Forecasting Michigan's Automotive Demand

Dev Agrawal
dagraw23@earlham.edu
Earlham College
Richmond, Indiana, U.S.A

Abstract

Michigan's automotive industry accounts for a substantial share of both state and national economic output, making accurate demand forecasting essential for production planning, inventory management, and labor allocation [1]. In this project, a comparison is proposed between the Adaptive Holt-Winters model and a hybrid model that combines linear regression of macroeconomic indicators with a backpropagation neural network (BPNN). Both models will be trained on Michigan's automotive data from 2005 to 2020 and evaluated on data from 2021 to 2025. Model performance will be compared using mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) to determine whether incorporating a linear regression of macroeconomic indicators into a BPNN yields measurable improvements in forecasting accuracy relative to the Adaptive Holt-Winters model for this regional market.
Keywords: Automotive demand forecasting, Adaptive Holt-Winters, Backpropagation Neural Network, Regression, Macroeconomic indicators

1 Introduction

The automobile industry in Michigan plays a key role in the state and national economy, requiring effective demand forecasting for manufacturing, stock control, and workforce planning. Many variables affect the demand for automobiles, including seasonal factors and macroeconomic factors such as GDP, employment figures, and changes in industrial policy. Thus, predicting automotive demand remains a difficult time-series and regression analysis task.

Traditionally, several approaches, ranging from standard exponential smoothing methods such as Holt-Winters forecasting to regression analysis, have been applied because they are easy to understand and perform well under stable conditions. Standard Holt-Winters models capture the underlying trends, seasonalities, and levels of the data and have performed extremely well at predicting car demand under stable market conditions. However, these traditional methods do not effectively capture nonlinear relationships and sudden structural changes driven by economic shocks.

To overcome these limitations, a new set of studies has been dedicated to investigating the potential of machine learning techniques, in particular artificial neural networks (ANNs), to account for nonlinearities in the demand dynamics. Even though ANNs tend to perform better than conventional regression analysis, they usually require large sample sizes and may struggle to model temporal structures without combining them with other methods.

Even though there have been many successful cases, the current literature mostly focuses on models using national or international datasets and pays little to no attention to regional forecasting. In particular, there is a lack of work evaluating hybrid forecasting models that incorporate both time-series structure and macroeconomic indicators in this context.

This project fills this void by comparing an Adaptive Holt-Winters Model, used as a benchmark for modeling seasonal and trend effects, with a hybrid approach that first performs linear regression on macroeconomic data and then uses a backpropagation neural network.

2 Related Work

2.1 Macroeconomic Factors in Automotive Demand

Purchasing an automobile is one of the most important and financially sensitive decisions made in an average household, which is why macroeconomic factors naturally influence the decision. GDP and unemployment rates provide a balanced view of a consumer's spending power, and both measures, along with fuel prices, have been shown to reduce forecasting error when included, compared to time-series approaches alone [2].

Michigan's economy is particularly sensitive to these dynamics, as the automotive industry accounts for a majority of the state's employment and output, further strengthening the influence of macroeconomic indicators in this regional context [1, 3]. The training window chosen for this project includes both the 2008 financial crisis and the COVID-19 shock, both of which are expected to be reflected in the macroeconomic features.

2.2 Regression and Time-Series Approaches

Regression and time-series analysis have been used to predict automobile demand over a long period, providing explanations for trends, seasonality, and economic influences. Regression analysis early on demonstrated that U.S. automobile demand was positively correlated with income, inversely correlated to prices, and negatively influenced by unemployment, and moreover, demonstrated that recessions inevitably pushed back the purchase of automobiles [4].

Regression is effective for identifying long-term drivers, but for short-term forecasting, exponential smoothing usually beats simple regression. An example of this can be seen on Indonesian automobile data, where double exponential smoothing achieved MAPEs of $3.2-3.4\%$ outperforming least-squares regression in that setting [5]. To further reduce the error, future work used genetic algorithms to tune the Holt–Winters parameters by optimizing smoothing coefficients for each unique brand (Toyota, Daihatsu, Suzuki) [6]. These results support the use of an Adaptive Holt–Winters model as a strong, interpretable baseline.

2.3 Adaptive Holt Winters Model

The standard Holt-Winters model is a classic time-series forecasting method that uses trends and seasonality by smoothing three components: level (l_t), trend (b_t), and seasonality (s_t). In its traditional formulation, the model relies on static smoothing parameters (α, β, γ) that remain constant. It is effective in stable environments but not in a volatile Michigan automotive market, especially during financial crises or supply shocks [5].

To address this limitation, recent literature has shifted toward Adaptive Holt-Winters variants, which vary the smoothing coefficients to respond to structural shifts. The components are updated at each step via the following equations:

$$l_t = \alpha_t(y_t - s_{t-m}) + (1-\alpha_t)(l_{t-1} + b_{t-1}) \tag{1}$$

$$b_t = \beta_t(l_t - l_{t-1}) + (1-\beta_t)b_{t-1} \tag{2}$$

$$s_t = \gamma_t(y_t - l_t) + (1-\gamma_t)s_{t-m} \tag{3}$$

where y_t represents observed automobile demand, m denotes the seasonal period, and $(\alpha_t, \beta_t, \gamma_t)$ are adaptive smoothing parameters that change over time. By allowing these parameters to change, the model becomes more responsive to sudden economic disruptions and nonlinear fluctuations in automobile demand [6].

In this project, the Adaptive Holt-Winters model will be implemented using Michigan automotive demand data from 2005 to 2020, enabling it to learn both stable seasonal cycles and major structural disruptions, such as the 2008 financial crisis and the COVID-19 pandemic. Because Michigan's economy is highly dependent on its automotive and related industries, significant variations in automobile demand are expected [3]. The adaptive and dynamic structure of the model will allow it to be used as a baseline for comparison with the hybrid model.

2.4 Neural Network Approaches

While the Adaptive Holt-Winters model performs well with complex data, artificial neural networks (ANNs) have become increasingly popular due to their ability to model the nonlinear relationships in automotive demand. Unlike conventional statistical models, which assume linearity and stationarity, ANNs do not make such assumptions, making them more appropriate for volatile markets [7]. Studies applying ANNs to automotive forecasting have consistently shown that they outperform conventional regression and smoothing approaches when multiple variables interact with demand [8].

Backpropagation Neural Networks (BPNNs), in particular, have become a dominant structure in demand forecasting problems due to their iterative weight-adjustment mechanism, which allows the network to learn from prediction errors across training epochs. Turkish automobile sales demonstrated that BPNNs produced lower forecasting error than traditional regression baselines [8]. In this project, a BPNN will be used as part of a hybrid forecasting framework to capture nonlinear relationships in automotive demand.

3 Data Collection

This project will draw from three data sources, all spanning January 2005 through December 2025. The training set will cover January 2005 through December 2020, and the test set covers January 2021 through December 2025.

3.1 Target Variable

Monthly Michigan Motor Vehicle Production totals will be sourced from the Michigan Economic Indicators report published by the Michigan Senate Fiscal Agency (SFA). The SFA draws its Michigan production figures from the Michigan Department of Treasury, Office of Revenue and Tax Analysis [9]. The data is published as a single PDF report per month, archived on the SFA website, and includes an unstructured table titled "Motor Vehicle Sales and Production Statistics". The target variable is the total number of Michigan Autos and Michigan Trucks produced each month, reported in thousands of units at seasonally adjusted annual rates.

Because the SFA publishes one PDF per month, data collection requires iterating over the archive of monthly reports from January 2005 through December 2025. A Python script using the pdfplumber library will open each PDF, locate the Motor Vehicle Sales and Production Statistics table, extract the Michigan Motor Vehicle Production row for Total, and write each observation to a CSV file indexed by month and year. During preparation, the script will handle formatting inconsistencies across years, such as changes in table layout or column ordering, and verify that no monthly observations are missing before the final CSV is passed into the model pipeline.

3.2 Macroeconomic Variables

Unemployment Rate

The Michigan state unemployment rate will be obtained from the Federal Reserve Bank of St. Louis FRED database, series named MIURN. The data is sourced from the Bureau of Labor Statistics State Employment and Unemployment Release [10]. It is available monthly and is not seasonally adjusted. MIURN covers January 1976 through the present, providing complete coverage of the 2005–2025 study window. The required window will be downloaded as a CSV file containing two columns: date and unemployment rate. During preparation, the series will be checked for missing values, and its date index will be aligned with the target variable to ensure consistency.

Michigan's GDP

The Real Gross Domestic Product for Michigan would be obtained from the Bureau of Economic Analysis Regional Economic Accounts [11]. This time series spans 2005–2025, with 84 quarterly entries during the research timeline, and will be retrieved as a CSV file from the BEA Interactive Data application. Because GDP is only available at quarterly frequency, each quarterly value will be forward-filled across its three corresponding months using Python before being merged with the monthly unemployment and production data, ensuring all three variables share a consistent monthly index prior to model training.

4 Building The Hybrid Model

In this project, a forecasting framework will be constructed that integrates multiple linear regressions with a backpropagation neural network (BPNN). The main reason for this approach is to separate the linear influence of macroeconomic indicators from the more complex nonlinear patterns present in the data. Prior studies have shown that macroeconomic variables such as GDP and unemployment significantly improve forecasting accuracy when incorporated into predictive models [2, 12].

In the first stage of the model, a multiple linear regression is performed using macroeconomic indicators as predictors of automobile demand. Let X_t represent a vector of macroeconomic variables at time t, including GDP, unemployment rate, and potentially fuel prices. The regression model produces an initial estimate of demand that captures the linear relationships between these economic indicators and automobile production.

In the second stage, the regression model residuals are used as inputs to a backpropagation neural network. The BPNN is designed to learn nonlinear dependencies that the regression component cannot capture. Through iterative weight updates using gradient descent, the network minimizes prediction error across training epochs, allowing it to model complex interactions [7, 8]. This framework enables the model to retain the strengths of both statistical and machine learning approaches.

Hybrid models like this are shown to outperform standalone models, particularly in volatile economic environments [13]. In Michigan's automotive market, where demand is influenced by both macroeconomic shocks and demand cycles, this hybrid model is expected to provide a more accurate forecast when compared to the Adaptive Holt-Winters model.

5 Evaluation

The performance of the Hybrid Model and the Adaptive Holt-Winters Model will be evaluated on the test dataset spanning January 2021 to December 2025. Three standard error metrics will be used: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE), as each metric captures a different dimension of forecasting error. A study on forecasting methods on vehicle sales data uses the same three metrics for direct comparison of different models [13].
MAE provides the average magnitude of error in the original units. It is computed by averaging the absolute differences between forecasted and actual values, treating all errors equally, regardless of direction. MAE is calculated by:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

where y_i is the actual value and $\hat{y}_i$ is the forecasted value. MAPE provides the average absolute percentage difference between forecasted and actual values. It is computed by averaging the absolute differences between forecasted and actual data, divided by the actual values, making it highly effective for comparing accuracy across varying volumes. MAPE is calculated by:

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{5}$$

where y_i is the actual value and $\hat{y}_i$ is the forecasted value.
RMSE is a standard metric for measuring the difference between predicted and observed values. It is computed by squaring the differences between forecasted and actual values, averaging those squares, and then taking the square root of the result. RMSE is calculated by:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

where y_i is the actual value and $\hat{y}_i$ is the forecasted value.

6 Visualization

Note: *The figures presented in this section are examples drawn from the internet. Final figures will be generated from Michigan automotive data upon model completion and added to the technical report.*
In order to represent the performance of the Hybrid Model and the Adaptive Holt-Winters Model, the following four graphs are generated.

Feature Importance Plot: This helps identify which macroeconomic factors matter most in the model, allowing us to focus on key predictors.

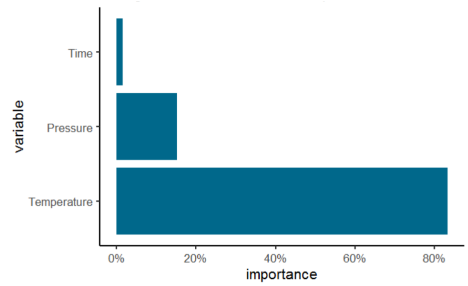

Figure 1: Feature Importance Plot

Time-Series Comparison: This provides a direct visual comparison between observed production and predicted values to assess how well the models track seasonal peaks and economic shocks.

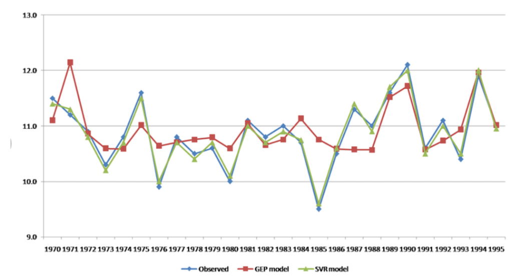

Figure 2: Actual vs. Predicted Production

Residual Analysis: A residual plot displays the differences between actual and predicted values to determine if the models are consistently over- or under-fitting.

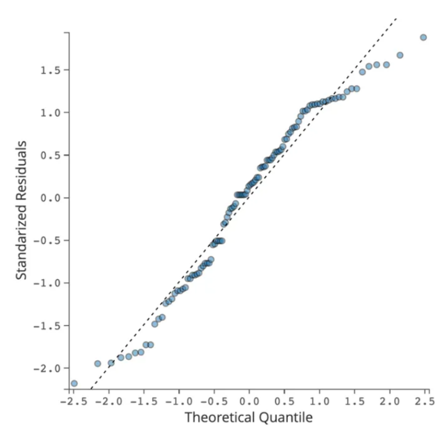

Figure 3: Residual Plot

Performance Metric Comparison: This comparison histogram summarizes the performance of both the Hybrid Model and the Adaptive Holt-Winters model across three key error metrics: MAE, MAPE, and RMSE. It allows for assessing which model maintains lower variance across the testing period.

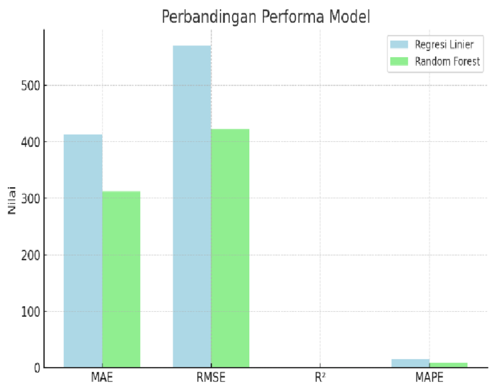

Figure 4: Performance Metric Comparison Histogram

7 Results

A model will be considered successful if it achieves a MAPE of below 10% and a failure if the MAPE is above 15% on the test dataset, which is consistent with acceptable performance thresholds in the automotive demand forecasting literature [7][8]. MAE will be used to report the average forecast error in the original production units. RMSE will be used as a secondary metric to assess which model has larger errors during high-variance months.

If the hybrid model outperforms the Holt-Winters baseline across all three metrics, this will be taken as evidence that a complex hybrid model outperforms a traditional time-series model. If performance is mixed, the approach taken in comparable studies will be followed, where no single metric is sufficient to draw a definitive conclusion [13][8].

8 Timeline

This project will be completed over a 15-week semester. The primary deliverables will be a technical report, a poster, and a project demonstration video. The schedule for primary deliverables is as follows:

Week 1: Project setup and initiation
Week 2: Technical Report (v0)
Week 3: Data Architecture Diagram (v0)
Week 4: Graphical Abstract (v0)
Week 5: Data Arch./Abstract (v1)
Week 6: Technical Report (v1)
Week 7: Data Arch./Abstract (v2) and Poster (v0)
Week 8: Project Demo Video (v0)
Week 9: Technical Report (v2)
Week 10: Poster (v1)
Week 11: Data Arch. & Graphical Abstract (v3)
Week 12: Project Demo Video (v1)
Week 13: Technical Report (v3)
Week 14: Poster & Demo Video (v2)
Week 15: Final version of all deliverables

As for the actual research to be conducted, all data will be preprocessed, extracted, and prepared in the first 6 weeks of the semester. The next 6 weeks will be dedicated to building both models. The final 3 weeks will be focused on completing the analysis, comparing model performance metrics, producing the visuals, and finalizing all presentation materials.

9 Works Cited

[1] Center for Automotive Research. (2013, April). *Contribution of the automotive industry to the economies of all fifty states and the united states* (tech. rep.). Alliance of Automobile Manufacturers.

[2] Sharma, R., & Sinha, A. K. (2012). Sales forecast of an automobile industry. *International Journal of Computer Applications*, *53*(12).

[3] Burton, J. T., et al. (2025, May). *The michigan economic outlook for 2025–2027: Executive summary* (tech. rep.) (Published May 16, 2025). Research Seminar in Quantitative Economics, University of Michigan.

[4] Kulash, D. J. (1971). *Forecasting long-run automobile demand* (tech. rep.). Jack Faucett Associates.

[5] Maulana, F. T., & Kamisutara, M. (2021). Demand forecasting of the automobile sales using least square, single exponential smoothing and double exponential smoothing. *Petra International Journal of Business Studies*, *4*(2), 122–130. https://doi.org/10.9744/IJBS.4.2.122-130

[6] Hani'ah et al. (2022). Genetic algorithms for holt winter exponential smoothing parameter optimization in indonesian car sales forecasting. *Proceedings of the International Conference on Mathematics, Geometry, Statistics, and Their Applications (ICMGSA 2021)*.

[7] Subramanian, K., Othman, M. B., Sokkalingam, R., & Thangarasu, G. (2020). A new approach for forecast sales growth in automobile industry. *International Journal of Scientific and Technology Research*, *9*(1), 2872–2875.

[8] Pekpazar, A., Kaya, G., & Cebi, F. (2019). Forecasting automobile sales in turkey with artificial neural networks. *International Journal of Business Analytics*, *6*. https://doi.org/10.4018/IJBAN.2019100104

[9] Michigan Senate Fiscal Agency. (2026). Michigan economic indicators [Monthly publication, accessed May 2026]. https://sfa.senate.michigan.gov/publications/econind/mei_mostrecent.pdf

[10] U.S. Bureau of Labor Statistics. (2026). Unemployment rate in michigan (MIURN) [Accessed May 2026]. https://fred.stlouisfed.org/series/MIURN

[11] U.S. Bureau of Economic Analysis. (2026). State quarterly gross domestic product summary, table SQGDP1: Michigan [Accessed May 2026]. https://apps.bea.gov/itable/?ReqID=70&step=1

[12] Homolka, L., Ngo, V. M., Pavelková, D., Le, B. T., & Dehning, B. (2020). Short- and medium-term car registration forecasting based on selected macro and socio-economic indicators in european countries [Article in volume 80(C)]. *Research in Transportation Economics*, *80*.

[13] Fleurke, S. (2017). Forecasting automobile sales using an ensemble of methods [Compares Holt-Winters, neural networks, GLM, ARIMA, and ensemble forecasting on vehicle-sales data using MAE, MAPE, RMSE, and MDA].