Forecasting Total Construction Spending in the United States

Ersun Erdem

San José State University

ECON 138

Dr. Sanchita Mukherjee

December 4, 2025

Introduction

Construction constitutes a significant proportion of the United States economy, making up roughly 4.4% of GDP, according to the U.S. Bureau of Economic Analysis (2025). So, forecasting the future of construction spending can help to understand spending trends within the overall economy, and is therefore of significant interest to economists. Furthermore, individuals and firms within the construction industry would like to know how total spending trends in construction might behave in the future, so that they can adjust their actions and expectations accordingly. Construction firms set their expectations of what projects might become available for bidding, depending upon the overall level of total construction spending. This will determine firm actions regarding hiring/firing and aggressiveness of bidding on projects: general contractors will try to aggressively set their estimates low if they project spending to be scarce, since there are fewer opportunities to generate revenue. Similarly, specialist contractors will likely set lower estimates for their work to general contractors in the same scenario, as again there are expected to be fewer revenue-generating opportunities.

The COVID-19 pandemic had mixed effects on the construction sector of the United States economy: the economic shutdown restricted spending in some types of construction, while lowered interest rates- due to greater economic concerns- increased spending in other areas of construction (Alsharef et al., 2021). This must be taken into account when forecasting any measure of construction spending, if the data include the COVID-19 pandemic.

This paper aims to forecast total construction spending in the United States over the next 5 years using an ARIMA and an Artificial Neural Network (ANN), and comparing those forecasts to Naïve and Seasonal Naïve (Snaïve) forecasts. The ANN model predicts a significant drop in total construction spending by July 2030 almost certainly, with spending predicted to be near $75 Billion, and between $25 Billion and $160 Billion with 90% confidence, compared to the current value of $195 Billion as of July 2025 U.S. Census Bureau, 2025. The ARIMA(4,1,1)(0,1,1)[12] predicts spending to be nearly as a Snaïve model predicts (which will necessarily hover around the Naïve model prediction) at roughly $200 Billion in July 2030. In fact, the ARIMA(4,1,1)(0,1,1)[12] tracks the Snaïve model very nearly throughout the five-year forecasting horizon.

Literature Review

There exists a robust set of literature regarding forecasting construction spending. Some articles forecast total construction spending and/or the construction cost index (CCI), typically due to interest from construction firms and economists alike (Ma et al., 2023; Sing et al., 2015; Williams, 1994; Zhang et al., 2024). Additionally, many articles forecast construction in particular subsets of construction alongside material and labor costs (Ma et al., 2023; Zhang et al., 2024), which is directly useful for construction firms who need to generate estimates of total project costs to create estimates for their clients.

Ma et al. (2023) conduct a literature review of the use of forecasting techniques in construction. Often, material prices are used as an indicator of total construction spending, as well as the construction cost index

(CCI.) The three most common methods for forecasting construction spending were Vector Error Correction Model (VECM), Artificial Neural Network (ANN), and ARIMA models, in decreasing order of frequency (Ma et al., 2023). ARIMA models were mostly used for short-term forecasts, VECM were used for both short-and-long-term forecasts, and ANN forecasts were used for forecasting the costs of individual projects, CCI, or material costs (Ma et al., 2023). No distinction was made as to what time frame exactly would constitute long-term versus short-term.

In line with the literature review by Ma et al. (2023), Sing et al. (2015) conduct a forecast of private-sector construction costs in Hong Kong using a VAR model. Using quarterly data on the cost of completed construction works, Hong Kong real GDP, best lending rate (BLR), producer price index (PPI), and vacancy rate (VR) from Q1 1989 to Q4 2013, Sing et al. (2015) find that a VAR model with up to 8 lags gives a sufficient forecast of construction in Hong Kong, but cannot generalize to other countries. The analysis required a first-difference of the log of real GDP, PPI, construction costs, BLR and VR to make the data stationary, signifying that construction data may require some work to make it appear stationary (Sing et al., 2015).

Similarly, Zhang et al. (2024) forecasts total construction costs as well as the construction costs for various type-specific construction, including private non-residential construction, private lodging construction, and health construction in the United States from January 1993 to December 2022. There is a weak correlation found between total construction spending and the construction cost index, and no correlation between total construction spending and the building cost index (BCI) (Zhang et al., 2024). Last, Zhang et al. (2024) find that their forecasts were more accurate pre-COVID-19 pandemic than post-pandemic, as the pandemic led to drastic increases in the standard errors of their forecasts.

The aforementioned phenomenon of COVID-19 distortions to construction spending is highlighted by Alsharef et al. (2021), who state mixed effects of the early COVID-19 pandemic on the United States construction industry. The negative effects on the construction industry were material shortages, delays in inspections and securing permits, reduced productivity, suspension of ongoing projects or delays of new projects, price increases, and an expected increase in litigation due to the above delays and increased costs (Alsharef et al., 2021). In the opposite direction, there were some unique benefits to the construction sector from the COVID-19 pandemic: increased demand for transportation, residential, and fast-track medical projects would have increased construction spending (Alsharef et al., 2021). Additionally, Alsharef et al. (2021) found that low interest rates increased construction spending during the pandemic, and the combination of massive unemployment across the board and the uniqueness of the pandemic as a societal event meant that construction firms gained a lot of talented workers from a variety of backgrounds and leveraged those varied backgrounds to audit their internal processes and improve their operational efficacy (Alsharef et al., 2021).

Finally, Williams (1994) conducts a forecast of changes in CCI using neural networks. Williams (1994) gathers weekly CCI data published in ENR magazine from July 1967 to December 1991, as well as data on the prime lending rate, or the short-term interest rate on business loans, and the number of housing starts. For comparison, Williams (1994) also forecasts using exponential smoothing and a linear regression model, and sets the data from September 1986 to December 1991 as the test set. Given these data, the neural network forecasting method was less accurate than both the exponential smoothing and the linear regression model. In particular, the neural network had issues predicting price decreases, though there were limited instances of CCI decreases in the training data (Williams, 1994).

Economic Model

This paper will consider five forecasting models: Naïve, Seasonal Naïve (Snaïve), Artificial Neural Network (ANN), and ARIMA. In particular, the auto.arima function yields a suggested ARIMA model of ARIMA(4,1,1)(0,1,1)[12] on the original data, so we will consider that ARIMA specification. Below are the mathematical representations of the four model specifications. It should be noted that y_t will represent the transformed time series, to make the data stationary, and x_t will represent the original time series.

The Naïve model is as follows:

$$\hat{x}_{t+h|t} = x_t$$

The Snaïve model can be represented as:

$$\hat{x}_{t+h|t} = x_{t+h-12(k+1)}, k = \lfloor \frac{h-1}{12} \rfloor$$

The ARIMA(4,1,1)(0,1,1)[12] model is represented by the following:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - B)(1 - B^{12})\hat{x}_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})\epsilon_t$$

Where B represents the backshift operator.

Finally, the ANN model can be represented by the following:

$$\hat{y}_t = g(\vec{y}_{t-1}) + \epsilon_t$$

Where g is a real-valued function such that $g : \mathbb{R}^{24} \to \mathbb{R}$ and $\vec{y}_{t-1}$ is the vector of the last 24 values of $y_t, \{y_{t-1}, ..., y_{t-24}\}$, as Artificial Neural Network is specified as a 24-12-1 structure.

Fundamentally, this paper seeks to understand how spending patterns within the construction industry might change over the next five years. Many infrastructure projects are publicly funded, with a significant subset coming from universities. With the current administration's desires to cut current federal spending- particularly with respect to universities- one might expect a significant reduction in construction spending, at least for the remainder of the current administration. Thus, we may expect to see reduced spending through the forecast horizon, perhaps with some increase in the fifth year.

Data & Analysis

The data for construction spending was sourced by the St. Louis Federal Reserve, Federal Reserve Economic Data, listed as TTLCON (U.S. Census Bureau, 2025). The data is monthly, spanning from January 1993 to July 2025. The time series is summarized below.

Table 1: Descriptive Statistics for Monthly Total Construction Spending in the U.S., TTLCON (Units, Millions of Dollars)

Minimum	30,264
1st Quartile	63,253
Median	80,703
Mean	90,324
3rd Quartile	107,374
Maximum	203,769
Standard Deviation	38,445
Number of Observations	391

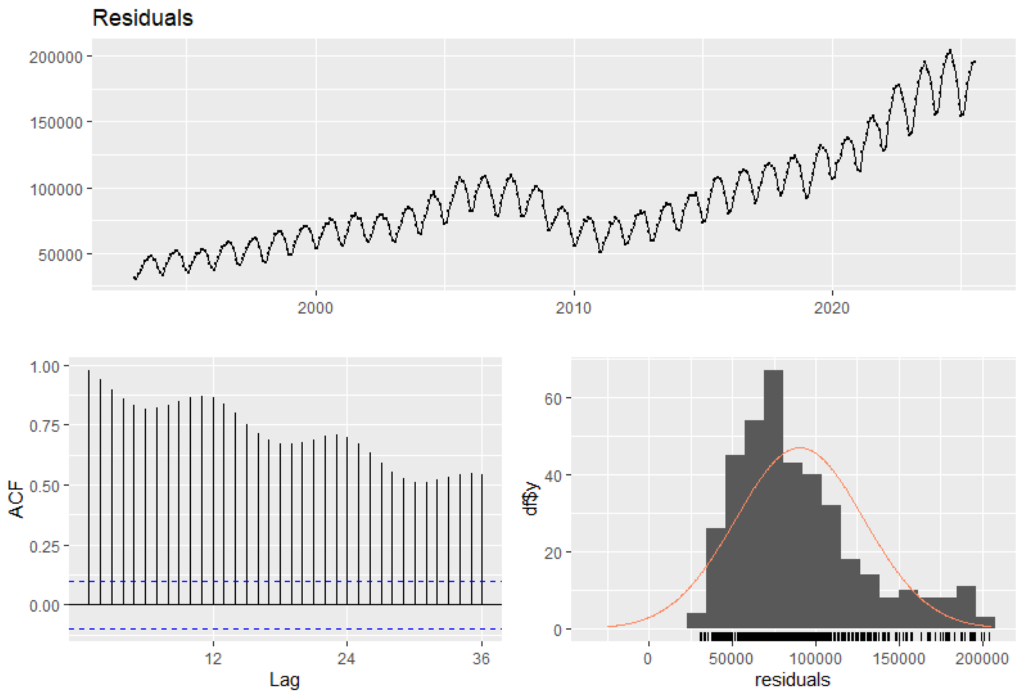

Figure 1: Residuals Plot of Total Construction Spending in the United States (TTLCON)

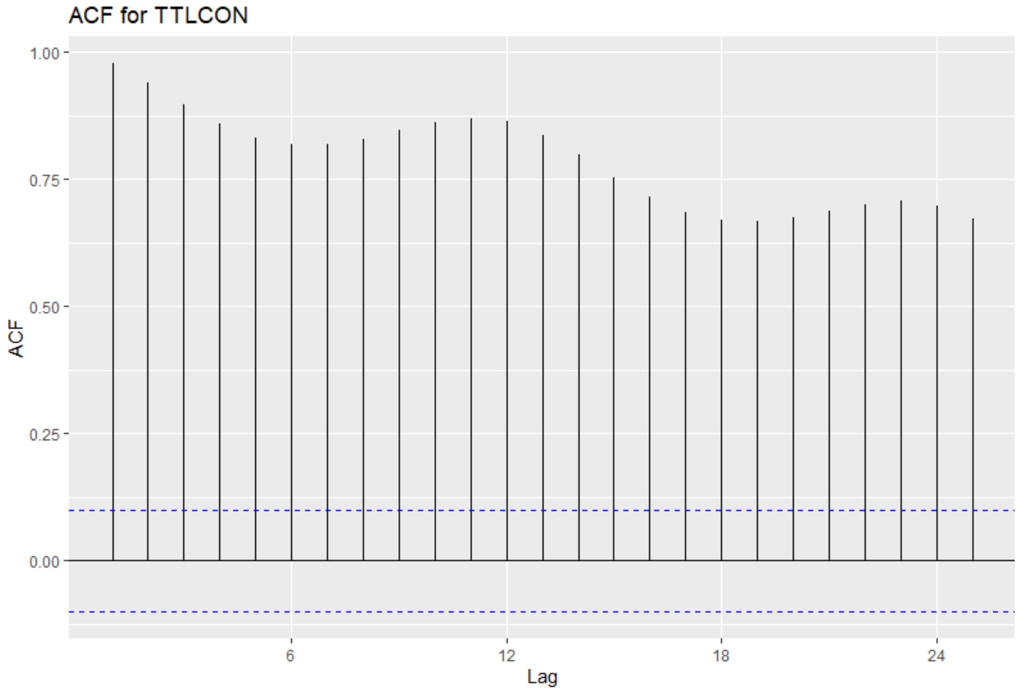

Figure 2: ACF of Total Construction Spending in the United States (TTLCON)

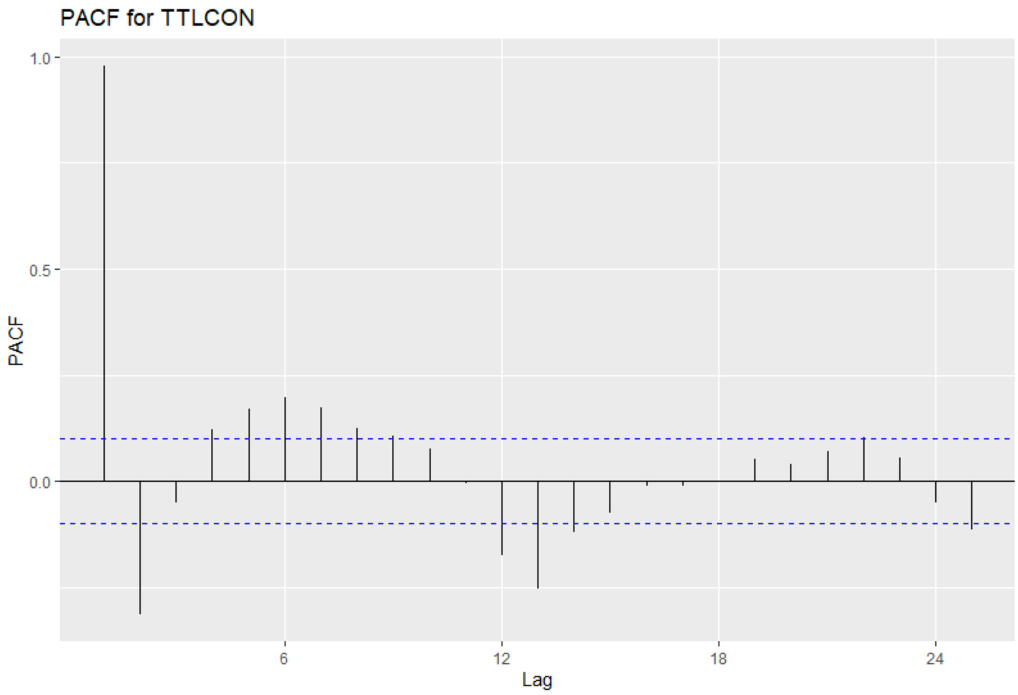

Figure 3: PACF of Total Construction Spending in the United States (TTLCON)

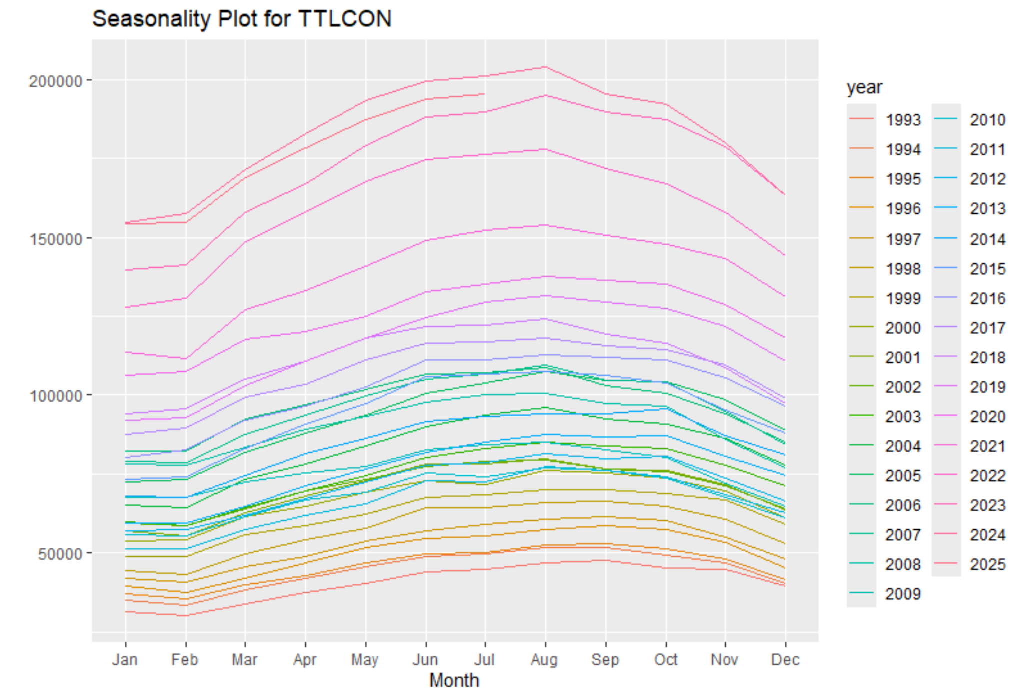

Figure 4: Seasonality of Total Construction Spending in the United States (TTLCON)

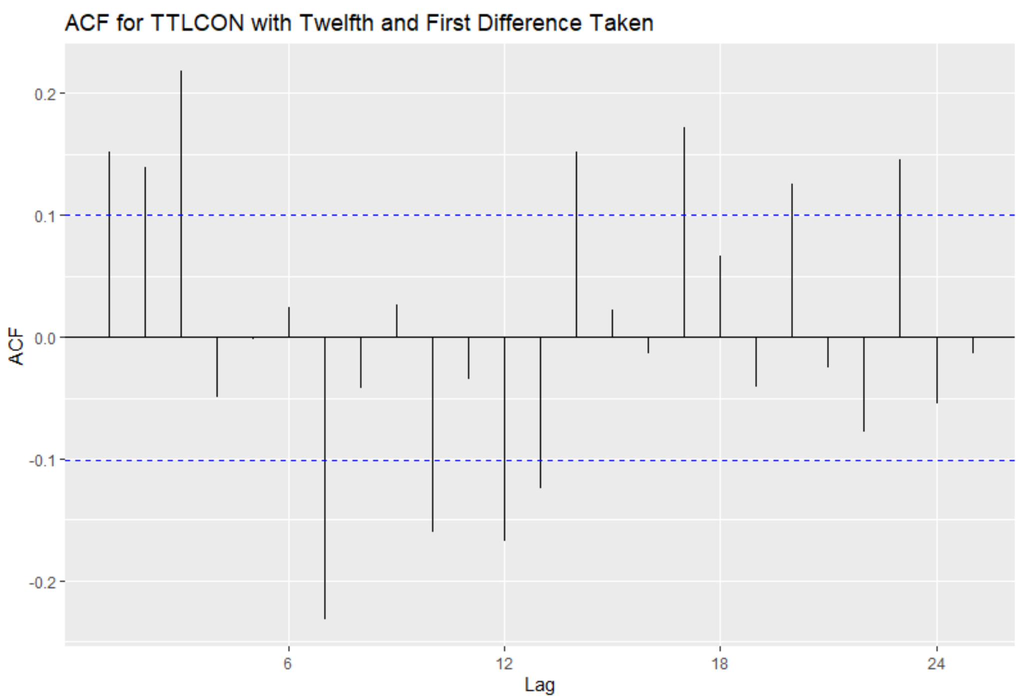

Figure 5: ACF of Total Construction Spending in the United States (TTLCON) with Twelfth and First Differences Taken

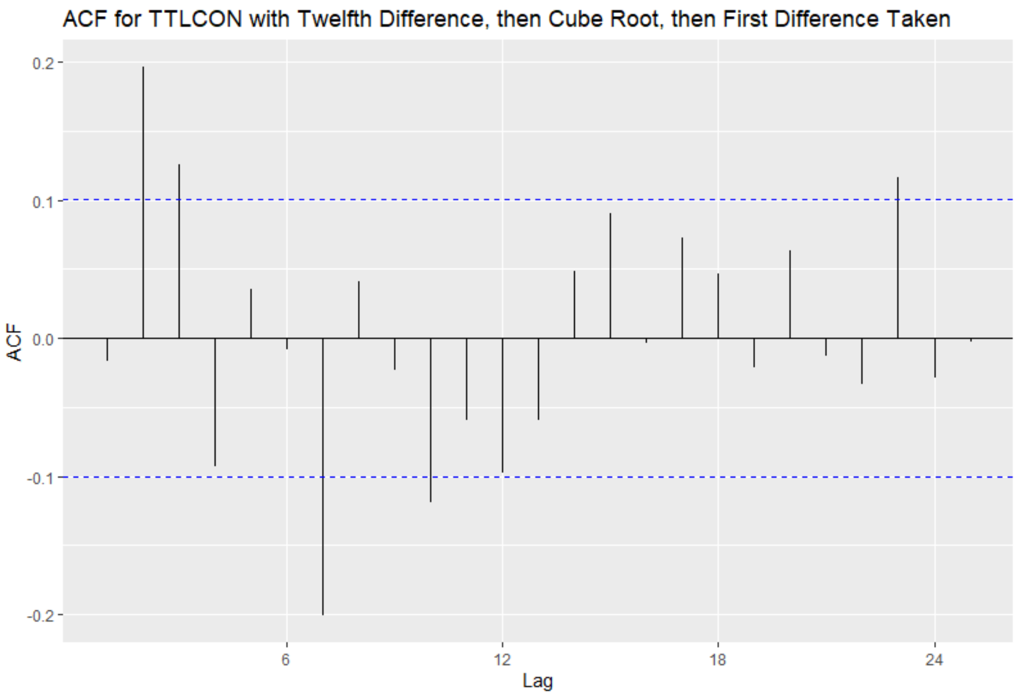

Figure 6: ACF of Total Construction Spending in the United States (TTLCON) with Twelfth Difference Taken, Then Cube Root, then First Differences Taken

The figures above highlight that the construction spending time series is non-stationary. In particular, Figure 4 shows that the data is clearly seasonal, and the ACF (Figure 2) shows the data has clear autocorrelation. We will take advantage of that significant autocorrelation of the original, untransformed data as the input to the ARIMA model. The PACF (Figure 3) suggests using a first difference, and the seasonality suggests using a 12th difference, to seasonally adjust the data. These two transformations did not make the data stationary, as evidenced by Figure 5.

Eventually, using the cube root of the seasonally-adjusted data, then taking a first-difference appeared to get closer to stationarity, as shown by Figure 6. Using a log instead of a cube root made the data very nearly look stationary, but the log function is undefined at 0, which completely ruins any back-transformations. Therefore, a cube root was used in its place. An Augmented Dickey-Fuller test of the transformed data yields a p-value of 0.01, allowing rejection of the null hypothesis that the data is non-stationary. Thus, we can conclude that the data is sufficiently close to stationarity for our analysis.

The transformation of our data is as follows, where x_t is the original data, and y_t is the transformed data, as stated above.

First, we take the natural log of the data, then seasonally adjust (i.e. take the twelfth difference, since the data is monthly,) and then take a first difference. The overall transformation occurs as follows:

$$y_t = (\ln(x_t) - \ln(x_{t-12})) - (\ln(x_{t-1}) - \ln(x_{t-13})) = \ln\left(\frac{x_t \cdot x_{t-13}}{x_{t-12} \cdot x_{t-1}}\right) \text{ for } t \geq 13$$

The back-transformation would be the inverse of the above function, with some reference point needed to fit the differences. Below, the transformed-and-then-back-transformed training and test data are mapped on top of the original data, to highlight that the back-transformation works as intended.

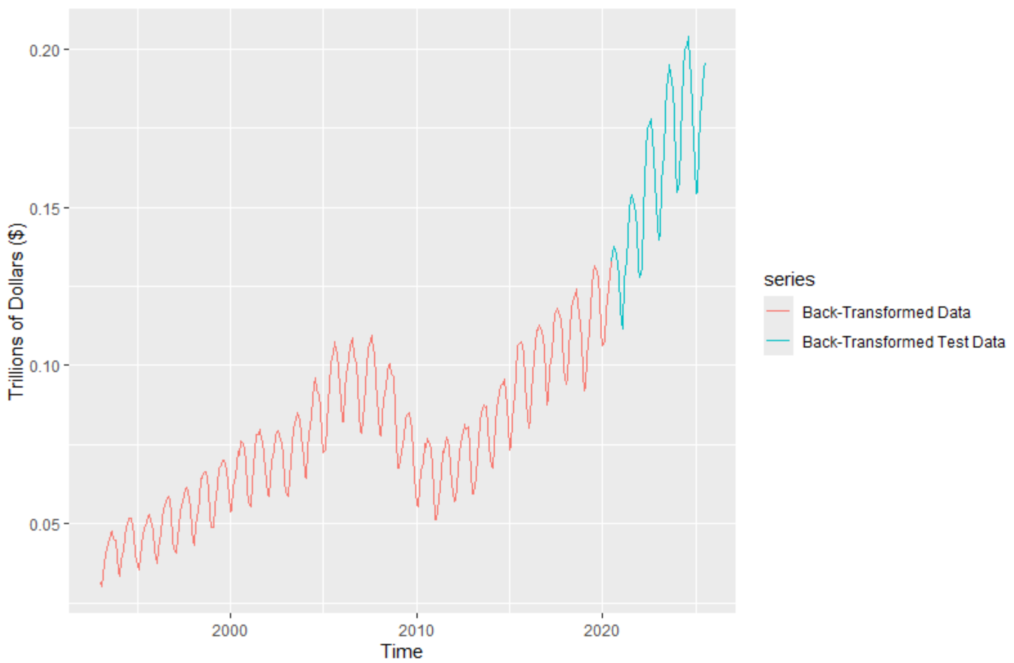

Figure 7: Back-Transformed Training and Test Data, Overlayed on Original Data

As we want to forecast 5 years into the future, we split the data first into a training and test set, where the test set is the last 5 years of spending data. Thus, the training data is all data through June 2020, and the test set is the data from July 2020 to July 2025. The models were fitted on the training data, then forecast for our horizon of 5 years. The ANN model was run 800,000 times in the forecast of the training set over the test period. The below table summarizes the findings with each model and the resultant RMSE on the test data.

Table 2: RMSE and MAE of Forecasts of TTLCON Training Data, Measured by Comparisons to the Test Data

Model	**RMSE**	**MAE**
Naïve	0.0377	0.0312
Snaïve	0.0448	0.0398
ARIMA(4,1,1)(0,1,1)[12]	0.0264	0.0229
ANN 24-12-1	0.0822	0.0725

The ARIMA(4,1,1)(0,1,1)[12] yielded the best predictive performance on the test data, with the lowest RMSE at 0.0264. The Naïve and Snaïve models provided more accurate forecasts of the test set than the ANN, which was the only of the four models to use the transformed data and then back-transform. The ANN, for context, had the highest RMSE, at a value of 0.0745. Figure 8 below plots the forecasts, as well as the test set, for visualization purposes.

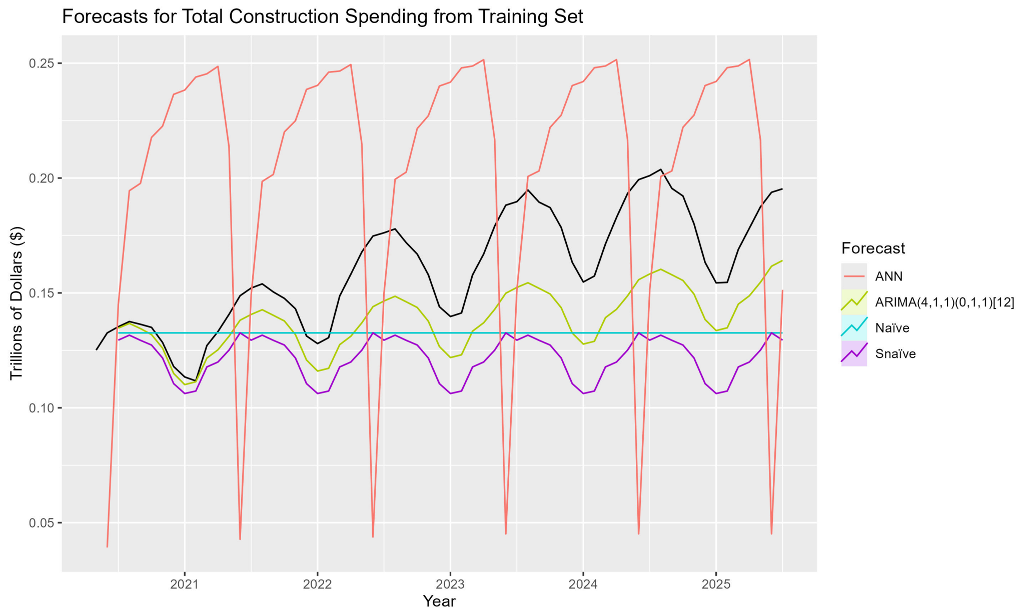

Figure 8: Forecasts from Training Data, with Test Set for Reference

Naïve and Snaïve Models

Figure 9 shows the plot of the Naïve and Snaïve forecasts of the original construction spending data for the next five years. These forecasts are relatively uninteresting, as they will necessarily forecast values at the most recent level (with Snaïve maintaining the seasonal pattern.) Figure 9 shows the Prediction Intervals (PI) for the Naïve and Snaïve forecasts, with the darker shade representing 10% and 90% for lower and upper values, respectively, and the lighter shade representing 2.5% and 97.5% for lower and upper values, respectively.

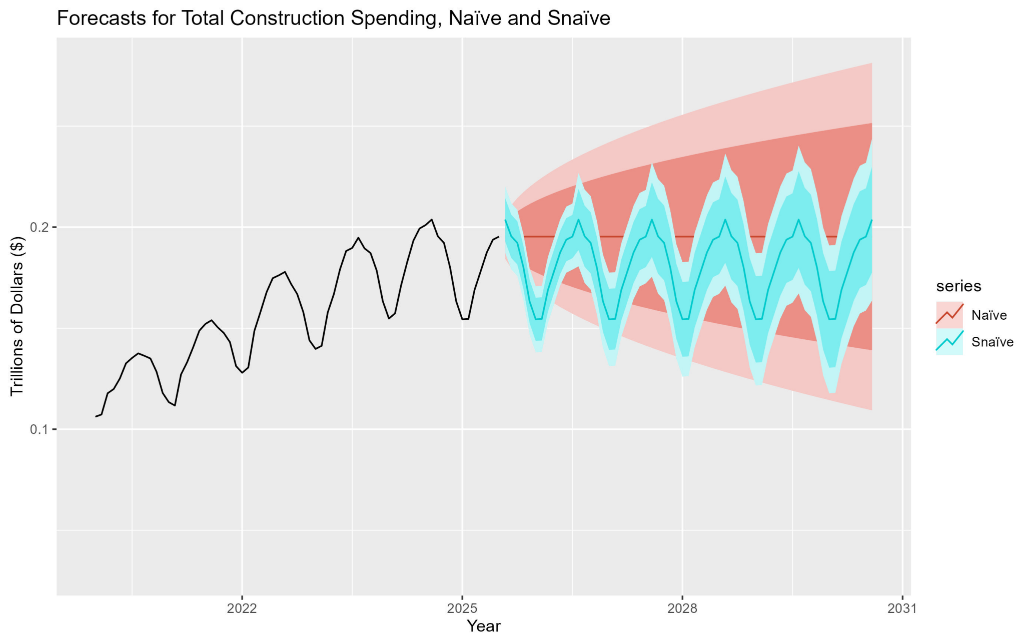

Figure 9: Forecasts from Naïve and Snaïve Models, with Prediction Intervals

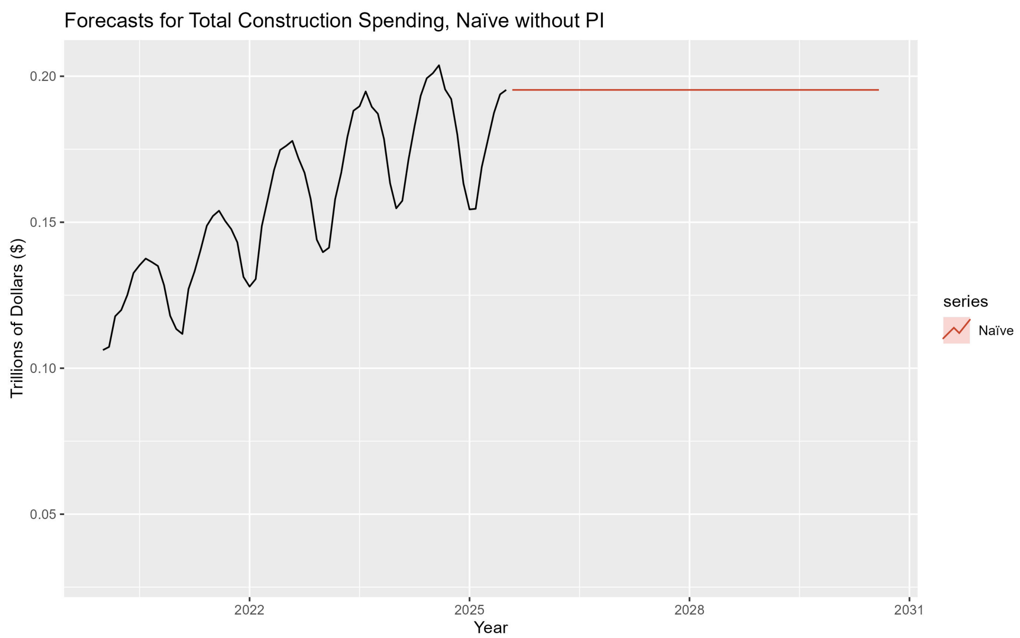

Figure 10: Forecasts from Naïve Model Only, no Prediction Intervals

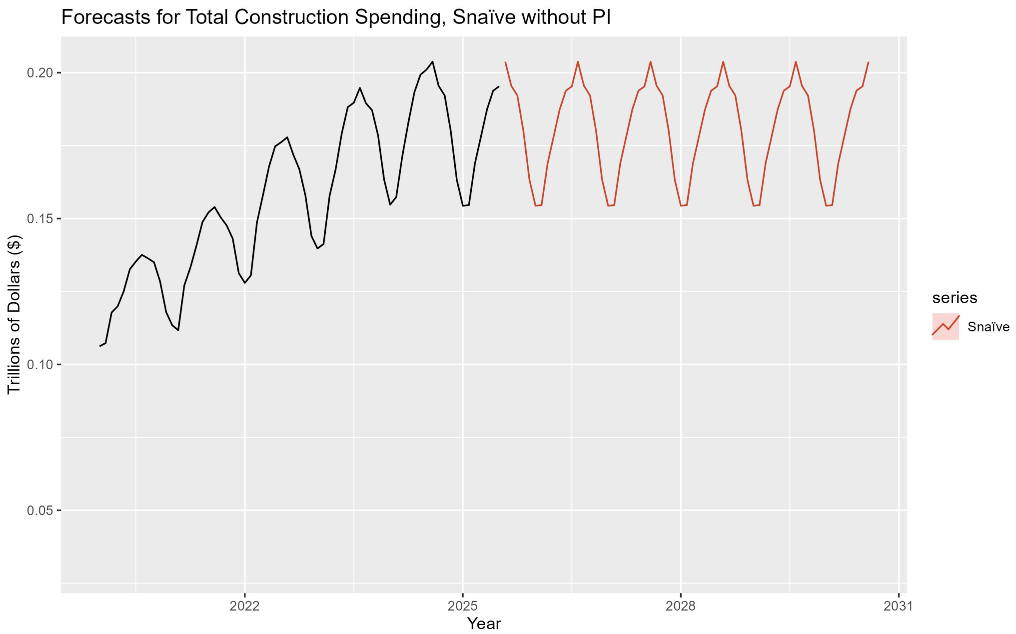

Figure 11: Forecasts from Snaïve Model Only, no Prediction Intervals

ARIMA(4,1,1)(0,1,1)[12] Model

Similarly to the Snaïve model, the ARIMA(4,1,1)(0,1,1)[12] model yielded seasonal oscillations about roughly the mean spending amount from 2024 through 2025, as evidenced by Figure 12 and Figure 13. The darker shaded region shown in Figure 12 is the 20% and 80% interval for lower and upper, respectively, while the lighter shaded region represents the 5% and 95% prediction intervals, respectively. In total, this ARIMA specification suggests that total construction spending in the United States will stay roughly between $200 Billion and $150 Billion over the next 5 years. Figure 12 suggests that, by July 2030, total construction spending in the United States is likely to be between $125 Billion and $275 Billion. Interestingly, ARIMA(4,1,1)(0,1,1)[12] very nearly tracks the Snaïve model, as shown in Figure 15.

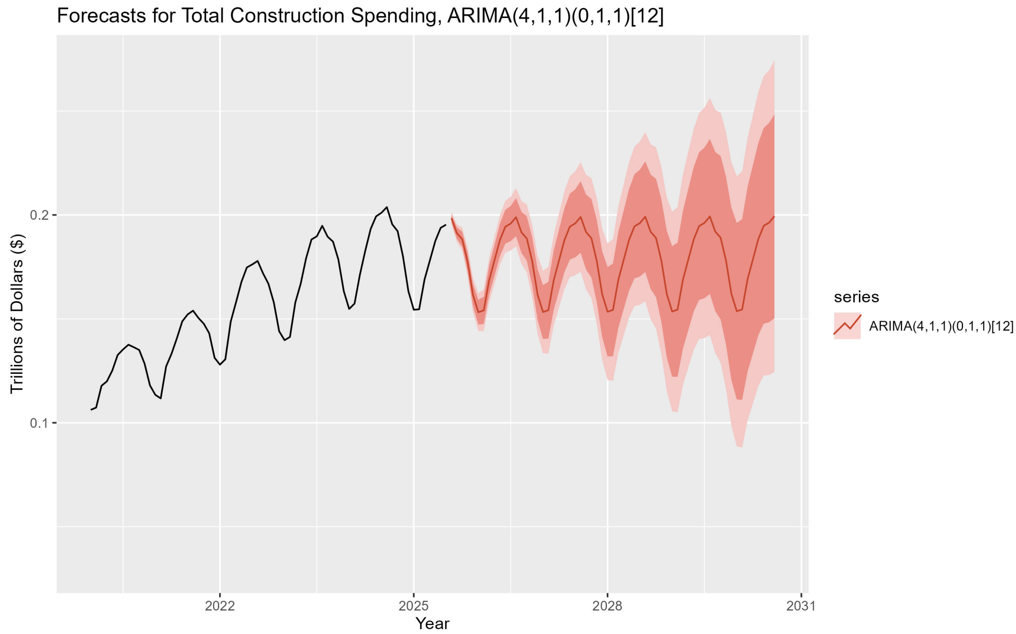

Figure 12: Forecasts from ARIMA(4,1,1)(0,1,1)[12] Model, with Prediction Intervals

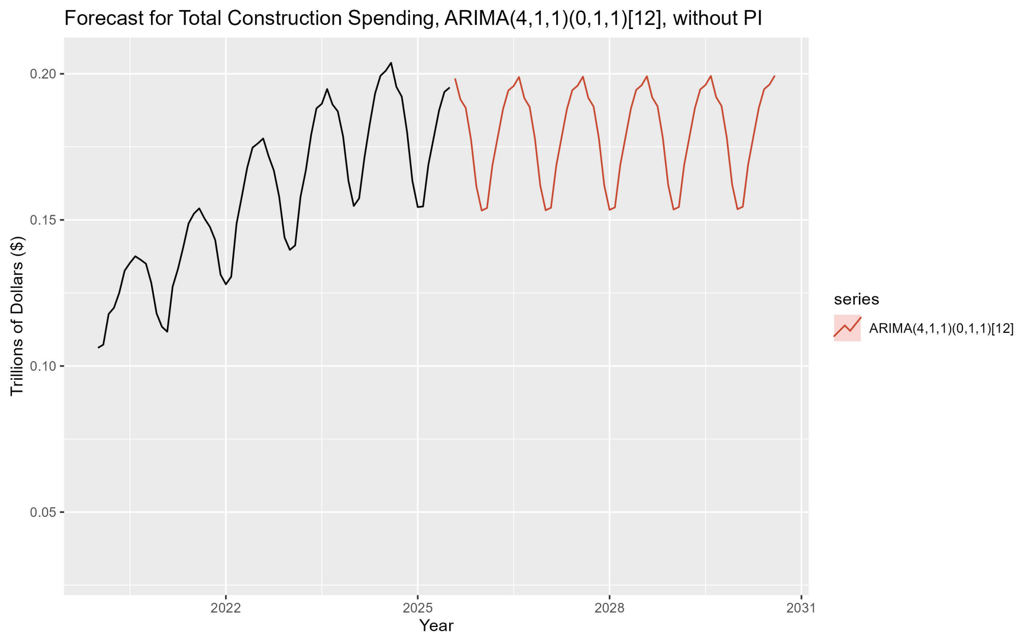

Figure 13: Forecasts from ARIMA(4,1,1)(0,1,1)[12] Model, without Prediction Intervals

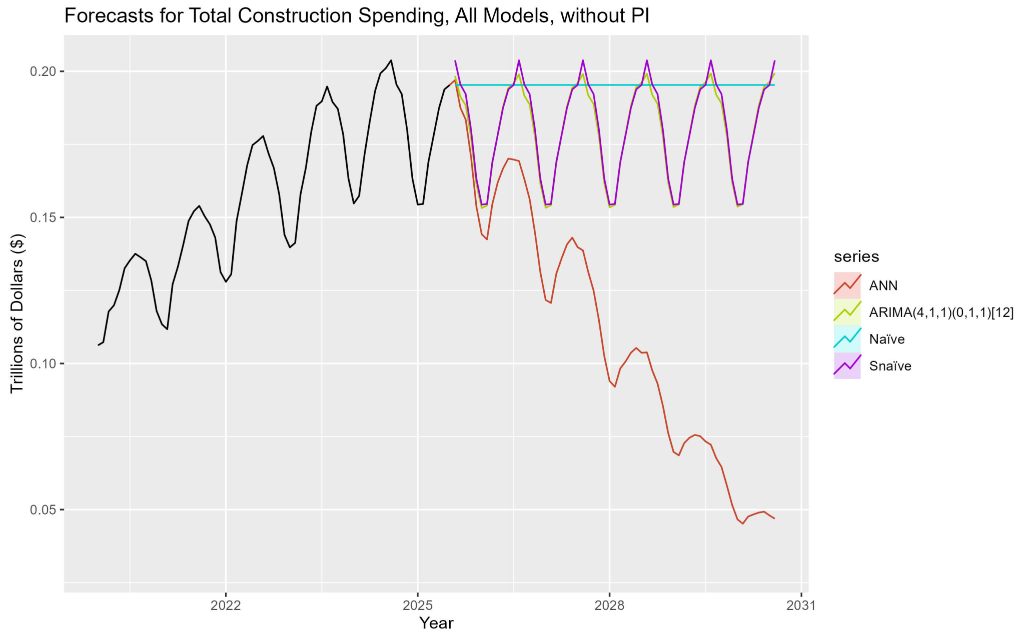

Figure 14: Forecasts from All Models, without Prediction Intervals

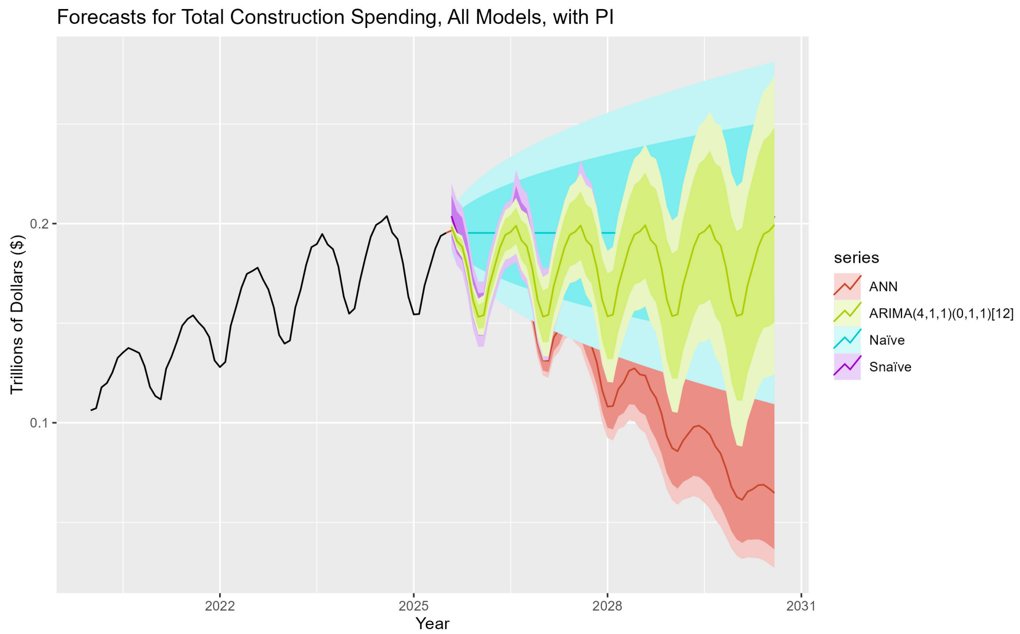

Figure 15: Forecasts from All Models, with Prediction Intervals

Artificial Neural Network Model

Finally, the ANN model was run for 800,000 iterations. The model predicts a significant decline in total domestic construction spending, with a mean value of roughly $65 Billion in mid-2030, compared to the current value of $195 Billion, as shown in Figure 18, with typical seasonal patterns. It appears that the ANN model captures the reduced trend around 2025, as well as the decline around the 2007-2008 recession and is weighing that decline significantly, compared to the significant upward trend from 2010 to 2024. Figure 17 shows the original ANN predictions from the transformed training data overlayed on the transformed data set.

Most strikingly, the prediction intervals in Figure 18 suggest that the Augmented Neural Network model predicts total construction spending will almost certainly decrease over the next five years, in stark contrast to the ARIMA, which suggests total construction spending is likely to remain around its current value.

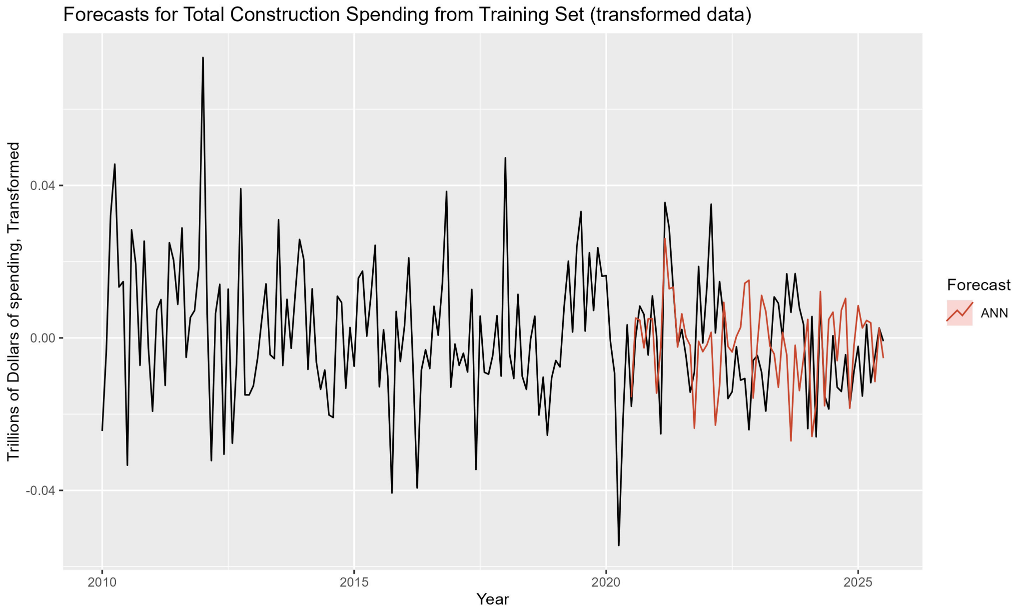

Figure 16: Forecasts from ANN Model on Training Set, pre-Back-Transformation and Transformed Data

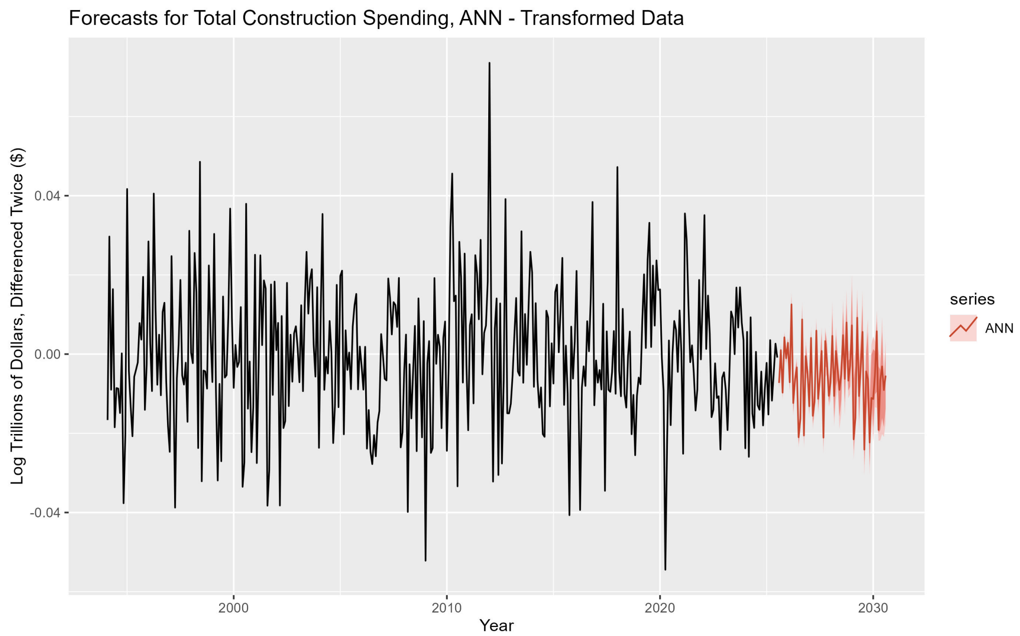

Figure 17: Forecasts from ANN Model, pre-Back-Transformation and Transformed Data

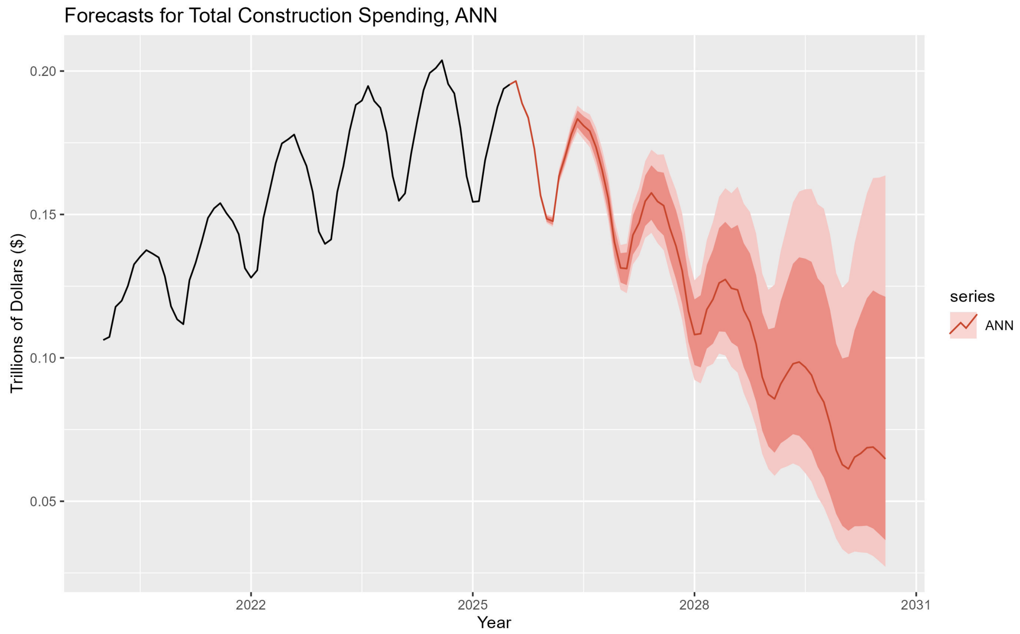

Figure 18: Forecasts from ANN Model, with Prediction Intervals

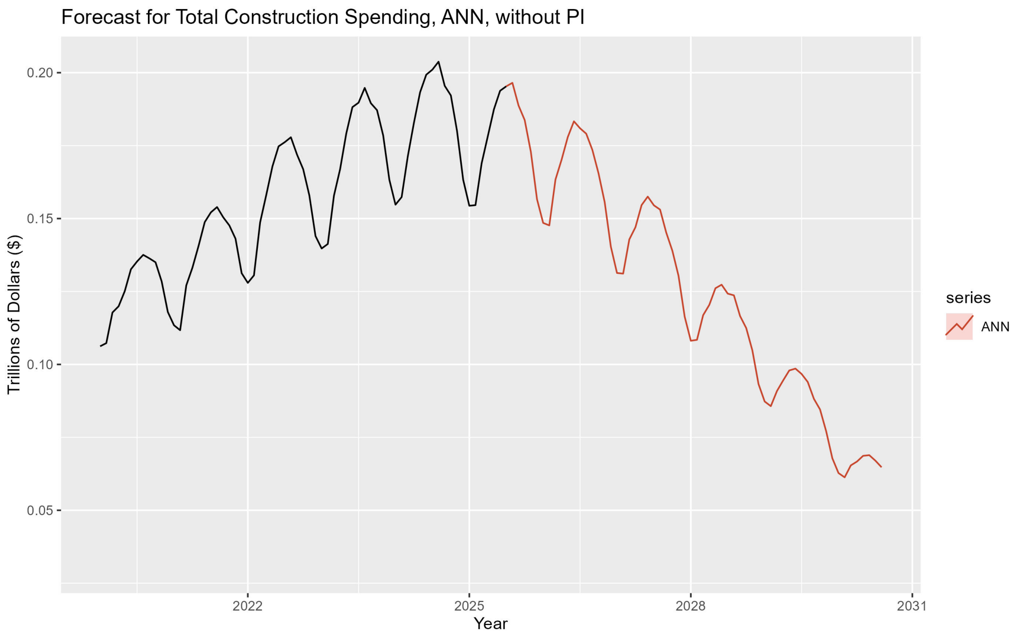

Figure 19: Forecasts from ANN Model, without Prediction Intervals

Conclusion

In conclusion, the ARIMA(4,1,1)(0,1,1)[12] model predicts total construction spending will remain nearly at current levels for the next five years, much in line with the simple Naïve and Snaïve models. The ARIMA(4,1,1)(0,1,1)[12] model predicts a value of $199 Billion, with the value between $124 Billion and $274 Billion with 95% confidence. The Naïve forecast continues spending at the July 2025 level of $195 Billion, with the value between $109 Billion and $281 Billion with 95% confidence in July 2030. The Snaïve forecast predicts total construction spending to be $204 Billion in July 2030, with the value between $164 Billion and $244 Billion with 95% confidence. Thus, the ARIMA prediction is much in line with the Naïve and Snaïve models.

The Artificial Neural Network, however, predicts a significant decrease in total construction spending over the next five years: the ANN predicts spending to be at $74 Billion in July 2030, with that value between $32 Billion and $155 Billion, with 95% confidence. This was an intriguing analytical exercise to try and forecast total construction spending, despite the complex transformations necessary to make total construction spending near-stationary for the purposes of an Artificial Neural Network (Ma et al., 2023; Williams, 1994; Zhang et al., 2024).

References

Alsharef, A., Banerjee, S., Jamil Uddin, S. M., Albert, A., & Jaselskis, E. (2021). Early impacts of the covid-19 pandemic on the united states construction industry. *International Journal of Environmental Research and Public Health*, *18*(4), 1559. https://doi.org/10.3390/ijerph18041559

Ma, X., Tam, V. W. Y., Le, K. N., & Osei-Kyei, R. (2023). A systematic literature review on price forecasting models in construction industry. *International Journal of Construction Management*, *24*(11), 1191–1200. https://doi.org/10.1080/15623599.2023.2241761

Sing, M. C., Edwards, D. J., Liu, H. J., & Love, P. E. (2015). Forecasting private-sector construction works: Var model using economic indicators. *Journal of Construction Engineering and Management*, *141*(11), 04015037.

U.S. Bureau of Economic Analysis. (2025). Value Added by Industry: Construction as a Percentage of GDP [VAPGDPC], retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/VAPGDPC

U.S. Census Bureau. (2025). Total Construction Spending: Total Construction in the United States [TTLCON], retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/TTLCON

Williams, T. P. (1994). Predicting changes in construction cost indexes using neural networks. *Journal of Construction and Engineering Management*, *120*(2), 306–320.

Zhang, X., Wang, Y., Xu, S., Yang, E., & Meng, L. (2024). Forecasting total and type-specific non-residential building construction spending: The case study of the united states and lessons learned. *Buildings*, *14*(5), 1317–1333. https://doi.org/10.3390/buildings14051317