Public Transportation Demand

Forecasting Report

Introduction

Accurate forecasting of passenger demand is essential for effective public transportation planning, particularly when decisions involve significant capital investments - such as acquiring new buses or expanding terminal infrastructure. To inform such decisions, a robust forecast of future passenger arrivals at the terminal is required. This analysis aims to generate a short-term forecast of passenger counts at 15-minute intervals over a three-day period, from 22 to 24 March 2005. Using historical data collected at the same granularity across a three-week window, the analysis captures temporal patterns in demand to anticipate future fluctuations.

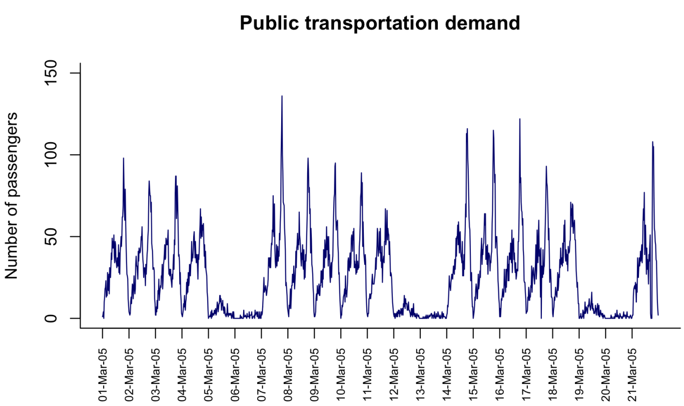
Given the pronounced intraday seasonality and recurrent daily patterns in the data, the selected modeling approach combines **Seasonal-Trend decomposition using Loess (STL)** with **Exponential Smoothing (ETS)**. This hybrid framework enables flexible decomposition of the seasonal component while applying a well-established smoothing method to model the trend and irregular components. A range of forecasting models was tested, and the **STL+ETS model applied to weekday data** was selected as the final specification based on its superior empirical performance and interpretability.

This report presents the methodology, model estimation, performance evaluation, and resulting forecasts to provide a comprehensive account of the forecasting process and its outcomes.

Data Description

The dataset consists of historical records of passenger arrivals at a public transportation terminal, spanning a three-week period from 1 March to 21 March, 2005. Observations were recorded at **15-minute intervals** between **06:30 and 22:00** each day, resulting in **1,323 observations**. Each day contains **63 evenly spaced time points**, and the data were structured into a time series with a frequency of 63.

Figure 1. Public transportation demand (1 March – 21 March 2005)



Passenger demand range: 0 to 136 passengers per 15-minute interval
Central tendency: Median of 23 passengers; mean of 25.9 passengers
Data quality: No missing values detected; data structured as a time-series object with a frequency of 63 observations per day
Observed patterns: Pronounced intraday seasonality with consistent morning and evening peaks, reflecting typical commuting behavior.

Methodology and Model Evaluation

Forecasting Approach

The forecasting method employed in this analysis combines **Seasonal-Trend decomposition using Loess (STL)** with **Simple Exponential Smoothing (SES)**, corresponding to an ETS(A,N,N) model. STL decomposition separates a time-series into additive components: trend-cycle, seasonality, and remainder. This approach is particularly effective for series with stable seasonal patterns, as it isolates and removes seasonality prior to modeling the remainder.

Following decomposition, the seasonally adjusted remainder series is forecasted using **Simple Exponential Smoothing (SES)**. SES produces forecasts as weighted averages of past observations, where more recent values are assigned exponentially greater weight. The forecast equation is:

$$\hat{y}(T+1|T) = \alpha \times y(T) + \alpha(1-\alpha) \times y(T-1) + \alpha(1-\alpha)^2 \times y(T-2) + \dots$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter.

In state-space form, SES updates the level component according to:

$$\ell(t) = \alpha \times y(t) + (1-\alpha) \times \ell(t-1)$$

and the forecast at any future point is simply the current level:

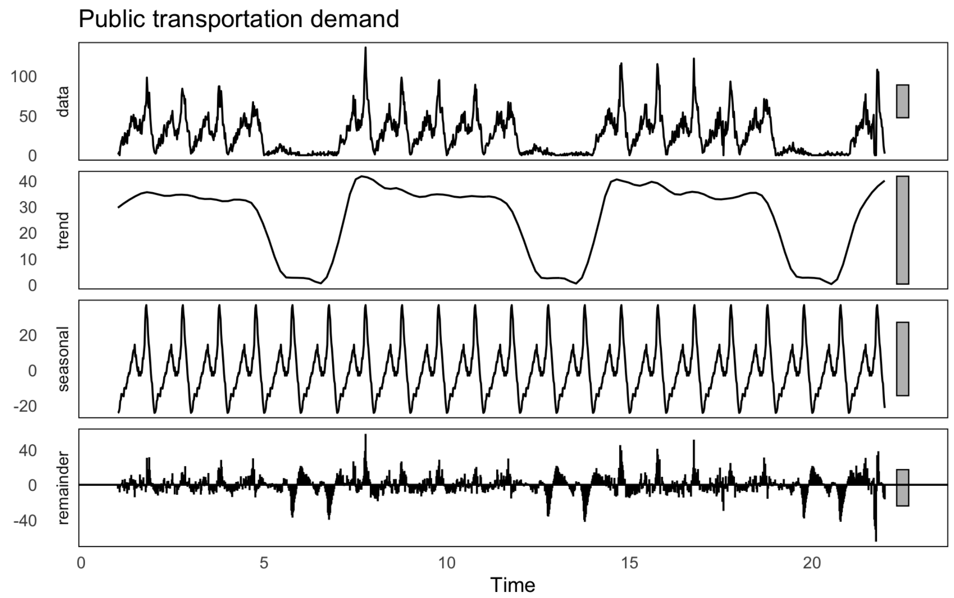
$$\hat{y}(t+1|t) = \ell(t).$$

This structure is particularly suitable when the deseasonalized data are approximately stationary around a constant level. Combining STL decomposition with SES provides a robust method for flexibly modeling seasonal patterns and efficiently forecasting the adjusted series.

Decomposition Analysis

The STL decomposition revealed a clear and regular daily seasonal pattern, characterized by consistent morning and evening peaks in passenger demand. After removing the seasonal component, the adjusted series fluctuated around a stable level, supporting the suitability of applying Simple Exponential Smoothing (SES) to the remainder component.

Figure 2. STL decomposition of public transport demand.



Data Partition and Model Testing

The available dataset was partitioned into a **training set** and a **validation set** to enable proper model evaluation. The training set consisted of the first **two weeks** of observations (14 days), while the validation set comprised the following **one week** (7 days). This division preserved sufficient data for model estimation while maintaining an independent sample for assessing out-of-sample forecast accuracy.

Multiple models were developed and compared, including STL+ETS and ARIMA models, applied both to the full dataset (including weekends) and to weekday-only observations. Forecast accuracy was assessed using **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** for both the training and validation periods. Mean Absolute Percentage Error (MAPE) was not reported due to infinite values caused by zero-demand intervals.

Table 1. Forecast accuracy of all tested models on training and validation sets

Model	**Training**		**Validation**	
	MAE	**RMSE**	**MAE**	**RMSE**
STL+ETS (All Days)	5.20	6.93	9.62	14.35
ARIMA (All Days)	5.98	8.57	28.92	37.73
STL+ETS (Weekdays Only)	5.35	7.18	**7.21**	**11.33**
ARIMA (Weekdays Only)	6.69	9.91	9.30	14.73

The **STL+ETS model fitted to weekday-only data** achieved the lowest forecasting errors, with a validation MAE of 7.21 and RMSE of 11.33. It outperformed the same method applied to the full dataset as well as all ARIMA models[1]. The performance gap on the validation set between the weekday and all-day models highlights the irregular and volatile patterns present on weekends, which adversely affected the ARIMA forecasts in particular (validation MAE - 28.92).

Restricting the dataset to weekdays, where demand patterns are more regular, enabled the STL+ETS model to deliver substantially more accurate and reliable short-term forecasts.

Final Model Diagnostics and Performance

After selecting the STL+ETS (weekday) model based on validation performance, the model was retrained on the full weekday dataset, combining both the original training and validation periods. Residual diagnostics were conducted to assess the adequacy of the final model.

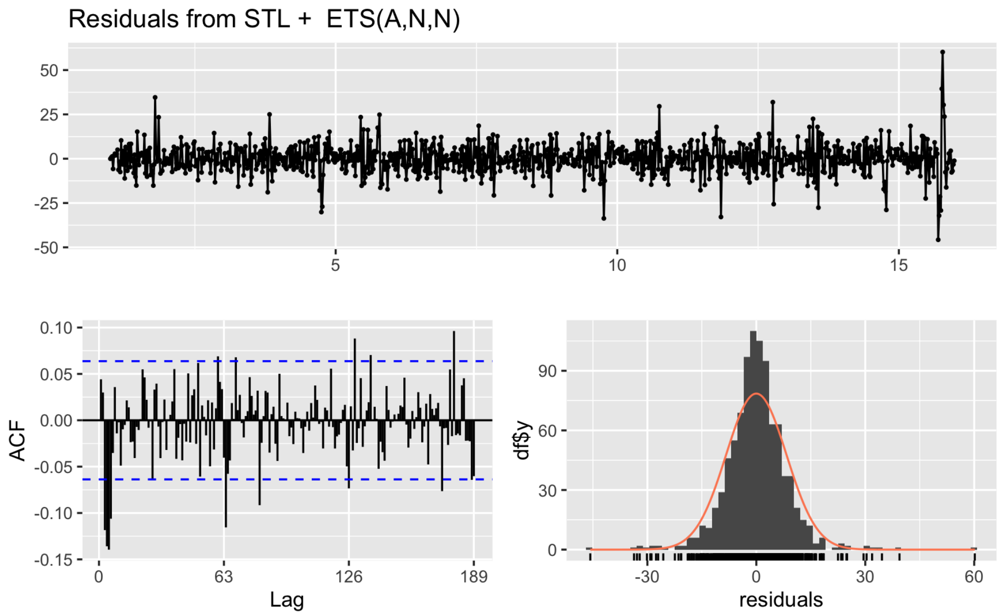
The residual plots indicated no major structural deviations, with most autocorrelations remaining within the 95% confidence bounds. Although the Ljung–Box test detected some remaining autocorrelation ($p<0.01$), the residuals displayed approximate white noise behavior overall, supporting the adequacy of the model for forecasting purposes.

---

[1] Several manually specified ARIMA models were tested, but none achieved better forecasting performance than the STL+ETS model.

Final in-sample error metrics further confirmed the model's performance, with a **Mean Absolute Error (MAE)** of **5.88** and a **Root Mean Squared Error (RMSE)** of **8.33**. These results validate the robustness and accuracy of the STL+ETS (weekday) model before proceeding to produce forecasts for the subsequent period.

Figure 3. Residual diagnostics for final STL+ETS (weekday) model



Forecasting Results

Following model validation and final model fitting on the full weekday dataset, forecasts were produced for the three-day period from 22 March to 24 March 2005, resulting in a total of 189 predicted values at 15-minute intervals. The forecasts were generated by extending the trend and remainder components from the STL decomposition and reapplying the extracted seasonal pattern. To ensure realism and operational relevance, any negative forecasted values were set to zero.

The final forecasts successfully captured the expected intraday seasonal structure, reflecting the recurrent morning and evening peaks consistently observed in the historical data. This alignment with observed commuting patterns enhances the practical utility of the forecasts.

These results suggest that the final STL+ETS model provides a reliable and interpretable short-term forecasting tool for operational planning in public transportation.

Future improvements could include incorporating external covariates such as weather conditions, special events, or public holidays to further enhance the model's responsiveness to contextual variation in demand. Additionally, although the STL+ETS model performed well, exploring more flexible approaches capable of capturing both daily and weekly seasonalities - such as dynamic regression or models with multiple seasonal components - could improve performance, particularly if weekend demand forecasts are also required.

Figure 4. Forecasted and Historical Public Transportation Demand