Enhancing Government Financial Planning-Time Series Forecasting of Personal Service Expenditures through Decomposition Analysis

Le' Sean Roberts (NYC Office of Management and Budget College Aide)

Summary

This development presents a comprehensive approach to time series forecasting applied to Financial Plan Statements, focusing specifically on Personal Service Expenditures within government agencies. Time series forecasting is of paramount importance in governmental financial planning, aiding decision-makers in allocating resources efficiently and effectively. By leveraging historical expenditure data, advanced statistical techniques, and forecasting models, this study aims to provide accurate predictions of future Personal Service Expenditures, enabling government entities to optimize budgetary allocations and enhance fiscal sustainability.Quarto enables you to weave together content and executable code into a finished document.

Introduction

Governmental financial planning plays a critical role in ensuring the effective allocation of resources to support public services and initiatives. Among the various components of financial planning, Personal Service Expenditures represent a significant portion of government budgets, encompassing expenditures related to employee compensation, benefits, and related personnel costs within government agencies. Forecasting Personal Service Expenditures is essential for governmental entities to anticipate future financial obligations, strategically plan budget allocations, and maintain fiscal responsibility.

Time series forecasting techniques offer valuable insights into the future trends and patterns of financial data, enabling decision-makers to make informed and proactive decisions. In the context of government financial planning, time series forecasting based on Financial Plan Statements provides a robust framework for predicting future expenditures, mitigating financial risks, and ensuring the sustainability of public finances.

This paper aims to explore the application of time series forecasting methodologies to Financial Plan Statements, focusing specifically on forecasting Personal Service Expenditures within government agencies. By leveraging historical expenditure data, statistical analysis, and advanced forecasting models such as ARIMA (AutoRegressive Integrated Moving Average) and exponential smoothing methods, this study seeks to develop accurate and reliable forecasts of future Personal Service Expenditures.

The outcomes of this research will contribute to enhancing the efficiency and effectiveness of governmental financial planning processes, enabling policymakers and budget managers to make data-driven decisions, optimize resource allocation strategies, and ensure the financial sustainability of government operations. Through the application of advanced time series forecasting techniques, this study aims to empower government entities to navigate complex fiscal environments and achieve their budgetary objectives effectively.

Data Cleaning & Wrangling

```r
library(readr)
Financial_Plan_Statements_Personal_Service_Expenditures_20240228 <-
  read_csv("C:/Users/verlene/Downloads/Financial_Plan_Statements_-_Personal_Service_Expend
```

```r
library(tidyverse)
glimpse(Financial_Plan_Statements_Personal_Service_Expenditures_20240228)
```

```
Rows: 1,404
Columns: 7
$ `PUBLICATION DATE`     <dbl> 20200928, 20200928, 20200928, 20200928, 2020092~
$ `FISCAL YEAR`          <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021,~
$ MONTH                  <chr> "July", "July", "July", "July", "July", "July",~
$ `CURRENT/YEAR-TO-DATE` <chr> "Current", "Current", "Current", "Current", "Cu~
$ AGENCY                 <chr> "Police", "Fire", "Correction", "Sanitation", "~
$ ACTUAL                 <dbl> 228, 85, 47, 63, 26, 41, 8, 26, 9, 42, 27, 20, ~
$ PLAN                   <dbl> 230, 83, 50, 66, 26, 43, 8, 26, 9, 34, 25, 22, ~
```

```r
Personal_service_expenditures <-
  Financial_Plan_Statements_Personal_Service_Expenditures_20240228 |>
  select(-`PUBLICATION DATE`)
# Showing first 6 rows.
head(Personal_service_expenditures)
```

```
# A tibble: 6 x 6
  `FISCAL YEAR` MONTH `CURRENT/YEAR-TO-DATE` AGENCY                 ACTUAL  PLAN
```

```
          <dbl> <chr> <chr>                  <chr>                   <dbl> <dbl>
1          2021 July  Current                Police                    228   230
2          2021 July  Current                Fire                       85    83
3          2021 July  Current                Correction                 47    50
4          2021 July  Current                Sanitation                 63    66
5          2021 July  Current                Admin. for Children's~     26    26
6          2021 July  Current                Social Services            41    43
```

Current versus Year-to-Date in Expenditure Reporting

"Current" and "year-to-date" are two different perspectives used in expenditure reporting, each providing valuable insights into an organization's financial status. Here's how they differ:

1. **Current Expenditures**:

    - "Current expenditures" typically refer to expenditures incurred within a specific period, often the current month or quarter. These expenditures represent the financial transactions that have occurred within the designated time frame.
    - Current expenditures provide a snapshot of the organization's financial activity within a short-term period, allowing stakeholders to assess recent spending patterns and make immediate decisions based on current financial data.

2. **Year-to-Date Expenditures**:

    - "Year-to-date expenditures" (YTD) encompass all expenditures incurred from the beginning of the fiscal year up to the current date. YTD expenditures accumulate over time, providing a cumulative view of spending since the fiscal year's commencement.
    - Year-to-date expenditures offer a broader perspective on financial performance, allowing stakeholders to track spending trends, monitor budget adherence, and compare current spending levels to those of previous periods or budget allocations.

In expenditure reporting, both current and year-to-date perspectives are essential for comprehensive financial analysis and decision-making:

- **Current Expenditures** help stakeholders understand recent spending patterns, identify emerging trends or anomalies, and make short-term adjustments to budget allocations or resource allocation decisions.
- **Year-to-Date Expenditures** provide context by showing cumulative spending over a longer period, enabling stakeholders to evaluate overall budget performance, assess spending against long-term goals or targets, and project future spending trends based on historical patterns.

Comparing current expenditures to year-to-date figures allows stakeholders to assess spending performance relative to budgetary constraints and make adjustments as necessary to achieve financial objectives.

Focusing on "year-to-date" spending for now. As well, as a test case choosing the NYC Police Department as the agency of interest.

```r
str(Personal_service_expenditures)
```

```
tibble [1,404 x 6] (S3: tbl_df/tbl/data.frame)
 $ FISCAL YEAR          : num [1:1404] 2021 2021 2021 2021 2021 ...
 $ MONTH                : chr [1:1404] "July" "July" "July" "July" ...
 $ CURRENT/YEAR-TO-DATE: chr [1:1404] "Current" "Current" "Current" "Current" ...
 $ AGENCY               : chr [1:1404] "Police" "Fire" "Correction" "Sanitation" ...
 $ ACTUAL               : num [1:1404] 228 85 47 63 26 41 8 26 9 42 ...
 $ PLAN                 : num [1:1404] 230 83 50 66 26 43 8 26 9 34 ...
```

```r
police_YTD_expenditure <- Personal_service_expenditures |>
  filter(`CURRENT/YEAR-TO-DATE` == "Year-to-Date", AGENCY == "Police")
#Showing first 6 rows
head(police_YTD_expenditure)
```

```
# A tibble: 6 x 6
  `FISCAL YEAR` MONTH     `CURRENT/YEAR-TO-DATE` AGENCY ACTUAL  PLAN
          <dbl> <chr>     <chr>                  <chr>   <dbl> <dbl>
1          2021 July      Year-to-Date           Police    228   230
2          2021 August    Year-to-Date           Police    784   761
3          2021 September Year-to-Date           Police   1154  1116
4          2021 October   Year-to-Date           Police   1516  1475
5          2021 November  Year-to-Date           Police   1880  1831
6          2021 December  Year-to-Date           Police   2283  2233
```

Adding a new variable to "MARGIN" to identify expenditure being above or falling short of budget.

```r
police_YTD_expenditure_new_view <- police_YTD_expenditure |>
  mutate(MARGIN = ACTUAL - PLAN)
# Showing first 6 rows.
head(police_YTD_expenditure_new_view)
```

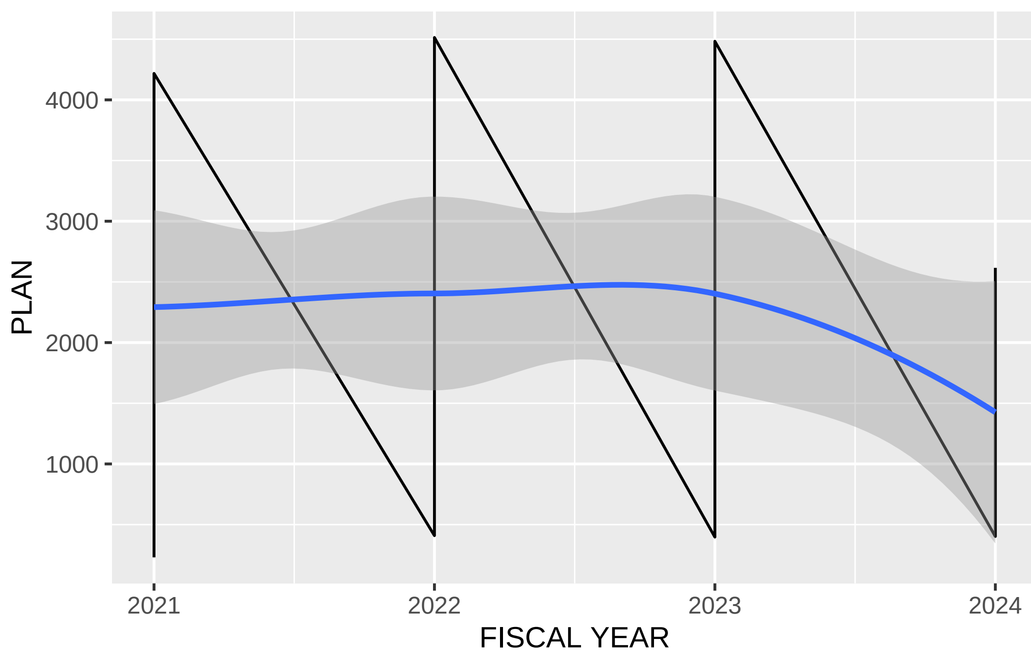
```
# A tibble: 6 x 7
  `FISCAL YEAR` MONTH     `CURRENT/YEAR-TO-DATE` AGENCY ACTUAL  PLAN MARGIN
          <dbl> <chr>     <chr>                  <chr>   <dbl> <dbl>  <dbl>
1          2021 July      Year-to-Date           Police    228   230     -2
2          2021 August    Year-to-Date           Police    784   761     23
3          2021 September Year-to-Date           Police   1154  1116     38
4          2021 October   Year-to-Date           Police   1516  1475     41
5          2021 November  Year-to-Date           Police   1880  1831     49
6          2021 December  Year-to-Date           Police   2283  2233     50
```

Time Series Analysis

Police Plan Expenditure time series:

```
police_plan_ts <- ggplot(police_YTD_expenditure_new_view,
                         aes(x = `FISCAL YEAR`, y = PLAN)) +
       geom_line() + geom_smooth()
police_plan_ts
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



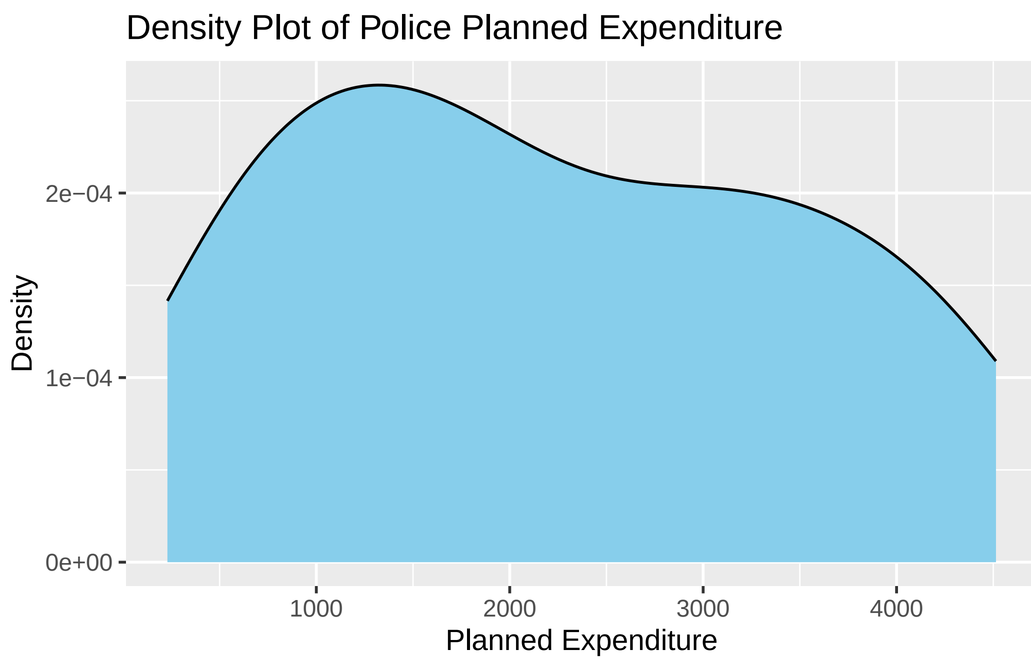
Some descriptive statistics:

```
police_plan_tseries <- police_YTD_expenditure_new_view$PLAN
summary(police_plan_tseries)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    230    1144    1993    2222    3246    4514 
```

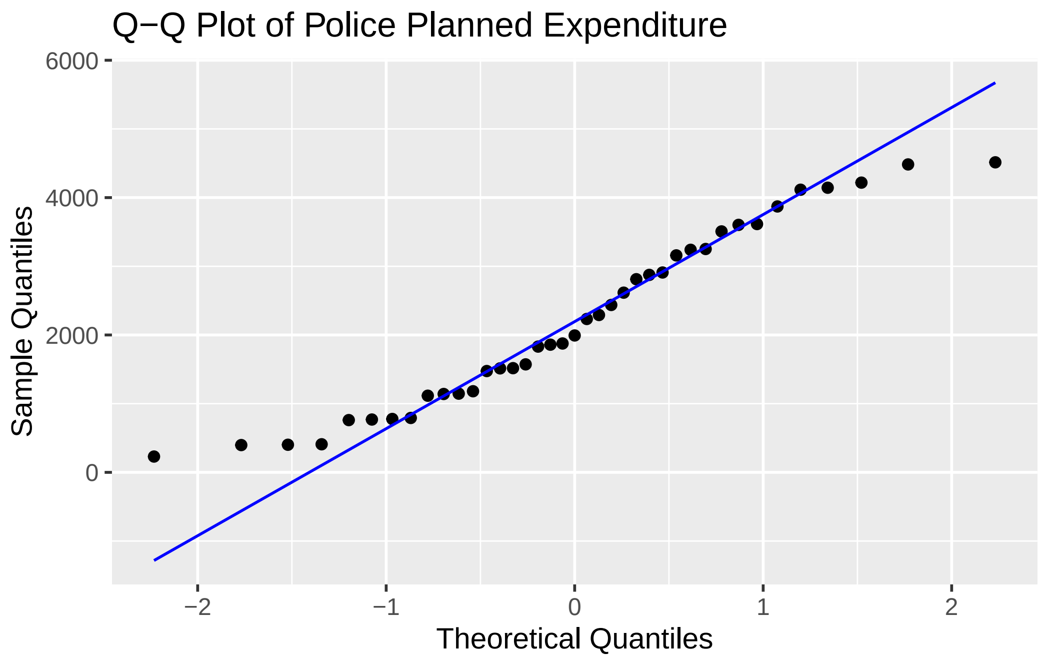
Density plot:

```
ggplot(police_YTD_expenditure_new_view, aes(x = PLAN)) +
  geom_density(fill = "skyblue", color = "black") +
  labs(x ="Planned Expenditure",
       y = "Density",
       title = "Density Plot of Police Planned Expenditure")
```




```
ggplot(police_YTD_expenditure_new_view, aes(sample = PLAN)) +
stat_qq() +
stat_qq_line(color = "blue") +
labs(x = "Theoretical Quantiles",
y = "Sample Quantiles",
```

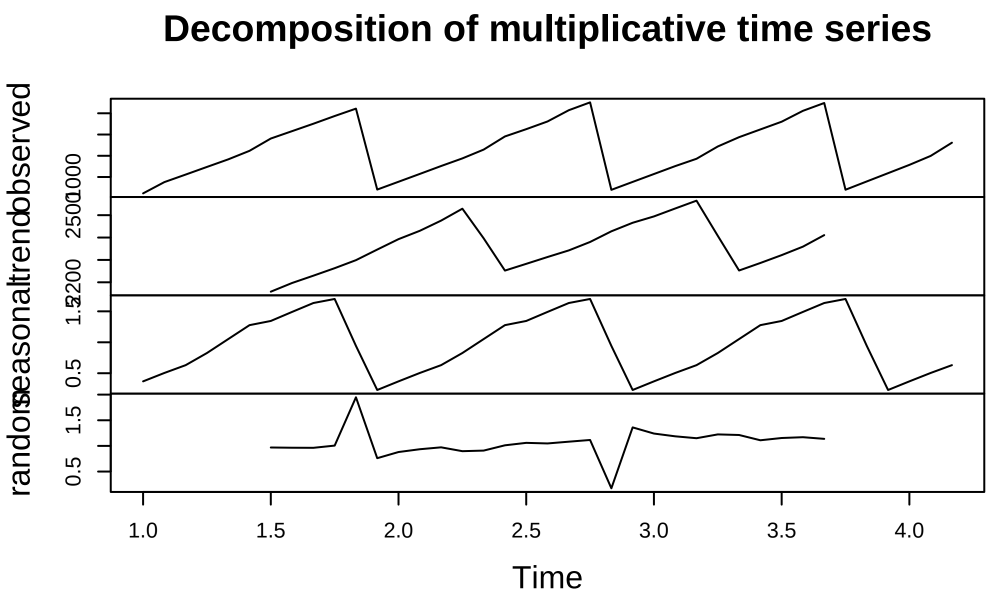
```
title = "Q-Q Plot of Police Planned Expenditure")
```



For the Q-Q plot the theoretical distribution by default is normal. Lack of conformity to the q-q line exhibits lack of normality in the data.

Decomposition of time series data is a statistical technique used to break down the observed series into its underlying components, typically including trend, seasonality, and residual components. This decomposition allows analysts to better understand the patterns and structure present in the data, facilitating forecasting, trend analysis, and anomaly detection.

```
# Recognizing 12 months in a year.
pp_data<-ts(police_plan_tseries, frequency = 12)
decompose_pp_data <- decompose(pp_data, "multiplicative")
plot(decompose_pp_data)
```



Decomposition of a multiplicative time series involves separating the original time series into three components: trend, seasonal, and residual (or error) components. These components provide insights into the underlying patterns and fluctuations present in the data. Here's how you can interpret the displays resulting from the decomposition:

1. **Original Time Series**:
    - The original time series plot shows the raw data over time. This display provides a visual representation of the overall pattern, including any trends, seasonal fluctuations, and irregularities present in the data.
2. **Trend Component**:
    - The trend component represents the long-term movement or directionality of the time series data. It captures the overall upward or downward trend over time, excluding seasonal and irregular fluctuations.
    - Interpretation: A rising trend indicates increasing values over time, while a declining trend suggests decreasing values. A flat trend suggests relatively stable values over the observation period.
3. **Seasonal Component**:

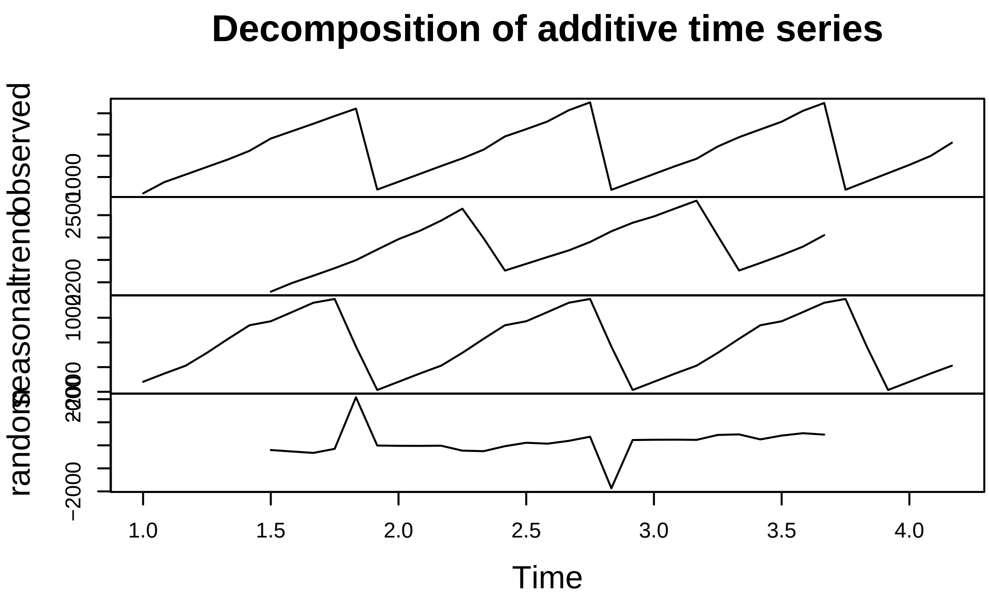
- The seasonal component captures regular patterns or fluctuations that occur within a specific period, such as daily, weekly, or monthly cycles. It represents the recurring patterns that occur at fixed intervals.
- Interpretation: Seasonal patterns can be observed as repetitive fluctuations around the trend line. Peaks and troughs in the seasonal component indicate periods of higher and lower values, respectively, recurring at regular intervals.

4. **Randomness**:

- Randomness refers to the absence of any discernible pattern or trend in a series of observations. In the context of residuals, randomness implies that the residuals exhibit no systematic behavior and are independent and identically distributed (i.i.d.) with a mean of zero and constant variance.
- Interpretation: If the residuals display randomness, it suggests that the decomposition model has effectively captured all systematic patterns in the data, leaving behind only random fluctuations that are inherent in the data.

Overall, interpreting the displays of decomposition helps to understand the underlying structure of the time series data, identify any systematic patterns or trends, and assess the effectiveness of the decomposition model in capturing these patterns. It provides valuable insights for forecasting, anomaly detection, and decision-making based on the temporal behavior of the data. From the display there isn't much to conclude concerning expenditure.

```r
# Recognizing 12 months in a year.
pp_data<-ts(police_plan_tseries, frequency = 12)
decompose_pp_data <- decompose(pp_data, "additive")
plot(decompose_pp_data)
```



Decomposition of an additive time series follows a similar principle to that of a multiplicative time series, but with the components added together rather than multiplied. In an additive decomposition, the original time series is separated into three components: trend, seasonal, and residual (or error). Here's how you can interpret the displays resulting from the decomposition of an additive time series:

1. **Original Time Series**:

   - As with multiplicative decomposition, the original time series plot shows the raw data over time. It provides a visual representation of the overall pattern, including any trends, seasonal fluctuations, and irregularities present in the data.

2. **Trend Component**:

   - The trend component in an additive decomposition represents the long-term movement or directionality of the time series data, similar to multiplicative decomposition. However, in additive decomposition, the trend is added to the seasonal and residual components.
   - Interpretation: A rising trend indicates increasing values over time, while a declining trend suggests decreasing values. A flat trend suggests relatively stable values over the observation period.

3. **Seasonal Component**:

- The seasonal component in an additive decomposition captures regular patterns or fluctuations that occur within a specific period, similar to multiplicative decomposition. However, in additive decomposition, the seasonal component is added to the trend and residual components.
- Interpretation: Seasonal patterns can be observed as repetitive fluctuations around the trend line, added to the overall level of the data. Peaks and troughs in the seasonal component indicate periods of higher and lower values, respectively, recurring at regular intervals.

4. **Randomness**:

- Randomness refers to the absence of any discernible pattern or trend in a series of observations. In the context of residuals, randomness implies that the residuals exhibit no systematic behavior and are independent and identically distributed (i.i.d.) with a mean of zero and constant variance.
- Interpretation: If the residuals display randomness, it suggests that the decomposition model has effectively captured all systematic patterns in the data, leaving behind only random fluctuations that are inherent in the data.

Overall, interpreting the displays of decomposition for an additive time series helps to understand the underlying structure of the data, identify any systematic patterns or trends, and assess the effectiveness of the decomposition model in capturing these patterns. It provides valuable insights for forecasting, anomaly detection, and decision-making based on the temporal behavior of the data. From the display there isn't much to conclude concerning expenditure.

Now to develop a forecast for the next 6 months.

```
library(forecast)
```

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

```
# Convert your data frame to a time series object.
# Assuming monthly data
police_plan_ts_data <- ts(police_plan_tseries, frequency = 12)

# Perform time series forecasting using auto.arima from the forecast package
fit <- auto.arima(police_plan_ts_data)

# Generate forecasts
```

```r
# Forecast for the next 6 months.
police_plan_forecast_result <- forecast(fit, h = 6)

# View the forecast result
print(police_plan_forecast_result)
```

```
      Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
Apr 4           2437  931.752 3942.248  134.9218 4739.078
May 4           2875 1369.752 4380.248  572.9218 5177.078
Jun 4           3240 1734.752 4745.248  937.9218 5542.078
Jul 4           3603 2097.752 5108.248 1300.9218 5905.078
Aug 4           4115 2609.752 5620.248 1812.9218 6417.078
Sep 4           4483 2977.752 5988.248 2180.9218 6785.078
```

```r
# Calculate standard errors, MAE, RMSE, RSE, AIC, and BIC
mae <- forecast::accuracy(police_plan_forecast_result)[1, "MAE"]
rmse <- forecast::accuracy(police_plan_forecast_result)[1, "RMSE"]
rse <- sqrt(sum((residuals(fit))^2) / length(police_plan_ts_data))
aic <- AIC(fit)
bic <- BIC(fit)

# View the forecast result
print(police_plan_forecast_result)
```


```
      Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
Apr 4           2437  931.752 3942.248  134.9218 4739.078
May 4           2875 1369.752 4380.248  572.9218 5177.078
Jun 4           3240 1734.752 4745.248  937.9218 5542.078
Jul 4           3603 2097.752 5108.248 1300.9218 5905.078
Aug 4           4115 2609.752 5620.248 1812.9218 6417.078
Sep 4           4483 2977.752 5988.248 2180.9218 6785.078
```


```r
# Print standard errors, MAE, RMSE, RSE, AIC, and BIC
print(paste("MAE:", mae))
```

```
[1] "MAE: 501.990333005853"
```

```r
print(paste("RMSE:", rmse))
```

```
[1] "RMSE: 977.285743167783"
```

```r
print(paste("RSE:", rse))
```

```
[1] "RSE: 977.285743167783"
```

```r
print(paste("AIC:", aic))
```

```
[1] "AIC: 460.329265388792"
```

```r
print(paste("BIC:", bic))
```

```
[1] "BIC: 461.625102254796"
```

The summary results of the forecast provide useful information about the forecasted values and the accuracy of the forecast. Let's interpret the summary results:
1. Point forecasts (Forecast): These are the estimated values for each forecast horizon (2021, 2022, 2023). These values represent the most likely outcome according to the forecasting model.

2. Prediction intervals (Lo 80, Hi 80, Lo 95, Hi 95): These intervals provide a range within which the actual future values are expected to fall with a certain level of confidence. The "Lo 80" and "Hi 80" columns represent the 80% prediction interval, while the "Lo 95" and "Hi 95" columns represent the 95% prediction interval. For example, the "Lo 80" and "Hi 80" columns provide a range within which 80% of the actual future values are expected to fall.

3. Mean absolute error (MAE): This is a measure of the average absolute difference between the forecasted values and the actual values. A lower MAE indicates better accuracy of the forecasts.

4. Root mean squared error (RMSE): This is a measure of the average magnitude of the forecast errors. It penalizes larger errors more heavily compared to MAE. Like MAE, a lower RMSE indicates better accuracy of the forecasts.

5. Residual standard error (RSE): This is an estimate of the standard deviation of the forecast errors. It provides a measure of the variability of the forecast errors around the mean.

6. AIC (Akaike Information Criterion): This is a measure of the goodness of fit of the forecasting model. Lower AIC values indicate better model fit, considering the trade-off between goodness of fit and model complexity.

7. The Bayesian Information Criterion (BIC): Also known as the Schwarz criterion, is a criterion for model selection among a finite set of models. It balances the trade-off between model fit and model complexity, penalizing models that are too complex. In the context of time series forecasting, lower BIC values indicate better models relative to the set of candidate models being considered. Therefore, when comparing different forecasting models, the one with the lowest BIC is generally preferred, as it achieves a good balance between model fit and complexity.

Reminder

In similar fashion to the application of expenditure plan data, time series analysis and forecasting can also be done for actual expenditure, and the for the margin case as well.

Conclusion

Time series decomposition serves as a fundamental tool in analyzing and understanding the underlying patterns within financial plan statements, particularly in forecasting Personal Service Expenditures for government agencies. This study has elucidated the importance of time series forecasting in governmental financial planning, emphasizing the critical role of accurate predictions in resource allocation, budgetary decision-making, and fiscal sustainability.

Through the application of decomposition techniques, such as trend, seasonality, and residual analysis, analysts can gain valuable insights into the temporal structure of financial data. By decomposing time series data into its constituent components, policymakers and budget managers can identify long-term trends, recurring seasonal patterns, and unexplained variations, enabling them to make informed decisions and develop effective budgetary strategies.

The introduction provided an overview of the significance of time series forecasting in governmental financial planning, highlighting the relevance of Personal Service Expenditures within government agencies and the need for accurate predictions to optimize budget allocations. Subsequently, the overview of time series decomposition elucidated the essential components involved in decomposing time series data, including trend, seasonality, and residuals, and the importance of understanding these components in financial analysis.

In conclusion, the integration of time series decomposition techniques into governmental financial planning processes offers significant benefits, including enhanced forecasting accuracy, improved resource allocation strategies, and better management of fiscal risks. By leveraging decomposition methodologies and advanced forecasting models, government entities can navigate complex financial landscapes, anticipate future expenditure trends, and ensure the long-term sustainability of public finances. This study underscores the value of time series analysis and decomposition in informing evidence-based decision-making and promoting financial resilience in government operations.

References

Mayor's Office of Management and Budget (OMB). (2024, February 13). *Financial plan statements - personal service expenditures: NYC open data*. Financial Plan Statements - Personal Service Expenditures | NYC Open Data. https://data.cityofnewyork.us/City-Government/Financial-Plan-Statements-Personal-Service-Expendi/nhtw-fjuj/about_data