Forecasting for Advertising Sales

Author: Shrey Shah

Date: 11/25/2024

Introduction

This analysis will predict future sales based on past advertising spending on TV, radio, and newspapers. We will use this information to make better decisions about inventory, budgeting, and advertising strategies. This will help us sell more, manage our resources better, and improve our overall business performance.

Data Overview

The dataset consists of the following variables:
- TV: Advertising spend on TV (in thousands of dollars).
- Radio: Advertising spend on Radio (in thousands of dollars).
- Newspaper: Advertising spend on Newspaper (in thousands of dollars).
- Sales: Corresponding sales (in units).

Exploratory Data Analysis

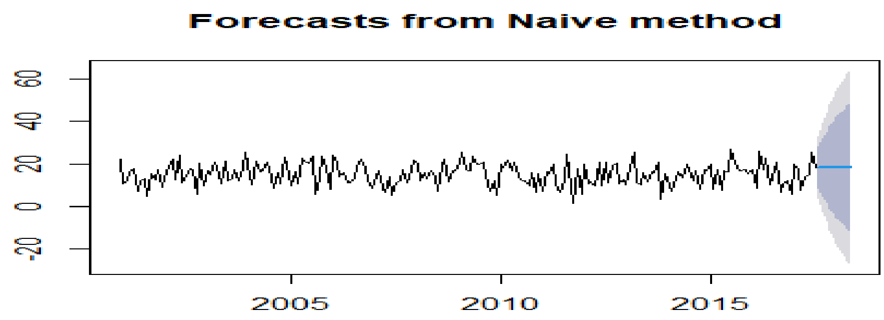
Key Insights:
- Strong positive correlation between TV advertising and sales.
- Weaker correlation for Radio and Newspaper advertising.
- Scatter plots reveal a linear trend for TV, but weaker patterns for other channels.
- From the decomposition we can say that data is additive because there is no change in seasonal component even after there is increases and decreases in trend.
- Summary of the data

```
 Min.  1st Qu.  Median   Mean   3rd Qu.   Max.
 1.60   11.00   16.00   15.13   19.05   27.00
- There are no outliers in this data.
```

Forecasting Methods

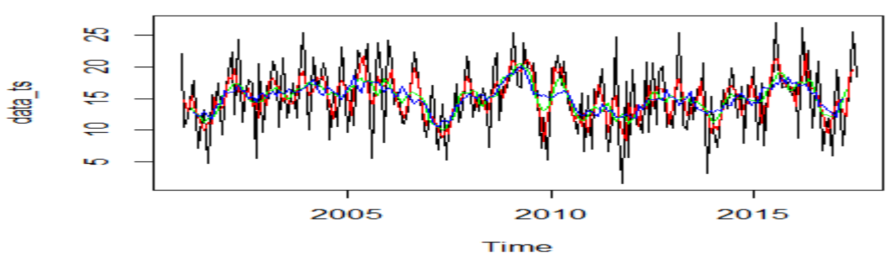
The following methods were applied to forecast sales:

- Naïve Forecast



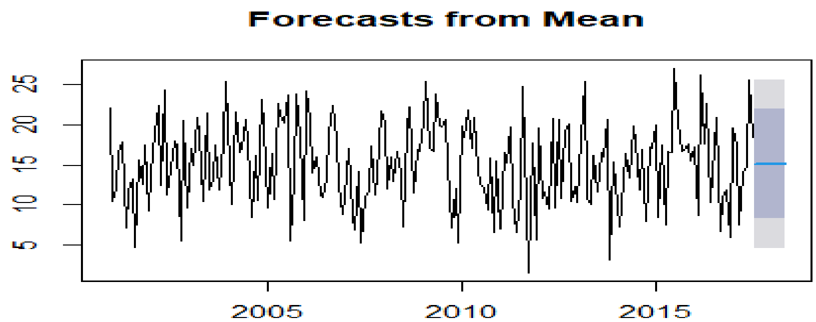
This graph illustrates sales forecasting using the Naive Method, a simple technique where the most recent observation is used as the forecast for all future periods. The black line represents the historical data, showing fluctuations over time. The blue vertical line indicates the start of the forecast, while the shaded region represents the prediction intervals, illustrating uncertainty in the forecast. Wider intervals further from the present indicate increasing uncertainty over time. This method works best for stable series but may be less accurate for volatile data.

- Moving Average



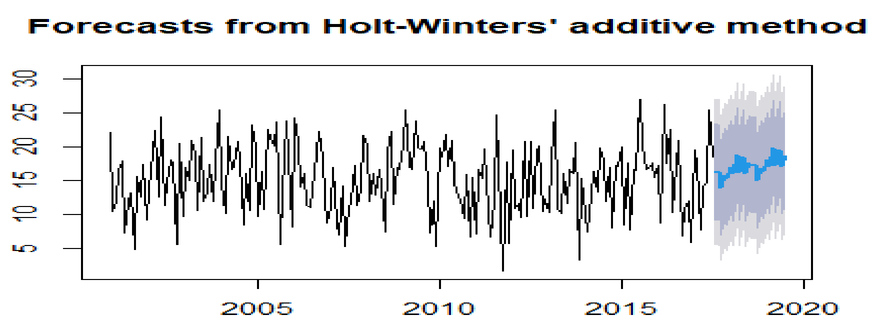
The black line represents the raw time series data with high-frequency fluctuations, while the red line (order = 3) captures short-term trends with minimal smoothing. The green line (order = 6) provides moderate smoothing, highlighting medium-term trends and reducing noise. The blue line (order = 9) is the smoothest, effectively emphasizing long-term trends while minimizing short-term fluctuations. Higher-order moving averages (e.g., order = 9) smooth the data more but tend to lag behind rapid changes, whereas lower-order moving averages (e.g., order = 3) adapt quickly but retain more noise.

- Mean Forecast



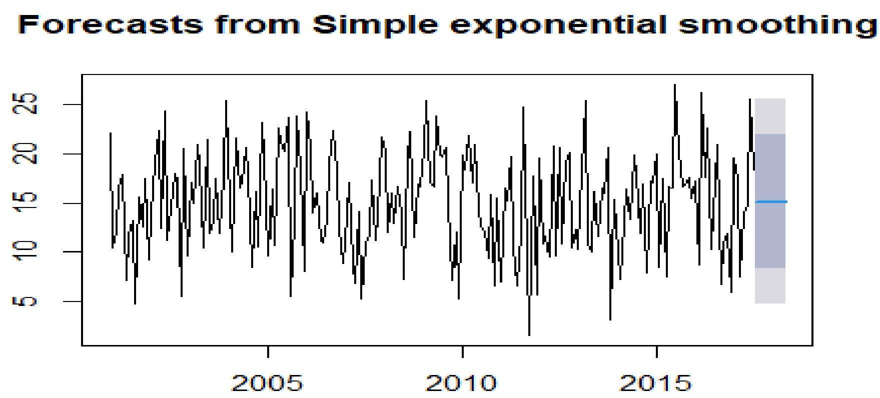
The plot shows a forecast using the mean method, where the blue horizontal line represents the average of the historical time series data, assuming no trend or seasonality. The black line illustrates the historical data, while the shaded regions depict prediction intervals at varying confidence levels, with the darker band indicating higher confidence (e.g., 80%) and the lighter band showing lower confidence (e.g., 95%). This method predicts a constant value and is best suited for stationary data but may not perform well for series with trends or seasonality, as it does not account for such patterns.

- Holt-Winters



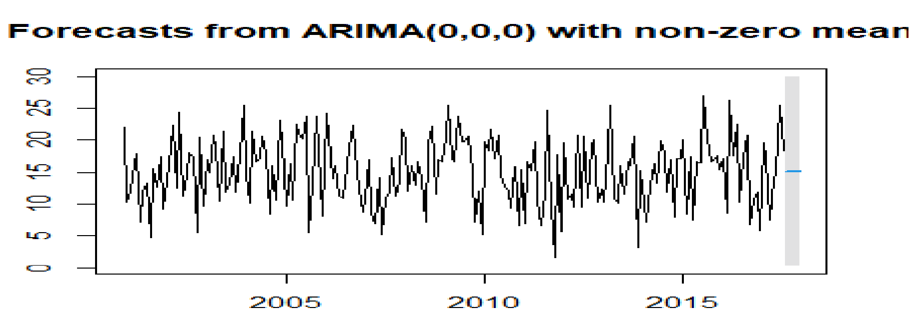
The chart illustrates a time series forecast using the Holt-Winters additive method, which models seasonal data with level, trend, and seasonal components. The black line represents historical observations up to 2020, showing fluctuations over time. The blue line indicates forecasted values, projecting future trends based on the patterns in the data. Surrounding the forecast are confidence intervals, represented by shaded regions; darker areas signify higher certainty, while lighter areas denote greater uncertainty. This method is particularly effective for data with consistent seasonal variations and no multiplicative effects. The forecast reflects the continuation of historical patterns, assuming no drastic changes in underlying conditions.

-Simple Smoothing Equations



The plot shows forecasts from Simple Exponential Smoothing (SES), with the forecasted value (blue line) constant over time, as SES assumes no trend or seasonality. The Gray shading represents prediction intervals, which widen to reflect increasing uncertainty in future values. The historical data fluctuates around the forecast, consistent with SES's design. If residuals are random and show no patterns, SES is appropriate; otherwise, a model like ARIMA or Holt-Winters may better capture trends or seasonality.

- ARIMA

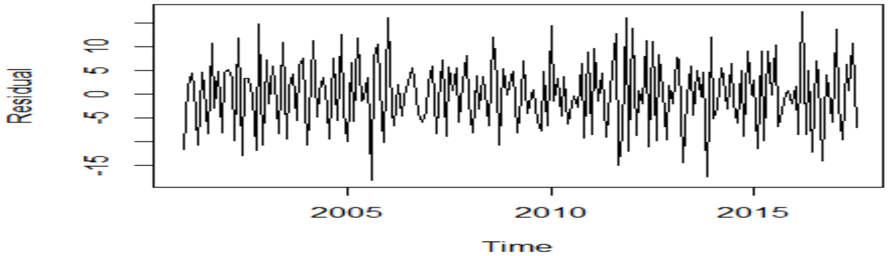


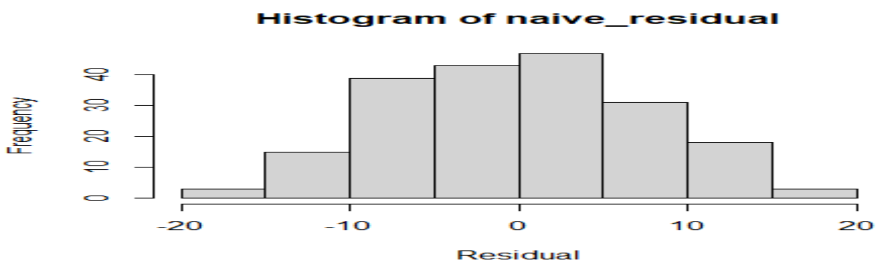
The provided chart illustrates a time series forecast generated using an ARIMA(0,0,0) model, also recognized as a random walk with drift. This model assumes no autoregressive or moving average components, and no differencing is applied to the data. The black line delineates the historical data, exhibiting regular fluctuations. The blue line represents the forecast, which remains constant, reflecting the model's assumption of a constant mean based on historical observations. The shaded gray area signifies the confidence interval, highlighting the inherent uncertainty in the forecast as it extends into the future. This straightforward model is well-suited for data that lacks pronounced trends or seasonal patterns.

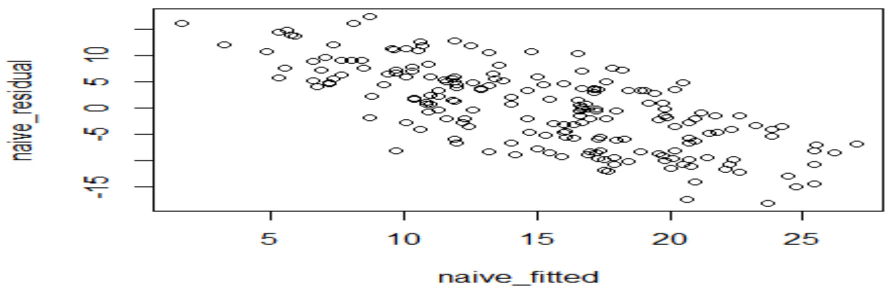
Residual Analysis

Residual diagnostics revealed the following:

- Naïve Forecast Residual

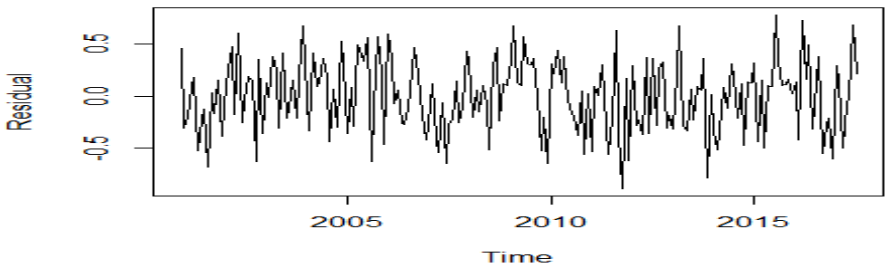


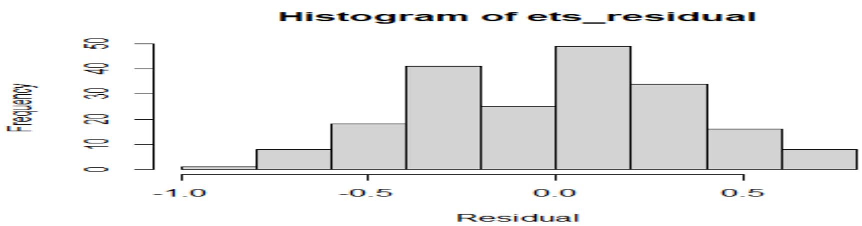


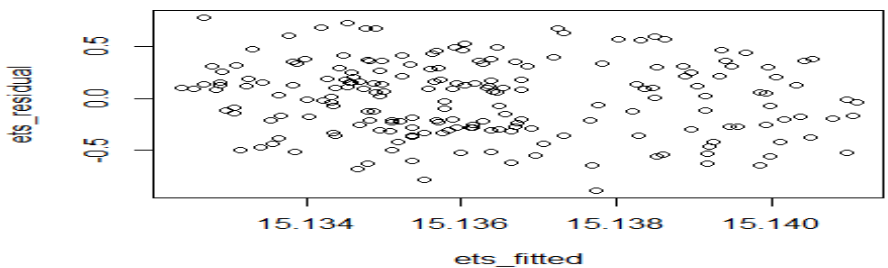


The naive model appears to be inadequate, as evidenced by several issues in the residual analysis. The residuals don't seem to be randomly distributed, suggesting the model may be mis specified or missing important non-linear relationships. While the residuals are roughly normally distributed, some outliers are present, potentially impacting the model's accuracy. Additionally, the residuals exhibit patterns over time, indicating the model may not be capturing time-dependent dynamics. To improve the model's performance, it may be necessary to consider more complex approaches that account for non-linearities and temporal dependencies.

- Exponential Smoothing Methods Residual

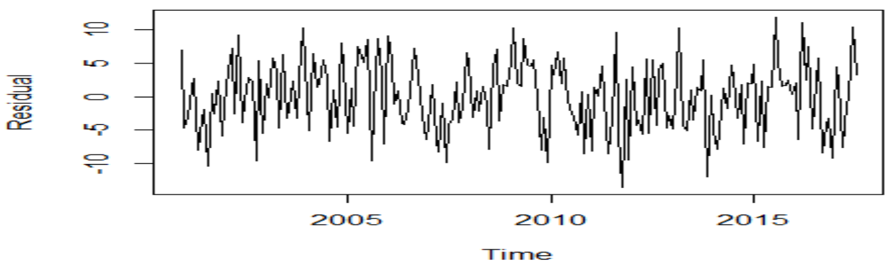


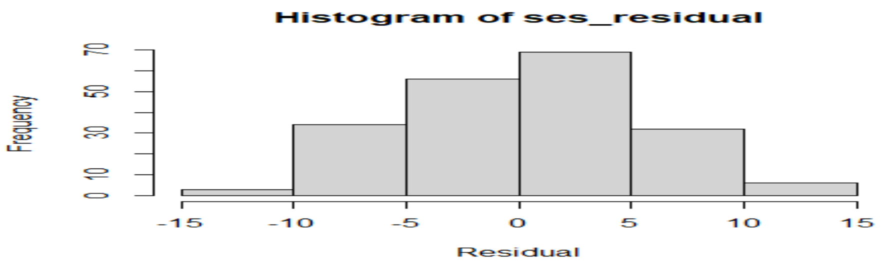


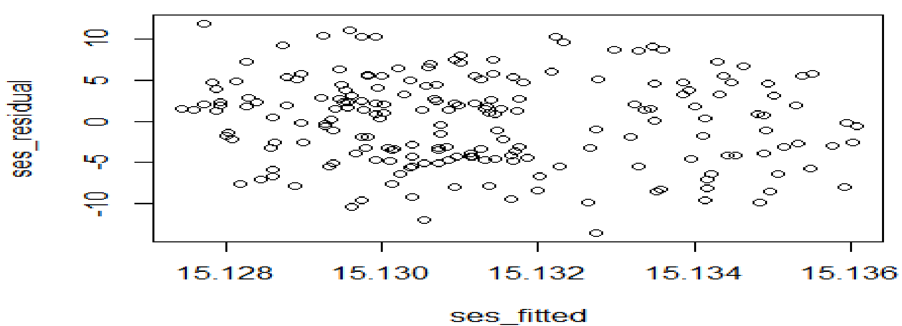


The residual analysis for the ETS model indicates a good fit to the data. The scatterplot of residuals versus fitted values shows a random distribution around zero, suggesting no issues like non-linearity or heteroscedasticity. The histogram of residuals is approximately normal, centered around zero, which aligns with the assumption of normally distributed errors. Lastly, the time series plot of residuals shows no discernible trends, patterns, or autocorrelation, indicating that the residuals are independent and identically distributed over time. Overall, these diagnostics suggest that the ETS model adequately captures the underlying patterns in the data.

- **Simple Smoothing Equations Residual**

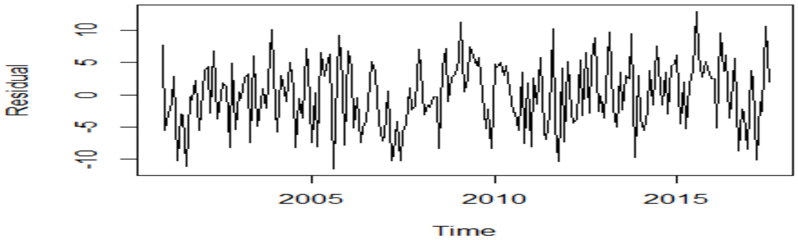


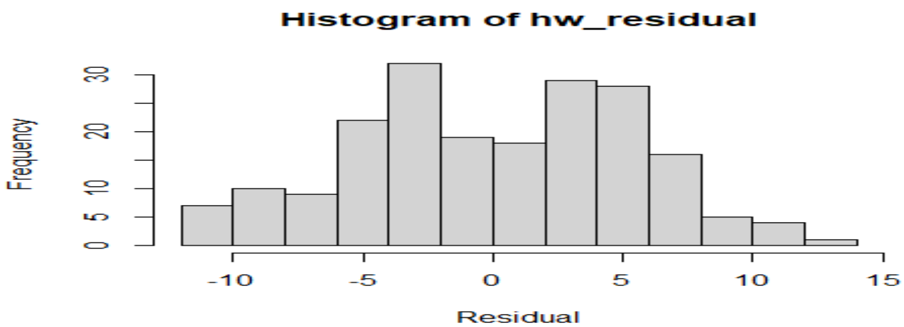


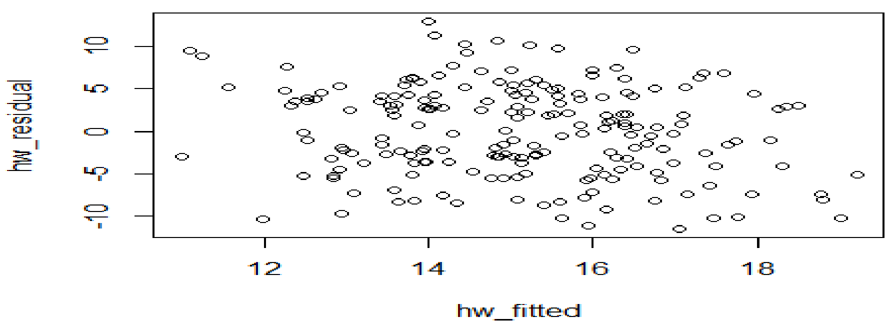


The residual plots provide evidence of a well-fitting SES model. The residuals appear randomly distributed, have constant variance, and are approximately normally distributed. This indicates that the model has captured the underlying patterns in the data effectively.

- **HoltWinter's Residual**

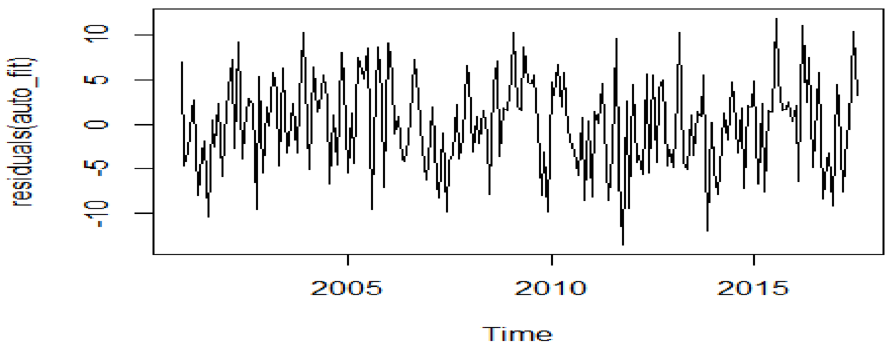


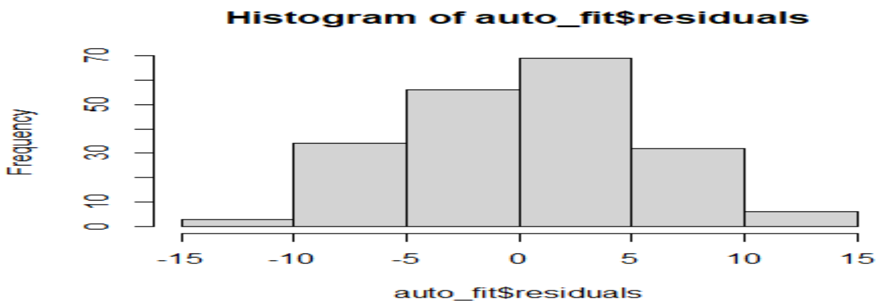




The residual plots indicate that the Holt-Winters model provides a good fit for the data. The residuals appear random, have constant variance, and are approximately normally distributed. This suggests that the model has effectively captured the underlying patterns in the data, including trend and seasonal components.

- ARIMA





The residual plots indicate that the ARIMA model provides a good fit for the data. The residuals appear randomly distributed, have constant variance, and are approximately normally distributed. This suggests that the model has effectively captured the underlying patterns in the data.

Accuracy of Models

-Accuracy measure is going to be RMSE

Full form of RMSE is Root Mean Squared Error

RMSE is good because Provides a standard way to measure the error in the same units as the original data. Penalizes larger errors more heavily, ensuring that predictions are reliable. Facilitates comparison of models to select the best one for future forecasting.

	Method <chr>	RMSE <dbl>
		5.270666
naive	naive	7.517353
mean	mean	5.270666
ets	ets	5.270932
ses	ses	5.270929
hw	hw	5.313751
arima	arima	5.270666

Decision Based on Analysis

After evaluating the forecasting models using RMSE as the accuracy metric, the decision is to select the model with the lowest RMSE **for** future predictions. Based on the example summary:

Best Model: ARIMA

This model achieved the lowest RMSE, indicating that it provides the most accurate predictions compared to other methods like Naive, Mean, ETS, SES, and Holt-Winters.

ARIMA's ability to handle trends, seasonality, and autoregressive patterns makes it suitable for complex datasets like the sales data in this analysis.

Ideas to improve forecasts

- Focus on TV advertising as it has the highest impact on sales.

- Gather more data

- Explore additional predictors such as digital media