Time Series Analysis of Retail Gas Prices in Boston

David Custer
Department of Mathematics and Statistics
Coastal Carolina University

1. Introduction and Decomposition

Over 20 million people travel through Boston every year. At some point, locals and visitors will need to stop for gas. Our goal is to model the Energy Information Administration’s collection of retail gas prices in Boston for all grades from January 2010 to December 2018 and predict future gas prices over the next five years until 2025. Modeling and forecasting data over time requires methods beyond regression. For example, residual assumptions such as constant variance and Normally distributed do not hold for time-dependent data. A linear relationship between the predictor and response rarely exists. Furthermore, the most significant issue is that autocorrelation between observations reduces the precision of mutual independence. We can employ time series analysis (a stochastic process with an indexed collection of random variables that randomly evolves) to perform through statistical inference. Our first step starts with plotting the data over time to look for seasonality, the periods of predictable cyclic patterns.

2. Autocorrelation and Differencing

• Differencing begins by subtracting the previous observation from the current observation.
  • A process made stationary by differencing is an integrated time series [1].
  • AR(p): an autoregressive process of order p with parameter $\phi$
  • MA(q): a moving average process of order q with parameter $\theta$
  • ARIMA(p, d, q): If d > 0, it is a non-stationary time series.
  • ARIMA(p, d, q): If d = 0, it is a stationary time series.
  • ARIMA(p, d, q): If d < 0, it is a trend-stationary time series.

3. Stationarity Testing and Invertibility

Stationarity Testing

• Aside from the sample ACF and PACF correlograms in the previous section, we can use formal tests for stationarity (a condition where the properties of the data are constant throughout).
  • A stationary process is a series that exhibits zero trends, no seasonality, and has constant variance.
  • KPSS: determines if the trend of a series or the level (observed) values are stationary
  • ADT: ascertain if a unit root is present in a series

Invertibility Requirements

• A series is invertible if the residuals represent a linear function of current and past observations.
  • Invertibility allows us to calculate the model’s parameters [3].
  • For an MA(q) process to be invertible, we need the roots of the MA characteristic polynomial to all exceed one in absolute value (or modulus, for this series).
  • For an AR(p) process to be invertible, all roots of the AR characteristic polynomial must lie inside the unit circle.

4. Order Specification and Parameter Estimation

The Extended Autocorrelation Function

• Now that we have enough evidence that our series is stationary, we can specify the order and estimate the parameters of our model.
  • Since the sample ACF and PACF correlograms show evidence that an AR or MA model is appropriate, we can use the EACF to determine the order of a model.
  • Our EACF on the left forms a wedge with its tip at the k left p left (0,1), suggesting an MA(1) model.
  • The EACF on the right is an example that would suggest an ARMA(1,1) process.

Akaike and Bayesian Information Criterion

• Given the output of the EACF and by the principle of parsimony, we will use an MA(1) model to process the gas prices.
  • The armafit function in R selects the order of the series via the Hyndman-Khandakar algorithm.

5. Residual Diagnostics and Model Suitability

Residuals

• After Differencing: sample ACF decays slowly, suggesting first differences should be taken, and the sample PACF cuts off after lag k = 1, and the sample PACF cuts off after lag k = 2

6. Forecasting

• For our forecasts and simulations, the left side consists of bootstrapped (resampled) residuals, unlike their Normally distributed counterparts on the right.
  • The prediction bands are smoother for Normally distributed residuals since resampled residuals may not always form a bell curve.
  • The forecast tails are at the top-consist of horizontal lines after 2020. This line is the moving average’s prediction of gas prices until December 2020, with a forecasted price of $2.77 per gallon.
  • When looking at the simulations, we notice that moving averages do not predict future values well. A notable drawback is their reliability on residuals from previous data, with each observation given equal weight. Moving averages also tend to overlook volatility in prices and cyclic patterns. An ARIMA process could forecast better but at the cost of a more complex model.

7. References