

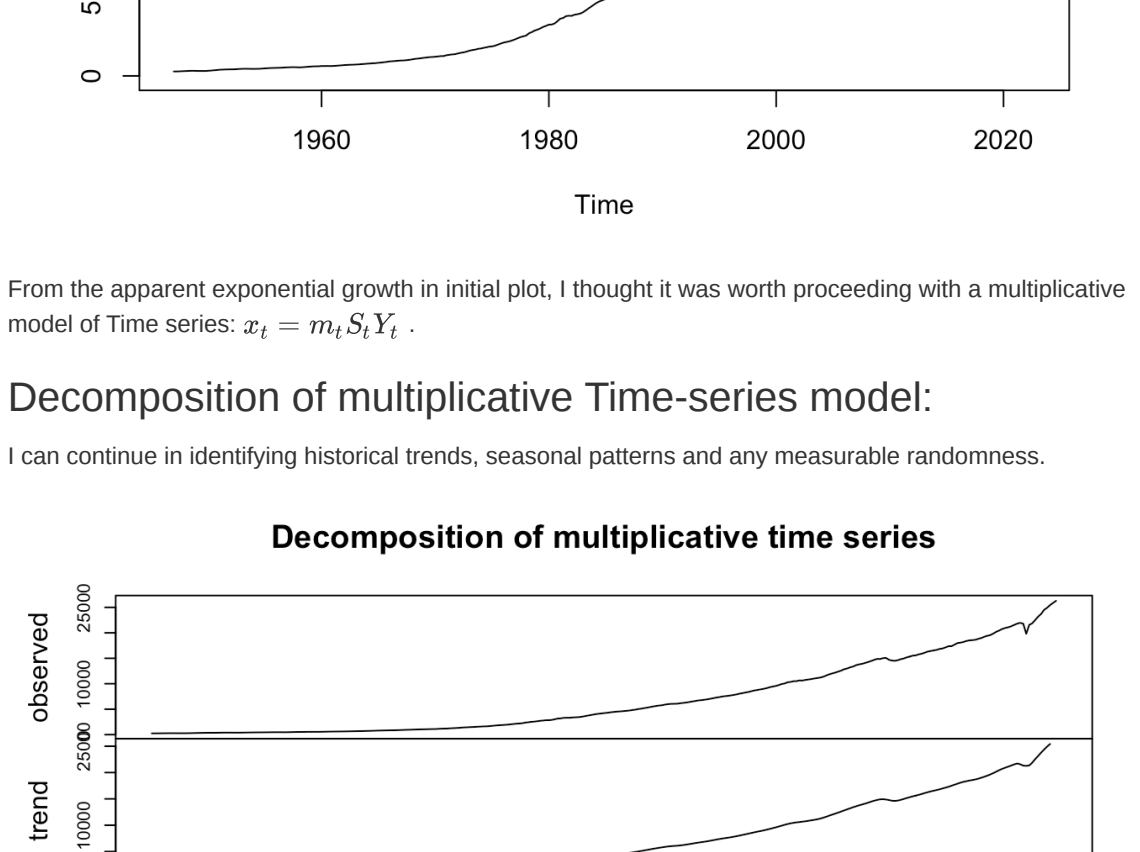
# U.S. Gross National Product Time-Series analysis

## Preface:

The motivation of this analysis is to understand the future behaviour of the time series data 'Quarterly U.S. GNP' (Gross National Product) from 1947-2023. To this end, I will use Meta's Prophet forecasting system to generate a prediction of up to this year's (2025) values. To verify the robustness of the forecast, I will compare the prophet prediction with a non-parametric forecasting method. To further understand the data, I will compare parametric methods to investigate to what extent the dependence of GNP growth is linear over time.

## Understanding the data:

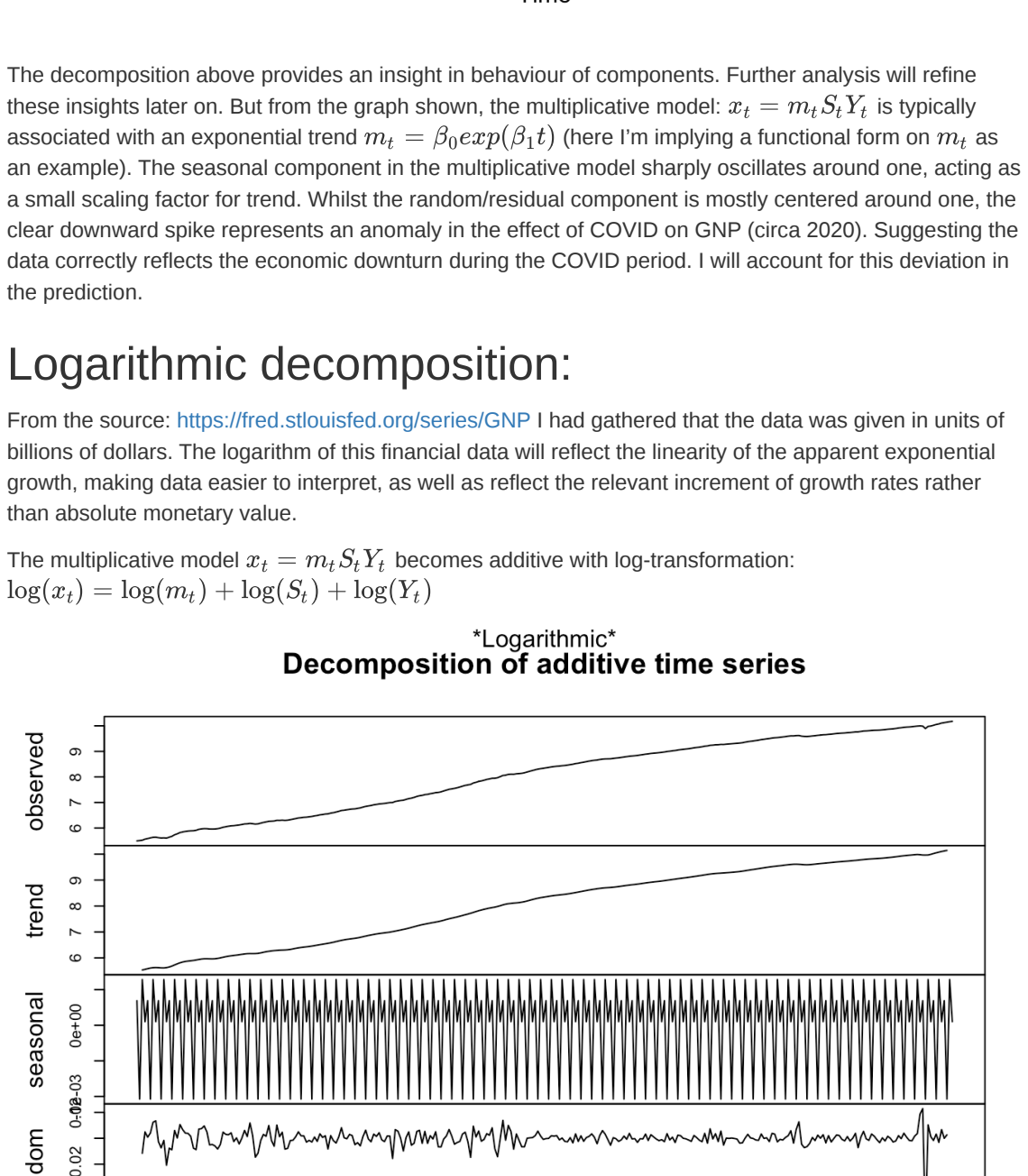
In R from library 'astsr', I have imported the quarterly US GNP data spanning from 1947-2023.



From the apparent exponential growth in initial plot, I thought it was worth proceeding with a multiplicative model of Time series:  $x_t = m_t S_t Y_t$ .

## Decomposition of multiplicative Time-series model:

I can continue in identifying historical trends, seasonal patterns and any measurable randomness.



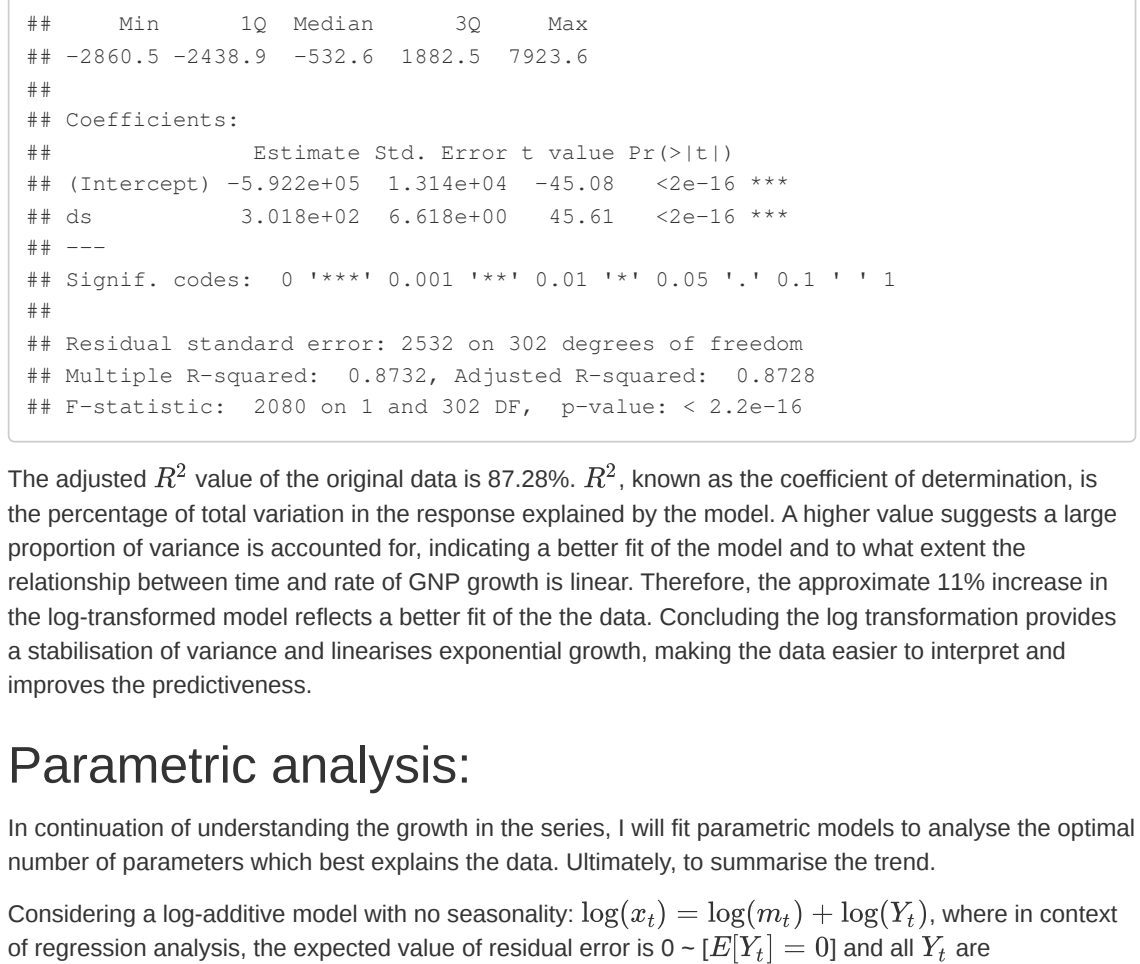
The decomposition above provides an insight in behaviour of components. Further analysis will refine these insights later on. But from the graph shown, the multiplicative model:  $x_t = m_t S_t Y_t$  is typically associated with an exponential trend  $m_t = \beta_0 \exp(\beta_1 t)$  (here I'm implying a functional form on  $m_t$  as an example). The seasonal component in the multiplicative model sharply oscillates around one, acting as a small scaling factor for trend. Whilst the random/residual component is mostly centered around one, the clear downward spike represents an anomaly in the effect of COVID on GNP (circa 2020). Suggesting the data correctly reflects the economic downturn during the COVID period. I will account for this deviation in the prediction.

## Logarithmic decomposition:

From the source: <https://fred.stlouisfed.org/series/GNP> I had gathered that the data was given in units of billions of dollars. The logarithm of this financial data will reflect the linearity of the apparent exponential growth, making data easier to interpret, as well as reflect the relevant increment of growth rates rather than absolute monetary value.

The multiplicative model  $x_t = m_t S_t Y_t$  becomes additive with log-transformation:

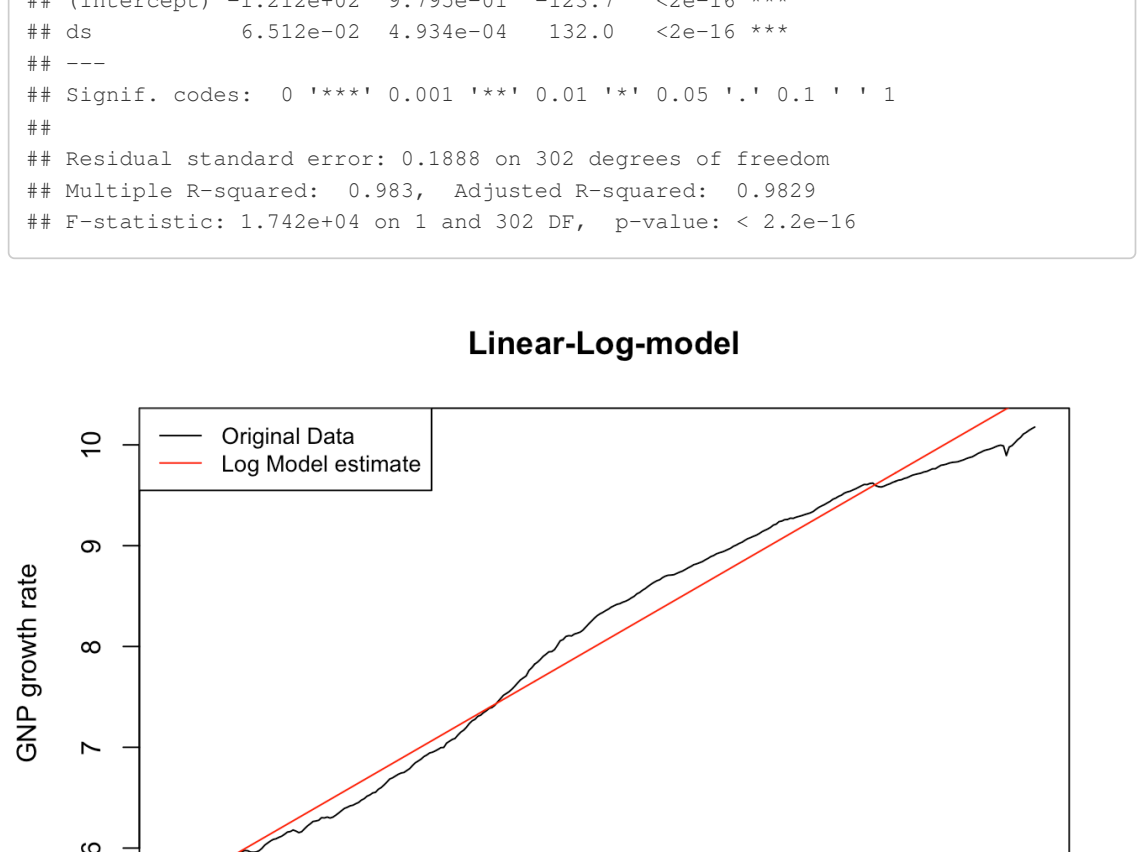
$$\log(x_t) = \log(m_t) + \log(S_t) + \log(Y_t)$$



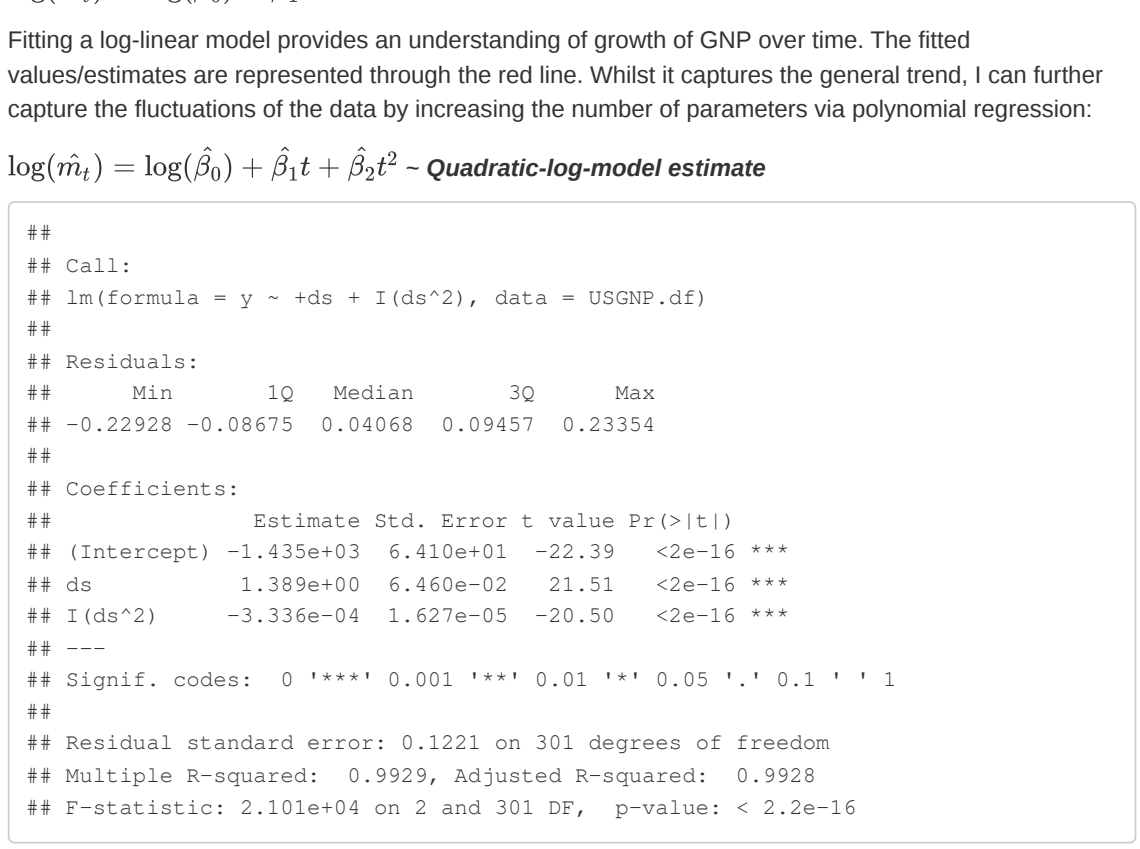
The log-transformation reflects a more linear form in trend, with seasonal and residual components now centered around zero. The continuation of using a log-additive model will be useful moving forward in parametric analysis.

## More on Logarithmic transformation:

Building on the findings of the logarithmic decomposition, I wanted definitive proof that a log transformation of the data would highlight the linearity of GNP growth over time. I would continue in defining a log-data-frame and log-linear model:



From the summary of the log-transformed model, we can see an adjusted  $R^2$  value of 98.29%. I can compare this to the original model's value by transforming the response back to reflect the exponential growth:



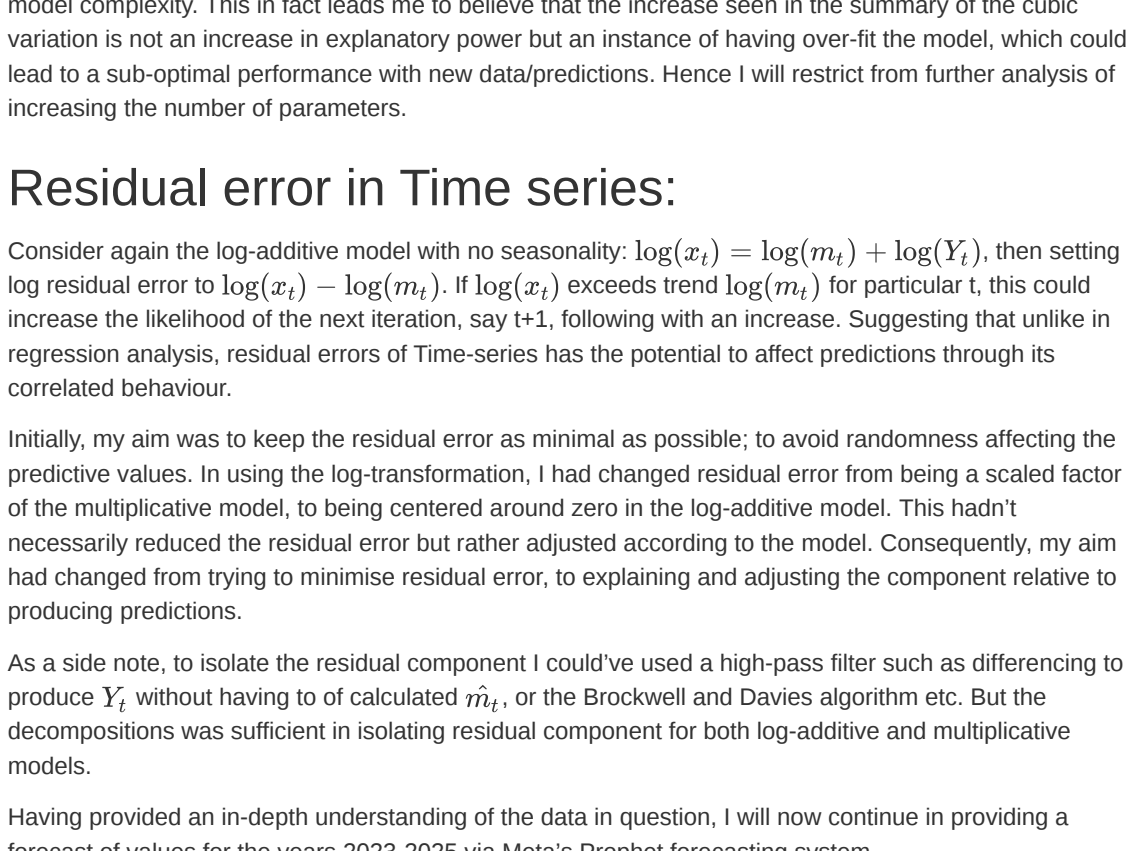
The adjusted  $R^2$  value of the original data is 87.28%.  $R^2$ , known as the coefficient of determination, is the percentage of total variation in the response explained by the model. A higher value suggests a large proportion of variance is accounted for, indicating a better fit of the model and to what extent the relationship between time and rate of GNP growth is linear. Therefore, the approximate 11% increase in the log-transformed model reflects a better fit of the data. Concluding the log transformation provides a stabilisation of variance and linearises exponential growth, making the data easier to interpret and improves the predictiveness.

## Parametric analysis:

In continuation of understanding the growth in the series, I will fit parametric models to analyse the optimal number of parameters which best explains the data. Ultimately, to summarise the trend.

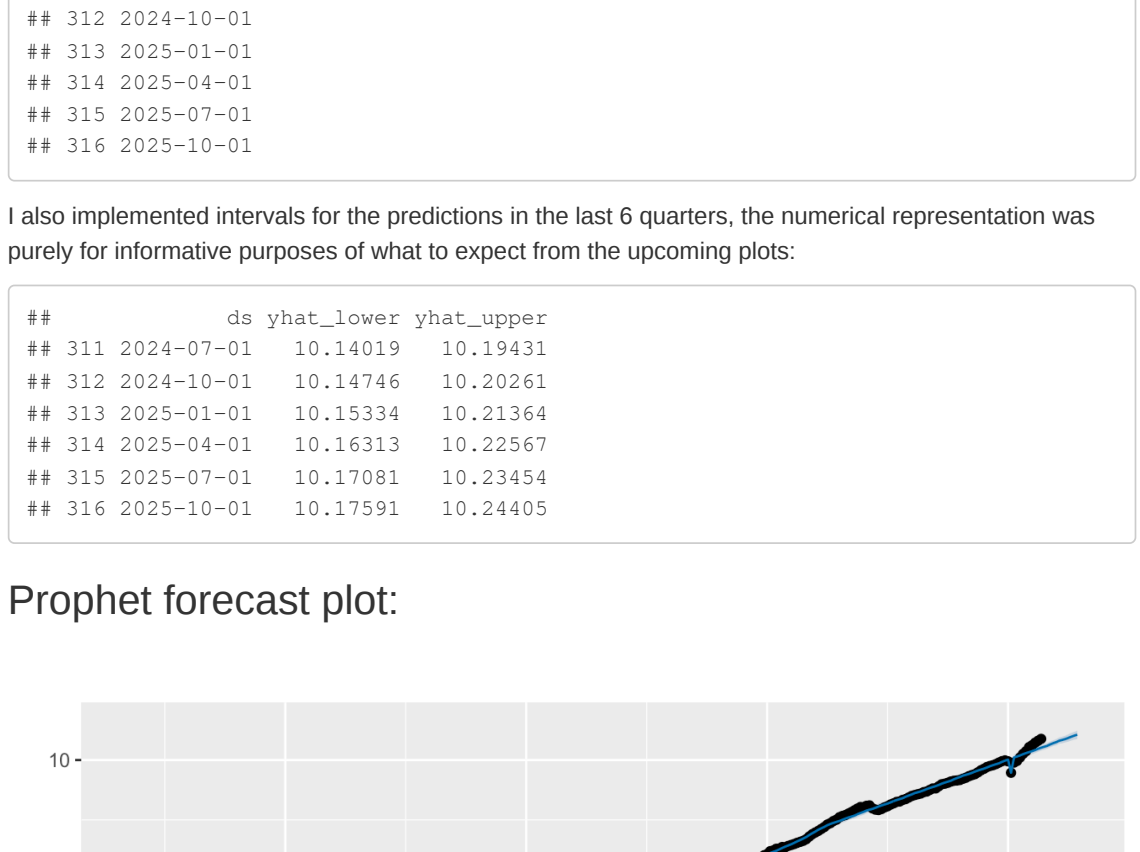
Considering a log-additive model with no seasonality:  $\log(x_t) = \log(m_t) + \log(Y_t)$ , where in context of regression analysis, the expected value of residual error is  $0 - [E[Y_t] = 0]$  and all  $Y_t$  are uncorrelated. Note that while the expected value equates to zero it doesn't necessarily rule out residual error, the model is still stochastic i.e. randomness inclusive.

In line with regression analysis, I assume  $\log(x_t) = \log(m_t)$  where  $m_t$  has functional form:  $m_t = \beta_0 \exp(\beta_1 t) \rightarrow \log(m_t) = \log(\beta_0) + \beta_1 t$ . Producing estimates of parameters  $\hat{\beta}_0$  and  $\hat{\beta}_1$  will be pivotal in fitting estimate of trend  $m_t$ .

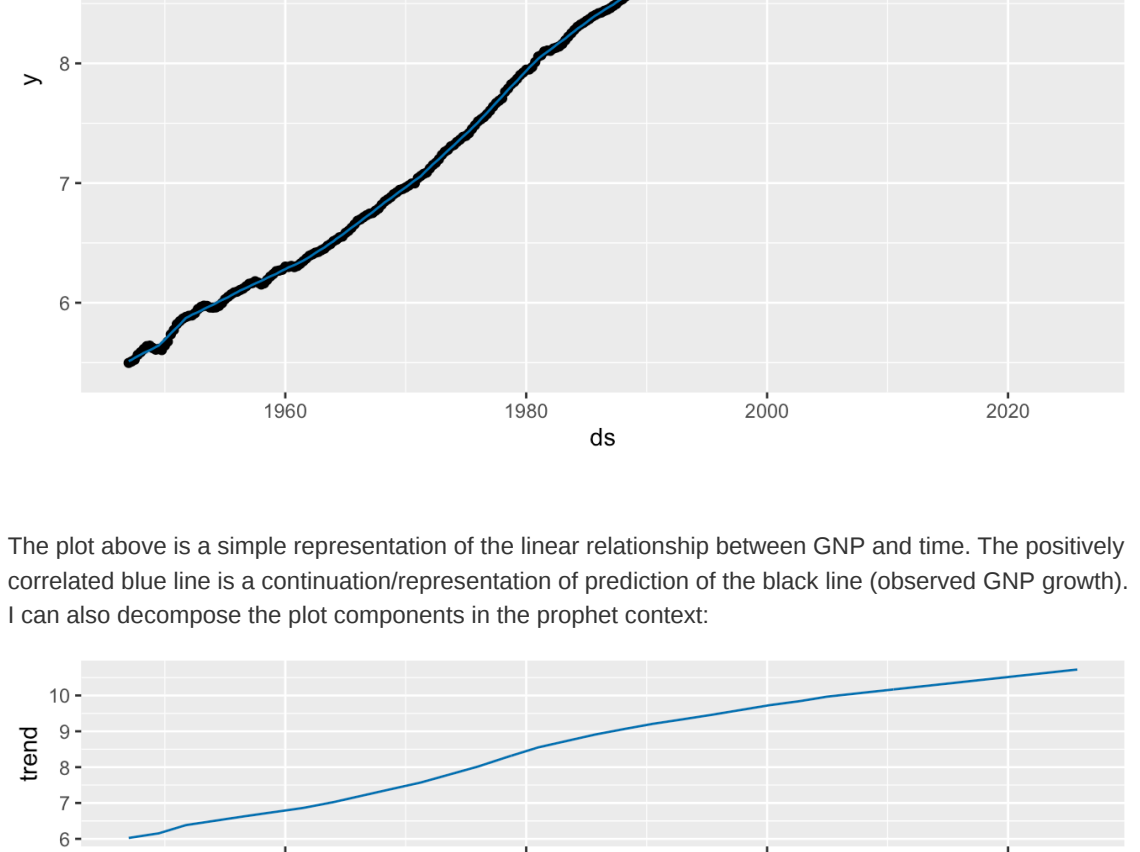


$$\log(\hat{m}_t) = \log(\hat{\beta}_0) + \hat{\beta}_1 t$$

Fitting a log-linear model provides an understanding of growth of GNP over time. The fitted values/estimates are represented through the red line. Whilst it captures the general trend, I can further capture the fluctuations of the data by increasing the number of parameters via polynomial regression:  $\log(\hat{m}_t) = \log(\hat{\beta}_0) + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3 - \text{Cubic-log-model estimate}$



$$\log(\hat{m}_t) = \log(\hat{\beta}_0) + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3 - \text{Cubic-log-model estimate}$$



Both quadratic and cubic log models increasingly capture the fluctuations of growth in the series graphically. However, from the summaries of models, there is a 0.01% decrease in adjusted  $R^2$  from linear to quadratic, suggesting the quadratic term lacks the explanatory power to justify the increase in model complexity. This in fact leads me to believe that the increase seen in the summary of the cubic variation is not an increase in explanatory power but an instance of having over-fit the model, which could lead to a sub-optimal performance with new data/predictions. Hence I will restrict from further analysis of increasing the number of parameters.

## Residual error in Time series:

Consider again the log-additive model with no seasonality:  $\log(x_t) = \log(m_t) + \log(Y_t)$ , then setting log residual error to  $\log(x_t) - \log(m_t)$ . If  $\log(x_t)$  exceeds trend  $\log(m_t)$  for particular  $t$ , this could increase the likelihood of the next iteration, say  $t+1$ , following with an increase. Suggesting that unlike in regression analysis, residual errors of Time-series has the potential to affect predictions through its correlated behaviour.

Initially, my aim was to keep the residual error as minimal as possible; to avoid randomness affecting the predictive values. In using the log-transformation, I had changed residual error from being a scaled factor of the multiplicative model, to being centered around zero in the log-additive model. This hadn't necessarily reduced the residual error but rather adjusted according to the model. Consequently, my aim had changed from trying to minimise residual error, to explaining and adjusting the component relative to producing predictions.

As a side note, to isolate the residual component I could've used a high-pass filter such as differencing to produce  $Y_t$  without having to calculate  $m_t$ , or the Brockwell and Davies algorithm etc. But the decomposition was sufficient in isolating residual component for both log-additive and multiplicative models.

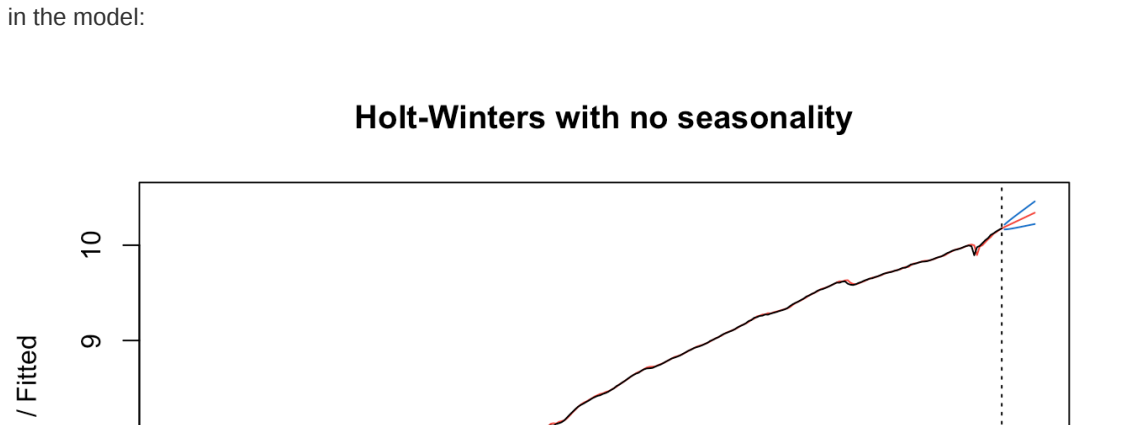
Having provided an in-depth understanding of the data in question, I will now continue in providing a forecast of values for the years 2023-2025 via Meta's Prophet forecasting system.

## Forecast via Meta's Prophet package:

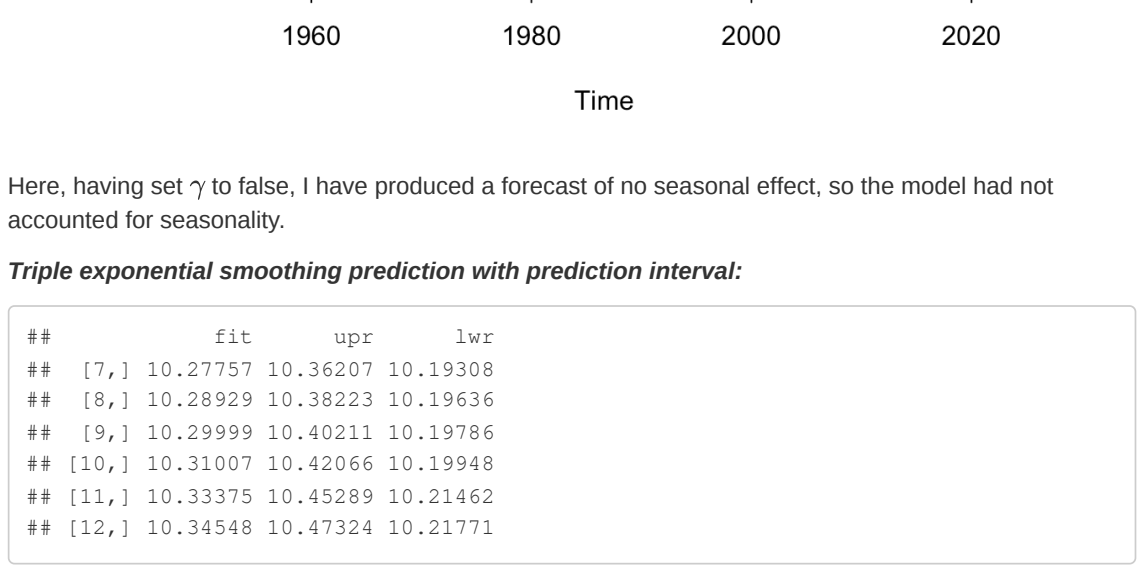
Importing the 'prophet' package from the library had provided the necessary functions to forecast future values of the log-transformed data. I also implemented the zoo package for confirming a quarterly time index.

After loading the necessary packages I again defined the data-frame with the log-transformation of GNP data. Then, fitted the prophet function to model the data-frame. The last step of initialisation was to create the future data-frame for 12 quarterly periods, Q1 of 2023 to Q4 of 2025, for which values I will forecast. I've also accounted for the impact of COVID so that the prophet function will take quarter 2 of 2020 as an anomaly.

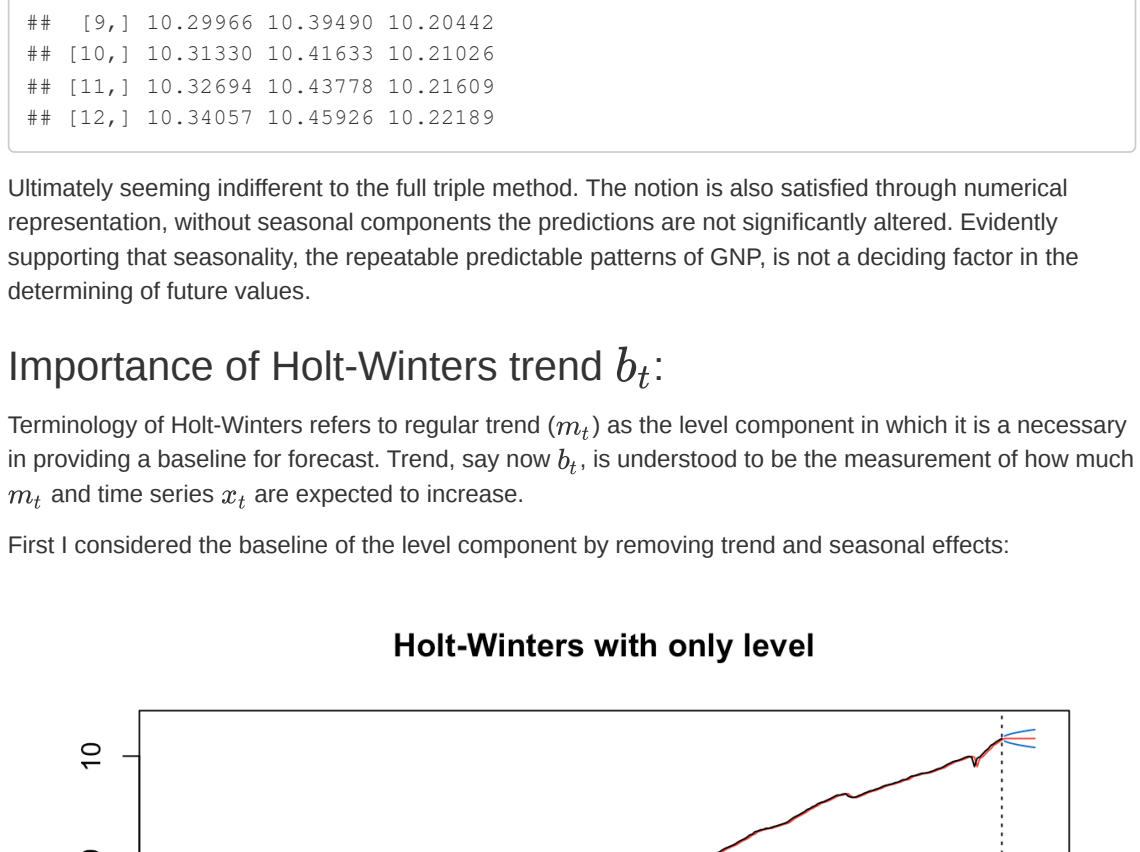
I can confirm the last 6 quarters of future dates:



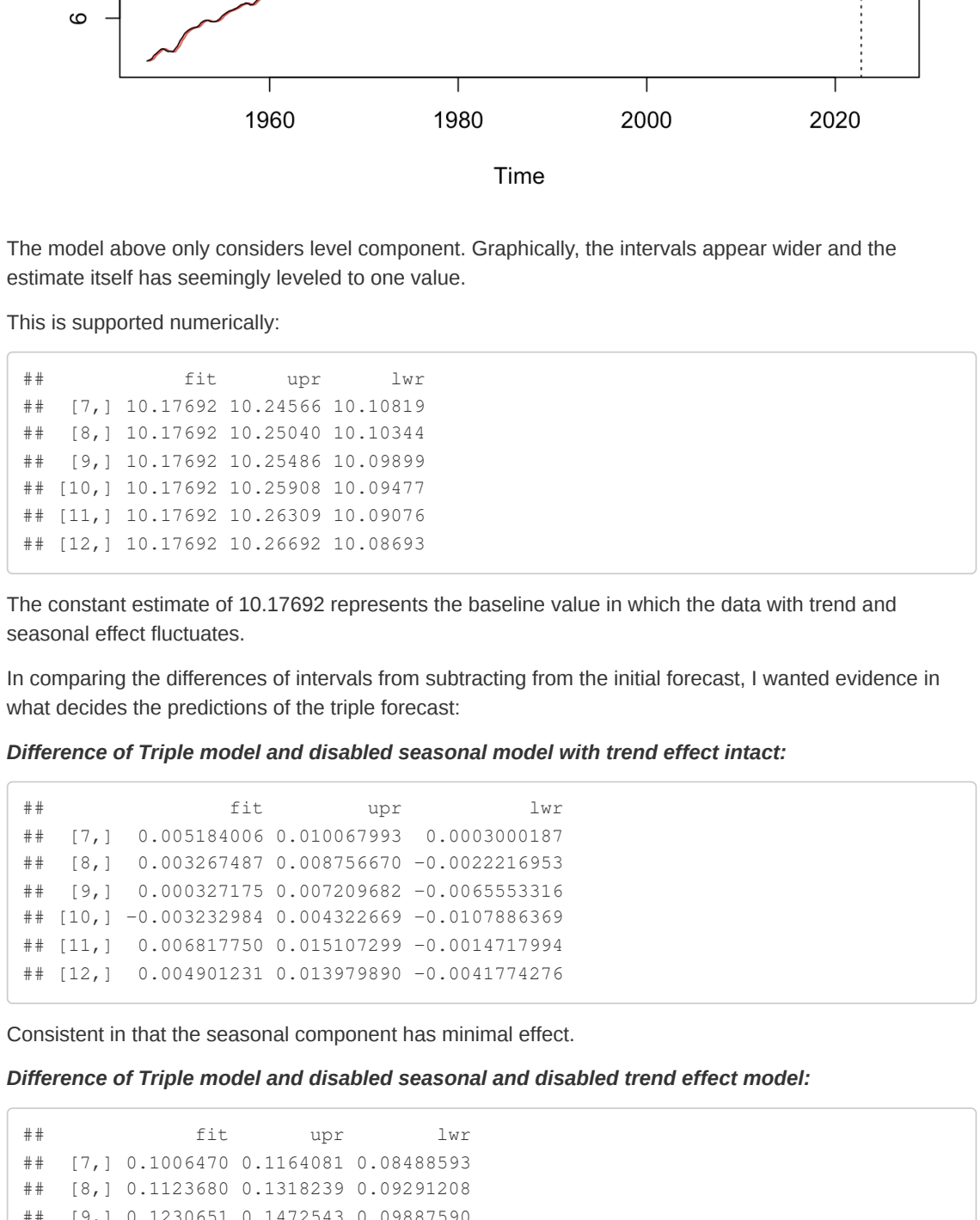
I also implemented intervals for the predictions in the last 6 quarters, the numerical representation was purely for informative purposes of what to expect from the upcoming plots:



## Prophet forecast plot:

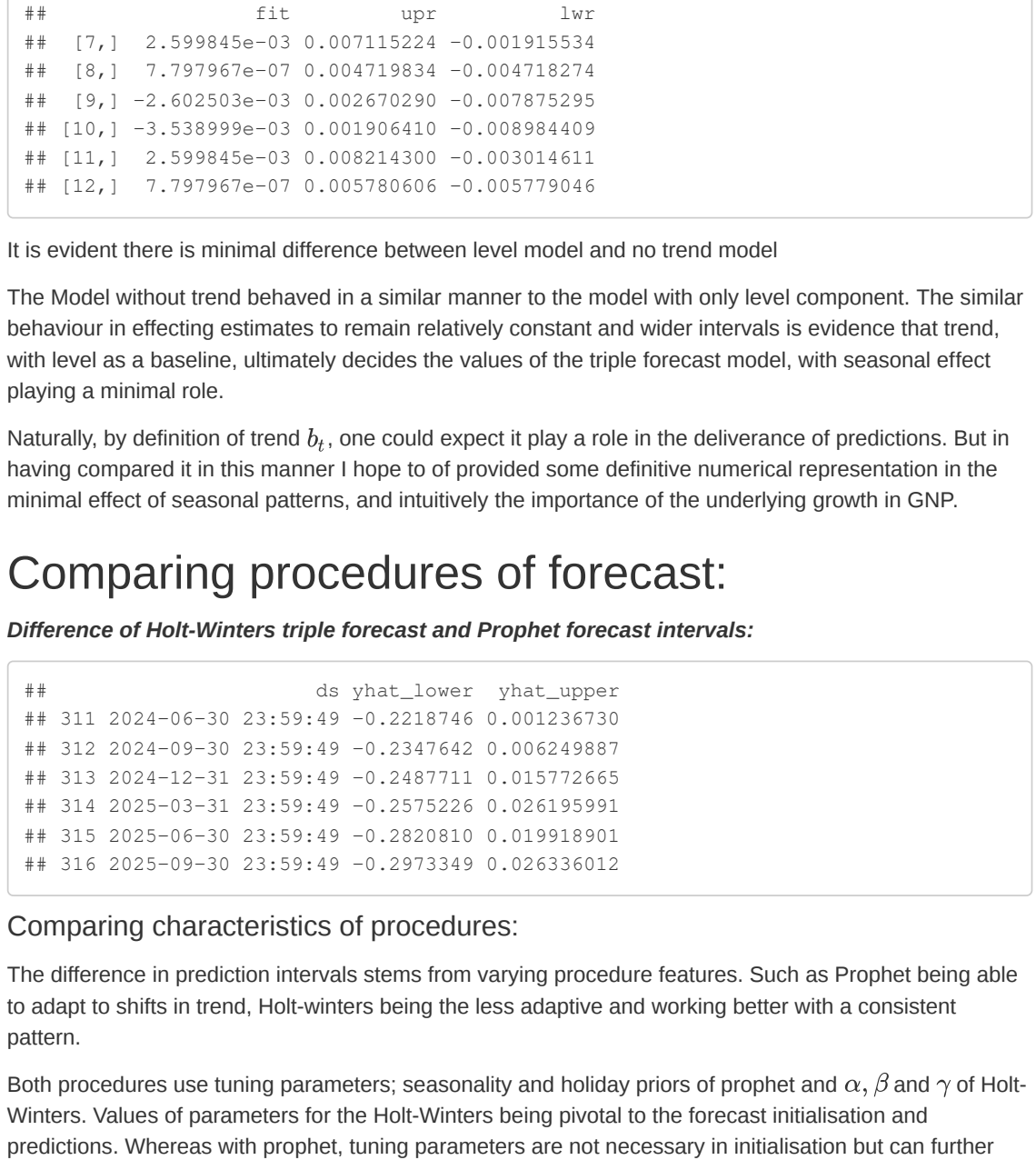


The plot above is a simple representation of the linear relationship between GNP and time. The positively correlated blue line is a continuation/representation of the black line (observed GNP growth) I can also decompose the plot components in the prophet context:



Displaying the predictive linear trend, along with the yearly seasonal component and the random component centered around zero. Accounting for the spike in the forecast initialisation was important, in that I've accounted for a significant deviation from the usual pattern, in an attempt to increase accuracy of the forecast.

## Interactive plot:



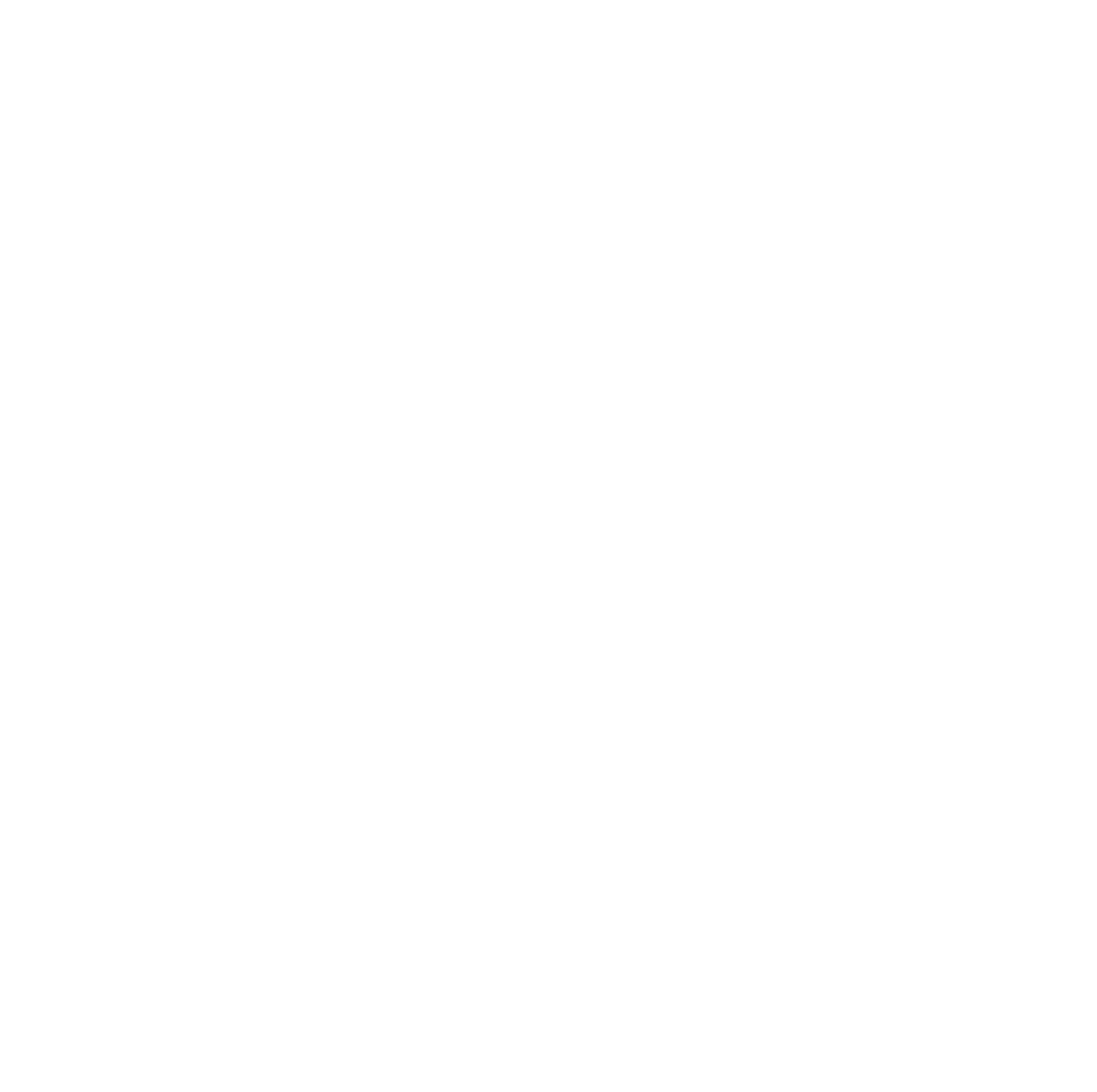
This interactive plot is a variation of the simple representation that allows for better exploration of the observed and fitted trend. For example, I can verify the predictive outcomes of the graph match that with the confidence interval produced earlier on of the last 6 quarters. Of Q4 2025, the prediction from the graph is 10.21 which coincides with the prediction interval of 10.17-10.24. Confirming the analysis is consistent in numerical and graphical representation.

## Non-parametric forecasting via Holt-Winters model:

My choice to include a comparison in the form of a non-parametric forecast is that unlike parametric methods, it does not impose a functional form on  $m_t$  with unknown parameters. This in turn, offers a more data driven shape for  $m_t$ .

## Understanding the seasonal component:

The seasonal component of the series refers to the regular, predictable patterns that repeat over specific periods. In either of the multiplicative or log additive models, it's important that the seasonal component effects the trend as minimal as possible. By minimising seasonal influence, the predictions become more stable and reflect the underlying pattern. Here, through triple exponential smoothing/Holt-Winters forecasting, I can show that the exclusion of the seasonal component does not significantly alter the predictions.



The graph above is the instance in which the model applies triple exponential smoothing, a process of averaging past values, to the series. Considering trend, level and seasonal components. Parameters  $\alpha$ ,  $\beta$  and  $\gamma$  control the degree of smoothing, setting parameters to zero or false effectively kills the component in the model:



Here, having set  $\gamma$  to false, I have produced a forecast of no seasonal effect, so the model had not accounted for seasonality.

## Triple exponential smoothing prediction with prediction interval:



## Disabling seasonal component:



Ultimately seeming indifferent to the full triple method. The notion is also satisfied through numerical representation, without seasonal components the predictions are not significantly altered. Evidently supporting that seasonality, the repeatable predictable patterns of GNP, is not a deciding factor in the determining of future values.

## Importance of Holt-Winters trend $\hat{b}_t$ :

Terminology of Holt-Winters refers to regular trend ( $m_t$ ) as the level component in which it is a necessary in providing a baseline for forecast. Trend, say now  $\hat{b}_t$ , is understood to be the measurement of how much  $m_t$  and time series  $x_t$  are expected to increase.

First I considered the baseline of the level component by removing trend and seasonal effects:



The model above only considers level component. Graphically, the intervals appear wider and the estimate itself has seemingly leveled to one value.

This is supported numerically:



The constant estimate of 10.17692 represents the baseline value in which the data with trend and seasonal effect fluctuates.

In comparing the differences of intervals from subtracting from the initial forecast, I wanted evidence in what decides the predictions of the triple forecast:

## Difference of Triple model and disabled seasonal model with trend effect intact:



Consistent in that the seasonal component has minimal effect.

## Difference of Triple model and disabled seasonal and disabled trend effect model:



Shows the wider interval may stem from removing the trend component.

## Now defining a model without trend:



## Difference of No trend model and level model:



It is evident there is minimal difference between level model and no trend model

The Model without trend behaved in a similar manner to the model with only level component. The similar behaviour in effecting estimates to remain relatively constant and wider intervals is evidence that trend, with level as a baseline, ultimately decides the values of the triple forecast model, with seasonal effect playing a minimal role.

Naturally, by definition of trend  $\hat{b}_t$ , one could expect it play a role in the deliverance of predictions. But in having compared in this manner I hope to of provided some definitive numerical representation in the minimal effect of seasonal patterns, and intuitively the importance of the underlying growth in GNP.

## Comparing procedures of forecast:

### Difference of Holt-Winters triple forecast and Prophet forecast intervals:



Comparing characteristics of procedures:

The difference in prediction intervals stems from varying procedure features. Such as Prophet being able to adapt to shifts in trend, Holt-winters being the less adaptive and working better with a consistent pattern.

Both procedures use tuning parameters: seasonality and holiday prior of prophet and  $\alpha$ ,  $\beta$  and  $\gamma$  of Holt-Winters. Values of parameters for the Holt-Winters being pivotal to the forecast initialisation and predictions. Whereas with prophet, tuning parameters are not necessary in initialisation but can further enhance the accuracy of the forecast.

Prophet's procedure is also based on an additive Time-series model in which capabilities range from fitting non-linear trends and working best with strong seasonal effects. Holt-Winters is a generalised method that does not impose an additive or multiplicative model.

In Relation to the GNP dataset, I believe both procedures had effectively showcased their respective features. Despite for example, the GNP dataset being minimal in seasonal effect; contrasting to the benefit of the Prophet forecast.

## Conclusion:</