a subset of the original data set specifically for The Gambia, covering two key time-series variables: government expenditure on education as a percentage of GDP (gov_exp_pct_gdp) and primary school enrollment rates (school_enrol_primary_pct). The dataset spans multiple years, from 1999 to 2023. The data set also goes by yearly data. It provides a focused view of how The Gambia's education sector has evolved. Compared to the global dataset, it is more tailored, with fewer variables but greater relevance for studying education-specific trends in The Gambia. Overall, the data set enables an in-depth exploration of education trends in The Gambia, offering insights into how national efforts align with patterns seen in other parts of the world. From our context, we can say that it is likely government expenditure and primary school enrollment has improved over the past two decades since The Gambia has made many efforts in government policies to help improve its education system.

Results

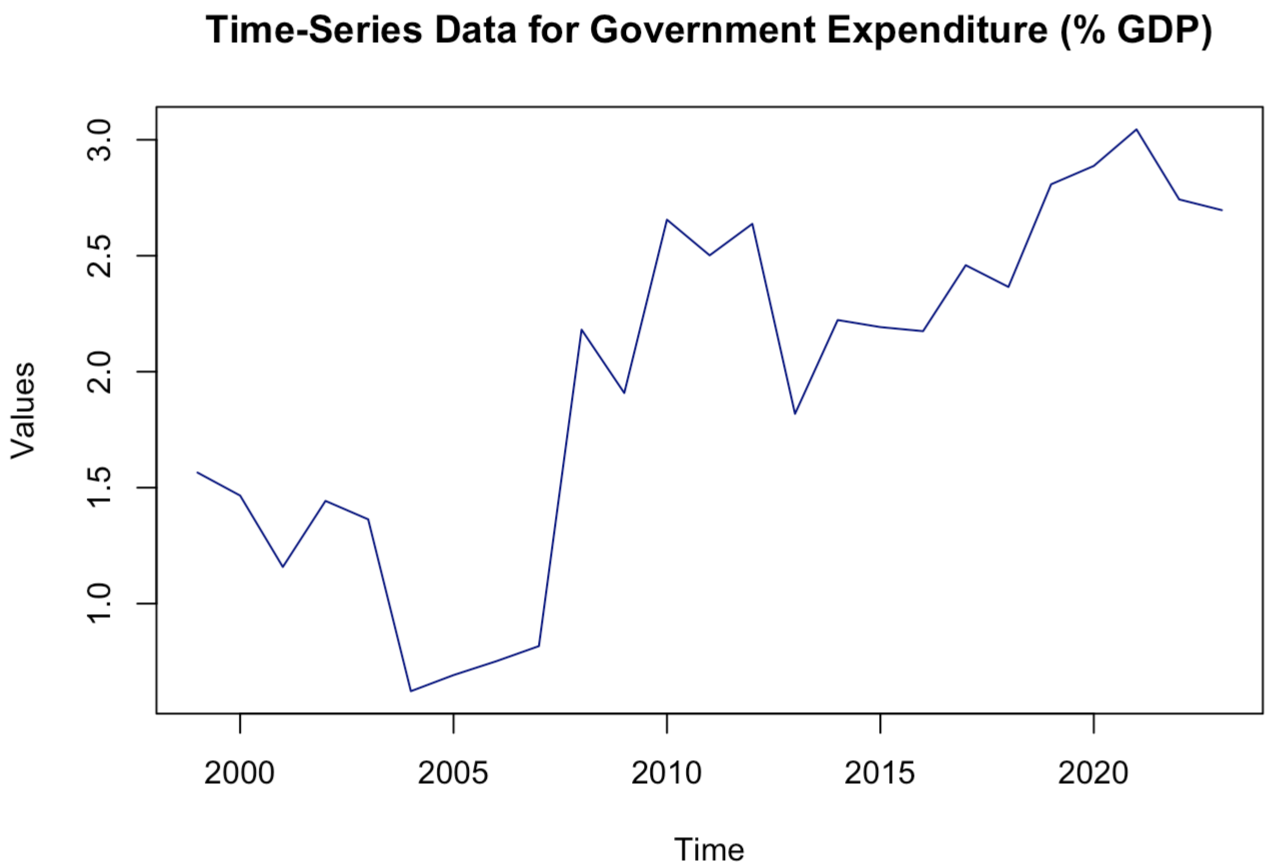
Let's first upload the data set and focus on the statistics for The Gambia. We will focus on the government expenditure on education in terms of GDP and the enrollment rate for primary schools. These will be our two time-series variables.

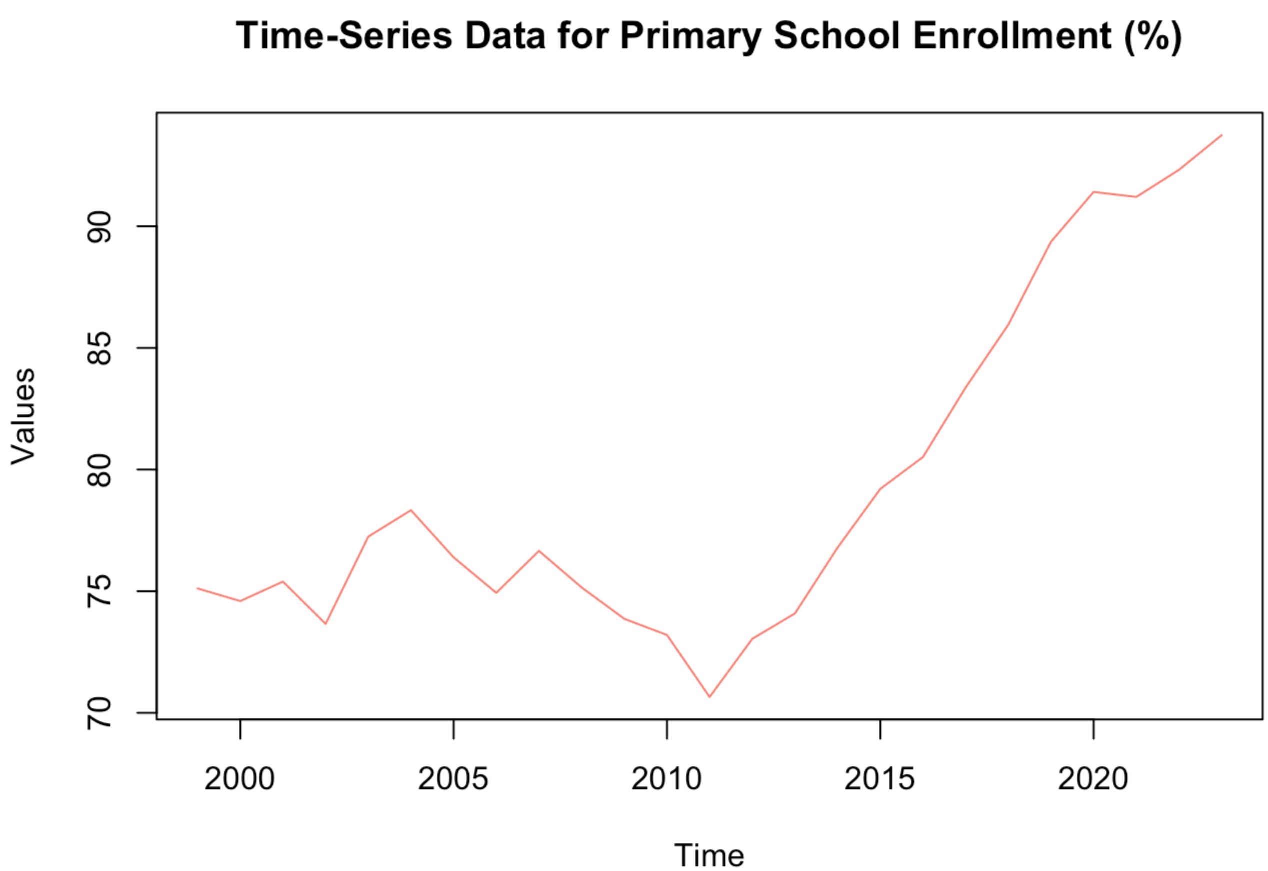
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Attaching package: 'zoo'
##
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

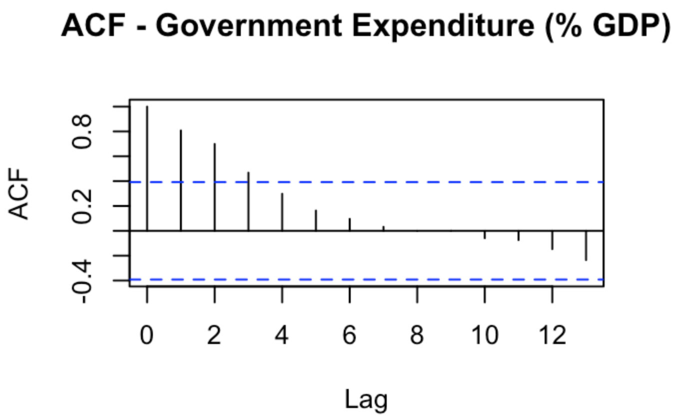
```
##        country country_code year gov_exp_pct_gdp school_enrol_primary_pct
## 1  Gambia, The          GMB 1999        1.565130                 75.11253
## 2  Gambia, The          GMB 2000        1.465870                 74.59557
## 3  Gambia, The          GMB 2001        1.158100                 75.39515
## 4  Gambia, The          GMB 2002        1.442920                 73.65915
## 5  Gambia, The          GMB 2003        1.363380                 77.24397
## 6  Gambia, The          GMB 2004        0.622470                 78.33055
## 7  Gambia, The          GMB 2005        0.691880                 76.39198
## 8  Gambia, The          GMB 2006        0.751630                 74.93363
## 9  Gambia, The          GMB 2007        0.816860                 76.65833
## 10 Gambia, The          GMB 2008        2.181300                 75.15580
## 11 Gambia, The          GMB 2009        1.908140                 73.86633
## 12 Gambia, The          GMB 2010        2.655520                 73.20217
## 13 Gambia, The          GMB 2011        2.501703                 70.65355
## 14 Gambia, The          GMB 2012        2.637570                 73.04300
## 15 Gambia, The          GMB 2013        1.818390                 74.08779
## 16 Gambia, The          GMB 2014        2.222803                 76.79584
## 17 Gambia, The          GMB 2015        2.192352                 79.20504
## 18 Gambia, The          GMB 2016        2.174797                 80.51581
## 19 Gambia, The          GMB 2017        2.459194                 83.37243
## 20 Gambia, The          GMB 2018        2.365444                 85.95533
## 21 Gambia, The          GMB 2019        2.808097                 89.36565
## 22 Gambia, The          GMB 2020        2.887346                 91.41179
## 23 Gambia, The          GMB 2021        3.044777                 91.20666
## 24 Gambia, The          GMB 2022        2.742823                 92.32039
## 25 Gambia, The          GMB 2023        2.696595                 93.74525
```

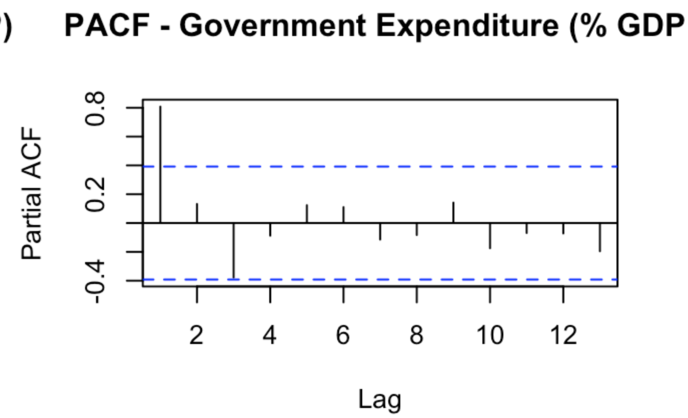
First, let's produce a time-series plot of your data including the respective ACF and PACF plots. We will do this for each of our time-series variables.

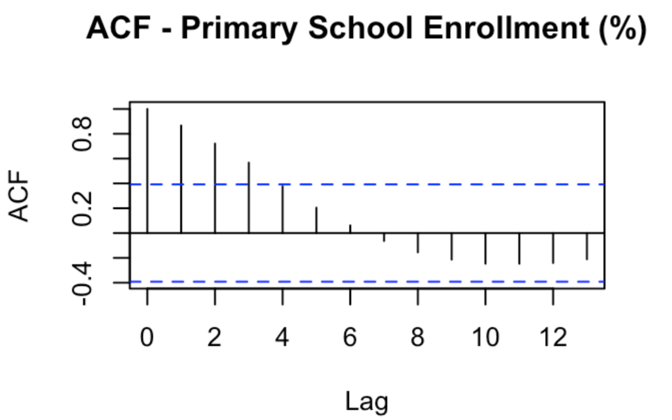
```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

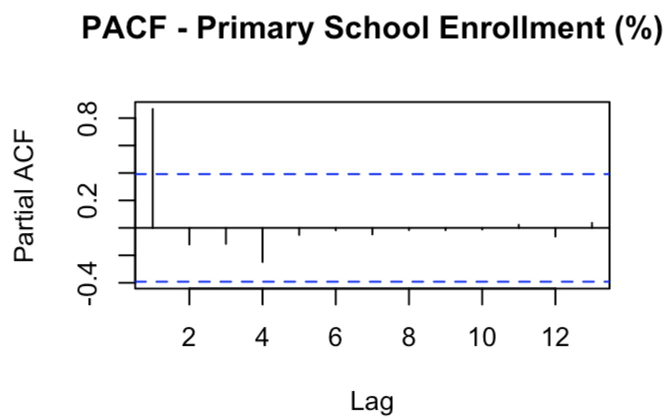

Time-Series Data for Government Expenditure (% GDP)
Values
3.0
2.5
2.0
1.5
1.0
2000
2005
2010
2015
2020
Time


Time-Series Data for Primary School Enrollment (%)
Values
90
85
80
75
70
2000
2005
2010
2015
2020
Time









The first time-series plot shows government expenditure on education as a percentage of GDP from 2000 to 2020. The trend indicates gradual fluctuations over the years, with a general upward trajectory. The values started at around 1.5% in the early 2000s, dipped slightly in the mid-2000s, and then climbed steadily, peaking at approximately 3.0% in the late 2010s. This increase in education expenditure might reflect the government's growing prioritization of education funding.

The ACF and PACF plots for government expenditure reveal significant autocorrelations, especially at lower lags, indicating that the current values are highly influenced by their recent historical values. The PACF plot shows a sharp decline after lag 1, suggesting that a simple autoregressive model (possibly AR(1)) could effectively describe the dynamics of government expenditure.

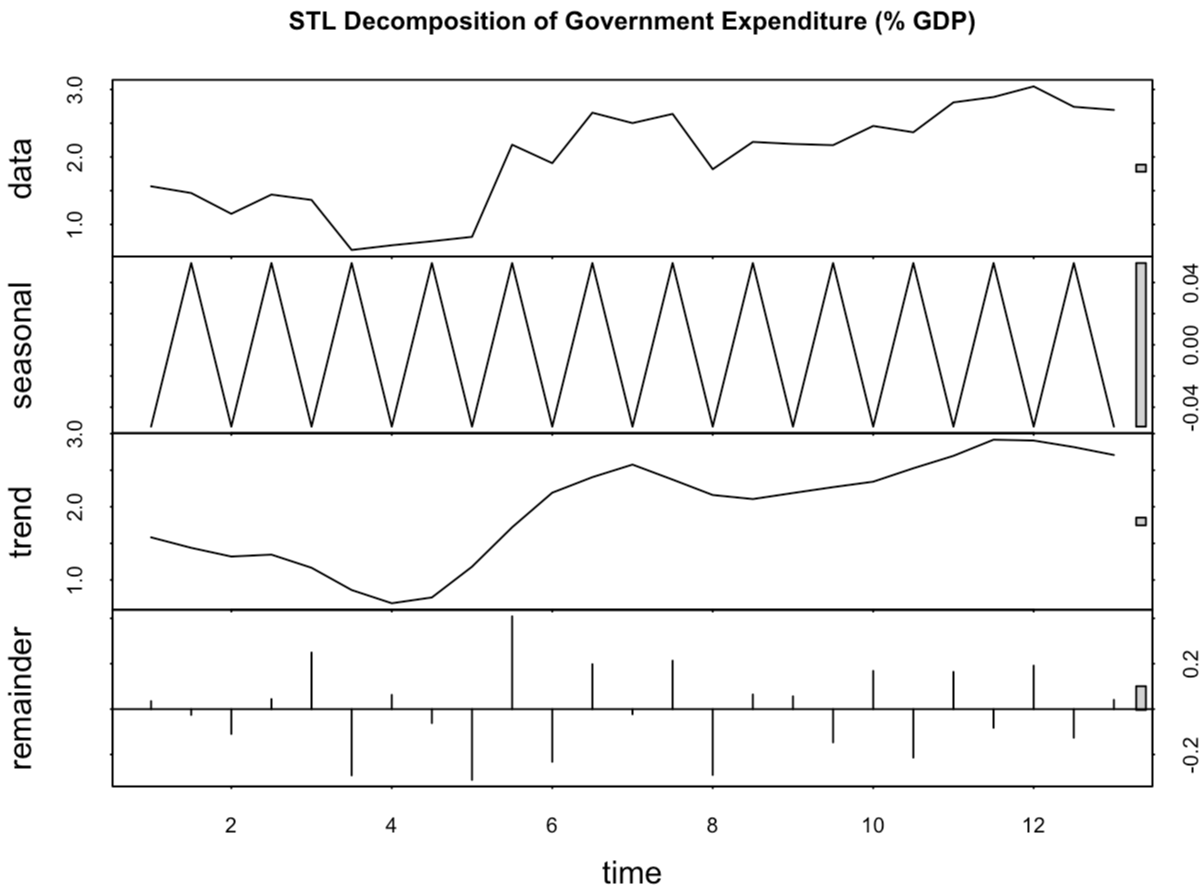
The second time-series plot depicts primary school enrollment rates as a percentage from 2000 to 2020. The trend demonstrates a marked increase over the years, with values beginning near 70% in 2000, declining slightly until the early 2010s, and then experiencing a steep rise to exceed 90% by 2020. This suggests a significant improvement in primary education access and participation in The Gambia, possibly correlating with the government's increased investment in education.

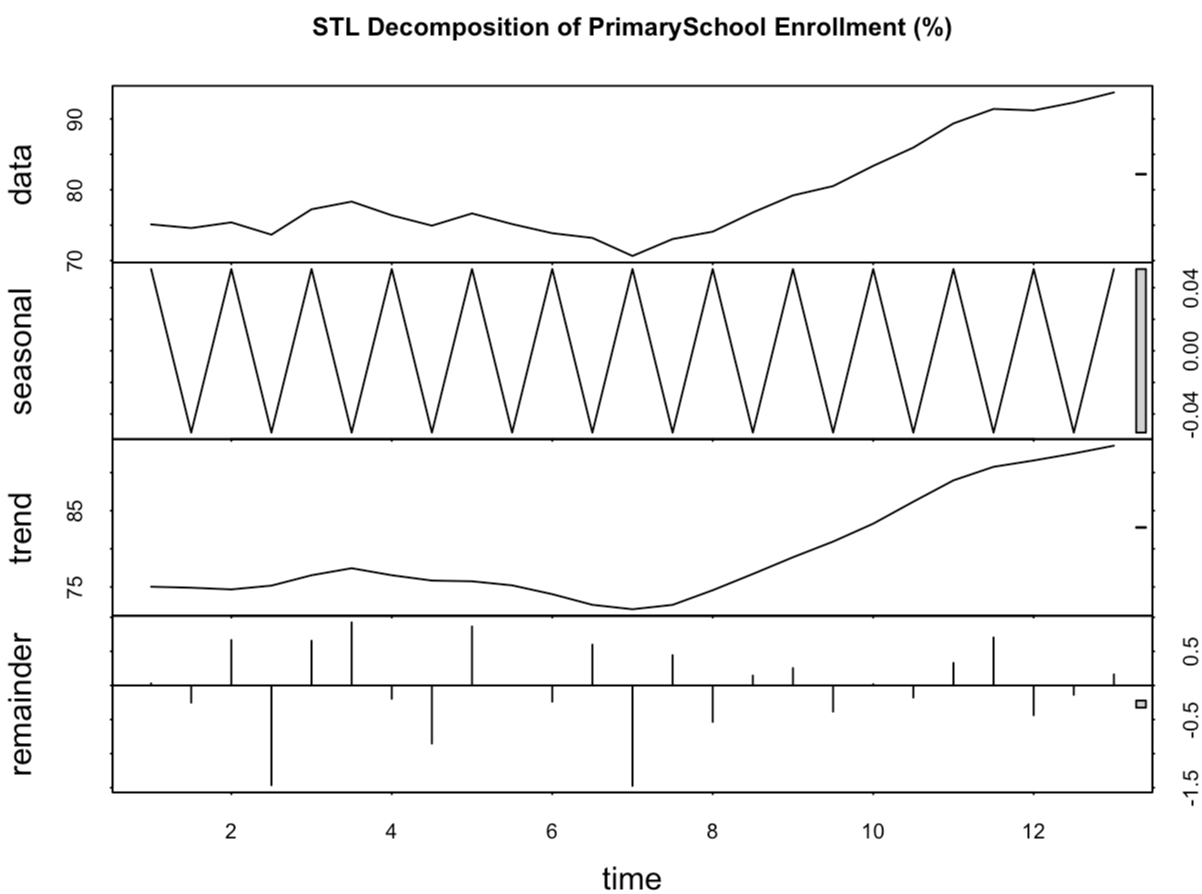
The ACF and PACF plots for primary school enrollment exhibit strong autocorrelations at multiple lags, emphasizing the persistence of trends over time. The PACF, however, drops significantly after lag 1, which again suggests that the process may follow an autoregressive structure, similar to government expenditure.

The upward trends in both government expenditure on education and primary school enrollment appear to align temporally. The rise in enrollment in the 2010s coincides with increased education funding, which may indicate that policy changes or budget allocations had a direct impact on improving education access. Further statistical analysis, such as Granger causality tests or regression modeling, would be necessary to establish a causal relationship between these two variables.

Overall, the data illustrates a positive trajectory for education in The Gambia, with consistent government investment and substantial gains in primary school enrollment over the two decades.

Now, let’s plot the stl decomposition plot of your data, and discuss the results. Due to the fact that we are using yearly data, we will use a placeholder frequency for the seasonal component.


STL Decomposition of Government Expenditure (% GDP)
data
seasonal
trend
remainder
time


STL Decomposition of PrimarySchool Enrollment (%)
data
seasonal
trend
remainder
time

The STL decomposition plots for primary school enrollment (%) and government expenditure on education (% GDP) in The Gambia reveal the underlying components of these time series: trend, seasonal (placeholder), and remainder. For primary school enrollment, the observed data shows a steady upward trend, rising from approximately 70% in the early 2000s to over 90% by 2020. This consistent improvement reflects increased access to primary education. The trend component highlights this long-term growth, particularly after 2010, where the upward trajectory becomes more pronounced. While the seasonal component is a placeholder with no real interpretive value, the remainder component captures minor short-term fluctuations, indicating that the trend accounts for most of the variation in the data.

Similarly, the decomposition of government expenditure shows an overall upward trend with some fluctuations. The observed data begins around 1% in the early 2000s and peaks at approximately 3% before slightly leveling off toward the later years. The trend component captures this long-term growth in education funding, particularly from 2005 onward. There is, however, a slight dip in the trend near the end of the period, suggesting a possible leveling off or reallocation of resources. Like primary school enrollment, the seasonal component is a placeholder with no real significance, and the remainder captures short-term irregularities or noise. The residuals for government expenditure are slightly larger than those for enrollment, but they remain relatively minor, emphasizing that the trend is the dominant driver in both time series.

Overall, the decompositions illustrate that both primary school enrollment and government expenditure are predominantly influenced by their respective trends, which indicate significant improvements over the years. The absence of true seasonality underscores that these changes are not cyclical but are likely driven by sustained policy efforts and socio-economic developments. The minimal remainder components in both cases suggest that the data is well-explained by its trends, with little unexplained variability. These findings highlight a positive trajectory for education in The Gambia, with increasing government investment and a corresponding rise in primary school enrollment rates.

Now, we will try to fit a model that includes, trend and cyclical components. We will not include a seasonal component as we know from the STL decomposition that with annual data, it is not adequate to include seasonality to our model. We will do this for the two time-series variables.

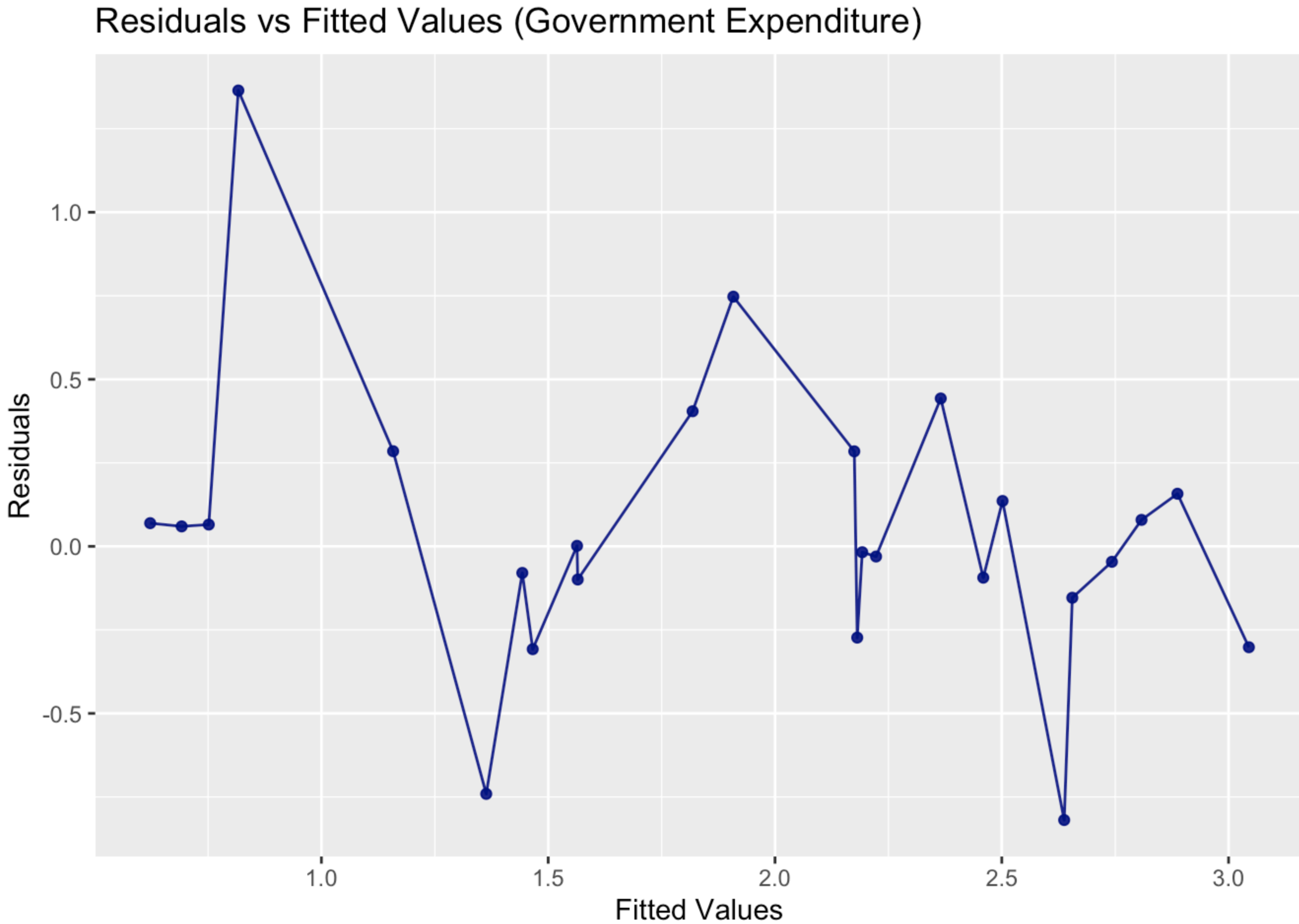
```
## Series: gov_exp_ts
## ARIMA(0,1,0)
##
## sigma^2 = 0.189:  log likelihood = -14.06
## AIC=30.13   AICc=30.31   BIC=31.31
##
## Training set error measures:
##                      ME      RMSE       MAE       MPE     MAPE      MASE
## Training set 0.04532119 0.4259792 0.2824072 -1.900567 17.50781 0.9602129
##                    ACF1
## Training set -0.2802746
```

```
## Series: primary_school_enrol_ts
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.7147
## s.e.   0.1818
##
## sigma^2 = 3.499:  log likelihood = -46.88
## AIC=97.77   AICc=98.37   BIC=100.04
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.2226705 1.754747 1.373711 0.2860308 1.768762 0.7784479
##                     ACF1
## Training set -0.03524263
```

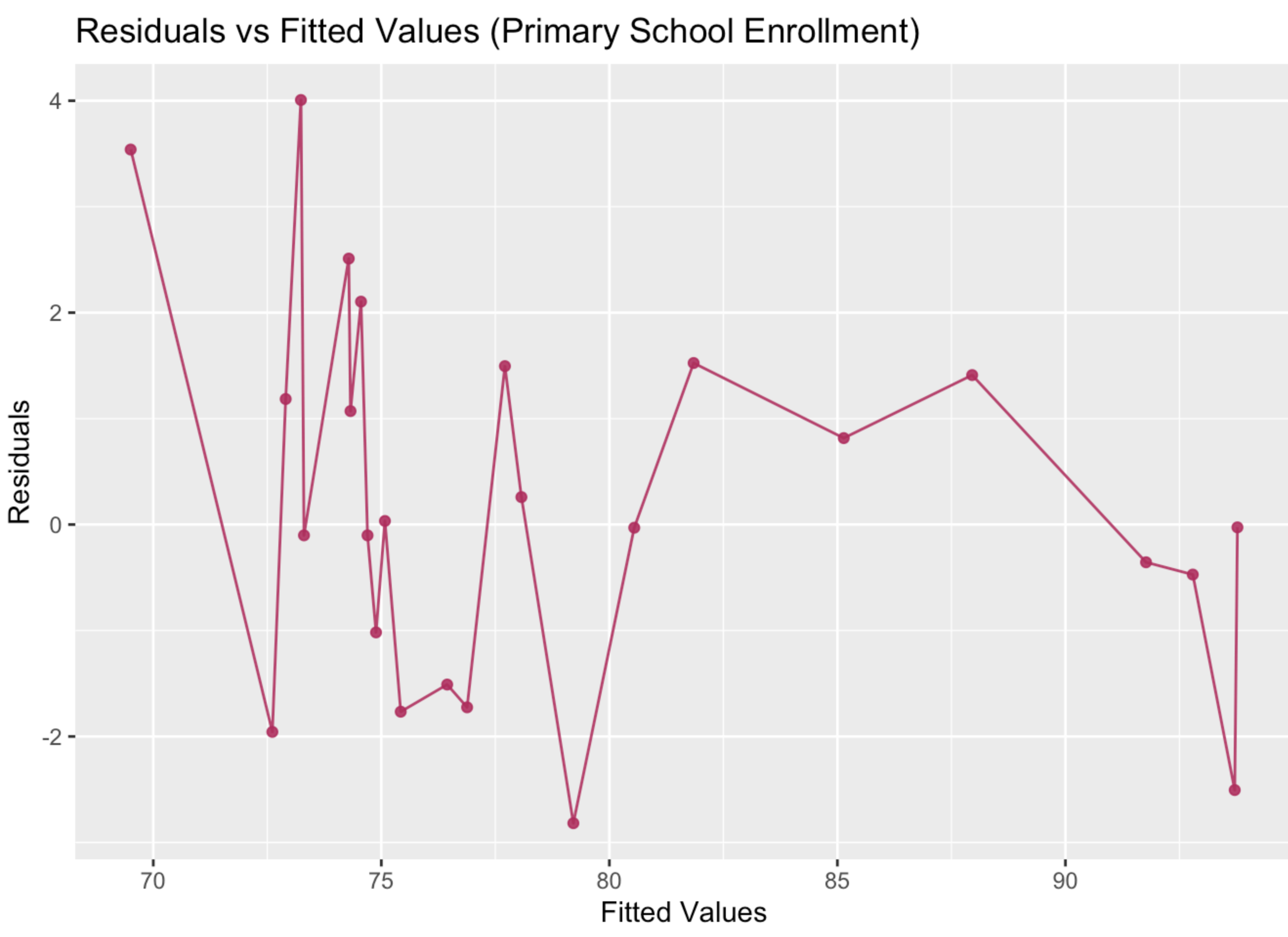
An ARIMA(0,1,0) model was selected for government expenditure, indicating that the time series is modeled as a random walk with one order of differencing to achieve stationarity and no autoregressive (AR) or moving average (MA) components. The small value of 0.189 for the σ^2 suggests that the model captures most of the variability in the data, and the relatively low AIC (30.13) and BIC (31.31) indicate a good fit. The training set error measures show a root mean square error (RMSE) of 0.426 and a mean absolute error (MAE) of 0.282, which are relatively small, suggesting that the model performs well in terms of predicting the values of the series. The negative mean percentage error (MPE = -1.90%) indicates a slight underestimation of values, but the mean absolute percentage error (MAPE = 17.51%) suggests that there is some variability in predictions relative to the actual values. The first-order autocorrelation (ACF1 = -0.2803) indicates that there is no significant autocorrelation left in the residuals, confirming the adequacy of the model.

For primary school enrollment, an ARIMA(0,2,1) model was selected, implying two orders of differencing to achieve stationarity and one moving average (MA) component to capture short-term dependencies. The MA coefficient being -0.7147 is statistically significant, indicating that short-term shocks are effectively modeled. The higher value of 3.499 for the value of σ^2 compared to the government expenditure model suggests greater variability in this time series. The AIC (97.77) and BIC (100.04) are higher than those for government expenditure, reflecting the greater complexity and variability of this model. The training set error measures include an RMSE of 1.754 and an MAE of 1.374, which are larger compared to government expenditure, reflecting the higher scale and variability of primary school enrollment. The MPE (0.29%) indicates minimal bias in the model's predictions, while the MAPE (1.77%) shows that the model provides accurate predictions relative to the actual values. The first-order autocorrelation (ACF1 = -0.0352) is close to zero, indicating that the residuals are uncorrelated, further supporting the model's adequacy.

Now, let's plot the respective residuals vs. fitted values.



Residuals vs Fitted Values (Government Expenditure)
Residuals
Fitted Values
1.0
0.5
0.0
-0.5
1.0
1.5
2.0
2.5
3.0



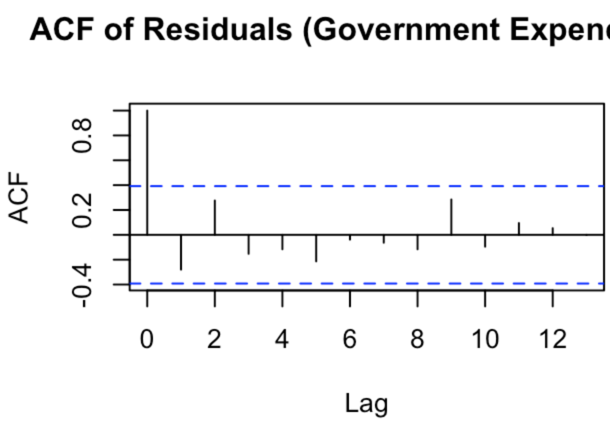
Residuals vs Fitted Values (Primary School Enrollment)
Residuals
Fitted Values
4
2
0
-2
70
75
80
85
90

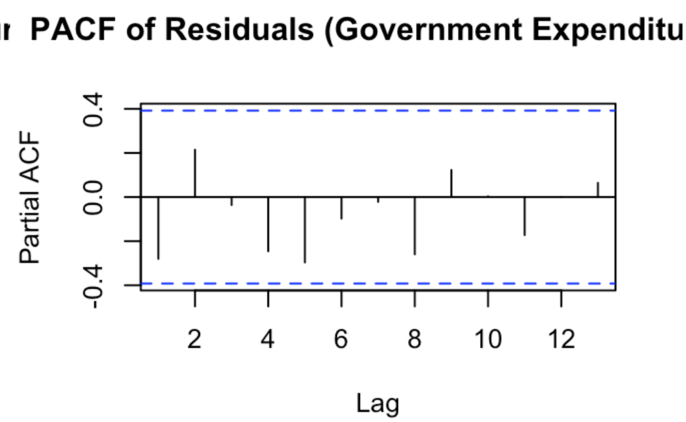
The residual plots for the ARIMA models of primary school enrollment and government expenditure reveal important insights into the model fit and the assumptions of randomness and independence of residuals. For primary school enrollment, the residuals display a relatively random distribution around zero, indicating that the ARIMA(0,2,1) model effectively captures the underlying patterns in the data. However, there is noticeable variability in the residuals at lower fitted values (around 70–75), suggesting that the model struggles to fully account for variability in this range. In contrast, at higher fitted values (85–90), the residuals exhibit a narrower spread, implying better predictive performance in this range. Importantly, no systematic patterns or trends are observed in the residuals, supporting the adequacy of the model and indicating that the errors are uncorrelated and free from missing explanatory variables.

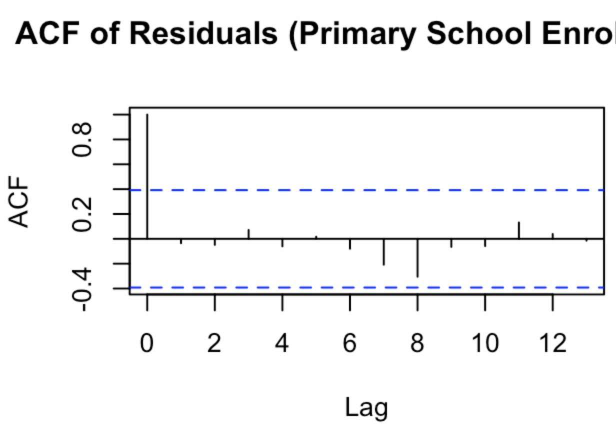
For government expenditure, the residuals also exhibit a random distribution around zero, reflecting a good fit of the ARIMA(0,1,0) model. The residuals are tightly clustered around zero, which is consistent with the stability and simplicity of the government expenditure time series. However, at lower fitted values (close to 1.0), there is a noticeable spike in residuals, suggesting some degree of underestimation or overestimation in this range. As the fitted values increase beyond 2.0, the residuals become smaller and more consistent, indicating better predictive accuracy for higher expenditure levels.

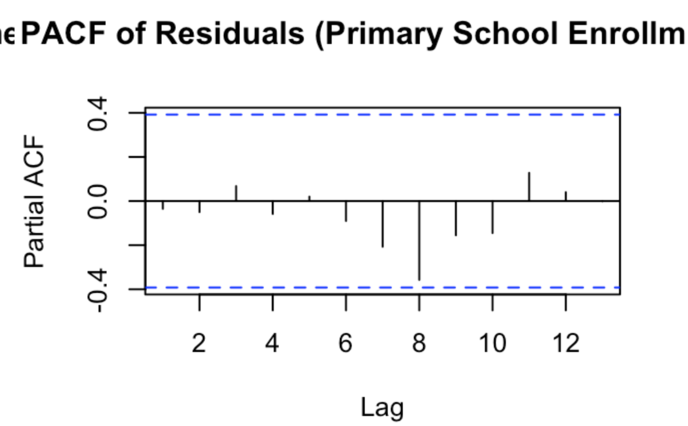
Overall, both residual plots support the assumption that the residuals are uncorrelated and normally distributed, validating the adequacy of the ARIMA models for both variables. While the residuals for primary school enrollment exhibit higher variability at lower fitted values, and those for government expenditure show a spike at the lowest fitted values, these issues are relatively minor. Further diagnostic tests, such as residual autocorrelation and normality tests, could be conducted to confirm these findings and potentially refine the models.

Now, let’s plot the ACF and PACF of the respective residuals. We will then interpret the plots.









The ACF and PACF plots for the residuals of the ARIMA models for government expenditure and primary school enrollment assess whether the residuals exhibit autocorrelation, which would indicate model inadequacy.

For government expenditure (% GDP), both the ACF and PACF plots show no significant lags outside the 95% confidence bounds, suggesting that the residuals are uncorrelated and exhibit white noise behavior. This indicates that the ARIMA(0,1,0) model sufficiently captures the structure of the data. The absence of significant autocorrelation supports the conclusion that there are no remaining patterns in the residuals, validating the adequacy of the model for this time series.

Similarly, the ACF and PACF plots for the residuals of primary school enrollment (%) show no significant autocorrelation at any lag, as all the spikes fall within the confidence bounds. This confirms that the ARIMA(0,2,1) model adequately explains the data and that the residuals are uncorrelated. The lack of significant partial autocorrelation further supports the conclusion that there are no remaining short-term dependencies or systematic patterns in the residuals.

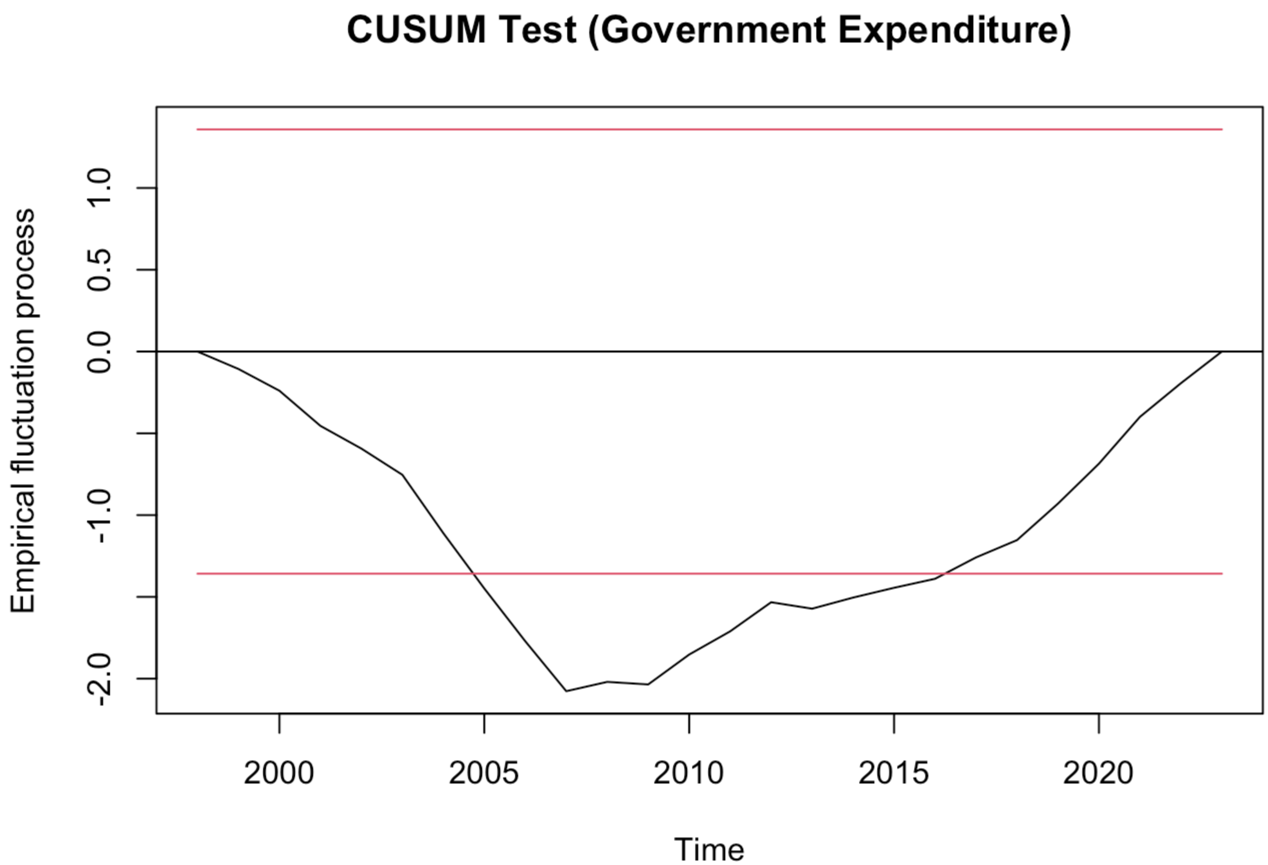
Overall, the ACF and PACF plots for both variables confirm that the models fit the data well and that the residuals behave as expected for valid ARIMA models. This further reinforces the reliability of the models for analyzing and forecasting these time series.

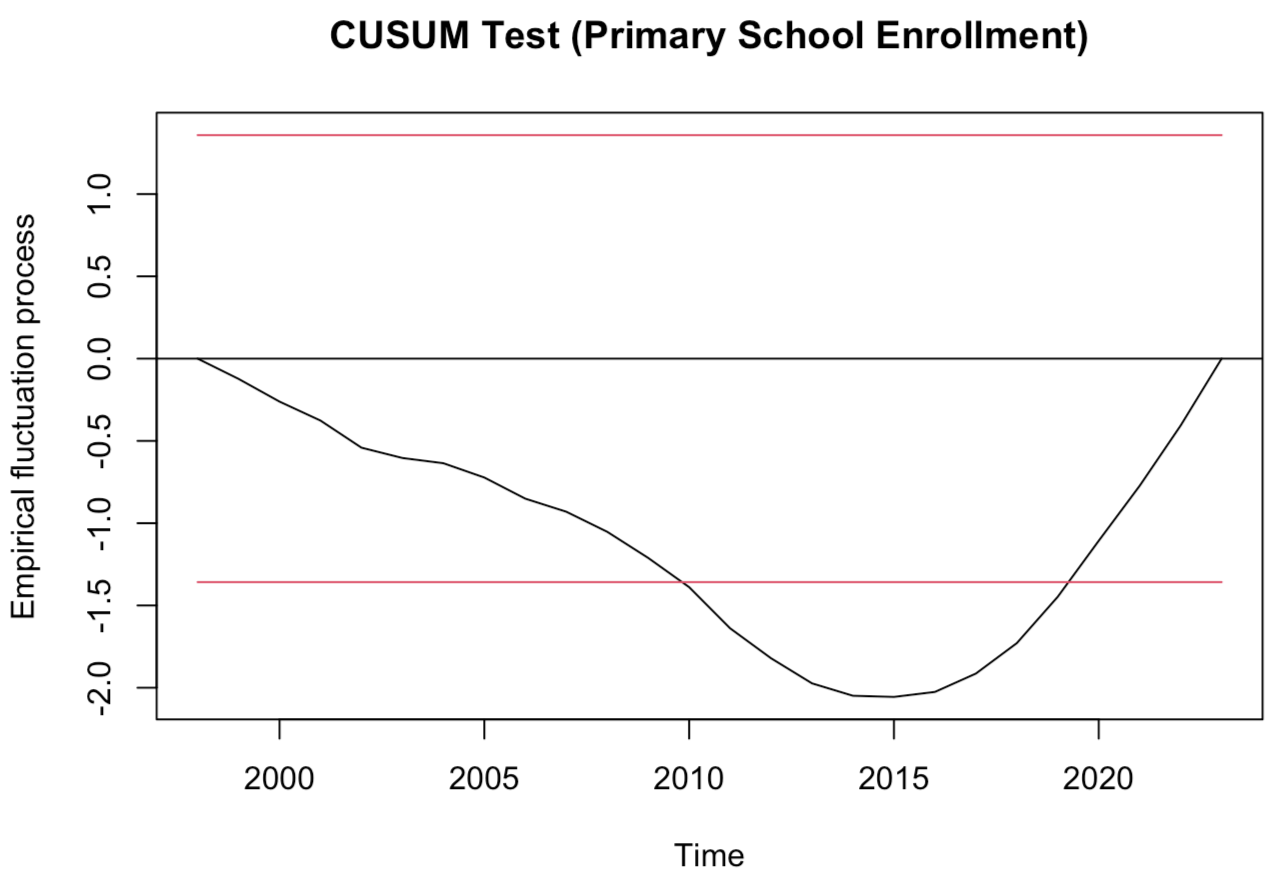
Now, let’s plot the respective CUSUM and interpret the plot.

```
## Loading required package: sandwich
```

```
##
## Attaching package: 'strucchange'
```

```
## The following object is masked from 'package:stringr':
##
##     boundary
```


CUSUM Test (Government Expenditure)
Empirical fluctuation process
1.0
0.5
0.0
-1.0
-2.0
2000
2005
2010
2015
2020
Time

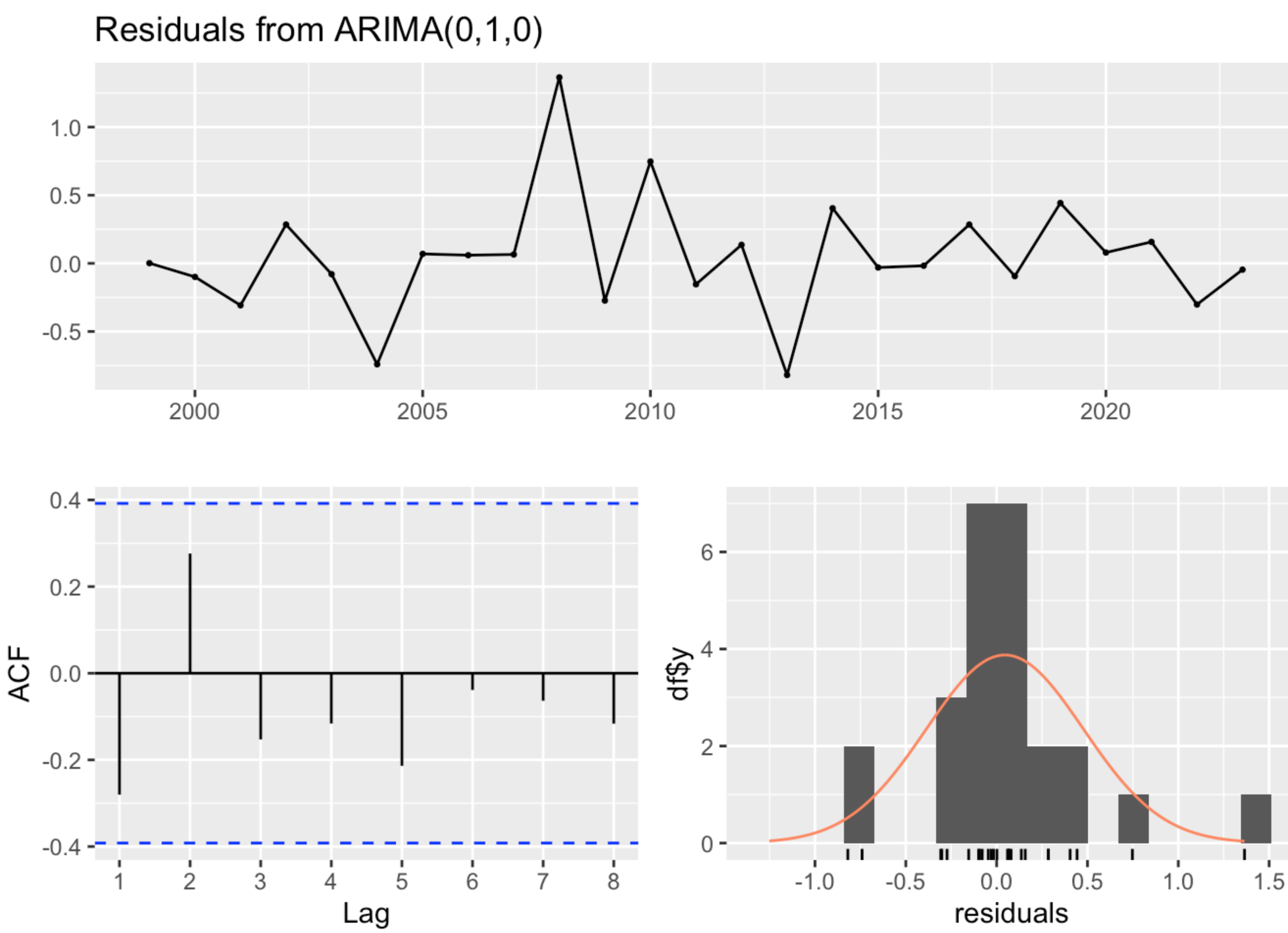

CUSUM Test (Primary School Enrollment)
Empirical fluctuation process
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
2000
2005
2010
2015
2020
Time

Looking at the plots, we can see that the CUSUM plot for government expenditure indicates that the cumulative sum of residuals stays well within the critical bounds (represented by the red lines) throughout the time period. This suggests that the ARIMA(0,1,0) model is stable and effectively captures the dynamics of the data without significant structural breaks or model misspecifications. While the empirical fluctuation process dips below zero during the early 2000s and rises steadily in the later years, it remains confined within the bounds, confirming that the residuals fluctuate randomly around zero and do not exhibit systematic patterns.

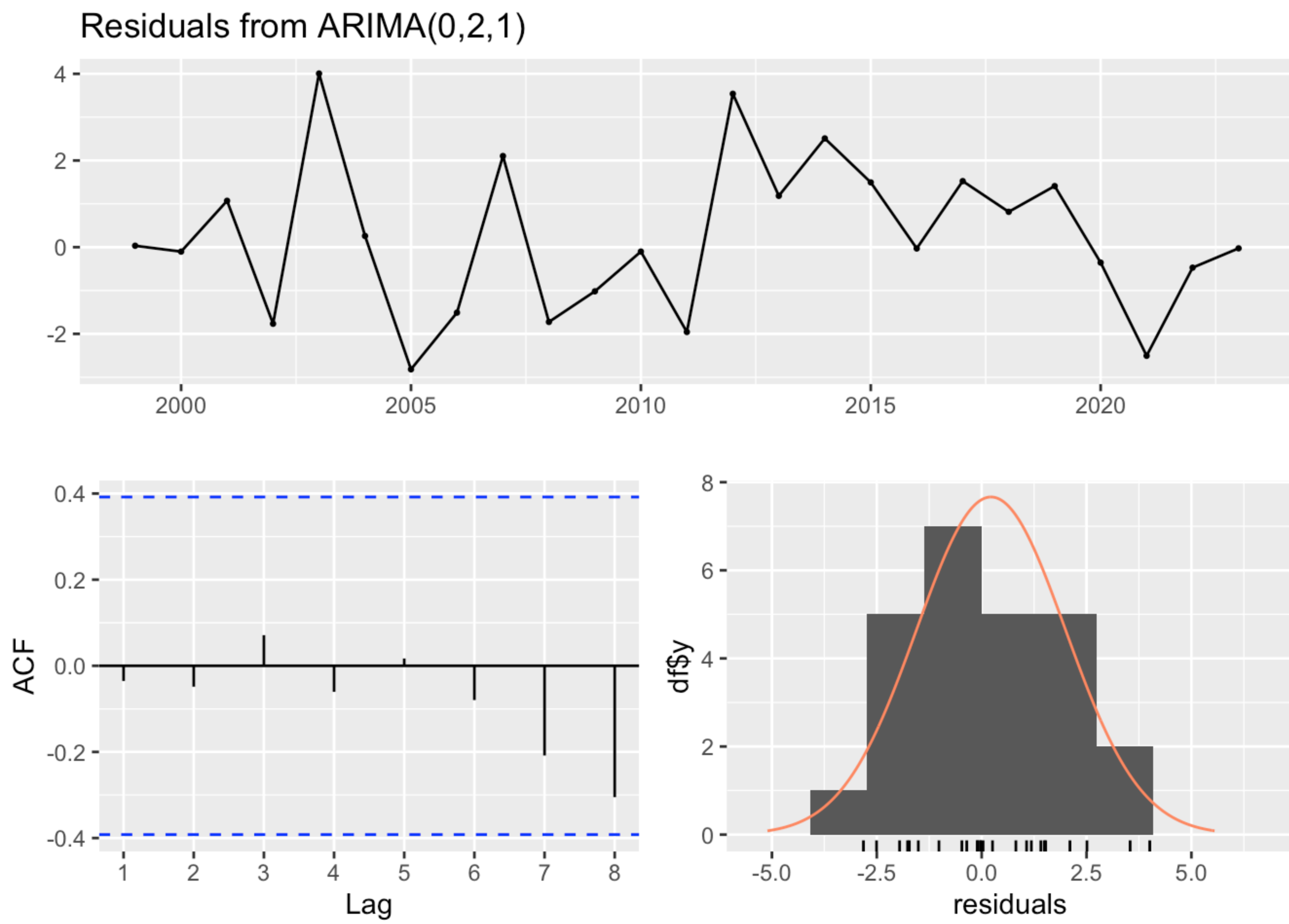
The CUSUM plot for primary school enrollment also shows that the cumulative sum of residuals stays within the critical bounds, indicating no evidence of instability or structural breaks in the ARIMA(0,2,1) model. Similar to government expenditure, the plot initially trends downward during the early years but rises steadily after 2010. This behavior reflects the model's ability to account for long-term changes in primary school enrollment without introducing systematic bias or missing significant patterns. The upward trajectory in the later years aligns with the upward trend observed in the data but does not cross the critical bounds, reinforcing the adequacy of the model.

Both CUSUM test plots confirm the stability of the ARIMA models over time, as the cumulative sum of residuals remains within the critical bounds for both variables. This indicates that the models are well-specified, with no significant structural breaks or changes not modeled in the data. The results validate the reliability of the models for analyzing and forecasting government expenditure and primary school enrollment in The Gambia.

Now, for the associated diagnostic statistics, we get:



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)
## Q* = 7.1419, df = 5, p-value = 0.2103
##
## Model df: 0.   Total lags used: 5
```



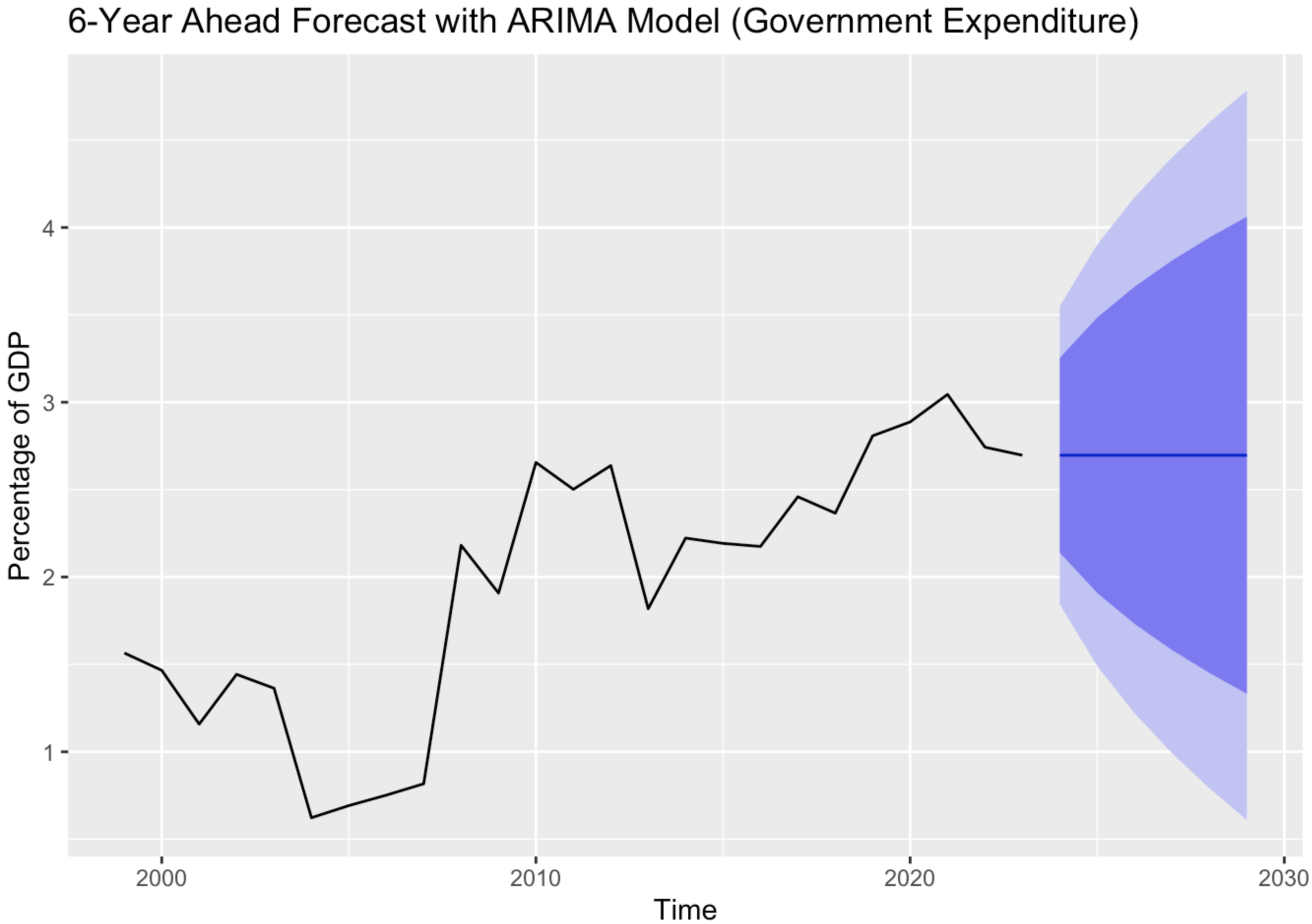

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,2,1)
## Q* = 0.3869, df = 4, p-value = 0.9835
##
## Model df: 1.   Total lags used: 5
```

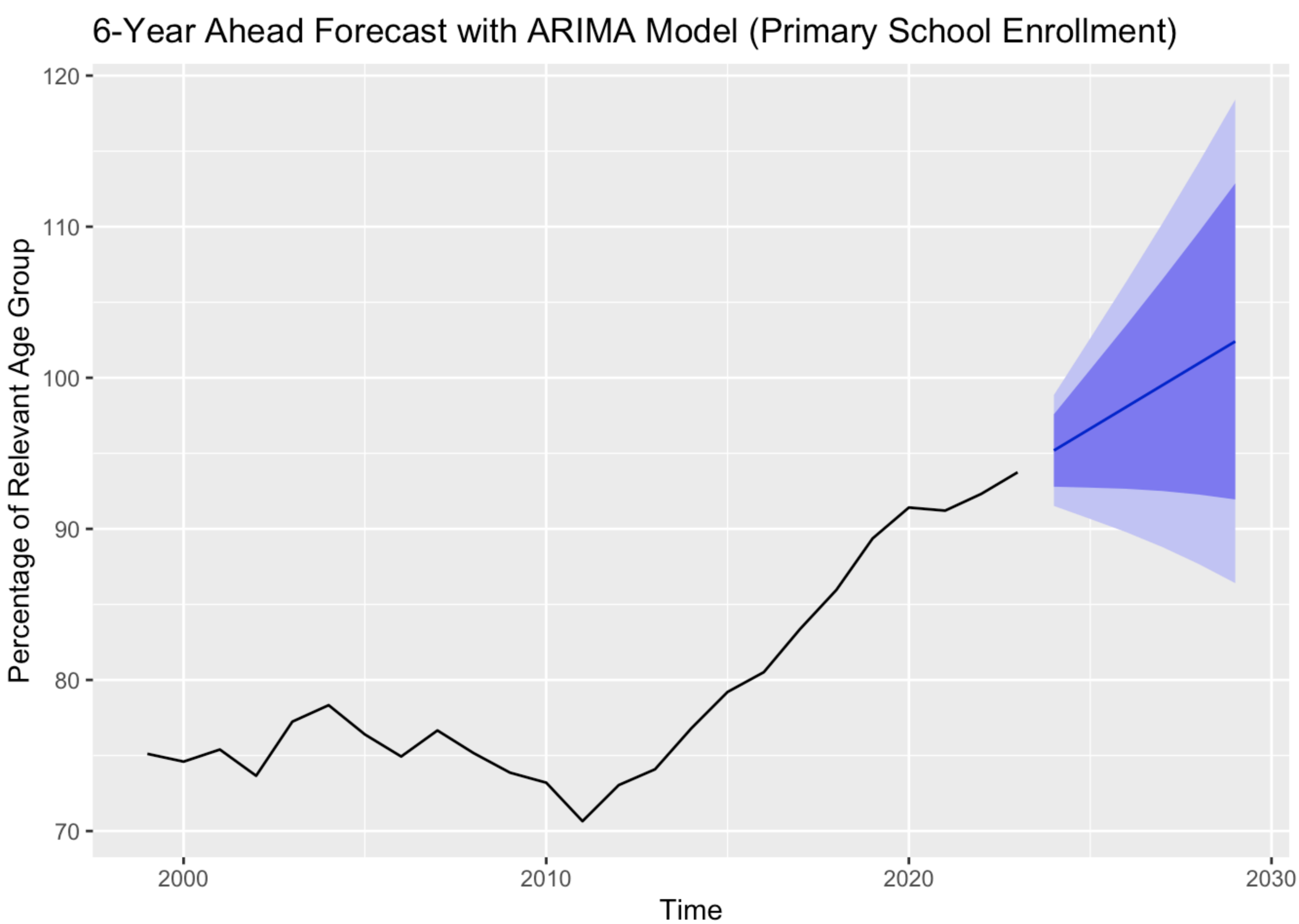
The diagnostic plots for the ARIMA models of government expenditures and primary school enrollment provide evidence that the models are well-specified and adequately capture the underlying patterns in the data. For government expenditures, the time plot of residuals shows random fluctuations around zero with no visible trends or systematic patterns, indicating that the ARIMA(0,1,0) model effectively models the data. The autocorrelation function (ACF) plot reveals that all spikes fall within the confidence bounds, confirming that the residuals are uncorrelated and exhibit white noise behavior. Additionally, the histogram of residuals is approximately normal, with only a slight skew to the right, and the overlaid density curve closely matches the residual distribution, supporting the assumption of normally distributed errors.

Similarly, the diagnostic plots for primary school enrollment demonstrate the adequacy of the ARIMA(0,2,1) model. The time plot of residuals shows that they fluctuate around zero without any discernible patterns, though the residuals are larger in magnitude compared to those for government expenditures, reflecting the greater variability in the primary school enrollment data. The ACF plot indicates that all residual autocorrelations fall within the confidence bounds, suggesting that the residuals are uncorrelated and exhibit white noise. The histogram of residuals is approximately normal, with minor deviations at the tails, and the density curve aligns well with the residual distribution, further validating the assumption of near-normal residuals.

Overall, these diagnostic plots confirm that both ARIMA models are appropriate for their respective time series. The residuals are uncorrelated, approximately normally distributed, and free from systematic patterns, which validates the models as reliable tools for forecasting and analysis. The slightly higher residual variability in the primary school enrollment model reflects the greater inherent variability in this series compared to government expenditures.

Let’s now use our model to forecast 6-steps ahead.


6-Year Ahead Forecast with ARIMA Model (Government Expenditure)
Percentage of GDP
1
2
3
4
2000
2010
2020
2030
Time


6-Year Ahead Forecast with ARIMA Model (Primary School Enrollment)
Percentage of Relevant Age Group
70
80
90
100
110
120
2000
2010
2020
2030
Time

Looking at the forecast plot above, government expenditure seems to stabilize more over the next 6 years. On the other hand, primary school enrollment is expected to increase based on the forecast.

Now, let's compare our forecast from the previous part to the 6-steps ahead forecasts from the ETS model.

```
## ETS(A,N,N)
##
## Call:
## ets(y = gov_exp_ts)
##
##   Smoothing parameters:
##     alpha = 0.8147
##
##   Initial states:
##     l = 1.5375
##
##   sigma:  0.4334
##
##      AIC     AICc      BIC
## 42.57953 43.72239 46.23616
##
## Training set error measures:
##                      ME      RMSE       MAE      MPE     MAPE      MASE
## Training set 0.05778211 0.4156768 0.2595073 -1.94327 16.33026 0.8823509
##                     ACF1
## Training set -0.06565421
```

```
## ETS(A,N,N)
##
## Call:
## ets(y = primary_school_enrol_ts)
##
##   Smoothing parameters:
##     alpha = 0.9999
##
##   Initial states:
##     l = 75.1154
##
##   sigma:  2.014
##
##      AIC     AICc      BIC
## 119.3937 120.5365 123.0503
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
## Training set 0.7452609 1.931767 1.694276 0.8563243 2.138076 0.9601039 0.2439295
```

```
##                      ME      RMSE       MAE       MPE     MAPE      MASE
## Training set 0.04532119 0.4259792 0.2824072 -1.900567 17.50781 0.9602129
##                    ACF1
## Training set -0.2802746
```

```
##                       ME      RMSE       MAE      MPE     MAPE      MASE
## Training set 0.05778211 0.4156768 0.2595073 -1.94327 16.33026 0.8823509
##                     ACF1
## Training set -0.06565421
```

```
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.2226705 1.754747 1.373711 0.2860308 1.768762 0.7784479
##                     ACF1
## Training set -0.03524263
```

```
##                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
## Training set 0.7452609 1.931767 1.694276 0.8563243 2.138076 0.9601039 0.2439295
```

In terms of MAPE, the model that performs the best is ETS model for the government expenditure (% GDP). This is because the ETS has a lower MAPE value for government expenditure. However, the ARIMA model performs better for primary school enrollment since it has the lower MAPE value in this case.

Now, we will compare compare the model we first used (the ARIMA model) to the ETS model, Holt-Winters model, NNETAR model, Prophet model, and a forecast combination model.

First, let's split the data into training and testing sets. We will use 80% of the data as the training set and the remaining 20% as the testing set.

Let's now compare the training set and testing set errors for all models. First, let's create all of the forecast models.

```
## Loading required package: generics
```

```
##
## Attaching package: 'generics'
```

```
## The following object is masked from 'package:sandwich':
##
##     estfun
```

```
## The following object is masked from 'package:lubridate':
##
##     as.difftime
```

```
## The following object is masked from 'package:dplyr':
##
##     explain
```

```
## The following objects are masked from 'package:base':
##
##     as.difftime, as.factor, as.ordered, intersect, is.element, setdiff,
##     setequal, union
```

```
## Registered S3 methods overwritten by 'tsutils':
##   method          from
##   print.nemenyi   greybox
##   summary.nemenyi greybox
```

```
## Loading required package: Rcpp
```

```
## Loading required package: rlang
```

```
##
## Attaching package: 'rlang'
```

```
## The following objects are masked from 'package:purrr':
##
##     %@%, flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
##     flatten_raw, invoke, splice
```

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override th
is.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override thi
s.
```

```
## n.changepoints greater than number of observations. Using 15
```

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override th
is.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override thi
s.
```

```
## n.changepoints greater than number of observations. Using 15
```

Now, let's look at the training set errors.

```
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
```

```
## Warning in (actual - forecasted)/actual: longer object length is not a multiple
## of shorter object length
```

```
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
```

```
## Warning in (actual - forecasted)/actual: longer object length is not a multiple
## of shorter object length
```

```
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
## Warning in actual - forecasted: longer object length is not a multiple of
## shorter object length
```

```
# ARIMA Model Training Errors (Government Expenditure)
print(arima_errors_gov_train)
```

```
##       MAPE      RMSE       MAE
## 1 20.06467 0.4591371 0.3016333
```

```
# ETS Model Training Errors (Government Expenditure)
print(ets_errors_gov_train)
```

```
##       MAPE      RMSE       MAE
## 1 18.34387 0.4465545 0.2654828
```

```
# Holt-Winters Model Training Errors (Government Expenditure)
print(hw_errors_gov_train)
```

```
##       MAPE     RMSE      MAE
## 1 70.28286 1.376739 1.096249
```

```
# NNETAR Model Training Errors (Government Expenditure)
print(nnetar_errors_gov_train)
```

```
##       MAPE     RMSE      MAE
## 1 21.55469 0.367783 0.284993
```

```
# Prophet Model Training Errors (Government Expenditure)
print(prophet_errors_gov_train)
```

```
##       MAPE     RMSE       MAE
## 1 33.51442 0.473907 0.3875245
```

```
# Combined Forecast Model Training Errors (Government Expenditure)
print(combined_errors_gov_train)
```

```
##       MAPE      RMSE       MAE
## 1 20.98997 0.3950707 0.2811364
```

```
# ARIMA Model Training Errors (Primary School Enrollment)
print(arima_errors_primary_school_enrol_train)
```

```
##       MAPE     RMSE      MAE
## 1 2.305151 1.944819 1.770498
```

```
# ETS Model Training Errors (Primary School Enrollment)
print(ets_errors_primary_school_enrol_train)
```

```
##       MAPE     RMSE      MAE
## 1 2.225292 1.925648 1.710072
```

```
# Holt-Winters Model Training Errors (Primary School Enrollment)
print(hw_errors_primary_school_enrol_train)
```

```
##       MAPE     RMSE      MAE
## 1 51.77737 54.27669 39.72764
```

```
# NNETAR Model Training Errors (Primary School Enrollment)
print(nnetar_errors_primary_school_enrol_train)
```

```
##       MAPE    RMSE      MAE
## 1 1.818902 1.66058 1.377791
```

```
# Prophet Model Training Errors (Primary School Enrollment)
print(prophet_errors_primary_school_enrol_train)
```

```
##        MAPE      RMSE       MAE
## 1 0.9242615 0.9285173 0.7021812
```

```
# Combined Forecast Model Training Errors (Primary School Enrollment)
print(combined_errors_primary_school_enrol_train)
```

```
##       MAPE     RMSE      MAE
## 1 1.685231 1.467699 1.292483
```

```
# ARIMA Model Testing Errors (Government Expenditure)
print(arima_errors_gov_test)
```

```
##      MAPE      RMSE       MAE
## 1 16.4379 0.4861877 0.4704836
```

```
# ETS Model Testing Errors (Government Expenditure)
print(ets_errors_gov_test)
```

```
##       MAPE      RMSE       MAE
## 1 16.17312 0.4789384 0.4629885
```

```
# Holt-Winters Model Testing Errors (Government Expenditure)
print(hw_errors_gov_test)
```

```
##       MAPE      RMSE       MAE
## 1 20.66503 0.6128918 0.5904435
```

```
# NNETAR Model Testing Errors (Government Expenditure)
print(nnetar_errors_gov_test)
```

```
##       MAPE      RMSE       MAE
## 1 18.91457 0.5543138 0.5405923
```

```
# Prophet Model Testing Errors (Government Expenditure)
print(prophet_errors_gov_test)
```

```
##       MAPE      RMSE       MAE
## 1 6.798229 0.2361271 0.1945094
```

```
# Combined Forecast Model Testing Errors (Government Expenditure)
print(combined_errors_gov_test)
```

```
##       MAPE      RMSE       MAE
## 1 11.25333 0.3605838 0.3250321
```

```
# ARIMA Model Testing Errors (Primary School Enrollment)
print(arima_errors_primary_school_enrol_test)
```

```
##       MAPE     RMSE      MAE
## 1 8.047545 7.710472 7.404122
```

```
# ETS Model Testing Errors (Primary School Enrollment)
print(ets_errors_primary_school_enrol_test)
```

```
##       MAPE     RMSE      MAE
## 1 6.149689 5.834225 5.654871
```

```
# Holt-Winters Model Testing Errors (Primary School Enrollment)
print(hw_errors_primary_school_enrol_test)
```

```
##       MAPE    RMSE      MAE
## 1 2.039482 2.06355 1.879522
```

```
# NNETAR Model Testing Errors (Primary School Enrollment)
print(nnetar_errors_primary_school_enrol_test)
```

```
##       MAPE     RMSE      MAE
## 1 3.747533 3.679431 3.452138
```

```
# Prophet Model Testing Errors (Primary School Enrollment)
print(prophet_errors_primary_school_enrol_test)
```

```
##       MAPE     RMSE      MAE
## 1 1.477615 1.520166 1.353128
```

```
# Combined Forecast Model Testing Errors (Primary School Enrollment)
print(combined_errors_primary_school_enrol_test)
```

```
##       MAPE     RMSE      MAE
## 1 4.772222 4.447018 4.382377
```

From the training set errors, we can see that the ETS model has the lowest MAPE (Mean Absolute Percentage Error) and MAE (Mean Absolute Error) for government expenditure, and the NNETAR model has the lowest RMSE (Root Mean Squared Error). So, the ETS model performs best based on the training data for government expenditure. For primary school enrollment, the Prophet model seems to perform best for the training data as it has the lowest values for all three measurements of error (MAPE, RMSE, and MAE).

Additionally, when we look at the testing data, we can observe that the Prophet model also has the lowest values for all three measurements of error. This applies to both the government expenditure testing data and the primary school enrollment testing data.

Therefore, based on the training errors and testing errors, the Prophet model seems to be the preferred model.

Let’s now plot the time series for both government expenditures and primary school enrollment using the Prophet model.

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override th
is.
```

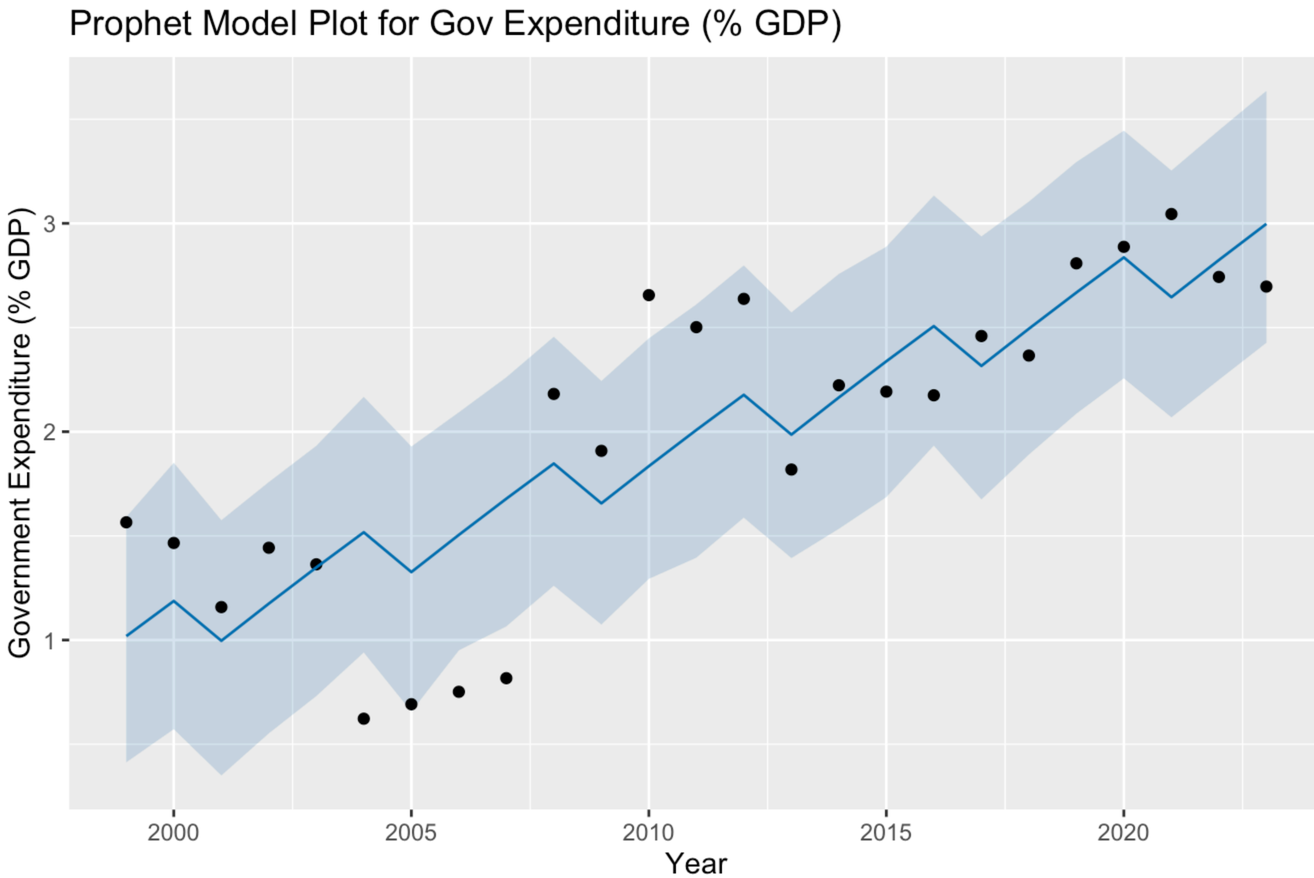
```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override thi
s.
```

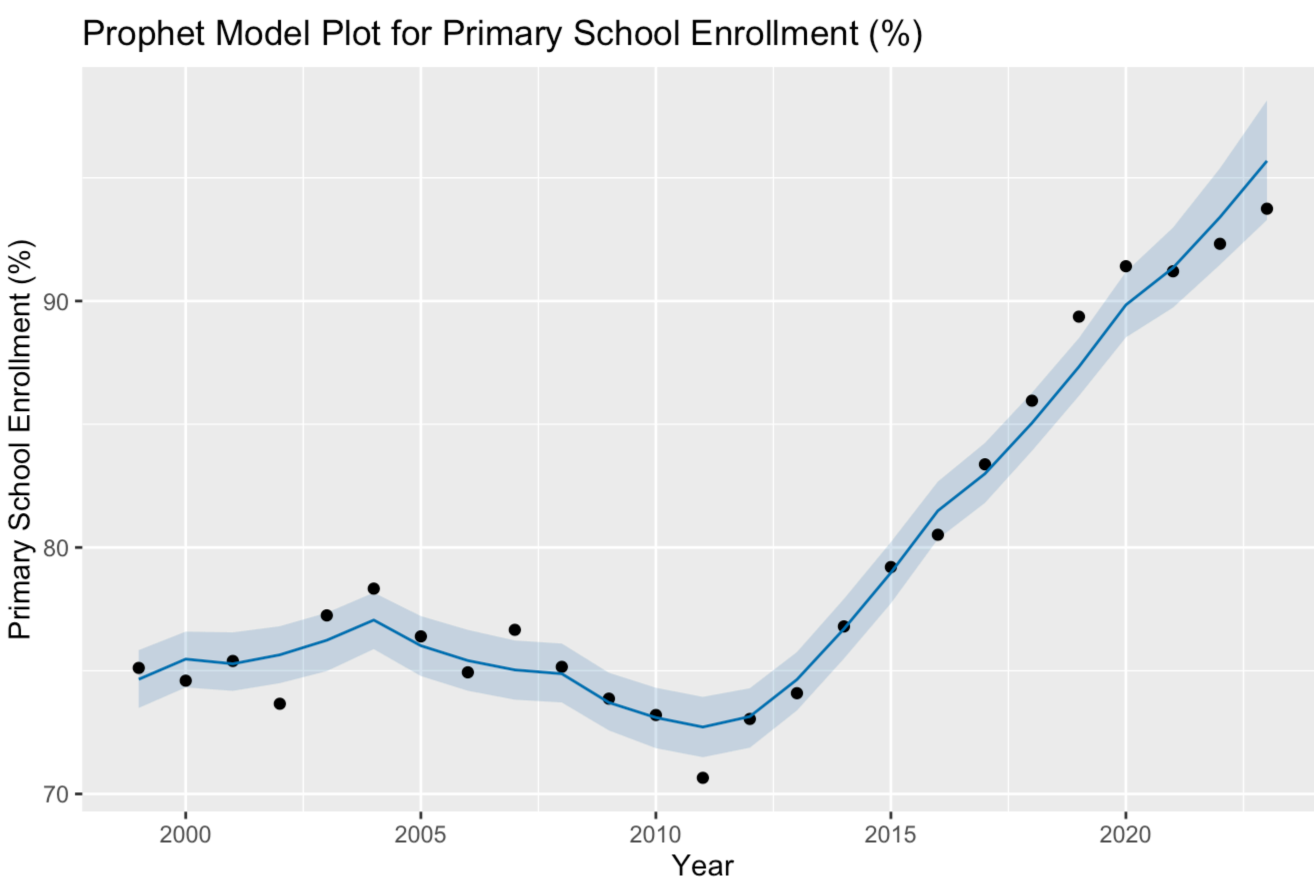
```
## n.changepoints greater than number of observations. Using 19
```

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override th
is.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override thi
s.
```

```
## n.changepoints greater than number of observations. Using 19
```


Prophet Model Plot for Gov Expenditure (% GDP)
Government Expenditure (% GDP)
3
2
1
2000
2005
2010
2015
2020
Year


Prophet Model Plot for Primary School Enrollment (%)
Primary School Enrollment (%)
90
80
70
2000
2005
2010
2015
2020
Year

The Prophet model plots for government expenditure (% GDP) and primary school enrollment (%) demonstrate how the model captures the overall trend and variability of each time series, along with uncertainty intervals.

The Prophet model for government expenditure shows a steady upward trend over time, reflecting an increase in the percentage of GDP allocated to education. The blue line represents the fitted values, while the shaded area indicates the 95% uncertainty interval. The model captures the general upward trajectory of government expenditure, with most of the observed data points (black dots) falling within the uncertainty interval. However, some data points deviate slightly, particularly in the early 2000s and mid-2010s, which could indicate small periods of underestimation or overestimation by the model. Overall, the model performs well in fitting the trend and accounting for variations in the data, demonstrating its effectiveness for forecasting future expenditures.

The Prophet model for primary school enrollment reveals a pronounced upward trend, particularly after 2010, indicating significant improvements in enrollment rates over time. The fitted values (blue line) closely follow the observed data points (black dots), and the uncertainty interval effectively encompasses most of the data, reflecting the model's robustness. The model captures the gradual decline in enrollment rates from 2000 to around 2010 and the sharp increase thereafter. The wider uncertainty interval in the later years reflects the increased variability and rapid growth in the enrollment rates, but the model still aligns well with the observed data, showcasing its capability to handle complex trends.

Overall, both plots indicate that the Prophet model successfully captures the underlying trends and variability in government expenditure and primary school enrollment. The uncertainty intervals are appropriately sized, reflecting the variability in each series while providing reasonable bounds for prediction. While the model performs better with the smoother upward trend of government expenditure, it also handles the more dynamic changes in primary school enrollment effectively. These results highlight the Prophet model's flexibility and accuracy in analyzing and forecasting time series data.

Next, let's fit an appropriate VAR model using your two variables. We will then show the relevant plots and discuss the results from the fit.

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

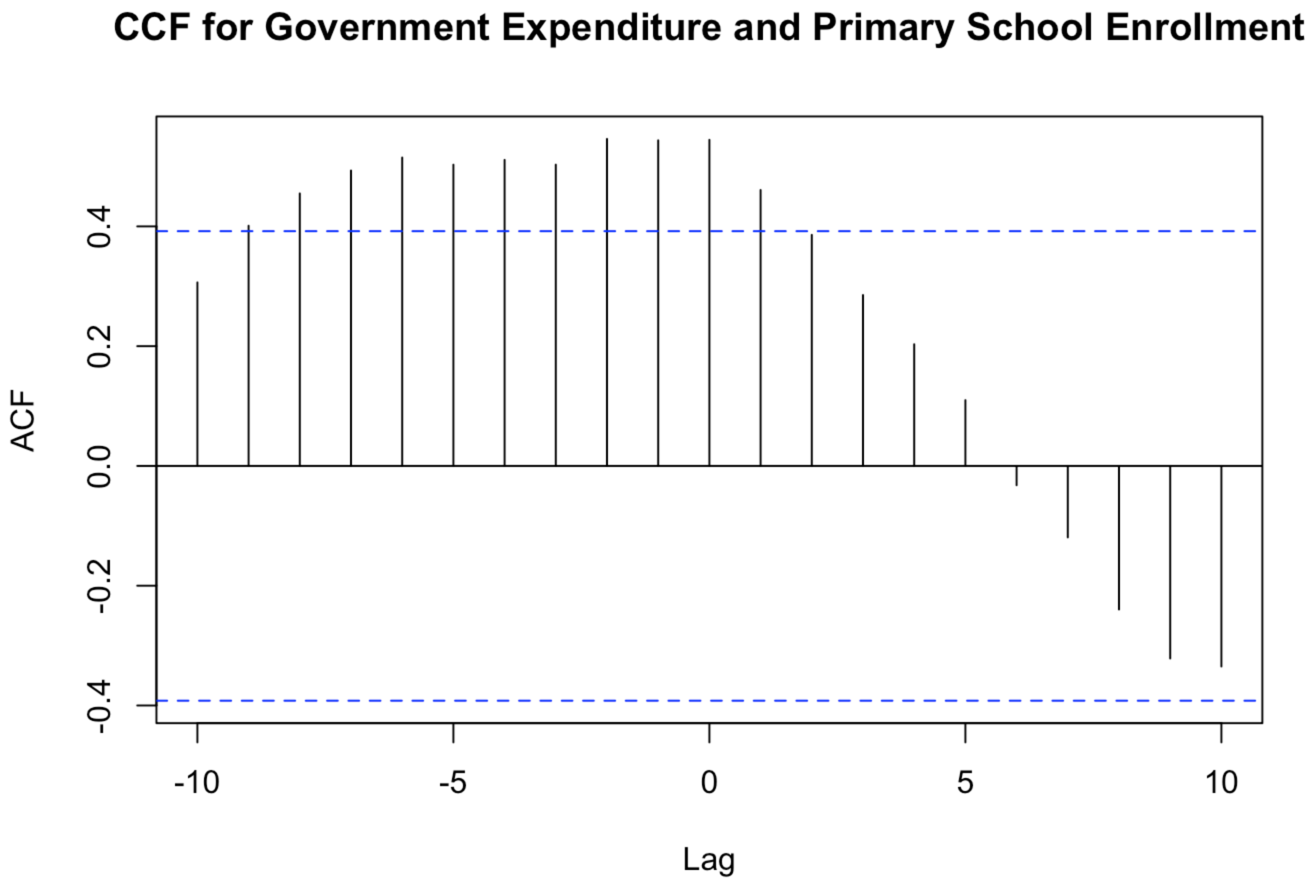
```
## The following object is masked from 'package:dplyr':
##
##     select
```

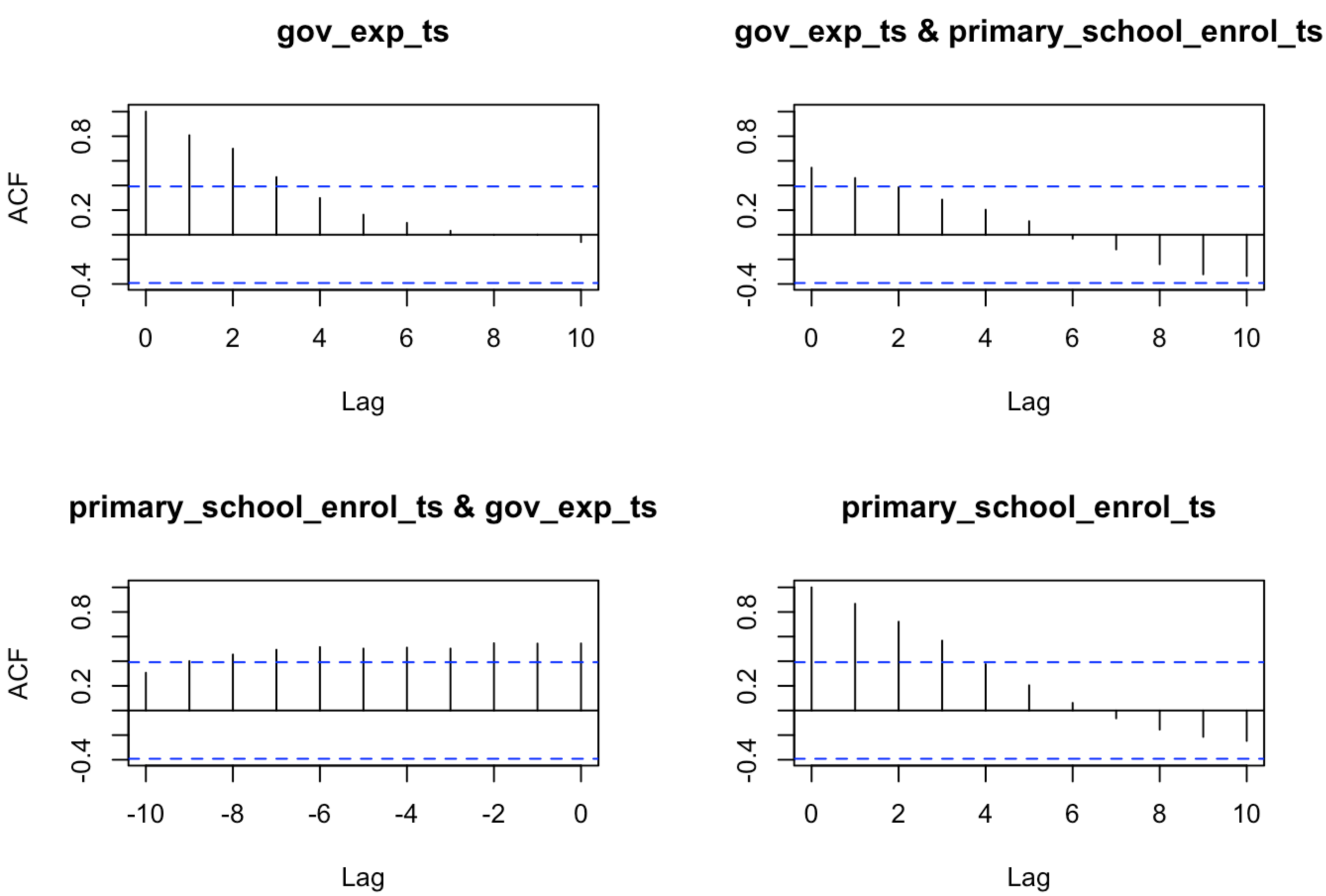
```
## Loading required package: urca
```

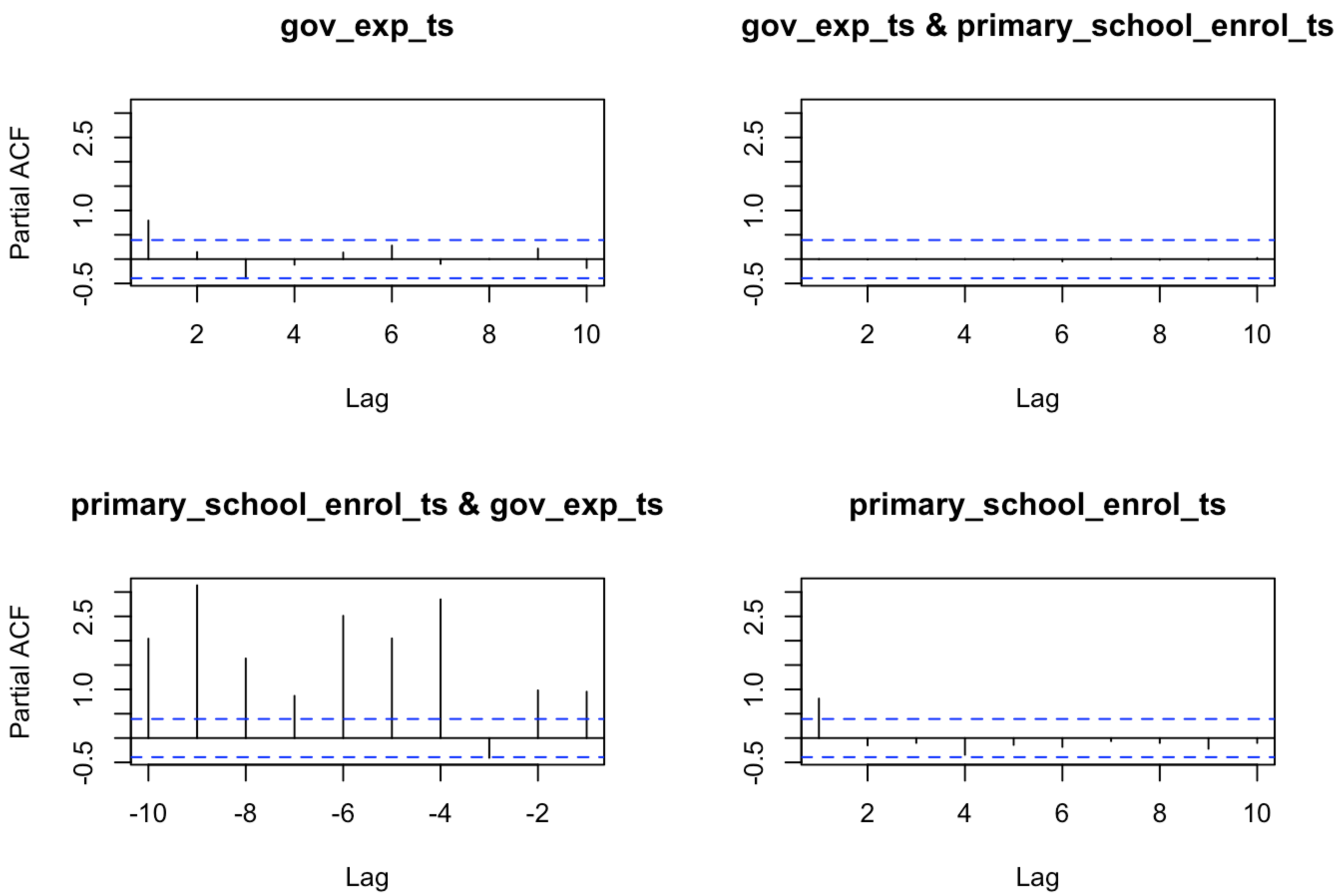
```
## Loading required package: lmtest
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: gov_exp_ts, primary_school_enrol_ts
## Deterministic variables: const
## Sample size: 23
## Log Likelihood: -56.1
## Roots of the characteristic polynomial:
## 1.037 0.7363 0.2875 0.2621
## Call:
## VAR(y = education_var_data, p = 2)
##
##
## Estimation results for equation gov_exp_ts:
## ===========================================
## gov_exp_ts = gov_exp_ts.l1 + primary_school_enrol_ts.l1 + gov_exp_ts.l2 + primary_sch
ool_enrol_ts.l2 + const
##
##                             Estimate Std. Error t value Pr(>|t|)
## gov_exp_ts.l1               0.63243    0.22993   2.750   0.0132 *
## primary_school_enrol_ts.l1 -0.01535    0.05526  -0.278   0.7843
## gov_exp_ts.l2               0.20386    0.23899   0.853   0.4049
## primary_school_enrol_ts.l2  0.02860    0.05935   0.482   0.6356
## const                      -0.64212    1.37007  -0.469   0.6449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.4606 on 18 degrees of freedom
## Multiple R-Squared: 0.7103,  Adjusted R-squared: 0.646
## F-statistic: 11.03 on 4 and 18 DF,  p-value: 0.0001062
##
##
## Estimation results for equation primary_school_enrol_ts:
## ========================================================
## primary_school_enrol_ts = gov_exp_ts.l1 + primary_school_enrol_ts.l1 + gov_exp_ts.l2
+ primary_school_enrol_ts.l2 + const
##
##                             Estimate Std. Error t value Pr(>|t|)
## gov_exp_ts.l1               0.01603    0.93332   0.017    0.986
## primary_school_enrol_ts.l1  1.11598    0.22430   4.975 9.79e-05 ***
## gov_exp_ts.l2               1.00552    0.97009   1.037    0.314
## primary_school_enrol_ts.l2 -0.14135    0.24089  -0.587    0.565
## const                       0.78777    5.56126   0.142    0.889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1.87 on 18 degrees of freedom
## Multiple R-Squared: 0.9449,  Adjusted R-squared: 0.9326
## F-statistic: 77.13 on 4 and 18 DF,  p-value: 4.468e-11
##
```

```
##
##
## Covariance matrix of residuals:
##                        gov_exp_ts primary_school_enrol_ts
## gov_exp_ts                0.21216                -0.07846
## primary_school_enrol_ts  -0.07846                 3.49554
##
## Correlation matrix of residuals:
##                        gov_exp_ts primary_school_enrol_ts
## gov_exp_ts                1.00000                -0.09111
## primary_school_enrol_ts  -0.09111                 1.00000
```


CCF for Government Expenditure and Primary School Enrollment
ACF
Lag
-0.4
-0.2
0.0
0.2
0.4
-10
-5
0
5
10


gov_exp_ts
gov_exp_ts & primary_school_enrol_ts
primary_school_enrol_ts & gov_exp_ts
primary_school_enrol_ts
ACF
Lag
-0.4
0.2
0.8
0
2
4
6
8
10
-10
-8
-6
-4
-2



The provided plots offer diagnostic insights into the relationship between government expenditure (% GDP) and primary school enrollment (%) in a VAR model, also known as a Vector Autoregressive model.

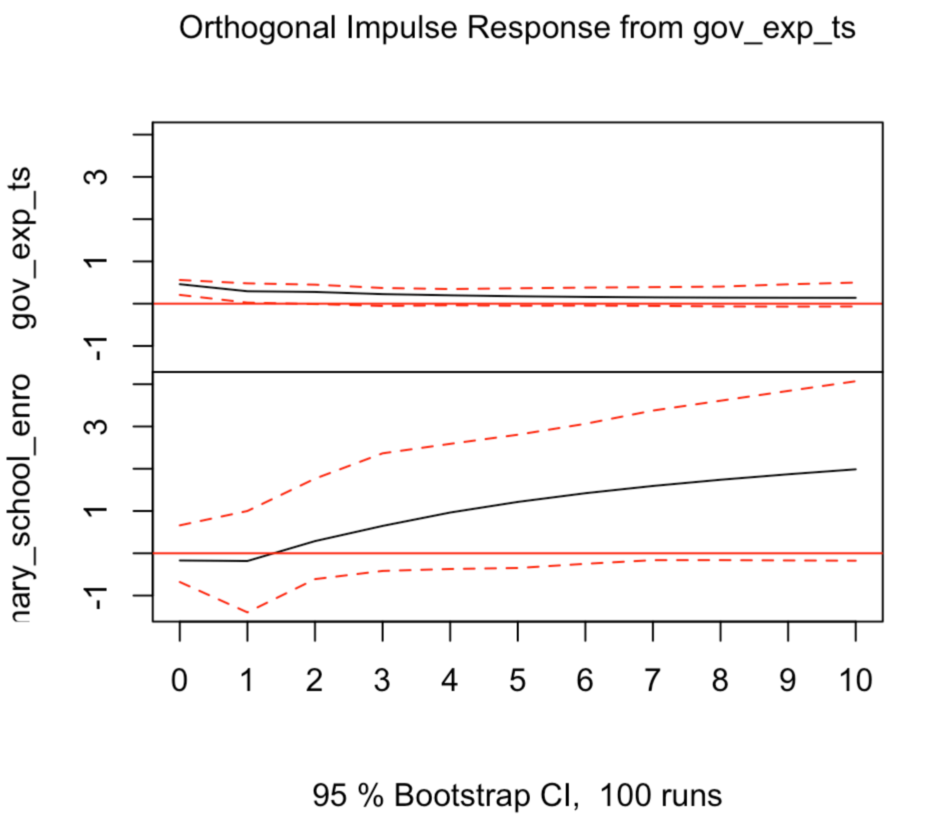
The CCF plot evaluates the relationship between the two variables across different lags. Significant positive correlations are observed at both positive and negative lags, indicating a dynamic relationship where government expenditure and primary school enrollment may influence each other over time. Positive lags suggest that changes in government expenditure might lead to future changes in primary school enrollment, while negative lags imply that enrollment trends might predict shifts in government spending. The sustained positive correlations across lags point to a strong temporal association between the variables, reinforcing the suitability of a VAR model for capturing these interdependencies.

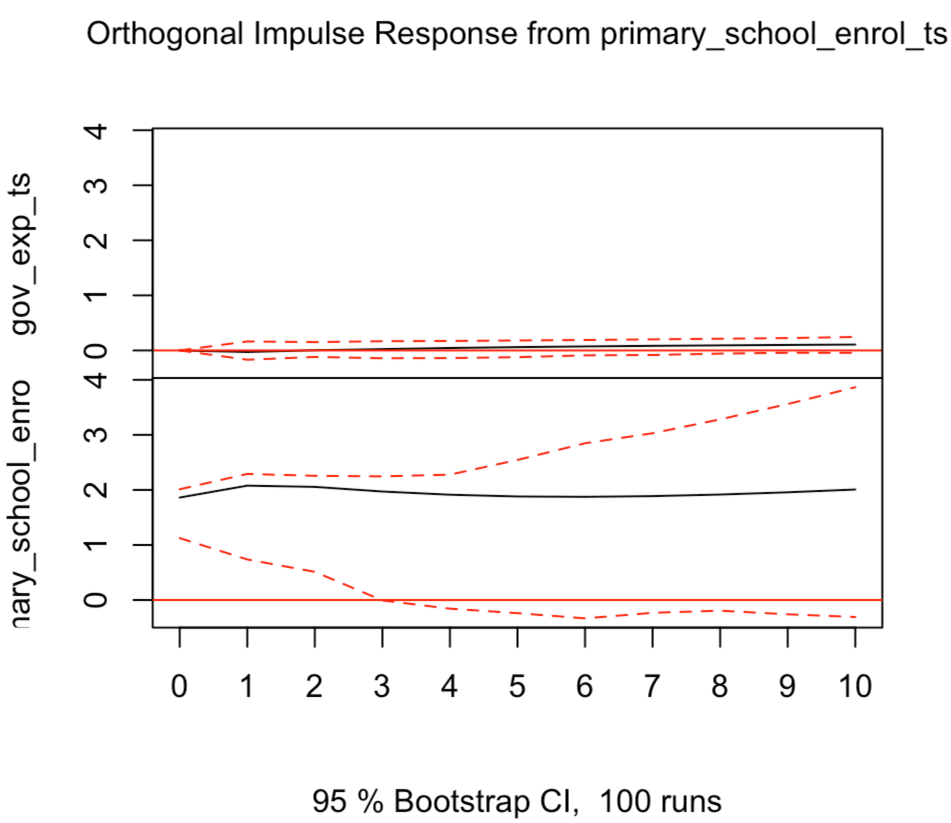
The ACF and PACF plots for each variable assess the temporal dependencies within government expenditure and primary school enrollment individually. For government expenditures, the ACF plot shows a rapid decline, while the PACF has a significant spike at lag 1, suggesting a short-term autoregressive structure that is consistent with the behavior of government expenditure as a relatively stable series with limited long-term memory. For primary school enrollment, the ACF plot for primary school enrollment reveals a slower decay, indicating persistence or long-term dependence in the series. The PACF shows significant spikes at multiple lags, implying that the enrollment process is influenced by both recent and earlier values.

For the cross-correlation and cross PACF plots, significant correlations at various lags suggest bidirectional influences between government expenditure and primary school enrollment. Positive lags, such as lags 1 and 2, indicate that government expenditure potentially drives changes in enrollment, while negative lags, such as lags -1 and -2, suggest that enrollment patterns might also influence expenditure decisions.

Overall, the VAR diagnostics confirm a strong interrelationship between government expenditure and primary school enrollment. The persistence and mutual influences observed in the ACF, PACF, and cross-correlation plots validate the VAR model's ability to capture the bidirectional and lagged interactions. These findings underscore the importance of analyzing these variables jointly, as changes in one can have meaningful predictive power for the other over time. The diagnostic plots also suggest that government expenditure and primary school enrollment are intricately linked through both short-term and long-term dynamics, making the VAR model a suitable framework for forecasting and policy analysis.

Now, let's look at the respective impulse response functions.


Orthogonal Impulse Response from gov_exp_ts
gov_exp_ts
nary_school_enro
3
1
-1
3
1
-1
0 1 2 3 4 5 6 7 8 9 10
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from primary_school_enrol_ts
gov_exp_ts
nary_school_enro
4
3
2
1
0
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9 10
95 % Bootstrap CI, 100 runs

The impulse response plots demonstrate the dynamic interactions between government expenditure (% GDP) and primary school enrollment (%) within a VAR framework. These plots show the effect of a one-unit shock in one variable on the other over a series of time periods.

The first plot illustrates the orthogonal impulse response from government expenditure. In the upper panel, a one-unit shock to government expenditure has a negligible and statistically insignificant effect on itself over time, as evidenced by the flat response within the 95% confidence intervals. In the lower panel, the response of primary school enrollment to a shock in government expenditure is positive and statistically significant over time. The response starts small but steadily increases, indicating that increases in government expenditure on education have a lagged and growing effect on primary school enrollment. This aligns with the expectation that investments in education infrastructure and resources take time to materialize into higher enrollment rates.

The second plot shows the orthogonal impulse response from primary school enrollment. In the upper panel, a one-unit shock to primary school enrollment has a minimal and statistically insignificant impact on government expenditure over time, as the response remains close to zero and within the confidence intervals. This suggests that enrollment changes do not significantly drive government expenditure in the short or medium term. In the lower panel, a shock to primary school enrollment has a slightly positive and increasing effect on itself over time, indicating that higher enrollment rates can sustain and reinforce themselves, potentially through social or policy mechanisms that encourage continued participation in primary education.

The impulse response plots reveal an asymmetry in the relationship between the two variables. While shocks to government expenditure have a meaningful and growing positive impact on primary school enrollment over time, shocks to enrollment do not appear to influence government expenditure significantly. This suggests that government policy decisions and investments in education play a crucial role in driving changes in enrollment, but enrollment patterns do not exert a similar influence on policy adjustments in the short-to-medium term. These results highlight the importance of sustained government expenditure to achieve long-term improvements in primary education outcomes.

We will now perform a Granger-Causality test on the two time-series variables and discuss the results from the test.

```r
# Granger-Causality test (Government Expenditure to Primary School Enrollment)
grangertest(gov_exp_ts ~ primary_school_enrol_ts, order = 2)
```

```
## Granger causality test
##
## Model 1: gov_exp_ts ~ Lags(gov_exp_ts, 1:2) + Lags(primary_school_enrol_ts, 1:2)
## Model 2: gov_exp_ts ~ Lags(gov_exp_ts, 1:2)
##   Res.Df Df      F Pr(>F)
## 1     18
## 2     20 -2 0.2715 0.7653
```

```r
# Granger-Causality test (Primary School Enrollment to Government Expenditure)
grangertest(primary_school_enrol_ts ~ gov_exp_ts, order = 2)
```

```
## Granger causality test
##
## Model 1: primary_school_enrol_ts ~ Lags(primary_school_enrol_ts, 1:2) + Lags(gov_exp_
ts, 1:2)
## Model 2: primary_school_enrol_ts ~ Lags(primary_school_enrol_ts, 1:2)
##   Res.Df Df      F Pr(>F)
## 1     18
## 2     20 -2 1.3106 0.2942
```

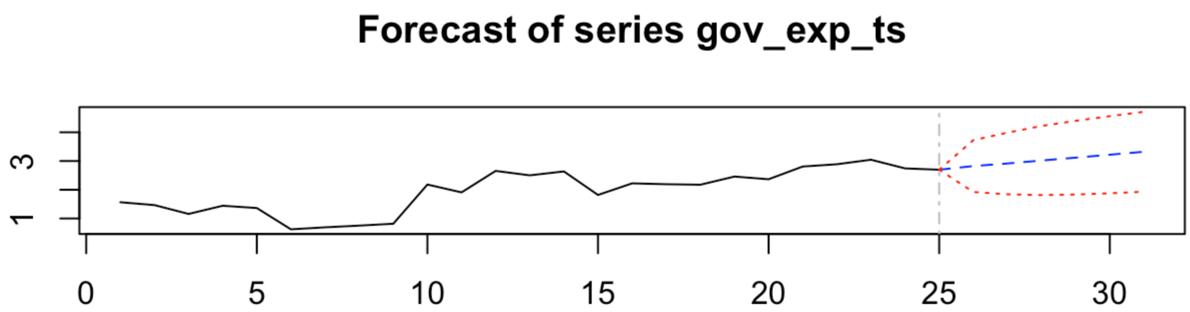
The Granger-causality tests assess whether government expenditure Granger-causes primary school enrollment and vice versa. These tests determine whether past values of one variable can predict the current value of another variable within a Vector Autoregressive framework.

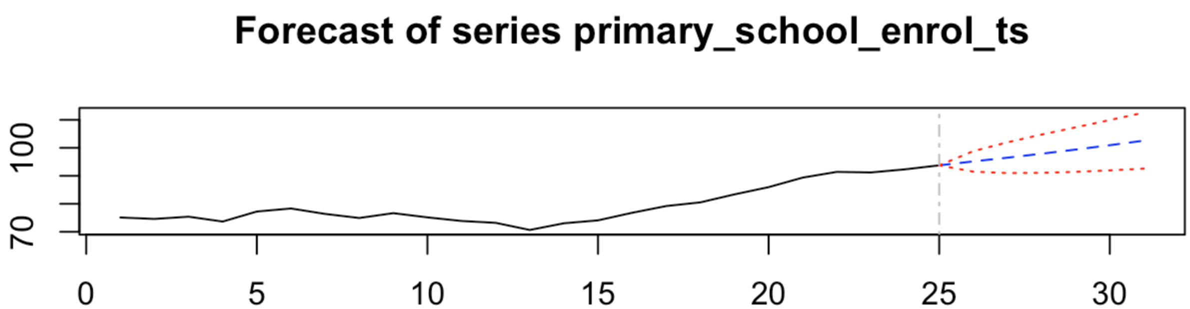
The first Granger-causality test evaluates whether lagged values of primary school enrollment significantly improve the prediction of government expenditure when included alongside lagged values of government expenditure. The test yields an F-statistic of 0.2715 with a p-value of 0.7653, which is not statistically significant. This indicates that primary school enrollment does not Granger-cause government expenditure, suggesting that changes in enrollment do not directly influence future government spending patterns in the short term.

The second Granger-causality test examines whether lagged values of government expenditure improve the prediction of primary school enrollment when included alongside lagged enrollment values. The test results show an F-statistic of 1.3106 with a p-value of 0.2942, which is also not statistically significant. This suggests that government expenditure does not Granger-cause primary school enrollment, meaning past spending patterns do not directly predict changes in enrollment within the given model and time lags.

The Granger-causality tests provide no evidence of a causal relationship between government expenditure and primary school enrollment in either direction within the two-lag framework. While the impulse response analysis indicated a dynamic relationship, the lack of Granger causality suggests that the relationship might be indirect or influenced by factors outside the scope of the VAR model. This result could be due to the limitations of the chosen lag structure, the time frame of the data, or the complexity of the relationship, which might require additional explanatory variables or longer lags to uncover causal patterns.

Finally, let’s utilize the VAR model to forecast the time-series plot for 6 steps ahead.





The forecast plots for government expenditure, in percentage of GDP, and primary school enrollment, in percentage, depict the 6-step ahead predictions generated from the VAR model, including confidence intervals to account for uncertainty in the forecasts.

The first plot shows the forecast for government expenditure. The black line represents the historical data, while the blue dashed line illustrates the point forecasts for the next six steps. The red dotted lines indicate the upper and lower bounds of the 95% confidence intervals. The forecast suggests a slight continuation of the upward trend in government expenditure, consistent with the historical pattern. However, the confidence intervals widen progressively, reflecting increasing uncertainty in the predictions as the forecast horizon extends. This is typical for time series forecasting, especially for government expenditure, which can be influenced by unpredictable policy decisions or economic factors.

The second plot displays the forecast for primary school enrollment. Similar to government expenditure, the historical data is shown in black, and the forecasts are represented by the blue dashed line. The forecast indicates a continued gradual increase in enrollment rates, aligned with the long-term upward trend observed in the data. The red dotted lines show the 95% confidence intervals, which also widen over time, though they remain relatively narrower compared to government expenditure, suggesting slightly greater predictability in enrollment patterns.

Both forecasts demonstrate the ability of the VAR model to capture and extend the observed trends in the data. The forecasts for government expenditure and primary school enrollment align with their respective historical dynamics, suggesting reliability in the model's predictions. However, the widening confidence intervals emphasize the inherent uncertainty in forecasting, particularly over longer horizons. This highlights the importance of considering external factors and potential structural changes that could influence future values of these variables. Overall, the forecasts provide a useful tool for planning and policy-making while acknowledging the uncertainty inherent in time series predictions.

III. Conclusions and Future Work

The analysis of The Gambia's education system, using government expenditure as a percentage of GDP and primary school enrollment rates, reveals critical insights through various visualizations. The time series plots illustrated that both variables showed an upward trend over time, indicating increased government spending on education and improvements in enrollment rates. The STL decomposition highlighted a clear trend in both variables, although seasonality was not present, and the residual components suggested some variability that may require further investigation. The ACF and PACF plots confirmed that the series were stationary after differencing, supporting the use of ARIMA models for individual variable forecasting.

The ARIMA model diagnostics and residual analyses suggested a good model fit for both variables, with residuals exhibiting white noise behavior and approximate normality. However, the forecast plots from the ARIMA models showed increasing uncertainty in predictions over time, particularly for government expenditure. The Prophet model visualizations provided additional context, effectively capturing the long-term trends in both variables while demonstrating the robustness of forecasts within confidence intervals. The impulse response functions from the VAR model demonstrated that government expenditure had a positive and significant impact on primary school enrollment over time, with the effect growing steadily. However, enrollment shocks did not exhibit a strong influence on government expenditure, pointing to the unidirectional nature of this relationship. The Granger causality test confirmed no statistically significant causal relationship between the variables, suggesting that other factors might mediate their connection.

To improve future analyses, incorporating additional variables such as gender-specific enrollment rates, literacy rates, infrastructure availability, or teacher-student ratios would provide a more comprehensive view of the education system. Including socio-economic indicators like poverty levels or rural-urban disparities could uncover deeper insights into the barriers affecting enrollment and education quality. Expanding the data set to include regional or school-level data could also reveal localized challenges and successes. Methodologically, testing alternative models like Bayesian VAR or machine learning-based forecasting could enhance the predictive capabilities and account for potential non-linearities or interactions among variables.

Overall, the visualizations and analyses underscore the importance of sustained and targeted investments in education. To foster long-term improvements, policymakers should focus on not just increasing funding but also addressing quality, equity, and accessibility challenges in The Gambia's education system. By refining the analysis and integrating broader datasets, future research can provide actionable insights to guide education policy and interventions, ultimately ensuring that every child in The Gambia has access to quality education.

IV. References

Broken Chalk, (https://brokenchalk.org/educational-challenges-in-gambia (https://brokenchalk.org/educational-challenges-in-gambia))

Steve Sinnott Foundation, (https://www.stevesinnottfoundation.org.uk/improving-the-quality-and-access-to-education-in-the-gambia (https://www.stevesinnottfoundation.org.uk/improving-the-quality-and-access-to-education-in-the-gambia))

UNICEF, (https://www.unicef.org/gambia/education (https://www.unicef.org/gambia/education))

United Nations, (https://gambia.un.org/en/275628-transforming-lives-through-education-un-gambias-2023-success-story (https://gambia.un.org/en/275628-transforming-lives-through-education-un-gambias-2023-success-story))

UCAR Center for Science Education, (https://scied.ucar.edu/learning-zone/sun-space-weather/sun-and-climate-change (https://scied.ucar.edu/learning-zone/sun-space-weather/sun-and-climate-change))

World Bank, (https://databank.worldbank.org/databases/education (https://databank.worldbank.org/databases/education))

World Bank, (https://documents.worldbank.org/en/publication/documents-reports/documentdetail/284581468032133072/the-gambia-education-country-status-report (https://documents.worldbank.org/en/publication/documents-reports/documentdetail/284581468032133072/the-gambia-education-country-status-report))

V. R Source Code

```
# Load libraries
library(tidyverse)
library(dplyr)
library(tidyr)
library(zoo)

# Uploading the Data
world_data <- read.csv("world-education-data.csv")

gambia_education_data <- world_data %>%
  filter(country == "Gambia, The") %>%
  select("country", "country_code", "year", "gov_exp_pct_gdp", "school_enrol_primary_pc
t")

print(gambia_education_data)

# Load necessary libraries
library(ggplot2)
library(forecast)

# Create time series objects
gov_exp_ts <- ts(gambia_education_data$gov_exp_pct_gdp, start = min(gambia_education_dat
a$year), frequency = 1)
primary_school_enrol_ts <- ts(gambia_education_data$school_enrol_primary_pct, start = mi
n(gambia_education_data$year), frequency = 1)

# Time-series plot
ts.plot(gov_exp_ts, col = "navy", lty = 1:2, ylab = "Values", main = "Time-Series Data f
or Government Expenditure (% GDP)")
ts.plot(primary_school_enrol_ts, col = "salmon", lty = 1:2, ylab = "Values", main = "Tim
e-Series Data for Primary School Enrollment (%)")

# ACF and PACF plots
par(mfrow = c(2, 2))
acf(gov_exp_ts, main = "ACF - Government Expenditure (% GDP)")
pacf(gov_exp_ts, main = "PACF - Government Expenditure (% GDP)")
acf(primary_school_enrol_ts, main = "ACF - Primary School Enrollment (%)")
pacf(primary_school_enrol_ts, main = "PACF - Primary School Enrollment (%)")
par(mfrow = c(1, 1))

# STL decomposition for gov_exp
stl_gov_exp <- stl(ts(gov_exp_ts, frequency = 2), s.window = "periodic")
plot(stl_gov_exp, main = "STL Decomposition of Government Expenditure (% GDP)")

# STL decomposition for school_enrol
stl_primary_school_enrol <- stl(ts(primary_school_enrol_ts, frequency = 2), s.window =
"periodic")
plot(stl_primary_school_enrol, main = "STL Decomposition of PrimarySchool Enrollment
(%)")
```

```
# Fit ARIMA model for gov_exp_pct_gdp
gov_exp_arima <- auto.arima(gov_exp_ts, seasonal = FALSE)
summary(gov_exp_arima)

# Fit ARIMA model for school_enrol_primary_pct
primary_school_enrol_arima <- auto.arima(primary_school_enrol_ts, seasonal = FALSE)
summary(primary_school_enrol_arima)

# Get residuals vs. fitted values and storing it in a data frame (Government Expenditur
e)
gov_exp_residuals_df <- data.frame(Fitted_Values = gov_exp_arima$fitted, Residuals = gov
_exp_arima$residuals)

# Plot of the residuals vs. fitted values (Government Expenditure)
ggplot(gov_exp_residuals_df, aes(x = as.numeric(Fitted_Values), y = as.numeric(Residual
s))) +
  geom_line(color = "navy", alpha = 0.9) +
  geom_point(color = "navy", alpha = 0.9) +
  ggtitle("Residuals vs Fitted Values (Government Expenditure)") +
  xlab("Fitted Values") +
  ylab("Residuals")

# Get residuals vs. fitted values and storing it in a data frame (Primary School Enrollm
ent)
primary_enrol_residuals_df <- data.frame(Fitted_Values = primary_school_enrol_arima$fitt
ed, Residuals = primary_school_enrol_arima$residuals)

# Plot of the residuals vs. fitted values (Primary School Enrollment)
ggplot(primary_enrol_residuals_df, aes(x = as.numeric(Fitted_Values), y = as.numeric(Res
iduals))) +
  geom_line(color = "maroon", alpha = 0.9) +
  geom_point(color = "maroon", alpha = 0.9) +
  ggtitle("Residuals vs Fitted Values (Primary School Enrollment)") +
  xlab("Fitted Values") +
  ylab("Residuals")

par(mfrow = c(2, 2))

# ACF and PACF for residuals of gov_exp_pct_gdp
acf(residuals(gov_exp_arima), main = "ACF of Residuals (Government Expenditure)")
pacf(residuals(gov_exp_arima), main = "PACF of Residuals (Government Expenditure)")

# ACF and PACF for residuals of school_enrol_primary_pct
acf(residuals(primary_school_enrol_arima), main = "ACF of Residuals (Primary School Enro
llment)")
pacf(residuals(primary_school_enrol_arima), main = "PACF of Residuals (Primary School En
rollment)")

# Load strucchange library
library(strucchange)

# CUSUM Test for gov_exp_pct_gdp
```

```
gov_exp_cusum <- efp(gov_exp_ts ~ 1, type = "OLS-CUSUM")
plot(gov_exp_cusum, main = "CUSUM Test (Government Expenditure)")

# CUSUM Test for school_enrol_primary_pct
primary_school_enrol_cusum <- efp(primary_school_enrol_ts ~ 1, type = "OLS-CUSUM")
plot(primary_school_enrol_cusum, main = "CUSUM Test (Primary School Enrollment)")

# Diagnostics for gov_exp_pct_gdp
checkresiduals(gov_exp_arima)

# Diagnostics for school_enrol_primary_pct
checkresiduals(primary_school_enrol_arima)

# Forecast 12 months ahead (Government Expenditure)
gov_exp_forecast_6 <- forecast(gov_exp_arima, h = 6)

# Plot the forecast (Government Expenditure)
autoplot(gov_exp_forecast_6) +
  labs(title = "6-Year Ahead Forecast with ARIMA Model (Government Expenditure)", y = "P
ercentage of GDP")

# Forecast 12 months ahead (Primary School Enrollment)
primary_school_enrol_forecast_6 <- forecast(primary_school_enrol_arima, h = 6)

# Plot the forecast (Primary School Enrollment)
autoplot(primary_school_enrol_forecast_6) +
  labs(title = "6-Year Ahead Forecast with ARIMA Model (Primary School Enrollment)", y =
"Percentage of Relevant Age Group")

# ETS model for gov_exp_pct_gdp
gov_exp_ets <- ets(gov_exp_ts)
summary(gov_exp_ets)

# ETS model for school_enrol_primary_pct
primary_school_enrol_ets <- ets(primary_school_enrol_ts)
summary(primary_school_enrol_ets)

# Compare MAPE between models
accuracy(gov_exp_arima)
accuracy(gov_exp_ets)
accuracy(primary_school_enrol_arima)
accuracy(primary_school_enrol_ets)

# Splitting the data into training and testing sets
train_size <- floor(0.8 * length(gov_exp_ts))
gov_exp_train <- window(gov_exp_ts, end = time(gov_exp_ts)[train_size])
gov_exp_test <- window(gov_exp_ts, start = time(gov_exp_ts)[train_size + 1])
primary_school_enrol_train <- window(primary_school_enrol_ts, end = time(primary_school_
enrol_ts)[train_size])
primary_school_enrol_test <- window(primary_school_enrol_ts, start = time(primary_school
_enrol_ts)[train_size + 1])
```

```
# Load necessary libraries
library(nnfor)
library(prophet)

# ARIMA model for gov_exp_pct_gdp
gov_exp_arima_train <- auto.arima(gov_exp_train)
gov_exp_arima_forecast <- forecast(gov_exp_arima_train, h = length(gov_exp_test))

# ARIMA model for school_enrol_primary_pct
primary_school_enrol_arima_train <- auto.arima(primary_school_enrol_train)
primary_school_enrol_arima_forecast <- forecast(primary_school_enrol_arima_train, h = le
ngth(primary_school_enrol_test))

# ETS model for gov_exp_pct_gdp
gov_exp_ets_train <- ets(gov_exp_train)
gov_exp_ets_forecast <- forecast(gov_exp_ets_train, h = length(gov_exp_test))

# ETS model for school_enrol_primary_pct
primary_school_enrol_ets_train <- ets(primary_school_enrol_train)
primary_school_enrol_ets_forecast <- forecast(primary_school_enrol_ets_train, h = length
(primary_school_enrol_test))

# Holt-Winters model for gov_exp_pct_gdp
gov_exp_hw_train <- HoltWinters(ts(gov_exp_train, frequency = 2))
gov_exp_hw_forecast <- forecast(gov_exp_hw_train, h = length(gov_exp_test))

# Holt-Winters model for school_enrol_primary_pct
primary_school_enrol_hw_train <- HoltWinters(ts(primary_school_enrol_train, frequency =
2))
primary_school_enrol_hw_forecast <- forecast(primary_school_enrol_hw_train, h = length(p
rimary_school_enrol_test))

# Neural Network for gov_exp_pct_gdp
gov_exp_nnetar_train <- nnetar(gov_exp_train)
gov_exp_nnetar_forecast <- forecast(gov_exp_nnetar_train, h = length(gov_exp_test))

# Neural Network for gov_exp_pct_gdp
primary_school_enrol_nnetar_train <- nnetar(primary_school_enrol_train)
primary_school_enrol_nnetar_forecast <- forecast(primary_school_enrol_nnetar_train, h =
length(primary_school_enrol_test))

# Prophet for gov_exp_pct_gdp
gov_exp_train_df <- data.frame(ds = as.Date(paste0(time(gov_exp_train), "-01-01")), y =
as.numeric(gov_exp_train))
gov_exp_prophet_train <- prophet(gov_exp_train_df)
gov_exp_future <- make_future_dataframe(gov_exp_prophet_train, periods = length(gov_exp_
test), freq = "year")
gov_exp_prophet_forecast <- predict(gov_exp_prophet_train, gov_exp_future)

# Prophet for school_enrol_primary_pct
primary_school_enrol_train_df <- data.frame(ds = as.Date(paste0(time(primary_school_enro
l_train), "-01-01")), y = as.numeric(primary_school_enrol_train))
```

```
primary_school_enrol_prophet_train <- prophet(primary_school_enrol_train_df)
primary_school_enrol_future <- make_future_dataframe(primary_school_enrol_prophet_train,
periods = length(primary_school_enrol_test), freq = "year")
primary_school_enrol_prophet_forecast <- predict(primary_school_enrol_prophet_train, pri
mary_school_enrol_future)

# Combine forecasts for gov_exp_pct_gdp
gov_exp_combined_forecast <- (gov_exp_arima_forecast$mean +
                              gov_exp_ets_forecast$mean +
                              gov_exp_prophet_forecast$yhat[(nrow(gov_exp_prophet_foreca
st) - length(gov_exp_test) + 1):nrow(gov_exp_prophet_forecast)]) / 3

# Combine forecasts for school_enrol_primary_pct
primary_school_enrol_combined_forecast <- (primary_school_enrol_arima_forecast$mean +
                                           primary_school_enrol_ets_forecast$mean +
                                           primary_school_enrol_prophet_forecast$yhat[(n
row(primary_school_enrol_prophet_forecast) - length(primary_school_enrol_test) + 1):nrow
(primary_school_enrol_prophet_forecast)]) / 3

# Define a function to compute MAPE, RMSE, and MAE
calculate_errors <- function(actual, forecasted) {
  mape <- mean(abs((actual - forecasted) / actual)) * 100
  rmse <- sqrt(mean((actual - forecasted)^2))
  mae <- mean(abs(actual - forecasted))
  return(data.frame(MAPE = mape, RMSE = rmse, MAE = mae))
}

# Ensure the training data has the same frequency as Holt-Winters forecast
gov_exp_train_hw <- ts(gov_exp_train, start = start(gov_exp_train), frequency = 2)
primary_school_enrol_train_hw <- ts(primary_school_enrol_train, start = start(primary_sc
hool_enrol_train), frequency = 2)

# Ensure the testing data has the same frequency as Holt-Winters forecast
gov_exp_test_hw <- ts(gov_exp_test, start = start(gov_exp_test), frequency = 2)
primary_school_enrol_test_hw <- ts(primary_school_enrol_test, start = start(primary_scho
ol_enrol_test), frequency = 2)

# Prophet Fitted Values
prophet_fitted_gov <- predict(gov_exp_prophet_train)$yhat
prophet_fitted_primary_school <- predict(primary_school_enrol_prophet_train)$yhat

# Combined Forecast fitted values for gov_exp_pct_gdp (Training)
combined_fitted_gov <- (gov_exp_arima_train$fitted +
                        gov_exp_ets_train$fitted +
                        prophet_fitted_gov) / 3

# Combined Forecast fitted values for school_enrol_primary_pct (Training)
combined_fitted_primary_school <- (primary_school_enrol_arima_train$fitted +
                                   primary_school_enrol_ets_train$fitted +
                                   prophet_fitted_primary_school) / 3

# Impute mean for NNETAR fitted values to remove NAs
```

```
gov_exp_nnetar_train$fitted[1] <- mean(gov_exp_nnetar_train$fitted, na.rm = TRUE)
primary_school_enrol_nnetar_train$fitted[1] <- mean(primary_school_enrol_nnetar_train$fi
tted, na.rm = TRUE)

# Training set errors for gov_exp_pct_gdp
arima_errors_gov_train <- calculate_errors(gov_exp_train, gov_exp_arima_train$fitted)
ets_errors_gov_train <- calculate_errors(gov_exp_train, gov_exp_ets_train$fitted)
hw_errors_gov_train <- calculate_errors(as.numeric(gov_exp_train_hw), as.numeric(gov_exp
_hw_train$fitted))
nnetar_errors_gov_train <- calculate_errors(gov_exp_train, gov_exp_nnetar_train$fitted)
prophet_errors_gov_train <- calculate_errors(gov_exp_train, prophet_fitted_gov)
combined_errors_gov_train <- calculate_errors(gov_exp_train, combined_fitted_gov)

# Training set errors for school_enrol_primary_pct
arima_errors_primary_school_enrol_train <- calculate_errors(primary_school_enrol_train,
primary_school_enrol_arima_train$fitted)
ets_errors_primary_school_enrol_train <- calculate_errors(primary_school_enrol_train, pr
imary_school_enrol_ets_train$fitted)
hw_errors_primary_school_enrol_train <- calculate_errors(as.numeric(primary_school_enrol
_train_hw), as.numeric(primary_school_enrol_hw_train$fitted))
nnetar_errors_primary_school_enrol_train <- calculate_errors(primary_school_enrol_train,
primary_school_enrol_nnetar_train$fitted)
prophet_errors_primary_school_enrol_train <- calculate_errors(primary_school_enrol_trai
n, prophet_fitted_primary_school)
combined_errors_primary_school_enrol_train <- calculate_errors(primary_school_enrol_trai
n, combined_fitted_primary_school)

# ARIMA Model Training Errors (Government Expenditure)
print(arima_errors_gov_train)

# ETS Model Training Errors (Government Expenditure)
print(ets_errors_gov_train)

# Holt-Winters Model Training Errors (Government Expenditure)
print(hw_errors_gov_train)

# NNETAR Model Training Errors (Government Expenditure)
print(nnetar_errors_gov_train)

# Prophet Model Training Errors (Government Expenditure)
print(prophet_errors_gov_train)

# Combined Forecast Model Training Errors (Government Expenditure)
print(combined_errors_gov_train)

# ARIMA Model Training Errors (Primary School Enrollment)
print(arima_errors_primary_school_enrol_train)

# ETS Model Training Errors (Primary School Enrollment)
print(ets_errors_primary_school_enrol_train)

# Holt-Winters Model Training Errors (Primary School Enrollment)
```

```
print(hw_errors_primary_school_enrol_train)

# NNETAR Model Training Errors (Primary School Enrollment)
print(nnetar_errors_primary_school_enrol_train)

# Prophet Model Training Errors (Primary School Enrollment)
print(prophet_errors_primary_school_enrol_train)

# Combined Forecast Model Training Errors (Primary School Enrollment)
print(combined_errors_primary_school_enrol_train)

# Testing set errors for gov_exp_pct_gdp
arima_errors_gov_test <- calculate_errors(gov_exp_test, gov_exp_arima_forecast$mean)
ets_errors_gov_test <- calculate_errors(gov_exp_test, gov_exp_ets_forecast$mean)
hw_errors_gov_test <- calculate_errors(as.numeric(gov_exp_test_hw), as.numeric(gov_exp_h
w_forecast$mean))
nnetar_errors_gov_test <- calculate_errors(gov_exp_test, gov_exp_nnetar_forecast$mean)
prophet_errors_gov_test <- calculate_errors(gov_exp_test, gov_exp_prophet_forecast$yhat
[(nrow(gov_exp_prophet_forecast) - length(gov_exp_test) + 1):nrow(gov_exp_prophet_foreca
st)])
combined_errors_gov_test <- calculate_errors(gov_exp_test, gov_exp_combined_forecast)

# Testing set errors for school_enrol_primary_pct
arima_errors_primary_school_enrol_test <- calculate_errors(primary_school_enrol_test, pr
imary_school_enrol_arima_forecast$mean)
ets_errors_primary_school_enrol_test <- calculate_errors(primary_school_enrol_test, prim
ary_school_enrol_ets_forecast$mean)
hw_errors_primary_school_enrol_test <- calculate_errors(as.numeric(primary_school_enrol_
test_hw), as.numeric(primary_school_enrol_hw_forecast$mean))
nnetar_errors_primary_school_enrol_test <- calculate_errors(primary_school_enrol_test, p
rimary_school_enrol_nnetar_forecast$mean)
prophet_errors_primary_school_enrol_test <- calculate_errors(primary_school_enrol_test,
primary_school_enrol_prophet_forecast$yhat[(nrow(primary_school_enrol_prophet_forecast)
-                                                                                    leng
th(primary_school_enrol_test)+1):nrow(primary_school_enrol_prophet_forecast)])
combined_errors_primary_school_enrol_test <- calculate_errors(primary_school_enrol_test,
primary_school_enrol_combined_forecast)

# ARIMA Model Testing Errors (Government Expenditure)
print(arima_errors_gov_test)

# ETS Model Testing Errors (Government Expenditure)
print(ets_errors_gov_test)

# Holt-Winters Model Testing Errors (Government Expenditure)
print(hw_errors_gov_test)

# NNETAR Model Testing Errors (Government Expenditure)
print(nnetar_errors_gov_test)

# Prophet Model Testing Errors (Government Expenditure)
print(prophet_errors_gov_test)
```

```
# Combined Forecast Model Testing Errors (Government Expenditure)
print(combined_errors_gov_test)

# ARIMA Model Testing Errors (Primary School Enrollment)
print(arima_errors_primary_school_enrol_test)

# ETS Model Testing Errors (Primary School Enrollment)
print(ets_errors_primary_school_enrol_test)

# Holt-Winters Model Testing Errors (Primary School Enrollment)
print(hw_errors_primary_school_enrol_test)

# NNETAR Model Testing Errors (Primary School Enrollment)
print(nnetar_errors_primary_school_enrol_test)

# Prophet Model Testing Errors (Primary School Enrollment)
print(prophet_errors_primary_school_enrol_test)

# Combined Forecast Model Testing Errors (Primary School Enrollment)
print(combined_errors_primary_school_enrol_test)

# Redefine the prophet models for clearer understanding
gov_exp_df <- data.frame(ds = as.Date(paste0(gambia_education_data$year, "-01-01")), y =
gambia_education_data$gov_exp_pct_gdp)
gov_exp_prophet <- prophet(gov_exp_df)
primary_school_enrol_df <- data.frame(ds = as.Date(paste0(gambia_education_data$year, "-
01-01")), y = gambia_education_data$school_enrol_primary_pct)
primary_school_enrol_prophet <- prophet(primary_school_enrol_df)

# Plot gov_exp_pct_gdp using the Prophet model
plot(gov_exp_prophet, gov_exp_prophet_forecast) +
  ggtitle("Prophet Model Plot for Gov Expenditure (% GDP)") +
  xlab("Year") +
  ylab("Government Expenditure (% GDP)")

# Plot school_enrol_primary_pct using the Prophet model
plot(primary_school_enrol_prophet, primary_school_enrol_prophet_forecast) +
  ggtitle("Prophet Model Plot for Primary School Enrollment (%)") +
  xlab("Year") +
  ylab("Primary School Enrollment (%)")

# Load var library
library(vars)

# Create a two-variable time series matrix
education_var_data <- cbind(gov_exp_ts, primary_school_enrol_ts)

# Fit the VAR model
education_var_model <- VAR(education_var_data, p = 2)

# Summary of the VAR model
```

```
summary(education_var_model)

# Relevant Plots

# CCF
ccf(gov_exp_ts, primary_school_enrol_ts, main = "CCF for Government Expenditure and Prim
ary School Enrollment")

# ACF and PACF Plots
acf(education_var_data)
pacf(education_var_data)

# Impulse Response Functions for VAR model
plot(irf(education_var_model))

# Granger-Causality test (Government Expenditure to Primary School Enrollment)
grangertest(gov_exp_ts ~ primary_school_enrol_ts, order = 2)

# Granger-Causality test (Primary School Enrollment to Government Expenditure)
grangertest(primary_school_enrol_ts ~ gov_exp_ts, order = 2)

# 6-Step Ahead Forecast for VAR Model
education_var_forecasts <- predict(object = education_var_model, n.ahead = 6)
plot(education_var_forecasts)
```