Cureus Journal of Business and Economics
https://doi.org/10.7759/s44404-025-00040-6

Original Article

Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods

Bhubaneswor Dhakal [1, 2]

1. *Statistics and Math, University of Canterbury, Christchurch, NZL*
2. *Support for Development New Zealand, a Charitable Organization, Christchurch, NZL*

Received: August 24, 2025 | Review began: September 03, 2025 | Review ended: February 23, 2026 | Published: March 24, 2026

Abstract

This study examined the dynamics of private car sales in New Zealand, a fully import-dependent and tariff-free economy with one of the highest car ownership rates globally. Using monthly data on small vehicle sales (cars, sport utility vehicles, and passenger vans) from 1998 to 2024 (324 observations) obtained from the Motor Industry Association, the study investigated market behavior and external shocks through both a deep learning approach, Long Short-Term Memory (LSTM), and an econometric framework, Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH). The study further examined methodological challenges associated with validating the LSTM model, with particular emphasis on the application of time-series k-fold cross-validation (TSCV) to data characterized by a long-term upward trend. Specifically, it assessed potential distributional inconsistencies between the train and test datasets of individual folds and showed reliability issues in k-fold cross-validation outcomes. Moreover, by construction, the TSCV framework systematically excludes important data segments from model estimation. This exclusion limits the ability of the validation procedure to provide a fully robust assessment of a model ultimately trained on the complete dataset. This study has also explained an underlying cause of negative out-of-sample R-square in k-fold cross-validation tests using a methodological framework and empirical evidence. Forecast performance comparisons between the EGARCH and LSTM models revealed statistically significant differences in their predictions. The study also identified statistically significant differences between models trained on split datasets and those trained on the full dataset. Results suggested that models trained on the full dataset provide more robust predictions and perform comparably to EGARCH, particularly at the beginning of the forecast horizon. The LSTM model effectively captured recent upward drifts and nonlinear patterns, generating point forecasts of expected sales and reflecting long-term trends. In contrast, the EGARCH model captured volatility persistence and mean reversion, offering forecasts that quantify uncertainty and risk. An integrated LSTM-EGARCH framework successfully captures both nonlinear and stochastic dynamics, producing forecasts that are more informative for policy formulation and managerial decision-making. From a policy perspective, the findings indicated that car sales in New Zealand have a long-term upward trend, potentially supported in part by tariff relaxation. However, sales display pronounced volatility clustering, strong mean reversion, and a high likelihood of extreme fluctuations in response to external shocks. Full import dependency is argued to be one contributing factor to this volatility. Sales declined sharply during the COVID-19 pandemic but rebounded rapidly, possibly reflecting strong institutional foundations. Forecasts for 2025-2027 suggest that sales growth is likely to continue, albeit at a marginal pace, approaching a potential peak level. In terms of overall lessons for other tariff-free, import-dependent economies, tariff relaxation, probably supported by favorable measures such as roads and effective institutional backups, can sustain car sales growth and support recovery, even amid market volatility and external shocks.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Categories: Trade and economic development, Transportation Management, Innovation Economics

Keywords: hyperparameter-tuning, forecast, import-dependent, data-splitting-issue, tarif-free, economic-policy, car-sale-business, business-volatility, out-of-sample-negative-r-square, cross-fold-validation

JEL Classifications: F13, L62, R40

Introduction

Access to personal vehicles significantly influences human mobility, economic activity, settlement patterns, industrial geography, recreational behavior, and broader lifestyle trends within a country (Mouratidis, 2025), (Ismael and Duleba, 2022), and (Li et al., 2024). In this sense, vehicle ownership can be rightly described as both an engine of economic development and a contributor to individual well-being (Mouratidis, 2025). The diverse uses of vehicles for work, leisure, and essential services distinguish them from many other goods, services, and even financial commodities in terms of their economic and social relevance (Ismael and Duleba, 2022). Despite this significance, some countries do not produce vehicles domestically, either due to resource constraints or a lack of comparative advantage in manufacturing. New Zealand exemplifies such a case: it relies entirely on imports to meet domestic demand for private vehicles. The elimination of import duties in 1998 effectively ended local vehicle production, shifting the country to a fully import-dependent car market (Pawson, 2014).

Remarkably, despite its reliance on imports, New Zealand has one of the highest rates of vehicle ownership in the world. According to the 2023 census, 91.4% of households own at least one motor vehicle, with a vehicle holding rate of 936 vehicles per 1,000 people (Stats NZ, 2024). Both car imports and ownership have increased substantially since the removal of import duties (Pawson, 2014). Vehicle sales have surged, particularly in the post-liberalization period and over the past decade. Private vehicles are of particular importance in New Zealand, where many businesses operate in localities with no or infrequent public transport services. Some of these businesses also run during nighttime hours. As a result, private vehicle ownership is often an essential requirement even for unskilled labor jobs.

Greater vehicle availability can enhance personal well-being and improve business productivity but also contribute to road congestion and increased greenhouse gas (GHG) emissions. GHG emissions from transport, mainly from car-related vehicles, have increased by 62.1% since 1990 (Stats NZ, 2022). Alongside emissions, managing the growing waste from end-of-life cars and their accessories has become another pressing challenge. Road congestion and parking shortages are also rising, making it harder for local governments to achieve their goal of creating cleaner and smarter cities. In densely populated regions, these vehicle management problems have forced people to spend a significant portion of their lives stuck on the road while commuting to and from workplaces or service centers (Stats NZ, 2024). These problems exacerbate critical policy challenges. Governments therefore require robust information on the future scenario of vehicle import, demand, and sales dynamics to address these challenges and maximize overall social welfare. Such foresight helps in identifying when significant increases in vehicle use are likely to occur, and allows for timely interventions or proactive actions, such as the expansion of public transport, demand management measures, or low-emission zones - before congestion and emissions worsen.

Studies on car sales phenomena over time can provide valuable insights and also support businesses in offering better customer services. The outputs can be useful mainly from two aspects. First, they explain the dynamics of car sales, which provide lessons on consumer behavior, market shifts, and the effects of past policies. Second, they provide a basis for forecasting future sales, which highlight potential future challenges or opportunities and help policymakers plan ahead to address them on time. As the saying goes, statistics provide probabilistic estimates and directional evidence, rather than definitive answers, particularly in policy decision-making contexts. However, rigorously analyzed data can still reveal meaningful patterns and trends in business phenomena (Citroen, 2011). Understanding car sales dynamics under zero-tariff conditions can also provide valuable lessons for other countries that are considering liberalizing vehicle imports to support economic growth and improve citizen well-being.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Considering its policy importance, a significant number of previous studies have forecasted car sales using different methods. For instance, Riker (Riker, 2024) used simulation techniques to evaluate the potential impacts of tariffs on car production and imports in the United States. Irfan (Irfan, 2020) applied ordinary least squares regression to explore the relationship between oil prices and car imports from 2000 to 2019. Shi et al. examined stakeholder views on used car imports in New Zealand (Shi et al., 2022). Ayetor et al. (Ayetor et al., 2021) provided descriptive insights into vehicle import trends in Africa, while Kisiała and Kudłak used multivariate analysis to assess the determinants of second-hand car imports into Poland (Kisiała and Kudłak, 2024). A number of promotional reports publish annual sales of cars worldwide (Fastener-World, 2022), (Fastener-World, 2024), and (Hall, 2025). The review of the extant literature shows that the forecast of private car sales has not been studied in sufficient depth to provide the knowledge needed to assist policy decision-making in fully liberalized markets. The key purpose of this study is to contribute to addressing this knowledge gap.

Prior research, such as Kim and Won (Kim and Won, 2018), has suggested the integration of econometric and machine learning techniques to get more precise information. The econometrics analysis methods have merits in providing information on disaggregating the effects of different phenomena on business trends, as well as producing forecasts. However, the time series data of many years carry many complexities, such as normality and stationarity, to fulfill the assumptions for econometrics analysis methods and demand high analysis costs (Iskhakov et al., 2020). Generating information for strategic decision-making requires long-term data, which often contains complex nonlinear patterns and long-range dependencies. Such data can also have hidden issues, such as structural breaks, which are more likely, especially in long-term car sales records. Machine learning methods are less sensitive to the data distribution problems and can provide predictive information (Athey and Imbens, 2019). If the business interest is only in forecast information with low cost, machine learning methods such as the Long Short-Term Memory (LSTM) neural network would be a better option (Mullainathan and Spiess, 2017). However, they generally do not yield interpretable parameters that explain underlying phenomena, as econometric methods do. In addition, the analytical approach that has proven robust enough to analyze sales series data of other sectors may not be replicable in this sector because the car market can behave differently due to its inherent intrinsic attributes. To the best of our knowledge, no recent studies have examined car sales forecasting in the New Zealand context. Furthermore, the application of LSTM and Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) models to car sales forecasting remains limited in the existing literature. Accordingly, this study evaluates the comparative and integrative forecasting performance of econometric and machine learning approaches in modelling car sales.

From a statistical perspective, the application of machine learning methods presents inherent methodological challenges related to model training and performance evaluation, particularly in evolving and complex historical time series contexts. These methods typically involve splitting the data into training and testing (or validation) sets, a practice used to evaluate the model's performance in predicting unseen data (An et al., 2021). This practice is especially critical in time series applications (Pimentel et al., 2025) (Kim and Won, 2018). Sales series dynamics are outcomes of the serial buildup of many latent factors. Randomly splitting time series data into training and test sets can result in a loss of important information, which would otherwise be preserved if the data were analyzed as a continuous series. To maintain temporal structure, time series studies typically apply chronological splits, where recent observations are reserved for testing and earlier data are used for training. Statisticians are concerned that such data-splitting approaches may reduce the explanatory power of a model by discarding part of the information. Since machine learning methods are predictive in nature, the issue of data splitting could have significant impacts on forecasting. This forecasting challenge based on split-data models has been poorly explained in the literature.

Some studies have suggested using k-fold validation to evaluate models even for time series data (Bates et al., 2024) (Bergmeir et al., 2018). The method is considered a gold standard benchmark. However, due to the systemic structure of k-fold validation and the distributional nature of time series data, the scientific appropriateness of this evaluation method is questionable. Theoretically, the distributional properties of training and testing datasets should be identical to ensure robust prediction (An et al., 2021) (Bates et al., 2024). Time series data with persistent upward trends over time may not have consistent distributional properties between training and test sets. The potential statistical weakness of this method in complex time series has not been well investigated. Moreover, many studies have reported negative out-of-sample

(OOS) R-square results in k-fold cross-validation tests. Negative OOS R-square is often interpreted as an indicator of poor predictive power (Bin-Inqiad et al., 2026), (Poldrack et al., 2020), (Konovalov et al., 2008), (Penido et al., 2022), and (Moles et al., 2009). However, the literature provides little empirical explanation for why these negative results occur or whether the model is truly poor. Practical challenges associated with training-test distributional shifts and the reliability of out-of-sample evaluation remain insufficiently examined.

Previous studies have proposed hybrid modeling frameworks that combine LSTM with GARCH-type specifications to better capture nonlinear dynamics and volatility structures in financial time series (Pimentel et al., 2025), (Kakade et al., 2022), and (Kim and Won, 2018). These studies demonstrate the effectiveness of hybrid models in applications such as market return volatility, stock price index volatility, and option pricing. Their evaluations primarily focus on improvements in in-sample fit statistics and comparative error metrics, highlighting the complementary strengths of statistical and deep learning approaches. However, few studies have examined the robustness of forecasting performance between these methods.

The above literature highlights an important knowledge gap in time series studies that warrants further investigation. The primary objective of this study is to forecast private-use light vehicle sales in a tariff-free economy, while the secondary objective is to examine the associated methodological issues. This study addresses this gap by employing a volatility-sensitive EGARCH-ARFIMA model alongside an LSTM model to forecast monthly private vehicle sales under policy liberalization conditions. In doing so, it contributes both to the applied forecasting literature and to the methodological understanding of car import dynamics under trade liberalization.

The purpose of this study is to address the knowledge gaps identified in the literature. To achieve this, the study focuses on the following specific objectives:

a. Show the dynamics of car sales in an open market economy context in New Zealand.

b. Explain the phenomena of volatility and shocks affecting the vehicle sales trajectory.

c. Forecast New Zealand's private car import dynamics for 2025 to 2027.

d. Compare the forecasting performance between econometric (EGARCH) and machine learning (LSTM) methods.

e. Evaluate the forecasting performance of machine learning models using split and full datasets.

f. Assess the methodological limitations of cross-fold validation and split-data forecasting approaches in modelling time-series data with persistent upward trends over time.

This study focused on EGARCH and machine learning models, as both are popular predictive models. The main methodological objective of this study is to compare full-data and split-data LSTM models. This study used an ordinary EGARCH model as a reference for the LSTM model and placed its nested model and other issues aside. The approaches are followed to avoid complexity in assessing the models.

Research Method

This study is based on time series-based analysis. Such an approach is widely recognized in business problem analysis and forecasting because many business variables evolve over time in systematic ways (Granger, 2014) (Hyndman and Athanasopoulos, 2018). By examining historical records of a single variable, such as sales, demand, or prices, analysts use past behavior to anticipate future outcomes and support decision-making. Observations that occur close together in time are typically related, meaning recent values contain useful information about near-term movements (Granger, 2014). Business data also often exhibit persistent patterns, including long-term trends and recurring seasonal cycles, which provide a reliable basis for projection when they remain stable. Time series methods further help distinguish meaningful signals from short-term random fluctuations, leading to clearer insights and more reliable forecasts. This approach is particularly valuable in practical business settings where external explanatory variables are unavailable, difficult to measure, or too complex to evaluate (Hyndman and Athanasopoulos, 2018).

Data

This study used light private vehicle sale count data that the New Zealand Transport Authority has compiled extensively, dating back to 1976. The data are available in an open database. The data account for both new and used cars and exclude motorcycles and other passenger vehicles weighing more than 3,500 kg. The study explicitly includes vehicles with private uses, which refer to cars, SUVs, and light passenger vans. This study used sale data of both used and new cars. Clean datasets, particularly those on used and new car imports, can be accessed through the websites of Stats New Zealand and the New Zealand Automobile Association (MIA, 2025) (Stats NZ, 2025). These monthly sales figures are somewhat different from monthly import data given by Stats NZ. The data are given in the appendices.

The government's decision to remove the 100% car import duty in 1998 effectively brought an end to nearly all car manufacturing industries in the country. Including data from 1975 did not meet the econometric assumption of "Nyblon stability" in the model, which is a requirement for employing the EGARCH analytical method. A significant structural break appeared in 1990 in the data distribution. In addition, the tariff policy of 1998 was another remarkable event. Accordingly, this study utilizes monthly import data from 1998 to 2024. It also utilizes monthly sale data from 1998 to 2024, which accounts for 27 years.

Analytical methods

Econometrics Model

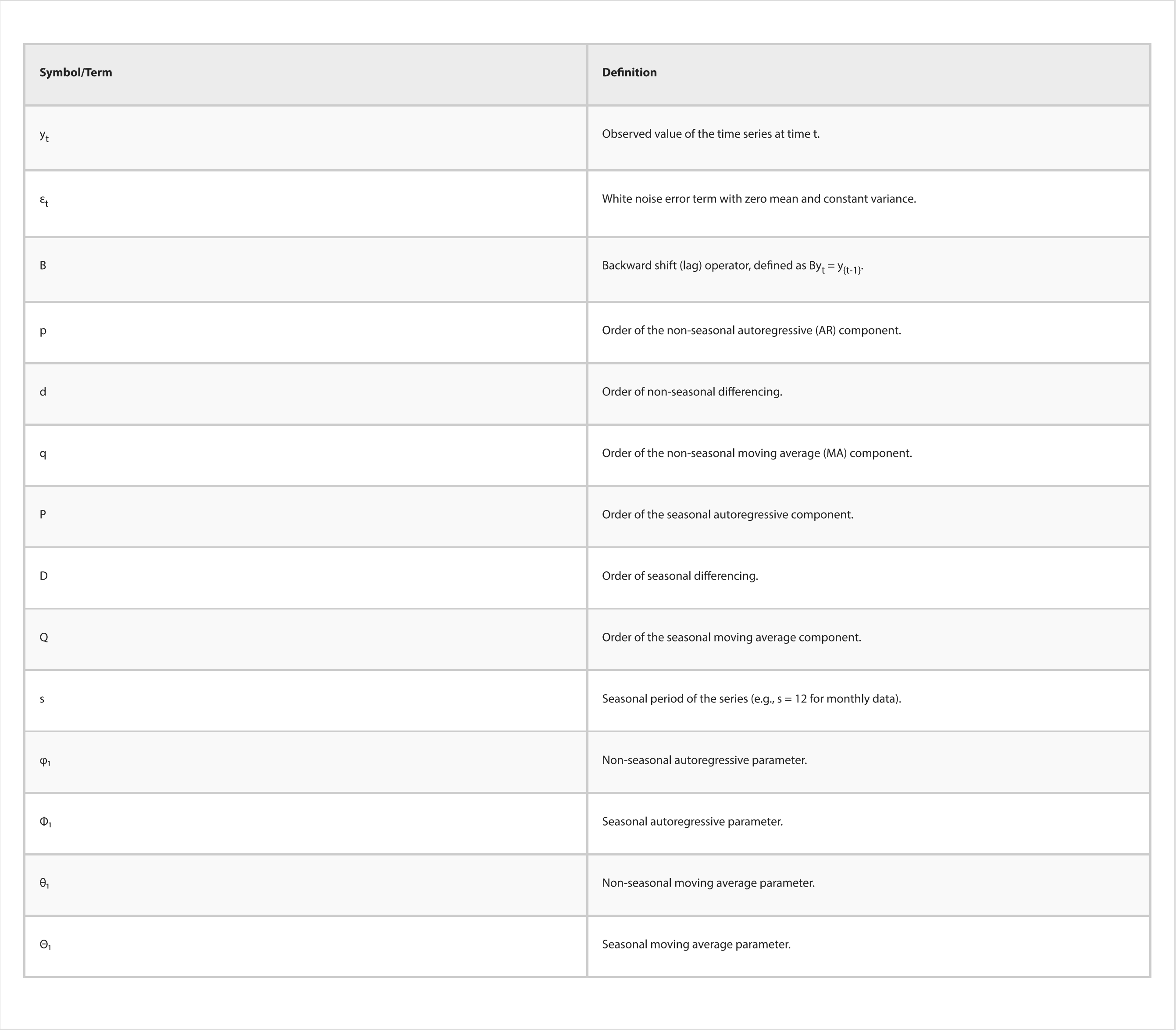
This study examined car import data using various GARCH models built on a Seasonal Autoregressive Integrated Moving Average (SARIMA) framework. While SARIMA captures recurring seasonal patterns through seasonal autoregressive, differencing, and moving-average terms, GARCH models account for time-varying volatility by modeling conditional variance as a function of past shocks and volatility. The Exponential GARCH (EGARCH) model was employed to address expected asymmetric and nonlinear volatility dynamics, which standard symmetric GARCH models cannot capture. EGARCH accommodates leverage effects, allowing positive and negative shocks of equal magnitude to have different impacts on future volatility, ensures strictly positive variance forecasts by modeling the log-variance without parameter constraints, and effectively captures volatility clustering and nonlinear persistence, features commonly observed in macroeconomic and financial time series. This model is explained hereafter based on a reference to a popular book (Asteriou and Hall, 2021).

This GARCH method accounts for both non-seasonal (p, d, q) and seasonal components in the autoregressive integrated moving average model. The terms in the parentheses are the order of the autoregressive component, integrated component, and moving average component, respectively. The s refers to the seasonal period.

$$(1-\phi_1 B)(1-\Phi_1 B^s)(1-B)(1-B^s)\,y_t = (1+\theta_1 B)(1+\Theta_1 B^s)\,\varepsilon_t \tag{1}$$

The equation consists of several terms: $(1-\phi_1 B)$ - Autoregressive (AR) component, $(1-\Phi_1 B^s)$- seasonal autoregressive component, $(1-B)$ - non-seasonal difference, $(1-B^s)$ - seasonal differencing operator, $(1+\theta_1 B)$ - moving average (MA) and $(1+\Theta_1 B^s)$- seasonal moving average. It also has the observed time series at time t, y_t, and a white noise error term, ε_t. In the equation, B is the backward shift (lag) operator, defined as $By_t = y_{t-1}$. The multiplier of the B are parameters. The SARIMA model does not account for volatility in the time series. Detailed definitions are given in Table 1 below.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6



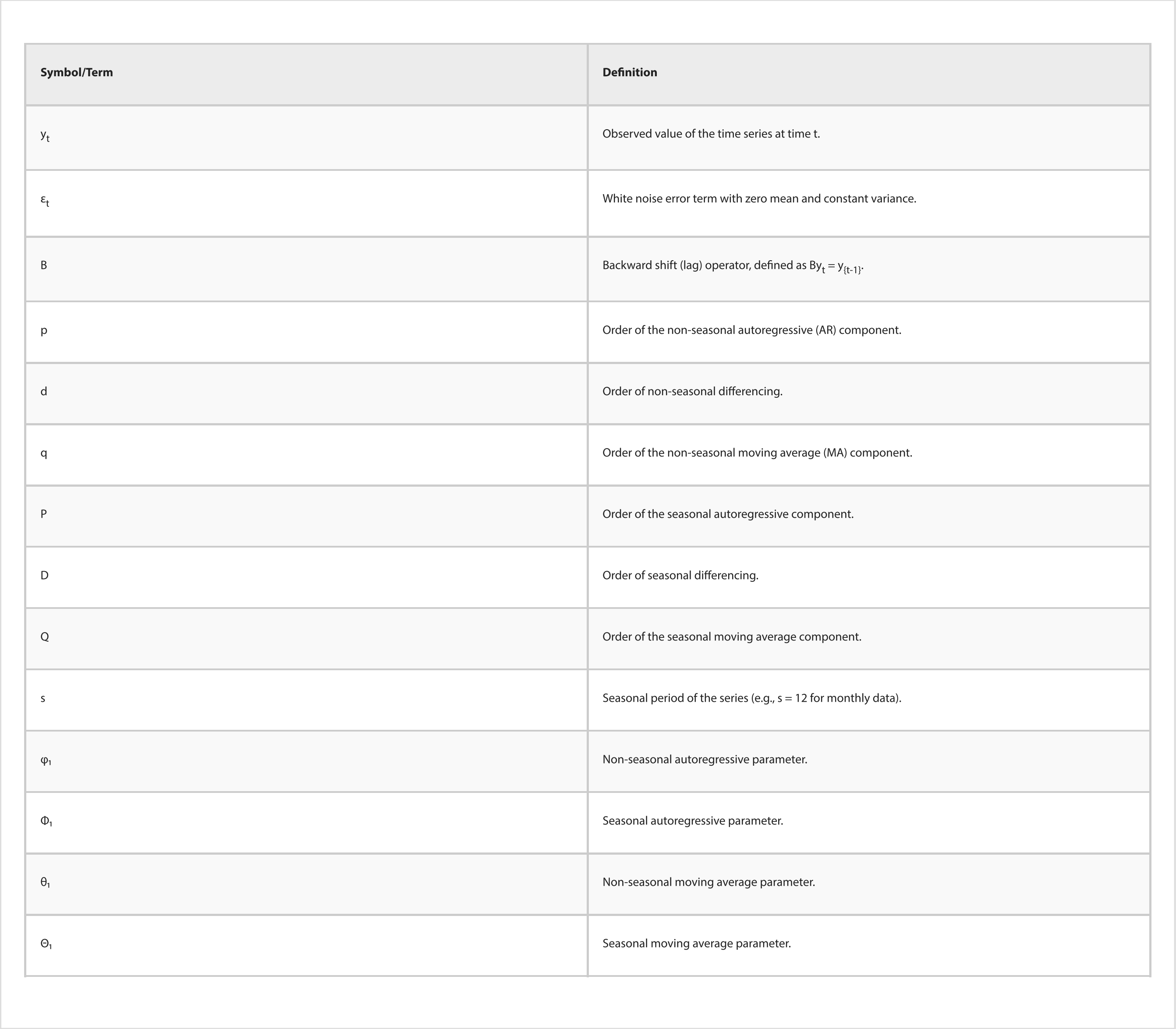
Symbol/Term	Definition
y_t	Observed value of the time series at time t.
ε_t	White noise error term with zero mean and constant variance.
B	Backward shift (lag) operator, defined as $By_t = y_{\{t-1\}}$.
p	Order of the non-seasonal autoregressive (AR) component.
d	Order of non-seasonal differencing.
q	Order of the non-seasonal moving average (MA) component.
P	Order of the seasonal autoregressive component.
D	Order of seasonal differencing.
Q	Order of the seasonal moving average component.
s	Seasonal period of the series (e.g., s = 12 for monthly data).
φ_1	Non-seasonal autoregressive parameter.
Φ_1	Seasonal autoregressive parameter.
θ_1	Non-seasonal moving average parameter.
Θ_1	Seasonal moving average parameter.

TABLE 1: Definition of Terms in the SARIMA Model

Source: Author

When volatility factors affect the data generation process, the GARCH method is proven to be better for modeling the time series data with high volatility in a substantial number of observations (Asteriou and Hall, 2021). The volatility phenomenon results in a heteroscedastic problem in the data. The conventional practice is to display the equations of time series outcomes exhibiting such phenomena as follows.

$$\begin{aligned} x_t &= \mu + a_t \\ a_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \\ \varepsilon_t &\sim \mathcal{N}(0,1) \end{aligned} \tag{2}$$



Symbol/Term	Definition
y_t	Observed value of the time series at time t.
ε_t	White noise error term with zero mean and constant variance.
B	Backward shift (lag) operator, defined as $By_t = y_{\{t-1\}}$.
p	Order of the non-seasonal autoregressive (AR) component.
d	Order of non-seasonal differencing.
q	Order of the non-seasonal moving average (MA) component.
P	Order of the seasonal autoregressive component.
D	Order of seasonal differencing.
Q	Order of the seasonal moving average component.
s	Seasonal period of the series (e.g., s = 12 for monthly data).
φ_1	Non-seasonal autoregressive parameter.
Φ_1	Seasonal autoregressive parameter.
θ_1	Non-seasonal moving average parameter.
Θ_1	Seasonal moving average parameter.

TABLE 2: Definition of Terms in Volatility Models (GARCH, EGARCH, and EWMA)

EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity; GARCH, Generalized Autoregressive Conditional Heteroskedasticity; EWMA, exponentially weighted moving average. Source: Author

Details of the terms are defined in Table 2. Here, x_t is a time series of the value of the case under study, with a GARCH model's mean, μ, and residual a_t at time t. The residual of the model, a_t, is assumed to exhibit volatility phenomena and contain some normally distributed error ε_t. The notation $P_v(0,1)$ refers to the error term, ε_t, which is normally distributed with zero mean and variance of 1. The terms p (number of lagged squared residuals) and q (number of lagged conditional variances) are the order of the autoregressive and volatility components. The terms α_i and β_j are parameters of the ARCH and GARCH, respectively. The GARCH model requires every coefficient of the dispersion equation to be nonnegative (Asteriou and Hall, 2021). However, volatility can result from both positive and negative shocks. The EGARCH model incorporates the leverage effect (ω) and relaxes the condition of non-negativity.

$$\ln \sigma_t^2 = \alpha_0' + \beta \ln \sigma_{t-1}^2 + \omega \left(\frac{\varepsilon_{t-1}}{\alpha_{t-1}} \right) + \gamma \left| \frac{\varepsilon_{t-1}}{\alpha_{t-1}} \right| \tag{3}$$

The above equation reflects the asymmetric impacts of negative and positive shocks of the same magnitude. When $\frac{\varepsilon_{t-1}}{\alpha_{t-1}} < 0$, it produces a shock effect of $\gamma - \omega$, and otherwise $\gamma + \omega$. This behavior captures real-life volatility phenomena.

The exponentially weighted moving average (EWMA) method also gives more weight to recent observations relative to those in the past (Asteriou and Hall, 2021). It reduces the influence of older observations exponentially, and calculates volatility using a rolling window technique.

$$\sigma_t^2 = \rho\,\sigma_{t-1}^2 + (1-\rho)\,\varepsilon_{t-1}^2, \quad 0 < \rho < 1 \tag{4}$$

However, a large value of ρ may compromise the potentially lasting effect of volatility.

Some precautionary modelling approaches were practiced to address data management. In one model, the data was split into training and testing sets. In another, the model was trained on the full dataset without a split. When the data were split, min-max scaling was applied separately to each subset to avoid information leakage from the test period into the training stage. The normalization approach is considered better for enhancing the efficiency and stability of the gradient descent optimization process. This independent normalization approach was motivated by clear differences in the distribution of observations between earlier and later periods. To further prevent data leakage during cross-fold validation, a gap of 24 observations was maintained between the training and test datasets. The observations in the test split dataset were the same as those in the final test set of the time-series cross-fold validation.

Machine Learning Method: Long-Short Term Memory (LSTM)

LSTM is a machine learning method. As a non-parametric estimation approach, it performs reasonably well even with complex and non-normal data, including cases with heteroscedasticity problems (Liu et al., 2025) (Szarek, 2023). Common ML methods for time series analysis include recurrent neural networks (RNN) and LSTM (Szarek, 2023). Our dataset contains over 324 months of observations. The RNN could not effectively handle complex, large-scale time series problems due to its limited memory capacity. Following the literature, the LSTM method is applied to address this issue, as its gating mechanism allows information to be retained in memory over long sequences. These gates also address processing challenges such as vanishing gradients, enabling stable training over extended sequences (Liu et al., 2025). The LSTM operates through the following gate architectures and mathematical processes.

The input gate operates in the following mathematical process, and determines what new information to store in the cell states:

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\hat{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

The forget gate operates in the following mathematical process and determines what information to discard from the cell state:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

The cell gate combines the forget and input gate outputs to update the cell state:

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \tag{8}$$

The output gate determines information from the updated cell state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \odot \tanh(c_t) \tag{10}$$

The terms $i_t,\ f_t,\ c_t,\ \text{and}\ o_t,$ respectively, refer to the input gate, forget gate, output gate, cell state (memory) at time state, t. Similarly, the terms $h_{t-1},\ w_x,\ b_x,\ \hat{c}_t$ represent, respectively, the hidden state from the previous time step, weight for neurons of the respective gates $i_t,\ f_t,\ c_t,\ \text{and}\ o_t,$ biases for the corresponding gates, and candidates for the

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

cell state. The term σ refers to a sigmoid activation function (to produce selector vectors with values between 0 and 1), whereas tanh refers to a hyperbolic tangent activation function (to generate candidate values and scale the cell state output).

LSTM-EGARCH Integrated (Hybrid) Forecasting Method

The hybrid forecasting approach combines a LSTM model with a stochastic EGARCH process to predict future sales. The LSTM model is first applied to historical sales data to capture the underlying trend and seasonal patterns, producing a base forecast denoted as $\hat{y}_t^{\mathrm{LSTM}}$:

$$\hat{y}_t^{\mathrm{LSTM}} = \mathrm{LSTM}(y_{t-1}, y_{t-2}, \ldots, y_{t-\mathrm{window}})$$

Here, $\hat{y}_t^{\mathrm{LSTM}}$ represents the expected sales at time t based on past observations. Although the LSTM captures the general trajectory, it does not account for random fluctuations or periods of increased volatility in the data. To incorporate this variability, a stochastic EGARCH model is applied to generate random deviations around the LSTM forecast. The stochastic component is computed as:

$$\epsilon_t = \phi\,\epsilon_{t-1} + \sigma_t z_t, \tag{11}$$

$$\log\left(\sigma_t^2\right) = \omega + \beta \log\left(\sigma_{t-1}^2\right) + \alpha\left(|z_{t-1}| - \mathbb{E}|z|\right) + \gamma z_{t-1}. \tag{12}$$

In these equations, ϵ_t represents the stochastic shock at time , σ_t^2 is the conditional variance, and $z_t$is a random draw from a Student-t distribution with degrees of freedom. The parameters $\omega, \alpha, \beta, \gamma$ control the memory, leverage, and persistence of volatility. This stochastic process ensures that the noise reflects realistic fluctuations, with larger shocks following periods of higher variability and smaller shocks during calmer periods.α, β, γ

The final hybrid forecast is obtained by adding the stochastic shocks to the LSTM forecast:

$$\hat{y}_t^{\mathrm{Hybrid}} = \hat{y}_t^{\mathrm{LSTM}} + \epsilon_t$$

This approach preserves the long-term trend identified by the LSTM while simultaneously incorporating realistic uncertainty derived from the EGARCH model. The conditional variance σ_t^2 can also be used to construct uncertainty bands around the hybrid forecast, providing a probabilistic range for future sales values. In practice, the LSTM determines the expected sales trajectory, while the EGARCH-derived stochastic component introduces the “bumps” and volatility observed in real-world data, resulting in a robust and informative forecast suitable for decision-making.

Model Evaluation Metrices

This study applied three commonly used error metrics to evaluate model accuracy: mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). The rationale for using the metrics in evaluating models of single time series studies are well explained in the literature (Abotaleb and Dutta, 2024). The MSE measures the average squared difference between actual and predicted values. It penalizes large deviations more heavily than small ones and works better for handling large errors. Its differentiable form also makes it well-suited for optimization during model training. The in-sample MSE can be derived with the following mathematical formula.

$$h_t = o_t \odot \tanh(c_t)$$

The MAE computes the average magnitude of errors without squaring them. The approach makes it more robust to outliers and easier to interpret. Its unit is the same as the unit of data. Therefore, it directly reflects the average size of prediction errors in practical terms. The following is the formula to derive in-sample MAE.

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

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

The RMSE is the square root of MSE. Therefore, it retains sensitivity to larger errors. It also expressing results in the original scale of the data. This property makes RMSE both interpretable and widely applicable in comparing the predictive accuracy of different models. The formula for in-sample RMSE is given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

The study suggests applying out-of-sample (OOS) MSE to evaluate the performance of cross-fold validation models. The OOS errors assess the model's predictive performance on data not used during estimation. Using consistent notation with the in-sample metrics, they are defined as follows:

$$\mathrm{MSE}_{\mathrm{OOS}} = \frac{1}{m}\sum_{j=1}^{m}\left(y_j^{\mathrm{OOS}} - \hat{y}_j^{\mathrm{OOS}}\right)^2$$

The formula for out-of-sample RMSE is as follows:

$$\mathrm{RMSE}_{\mathrm{OOS}} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(y_j^{\mathrm{OOS}} - \hat{y}_j^{\mathrm{OOS}}\right)^2}$$

The formula for out-of-sample MAE

$$\mathrm{MAE}_{\mathrm{OOS}} = \frac{1}{m}\sum_{j=1}^{m}\left|y_j^{\mathrm{OOS}} - \hat{y}_j^{\mathrm{OOS}}\right|$$

where y_j^{OOS} is the actual value, $\hat{y}_j^{\mathrm{OOS}}$ is the forecasted value for the out-of-sample period, and m is the number of out-of-sample observations.

Test of Mean Differences Between Two Models

This study used the two-sample t-test for unequal variances. The test is often called Welch's t-test. Schors stated that it compares the means of two groups without assuming equal variances, which makes it more reliable when sample spreads differ (Ruxton, 2006). It standardizes the mean difference using a variance estimate and tests whether that difference could plausibly arise under the null hypothesis of equal means. The result can be derived from the following formula.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}}$$

where

$\bar{X}_1, \bar{X}_2$ are the sample means,

s_1^2, s_2^2 are the sample variances,

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

n_1, n_2 are the sample sizes,

and ν is the approximate degrees of freedom.

Out-of-sample R-squared

In the k-th fold of the cross-validation procedure, the LSTM model is trained using only the training data. This training process produces the estimated parameters of the model, denoted by $\hat{\theta}^{(k)}$. These parameters include all weights and biases obtained from the trained model.

The trained model is then used to generate forecasts for each observation i in the corresponding test (out-of-sample) dataset:

$$\hat{y}_i^{(k)} = f(x_i; \hat{\theta}^{(k)}),$$

where $\hat{\theta}^{(k)}$ are the parameters estimated from the trained model, and x_i is the input from the test dataset of fold k.

Using these forecasts, the average out-of-sample coefficient of determination R^2_{OOS} is calculated as:

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{k=1}^{K} \sum_{i=1}^{m_k} \left(y_i^{(k)} - \hat{y}_i^{(k)} \right)^2}{\sum_{k=1}^{K} \sum_{i=1}^{m_k} \left(y_i^{(k)} - \bar{y}^{\text{OOS}} \right)^2},$$

where $y_i^{(k)}$ is the actual observed value in the test dataset,
$\hat{y}_i^{(k)}$ is the forecasted value,
m_k is the number of observations in the test dataset of fold K,
K is the total number of folds, and
$\bar{y}^{\text{OOS}}$ is the mean of all actual values in the out-of-sample test datasets.

The formula for out of sample R-square for a single fold is as follows:

$$R^2_{\text{OOS},k} = 1 - \frac{\sum_{i=1}^{m_k} \left(y_i^{(k)\mathcal{M}} - \hat{y}_i^{(k)} \right)^2}{\sum_{i=1}^{m_k} \left(y_i^{(k)\mathcal{M}} - \bar{y}^{(k)} \right)^2}$$

where

$$\bar{y}^{(k)} = \frac{1}{m_k} \sum_{i=1}^{m_k} y_i^{(k)}$$

Data Preprocessing

The original data represented the count of vehicles. The absolute values were transformed into their natural logarithm format to linearize relationships, normalize growth or trends, and mitigate undesirable attributes (e.g., skewness and heteroscedasticity) in the data. Initially, the outlier observations were identified and reset to the acceptance limit. Values of some observations near the end of the time series were notably high. These observations had the potential to influence measures of central tendency, generally leading to overestimated variability. Forecasts based on such data could deviate significantly from the true trend or seasonality, resulting in misleading conclusions. The Median Absolute Deviation (MAD) test identified seven outlier observations in the dataset. The values of outlier observations were trimmed due to the normalization requirements in econometric methods.

For the EGARCH methods, the augmented Dickey-Fuller tests were carried out to obtain stationary data. The data were found to be stationary with the first difference. When the data were modified using the first difference, the distribution changed considerably, and the test identified 13 outlier observations in the first-differenced dataset. Despite the EGARCH being specifically designed to handle volatility, including extreme values (e.g., market crashes or sudden spikes), the model instability problem remained highly significant in all sub-models of EGARCH (Carnero et al., 2012). Therefore, these outliers were also treated with maximum tolerable values. Despite its inherent robustness in handling outlier observations, this study used outlier-treated data in the LSTM method for comparing the robustness of its results with the EGARCH method.

The log-transformed data were further standardized using the min-max normalization method to address potential analysis problems for the LSTM model. This transformation scaled the data to a range between 0 and 1. The normalization approach is considered better for enhancing the efficiency and stability of the gradient descent optimization process. The model was trained using the entire dataset without splitting it into separate training and test sets. This approach was chosen due to the distribution patterns and values of observations in the later data being obviously different from those in the earlier periods.

LSTM Model Parameter Specification

In the case of LSTM, the main hyperparameters were determined by simulating with the Optuna library. Some of them were further tuned to fit the data. The hyperparameters were tuned based on criteria such as consistency between training and validation loss, cross-validation performance, and smoother (less noisy) training and validation loss curves. The values of hyperparameters and other elements in the LSTM architecture set up for training the model are listed in Table *3* below.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Component	Specification	Function (Concise Role in Model)
Model Type	Stacked LSTM (Univariate Time Series)	Captures nonlinear temporal dependencies via layered memory cells
Target Variable	SaleIn	Variable being forecasted
Input Shape	(24, 1)	24 past observations used as input sequence
Window Size	24	Defines historical look-back length
Forecast Horizon	1-step ahead (recursive)	Predicts next period iteratively
Forecast Strategy	Recursive (free-running after training)	Uses prior predictions as future inputs
Data Frequency	Monthly	Temporal resolution of observations
Temporal Order Preserved	Yes (shuffle = FALSE)	Maintains sequence dependence
Output Layer	Dense(1), Linear activation	Maps hidden state to continuous forecast
Batch Size	32	Number of samples per gradient update
Epochs	164	Training iterations over full dataset
Optimizer	Adam	Adaptive gradient-based weight update
Learning Rate	0.007	Step size for parameter updates
Gradient Clipping	clipnorm = 1.5770	Prevents exploding gradients
Kernel Constraint Value	2.1186	Upper bound on weight magnitude
Regularization (shared)	L1 = 2.1664e-05, L2 = 0.00178285	Controls overfitting via penalty terms
Kernel Constraint	MaxNorm	Limits weight norm for stability
Number of LSTM Layers	3	Enables hierarchical feature extraction
LSTM Units in order of layer	(96, 96, 49)	Memory capacity per layer
Activation Functions in order of layer	(tanh, tanh, tanh)	Controls nonlinear state transitions
Return Sequences in order of layer	(TRUE, TRUE, FALSE)	Passes sequence output to next layer
Dropout Rates in order of layer	(0.111, 0.0672, —)	Reduces overfitting via neuron deactivation
Recurrent/Bias Constraint	None	No restriction on recurrent weights

Loss Function	Mean Squared Error (MSE)	Penalizes large forecast errors
Evaluation Metric	Mean Absolute Error (MAE)	Measures average prediction error
Callbacks Used	Terminate On NaN	Stops training on invalid loss
Early Stopping	Not used	Training not halted by validation loss
Stateful LSTM	No	Hidden states reset each batch
Scaling Method	MinMaxScaler	Normalizes input range
Forecast Period	Jan-2022 → Dec-2027 (recursive)	Out-of-sample prediction span

TABLE 3: LSTM Model Configuration and Hyperparameters

Source: Author. LSTM, Long Short-Term Memory

Two modeling strategies, particularly in the LSTM method, are used in this study. In the first, the data are split into training and testing sets. In the second, the model is trained on the full dataset without a split. When the data are split, min-max scaling is applied separately to each subset to avoid information leakage from the test period into the training stage. The normalization approach is considered better for enhancing the efficiency and stability of the gradient descent optimization process. This approach is motivated by clear differences in the distribution of observations between earlier and later periods. The observations in the test split dataset were the same as those in the final test set of the time-series cross-fold validation. The cross-fold validation was done on entire data set. To further prevent data leakage during cross-fold validation, a gap of 24 observations (2 years) was maintained between the training and test datasets. The test size for each validation fold was 36 months. The last fold test outputs represent the result of ordinary one-fold validation test of the time series LSTM model.

Forecasting methods

This study employed the recursive forecasting approach using inputs from both LSTM and EGARCH models. Although the approach is susceptible to error accumulation, it offers the advantage of requiring only a single model and naturally capturing the temporal dependencies in the data. Its prediction can be more precise in the early to medium period of the forecasting horizon than that of the direct multistep method (Hamzaçebi et al., 2009). As a result, the LSTM model captured complex nonlinear patterns and sequential dependencies in the data. The EGARCH model also captured volatility clustering and asymmetry in fluctuations. The method emphasizes recent observations by giving them higher weights (Livieris and Pintelas, 2022) (Hamzaçebi et al, 2009). This approach facilitated showing potential forecasting flaws of split data model over full data model. The LSTM-EGARCH hybrid model was constructed by integrating the one-step-ahead forecasts generated by the LSTM model with the stochastic volatility parameters estimated from the EGARCH model.

Data analysis tools

In terms of software use, the data for the EGARCH model were analyzed on R Studio, whereas the data for LSTM model were analyzed in Anaconda Jupiter notebook (python platform). Task-specific libraries such as python's TensorFlow that provides tools to define LSTM architectures, preprocess and manage data, train and optimize the model, evaluate performance, and deploy the trained model into production, were used. Microsoft Excel was used to carry out t-tests and plot the comparative results of forecasts from EGARCH and LSTM.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Results And Discussion

Descriptive results

The descriptive statistics are given in Table *4*. The results show that the log-transformed sales series is tightly distributed, with the mean (9.707) and median (9.723) lying close to each other, which indicates a largely symmetric distribution. The standard deviation (0.222) reflects limited variability, and the maximum value (10.136) suggests the absence of extreme upper-end observations. During data preprocessing, extreme low observations were truncated to a tolerable threshold. As a result, the mode equals the minimum value (9.039) because several observations take this imposed lower bound. This outcome reflects a data-handling choice rather than a genuine modal concentration of the underlying continuous process and should be interpreted accordingly in subsequent modeling and inference.

Parameters	Values
Observations	324
Mean	9.71
Median	9.72
Mode	9.04
Standard deviation	0.22
Minimum observation value	9.04
Maximum observation value	10.14

TABLE 4: Descriptive Statistics of Car Sales

Source: Author

Dynamics of car sales across the time in New Zealand since 1998 to 2024

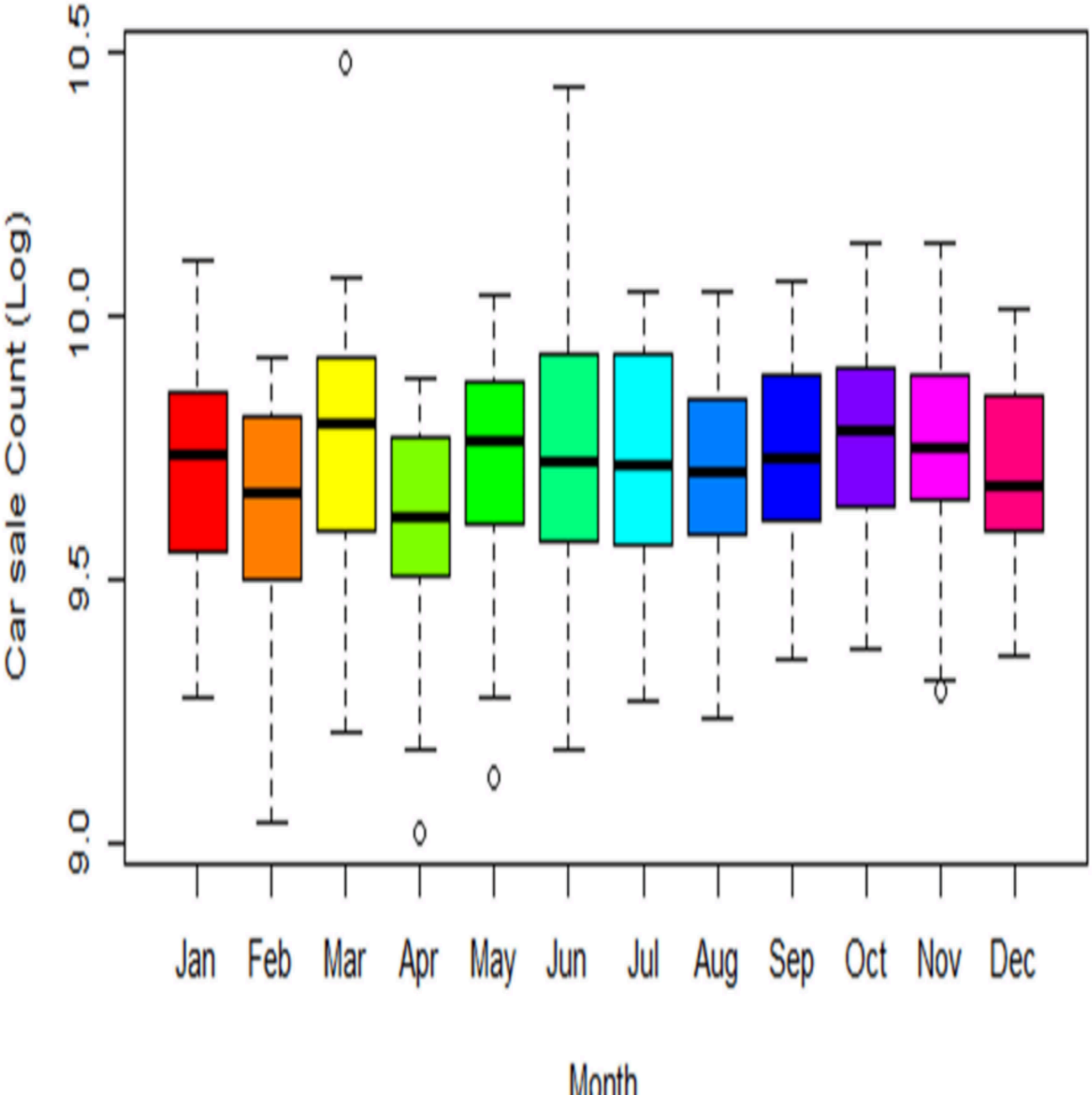


FIGURE 1: Monthly Trend of the Car Sales

Source: Author

Figure 1 shows the average trend of monthly car sales from 1998 to 2024. The sales were slightly higher in March, June, and July. The tolerable level of outlier sales is also marked on the plot. The trend of car sales in New Zealand from 1998 to 2024 is illustrated in Figure 2. The number of car sales greatly varies with time. For example, the total monthly car sales were 10,914 in January 1998, 8,912 in April 2009, 25,332 in October 2017, 1,319 in April 2020, and 14,349 in December 2024. The notable declines in 2009 and 2020 correspond to the economic recession and the COVID-19 pandemic, respectively. Interestingly, record-high sales were observed in the post-COVID period, such as 35,493 in March 2022 and 33,970 in June 2023. The trend indicates that car sales have become more volatile in recent years, although the overall long-term trajectory shows an increase.

FIGURE 2: Visualization of Historical Car Sale Trend 1998-2024

Source: Author

Model test performances

Test of Stationary Property of Data

Stationarity property is required for time series data to be valid in EGARCH models. The Dickey-Fuller (DF) test statistic for the original data was −2.2005 with a p-value of 0.492. Since the p-value exceeded the 0.05 (5%) significance level, the series was deemed non-stationary. The same test applied to the first-differenced data yielded a DF statistic of −9.1665 with a p-value of 0.01. As the p-value was below the 0.05 threshold, the first-differenced series was found to be stationary. The distribution of the first-differenced data over the historical period is shown in Figure 3. This distribution pattern is consistent with that of the original data, with volatility skewed toward recent observations or years.

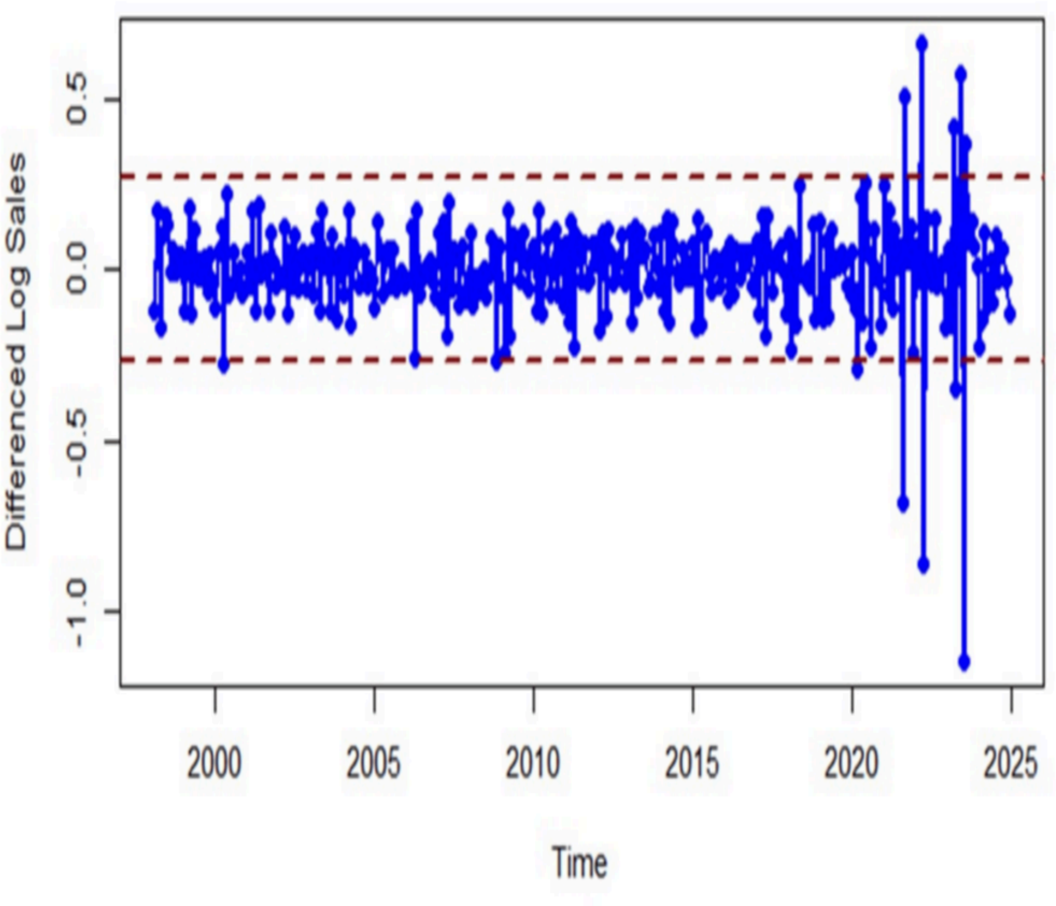


FIGURE 3: Distribution of the First Difference Data

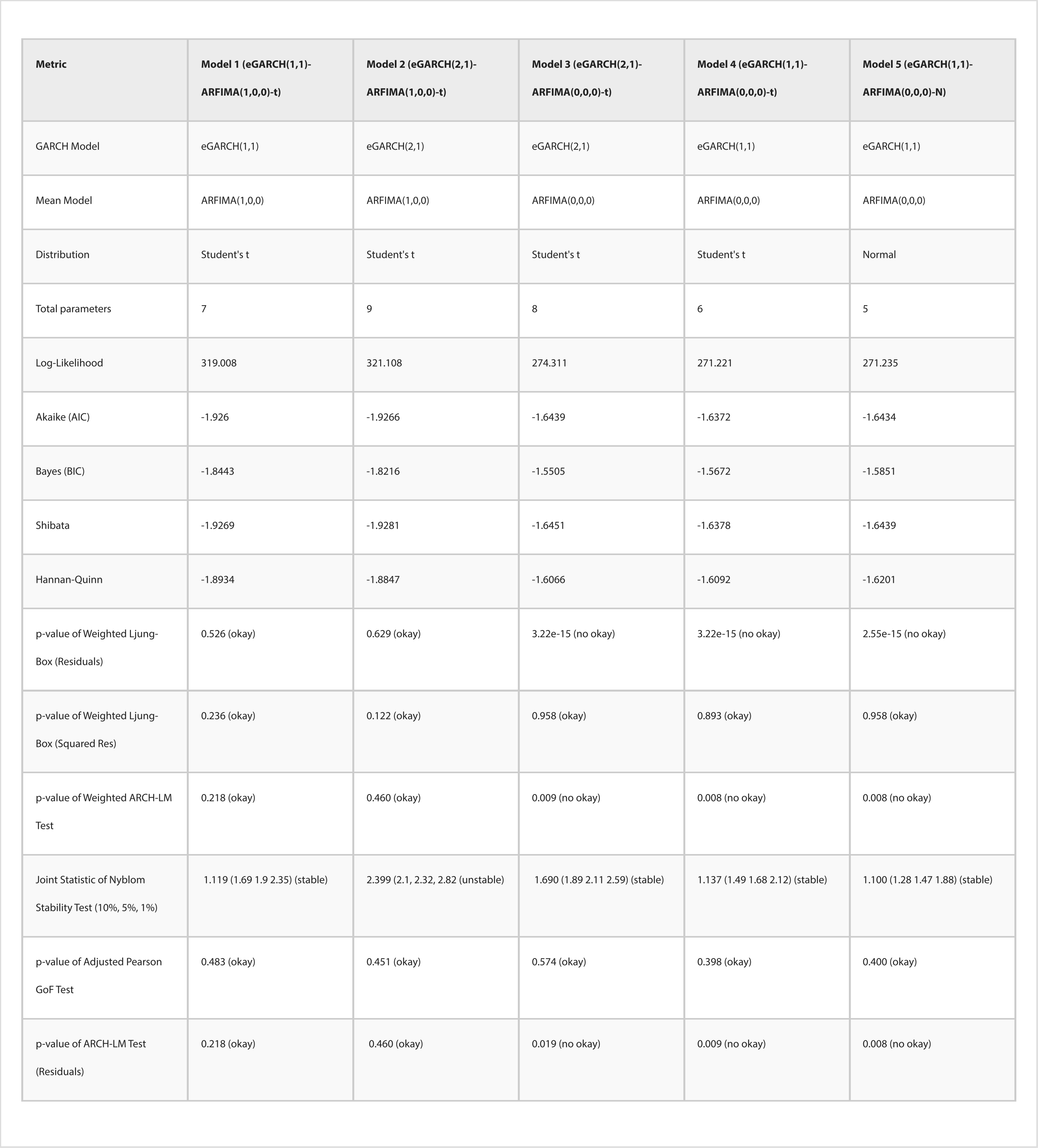
Source: Author

EGARCH Model Test Statistics

Many forms of the EGARCH model were screened. The model test statistics of some robust models among the many candidates are given in Table 5. Models 3-5 show significant ARCH effects (indicated by low ARCH LM p-values) and residual autocorrelation. These statistics make them less reliable. Despite good statistical properties, the Nyblom Stability

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Test showed model 2 to be unstable. This suggests structural breaks (affected by many phenomena) that invalidate the model's assumptions. Forecasting and inference require stable parameters, such as volatility persistence and asymmetry that are consistent over time. Model 1 satisfied these conditions; it is also more parsimonious (meaning it is a simpler model with fewer parameters).



Metric	Model 1 (eGARCH(1,1)-ARFIMA(1,0,0)-t)	Model 2 (eGARCH(2,1)-ARFIMA(1,0,0)-t)	Model 3 (eGARCH(2,1)-ARFIMA(0,0,0)-t)	Model 4 (eGARCH(1,1)-ARFIMA(0,0,0)-t)	Model 5 (eGARCH(1,1)-ARFIMA(0,0,0)-N)
GARCH Model	eGARCH(1,1)	eGARCH(2,1)	eGARCH(2,1)	eGARCH(1,1)	eGARCH(1,1)
Mean Model	ARFIMA(1,0,0)	ARFIMA(1,0,0)	ARFIMA(0,0,0)	ARFIMA(0,0,0)	ARFIMA(0,0,0)
Distribution	Student's t	Student's t	Student's t	Student's t	Normal
Total parameters	7	9	8	6	5
Log-Likelihood	319.008	321.108	274.311	271.221	271.235
Akaike (AIC)	-1.926	-1.9266	-1.6439	-1.6372	-1.6434
Bayes (BIC)	-1.8443	-1.8216	-1.5505	-1.5672	-1.5851
Shibata	-1.9269	-1.9281	-1.6451	-1.6378	-1.6439
Hannan-Quinn	-1.8934	-1.8847	-1.6066	-1.6092	-1.6201
p-value of Weighted Ljung-Box (Residuals)	0.526 (okay)	0.629 (okay)	3.22e-15 (no okay)	3.22e-15 (no okay)	2.55e-15 (no okay)
p-value of Weighted Ljung-Box (Squared Res)	0.236 (okay)	0.122 (okay)	0.958 (okay)	0.893 (okay)	0.958 (okay)
p-value of Weighted ARCH-LM Test	0.218 (okay)	0.460 (okay)	0.009 (no okay)	0.008 (no okay)	0.008 (no okay)
Joint Statistic of Nyblom Stability Test (10%, 5%, 1%)	1.119 (1.69 1.9 2.35) (stable)	2.399 (2.1, 2.32, 2.82 (unstable)	1.690 (1.89 2.11 2.59) (stable)	1.137 (1.49 1.68 2.12) (stable)	1.100 (1.28 1.47 1.88) (stable)
p-value of Adjusted Pearson GoF Test	0.483 (okay)	0.451 (okay)	0.574 (okay)	0.398 (okay)	0.400 (okay)
p-value of ARCH-LM Test (Residuals)	0.218 (okay)	0.460 (okay)	0.019 (no okay)	0.009 (no okay)	0.008 (no okay)

TABLE 5: Test Results of Different Modalities of EGARCH Models

Source: Author

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Figure *4* shows the plots of the autocorrelation and partial autocorrelation functions of both the standardized residuals and their squares. The plots show no significant spikes in their distributions. The results suggest the model has adequately captured the linear dynamics and volatility clustering in the series. However, the normality diagnostic tests show the standardized residuals from the EGARCH (1,1) model are not adequately normally distributed. The QQ plot against the normal distribution and the histogram both indicate heavier tails than the Gaussian case. The heavy-tailed distribution implies a higher likelihood of extreme shocks, which explains the model's tendency to underestimate risk during periods of large fluctuations. When compared them against a Student's t-distribution with estimated degrees of freedom around 6, the residuals align more closely with the theoretical line. The results suggest the data are better fitted with skewed-t specification.

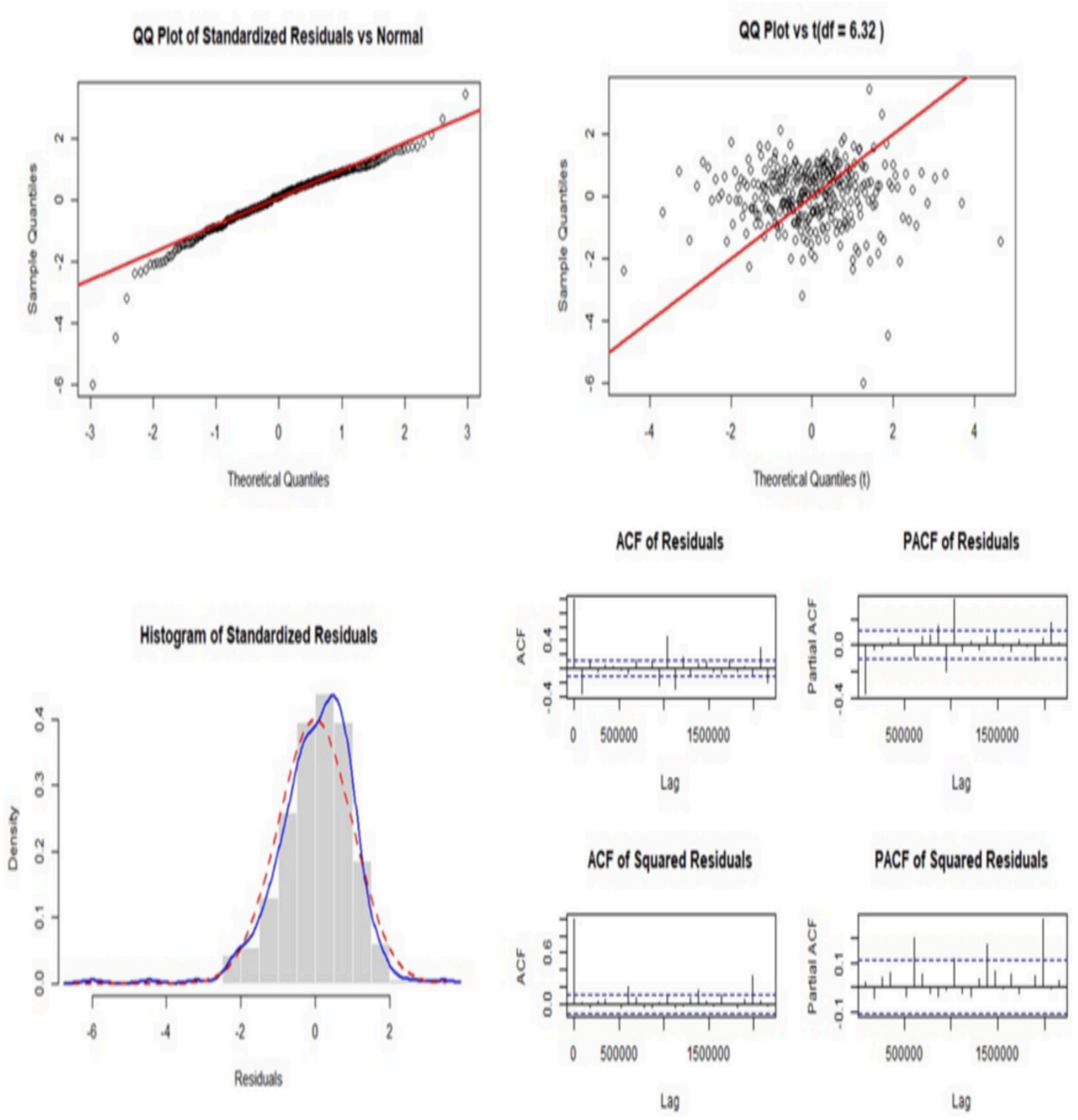


FIGURE 4: The Error Diognostic Test Results of the Model 1 (EGARCH(1,1)-ARFIMA(1,0,0)-t)

Source: Author

In summary, the error diagnostic results showed that the data validly fitted with the EGARCH (1,1) model with a skewed-t distribution. In terms of limitations, the residuals have some departure in the extreme tails. It means that rare and extreme shocks may not be fully captured. Such phenomena are common in time-series models of financial and economic data. The issue needs to be considered while interpreting forecasts.

LSTM Model Performance

Figure *5* shows the Training error dynamics of the full data model. The training curves show a sharp reduction in both MSE and MAE during the early epochs, indicating that the model quickly learns the core temporal structure of the series. After this initial phase, the loss stabilizes and enters a long plateau, suggesting convergence to a stable solution rather than continued fitting to noise. The absence of late-epoch divergence or rising error reflects numerical stability despite the learning rate and lack of early stopping. Small MAE fluctuations are present but remain bounded, which is consistent with recursive sequence learning and dropout regularization. There are no visible signs of overfitting or instability. Overall, the model appears well-converged, with later epochs reinforcing an already learned representation rather than materially changing performance.

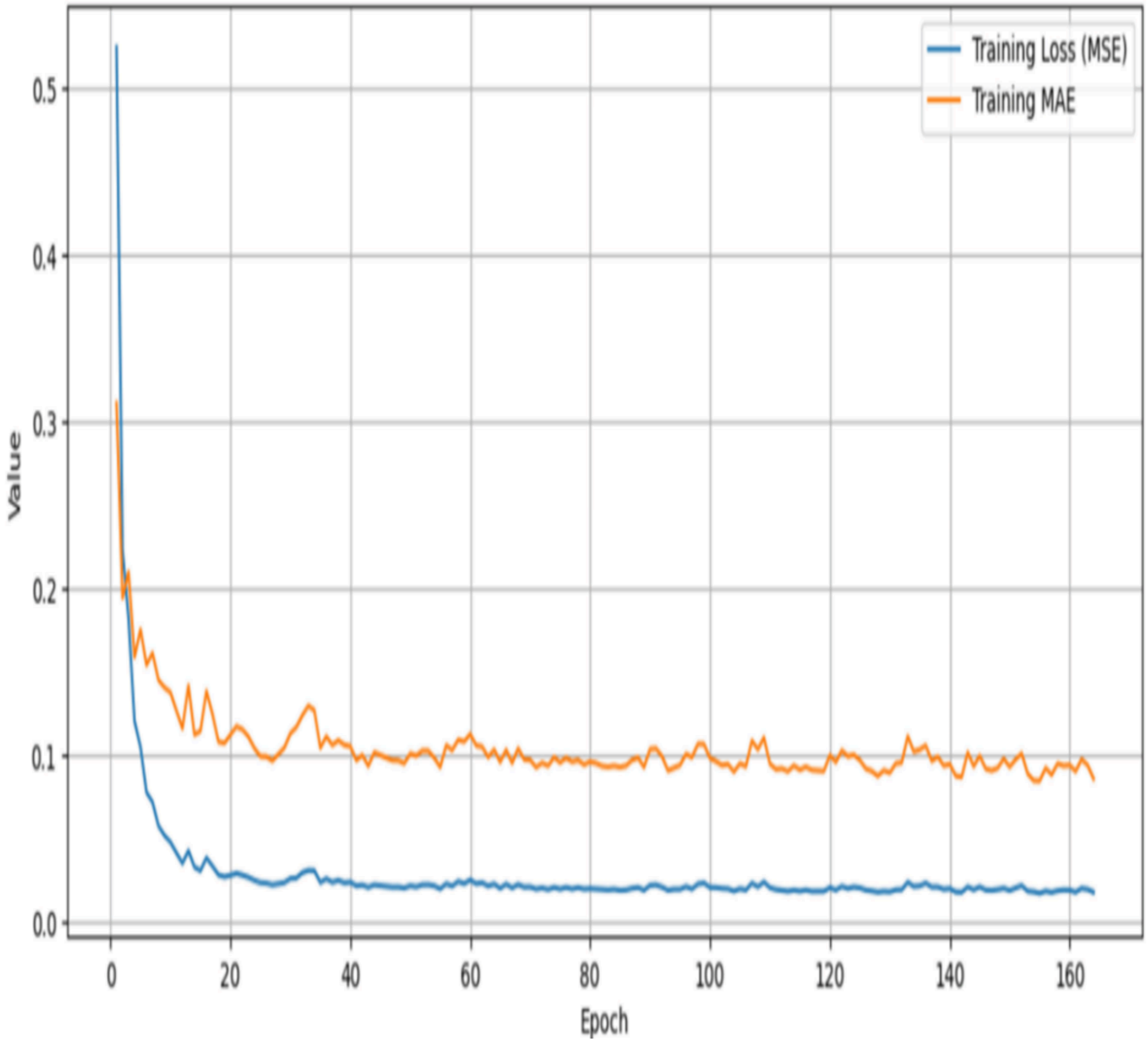


FIGURE 5: Training Error Dynamics of the Full-Data-Based Model

Source: Author

Time Series Cross-Fold Validation Results

This study evaluated the issues of TSCV in different dimensions. The results of 7-fold TSCV with both rolling-over and expanding window approaches are presented in Table 6. In the rolling window setting, each model was trained on only 36 months of data with a 24-month gap before a 36-month test period. The results show substantial instability across folds and clear differences between the rolling and expanding window approaches. Their performances varied sharply across time. Some folds produced moderate errors, but others-particularly Folds 2, 4, and 7-recorded very large negative R^2 values. We got similar errors for the training model set on bigger observation period (48 months) and consequently a smaller number of test folds (5 folds). The results are provided in the appendices. The results indicate that the model failed to outperform a simple mean benchmark during those periods. The expanding window approach, which retained all observations from January 1998 up to each fold's training end, performed better in a few cases (notably Fold 3, where

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

R^2 turned slightly positive), yet it also showed large variability and several strongly negative R^2 values. On average, the rolling window produced lower mean RMSE and MAE than the expanding window, although both approaches had high dispersion across folds. These results provided evidence of how sensitive forecasting performance is to the specific out-of-sample period. Even though data are available up to 2024, the model under validation was trained only up to 2019. This was due to the enforced 24-month gap and the need to preserve a genuinely out-of-sample test set. Importantly, the unused recent data - particularly those covering structurally different periods - appear to influence forecasts more than the historical data used for training, highlighting the practical challenges of time series cross-validation under real-world conditions. The latest data were unavailable for training due to the requirement for an OOS test. However, the data can carry more crucial information for the accuracy of forecasts than the data used in training the models.

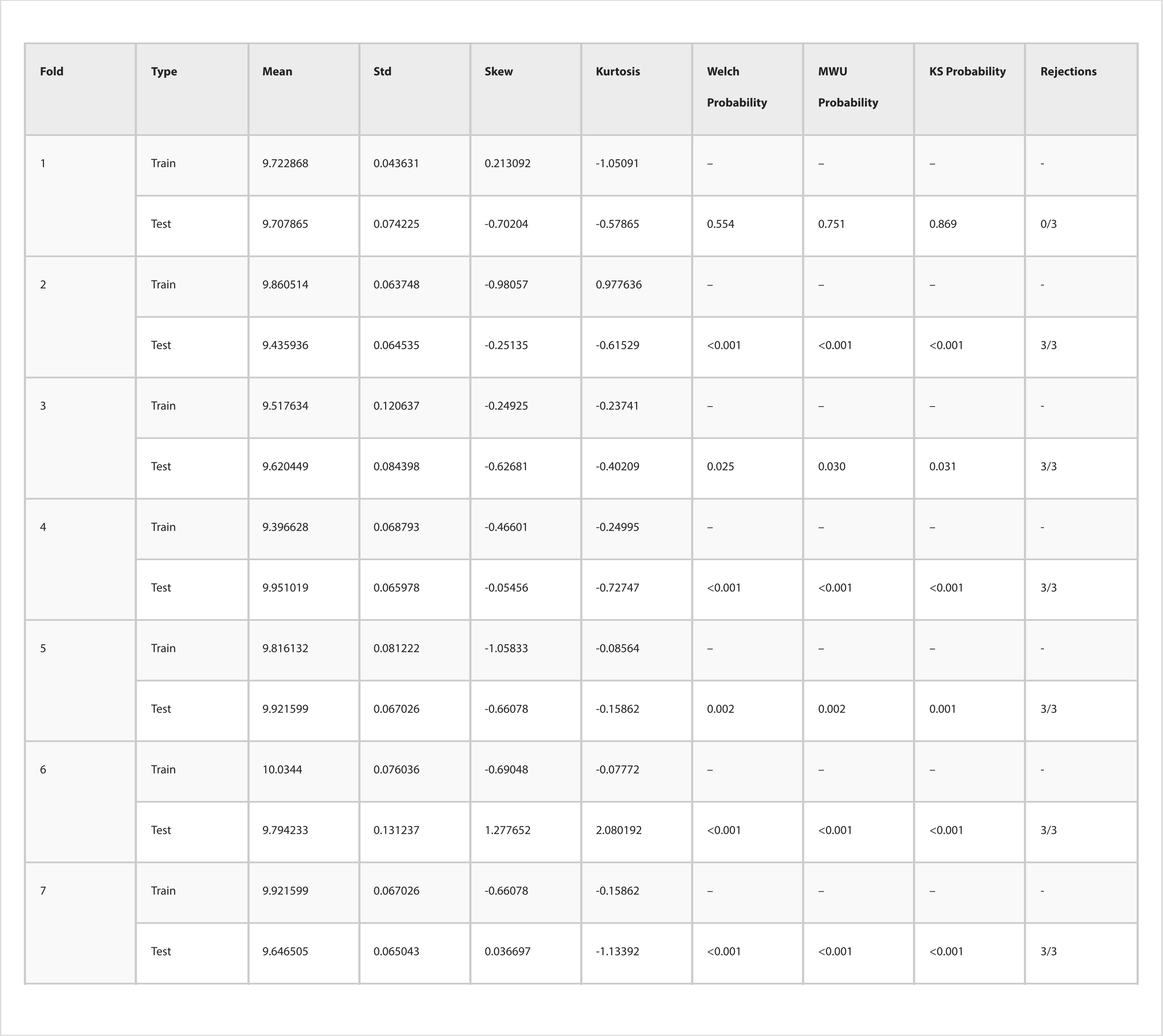
CFV Type	Train Start	Train End	Gap Start	Gap End	Test Start	Test End	RMSE	MAE	R^2	Fold
Rolling window	Jan-00	Dec-02	Jan-03	Dec-04	Jan-05	Dec-07	0.1021	0.0772	-1.065	1
Expanding-window	Jan-98	-	-	-	-	-	0.1027	0.0924	-1.0897	
Rolling window	Jan-03	Dec-05	Jan-06	Dec-07	Jan-08	Dec-10	0.2927	0.284	-21.436	2
Expanding-window	Jan-98	-	-	-	-	-	0.1971	0.1875	-9.1769	
Rolling window	Jan-06	Dec-08	Jan-09	Dec-10	Jan-11	Dec-13	0.1797	0.1673	-3.9438	3
Expanding-window	Jan-98	-	-	-	-	-	0.074	0.056	0.1617	
Rolling window	Jan-09	Dec-11	Jan-12	Dec-13	Jan-14	Dec-16	0.5444	0.5408	-73.2772	4
Expanding-window	Jan-98	-	-	-	-	-	0.1409	0.1271	-3.9753	
Rolling window	Jan-12	Dec-14	Jan-15	Dec-16	Jan-17	Dec-19	0.0646	0.0516	-0.0144	5
Expanding-window	Jan-98	-	-	-	-	-	0.0977	0.0844	-1.3174	
Rolling window	Jan-15	Dec-17	Jan-18	Dec-19	Jan-20	Dec-22	0.1499	0.1205	-0.4234	6
Expanding-window	Jan-98	-	-	-	-	-	0.1596	0.1272	-0.6132	
Rolling window	Jan-17	Dec-19	Jan-20	Dec-21	Jan-22	Dec-24	0.2821	0.2752	-19.5256	7
Expanding-window	Jan-98	-	-	-	-	-	0.145	0.1228	-4.4207	
Rolling window	-	-	-	-	-	-	0.2471 (0.1633)	0.2197 (0.169)	-17.0244 (26.40)	Mean (Std)

Expanding-window	-	-	-	-	-	-	0.1310 (0.0421)	0.1139 (0.0419)	-2.9188 (3.2468)	

TABLE 6: Seven-Fold Time-Series Cross-Validation Results for Rolling and Expanding Window Schemes (Out of Sample RMSE, MAE, and R^2)

Source: Author. MAE, mean absolute error; RMSE, root mean squared error

The cross-validation results showed consistently negative R^2 across nearly all folds. The result formally indicates performance worse than a naïve constant-average prediction (Hawinkel et al., 2024). The model quality is a subject of verification. Technically, the result happens when the numerator (model errors) exceeds the denominator (errors of the mean forecast). In terms of the root cause, the training and validation data sets might not follow the assumed distribution. Standard validation frameworks typically assume that the training and test samples follow the same underlying distribution (Bates et al., 2024). Table *7* presents the distributional diagnostics for the training and validation samples across individual folds under the rolling-window cross-validation method. The results indicate that, in most folds, the distributional properties of the training and test sets differ materially. This conclusion is supported jointly by the Welch's t-test (equality of means), the Mann-Whitney U test (equality of central tendency), and the Kolmogorov-Smirnov (KS) test (equality of overall distributions). The first test is parametric, and the others are non-parametric tests. Specifically, all folds except the first one showed statistically significant differences between training and validation samples. Large shifts in central tendency and dispersion - most prominently in Fold 4 (mean: 9.3966 vs. 9.9510) and Fold 2 (mean: 9.8605 vs. 9.4359) - coincided with the most pronounced deterioration in predictive performance. On these folds, all three tests consistently reject the null hypothesis of distributional equivalence at conventional significance levels. In the estimation process, the training dataset of each fold was normalized independently to avoid data leakage. In time series forecasting, such discrepancies are common because real-world series often evolve over time due to trends, changing volatility, or structural breaks.



Fold	Type	Mean	Std	Skew	Kurtosis	Welch Probability	MWU Probability	KS Probability	Rejections
1	Train	9.722868	0.043631	0.213092	-1.05091	–	–	–	-
	Test	9.707865	0.074225	-0.70204	-0.57865	0.554	0.751	0.869	0/3
2	Train	9.860514	0.063748	-0.98057	0.977636	–	–	–	-
	Test	9.435936	0.064535	-0.25135	-0.61529	<0.001	<0.001	<0.001	3/3
3	Train	9.517634	0.120637	-0.24925	-0.23741	–	–	–	-
	Test	9.620449	0.084398	-0.62681	-0.40209	0.025	0.030	0.031	3/3
4	Train	9.396628	0.068793	-0.46601	-0.24995	–	–	–	-
	Test	9.951019	0.065978	-0.05456	-0.72747	<0.001	<0.001	<0.001	3/3
5	Train	9.816132	0.081222	-1.05833	-0.08564	–	–	–	-
	Test	9.921599	0.067026	-0.66078	-0.15862	0.002	0.002	0.001	3/3
6	Train	10.0344	0.076036	-0.69048	-0.07772	–	–	–	-
	Test	9.794233	0.131237	1.277652	2.080192	<0.001	<0.001	<0.001	3/3
7	Train	9.921599	0.067026	-0.66078	-0.15862	–	–	–	-
	Test	9.646505	0.065043	0.036697	-1.13392	<0.001	<0.001	<0.001	3/3

TABLE 7: Distribution Properties of Train and Test Dataset of Each Fold of Cross-validation

Welch: Welch's t-test for equality of means. MWU: the Mann–Whitney U test for equality of central tendency. KS: the Kolmogorov–Smirnov test for equality of overall distributions. Source: Author

When the test distribution diverges materially from the training distribution, the interpretation of performance metrics becomes inherently more complex. Standard error measures such as RMSE and MAE will mechanically increase in the presence of higher volatility or pronounced mean shifts, even if the underlying model structure remains unchanged. Similarly, they may become negative, not necessarily because the model is fundamentally misspecified, but because it is evaluated under a distributional regime that differs from the one on which it was trained. In such circumstances, deteriorating forecast accuracy may primarily reflect structural instability between the training and test periods rather than intrinsic deficiencies in model specification. The model's difficulty arises from its limited capacity to extrapolate into a new regime characterized by altered location, scale, or higher-order moments. Moreover, the conventional interpretation of a negative value - that the model performs worse than a naïve mean benchmark - becomes less straightforward when the benchmark itself is computed under a substantially different distributional environment.

TSCV was performed on both level and first-differenced data to examine whether the large forecast errors, negative values, and observed distributional differences were primarily caused by non-stationery property of data. Table 8 results show that the RMSE and MAE values are strikingly similar across all folds, whether the model is trained on level data or on first-differenced data. The R^2 values show almost no change between the two approaches, which suggests that the LSTM captures roughly the same proportion of variance regardless of how the data are pre-processed. The fold-by-fold trends are also consistent: Fold 3 consistently performs the worst, while Folds 4 and 5 show noticeably better results. There are minor differences in the numerical values, for example, in Fold 1 the RMSE changes from 0.3726 for the level data to 0.3598 for the differenced data. This is expected, as first-differencing naturally alters the scale and variance of the series. Slight differences in R^2 mostly reflect the impact of differencing on the overall mean level rather than the model's predictive ability.

Fold	RMSE (Level)	RMSE (Difference)	MAE (Level)	MAE (Difference)	R^2 (Level)	R^2 (Difference)
1	**0.3726**	**0.3598**	**0.3597**	**0.3488**	**-9.4089**	**-8.7052**
2	0.2473	0.2545	0.2390	0.2464	-14.1567	-15.0459
3	0.5376	0.5379	0.5338	0.5342	-71.4185	-71.5018
4	0.3171	0.3154	0.2160	0.2156	-0.8440	-0.8239
5	0.3000	0.3019	0.2935	0.2954	-22.2153	-22.5073
Mean	0.3549	0.3539	0.3284	0.3281	-23.6087	-23.7168
Std	0.1026	0.1030	0.1231	0.1147	28.1428	28.0483

TABLE 8: Out-of-Sample Forecast Performance Using Level (Non-Stationary) and Differenced (Stationary) Data

Source: Author. RMSE, root mean squared error; MAE, mean absolute error

These similarities indicate that the LSTM is already capable of learning trends and levels directly, so applying first-differencing does not offer any clear advantage. The interpretation is that using the level data alone is sufficient. The model can capture trends, seasonal patterns, and absolute levels without needing stationary transformations, and first differencing does not improve prediction accuracy. The patterns in performance also persist across folds. The high error in Fold 3 points to a period of structural shift or unusual volatility, which appears in both the level and differenced datasets. Unlike classical time series models such as ARIMA or OLS, which require stationarity to perform well, the LSTM is able to handle non-stationary levels directly through its memory cells.

Performances of LSTM and EGARCH Models

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Metrics	LSTM Models		EGARCH	
	Training Model on Full Data	Training Model on Split Data	Full data	
MAP%	0.91	0.93	0.98	
MAE	0.0876	0.0894	0.0948	
RMSE	0.1238	0.1272	0.1352	
R^2	0.6848	0.6857	0.68 (Adj:0.67)	

TABLE 9: Comparative In-sample Metrices of Models

Source: Author. LSTM, Long Short-Term Memory; EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity; RMSE, root mean squared error; MAE, mean absolute error

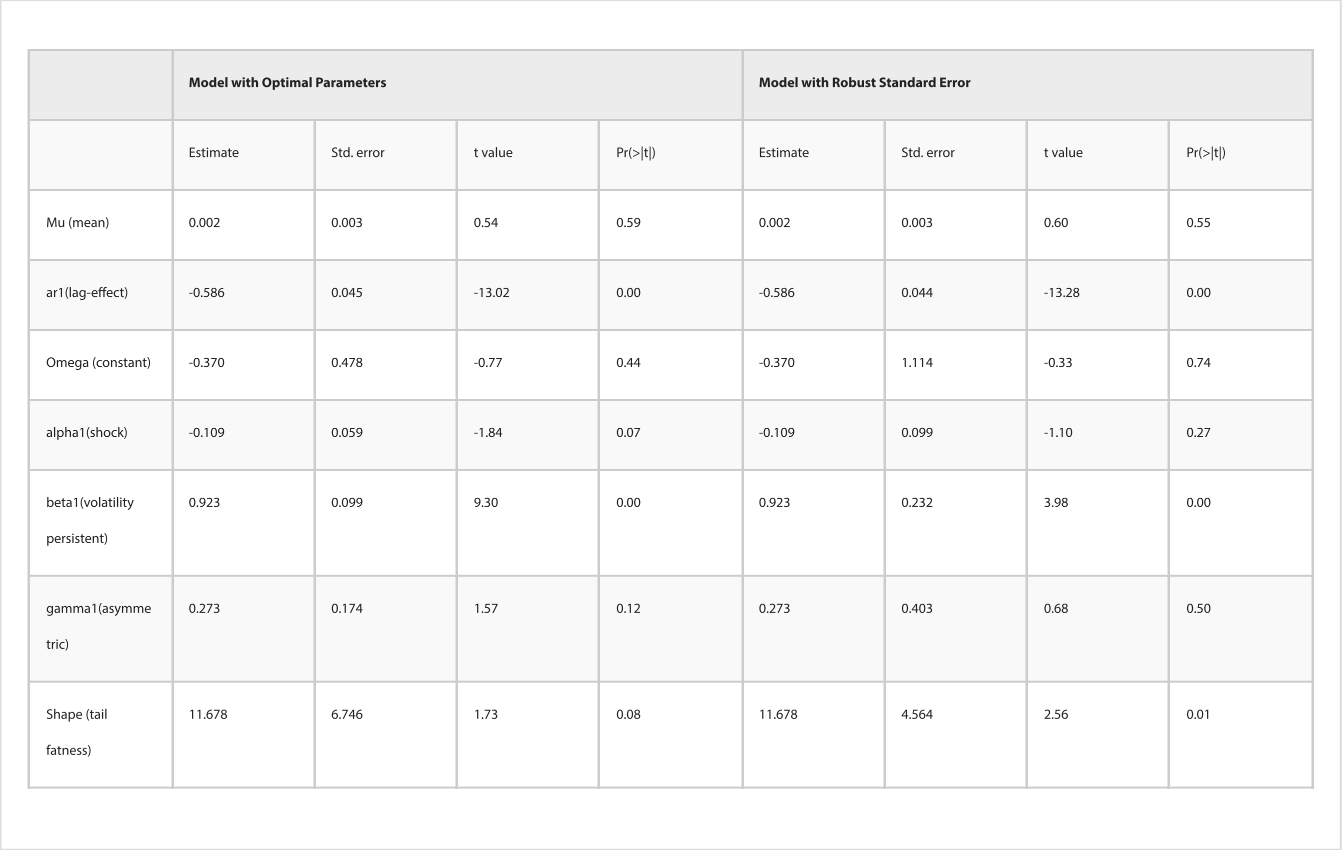
Table *9* reports the in-sample performance metrics of the LSTM and EGARCH models. The LSTM trained on the full dataset attains an R^2 of 0.6848, with an RMSE of 0.1238, an MAE of 0.0876, and a MAPE of 0.91%. When trained on the split dataset, the LSTM yields a marginally higher R^2 of 0.6857, alongside very similar error measures (RMSE = 0.1272 and MAE = 0.0894). Overall, the differences in predictive errors are small and not practically meaningful between the LSTM split data model and full data model. For the convenience of presenting the results, the split data model was trained on observations from 1998 to 2021, and the test data were allocated to the period 2022-2024. the gap between test and training sets was ignored. This is unsurprising, as the full-sample model includes only 36 additional observations out of a total of 324. The errors are also average values, which may obscure meaningful differences in individual forecasted values. Notably, despite relying on different data scaling procedures and estimation frameworks, the LSTM and EGARCH models produce error magnitudes that are broadly comparable.

The main issue here is whether split data model has explanatory power better than full data model or vice versa. The split data model has not captured the full distribution and recent dynamics of the dataset, particularly trends observed in the last 36 months (testing period). LSTM models naturally give more weight to recent observations in the input sequence because the memory (cell state) prioritizes information closer to the forecast target, while older inputs gradually lose influence. Consequently, the full data model incorporates these recent phenomena, producing forecasts that better reflect real-world behavior and trends. Comparable metrics with the EGARCH model and the validation results further confirm that the full data model is more reliable for forecasting applications in evolving time series.

Phenomena of Volatility and Shocks Affecting the Vehicle Sale Trajectory

Table *10* shows parameters of the EGARCH model that were estimated using optimal parameters (standard errors under ideal assumptions) and robust standard errors (adjusted for potential heteroskedasticity and model misspecification) of the (eGARCH(1,1)-ARFIMA(1,0,0)-t) model. The mean parameter (μ) is positive but statistically insignificant. This suggests that the average return is not different from zero after accounting for volatility dynamics. The autoregressive term (AR(1)) is negative and highly significant ($t = -13.02$, $p < 0.001$). The statistics indicate strong mean-reversion in the return process. This implies that a positive shock in returns is likely to be followed by a negative adjustment, consistent with short-term correction behavior in financial markets. The constant term (ω) is statistically insignificant, both under conventional and robust standard errors. The statistics suggest that the long-term baseline level of volatility is not distinguishable from zero when shocks and persistence are accounted for. The shock parameter (α_1), which measures the immediate impact of innovations on volatility, is negative and only marginally significant under conventional errors (p =

0.07). This weak evidence suggests that shocks may reduce volatility in the short run, although the effect is not robust across specifications. The persistence parameter (β_1) is large and highly significant ($\beta_1 \approx 0.92$, $p < 0.001$). These results indicate strong persistence in conditional volatility. This result suggests that volatility shocks have long-lasting effects, which is consistent with the well-documented volatility clustering phenomenon in financial returns. The asymmetry parameter (γ_1), which captures the leverage effect, is positive but statistically insignificant. This implies that there is no strong evidence of asymmetric volatility responses to positive and negative shocks in the data. Finally, the shape parameter, which governs the tail thickness of the distribution, is significant ($p = 0.01$ under robust errors). The estimated value (≈ 11.7) indicates heavier tails than the normal distribution. The result reflects the presence of extreme events in the return series. Even though the asymmetry parameter γ_1 is only marginally significant ($p \approx 0.12$), the strong persistence of volatility captured by β_1 and the ability to model fat tails indicate that the EGARCH specification provides a better description of the data compared to a standard GARCH model.



	Model with Optimal Parameters				**Model with Robust Standard Error**			
	Estimate	Std. error	t value	Pr(>|t|)	Estimate	Std. error	t value	Pr(>|t|)
Mu (mean)	0.002	0.003	0.54	0.59	0.002	0.003	0.60	0.55
ar1(lag-effect)	-0.586	0.045	-13.02	0.00	-0.586	0.044	-13.28	0.00
Omega (constant)	-0.370	0.478	-0.77	0.44	-0.370	1.114	-0.33	0.74
alpha1(shock)	-0.109	0.059	-1.84	0.07	-0.109	0.099	-1.10	0.27
beta1(volatility persistent)	0.923	0.099	9.30	0.00	0.923	0.232	3.98	0.00
gamma1(asymmetric)	0.273	0.174	1.57	0.12	0.273	0.403	0.68	0.50
Shape (tail fatness)	11.678	6.746	1.73	0.08	11.678	4.564	2.56	0.01

TABLE 10: Parameters of eGARCH(1,1)-ARFIMA(1,0,0)-t model

Source: Author

The signs of the significant estimates align with expected patterns: the negative AR(1) confirms short-term mean reversion, where high import increases are followed by corrections, and the large positive β_1 confirms strong volatility persistence, indicating that high-volatility periods cluster over time. The finite shape estimate further supports the presence of fat tails, meaning extreme import changes occur more frequently than predicted by a normal distribution. While the positive γ_1 indicates potential asymmetric volatility, its non-significance ($p \approx 0.12$) confirms the effect is not statistically robust. The near-zero, non-significant μ is consistent with the weak or absent predictable drift often observed in EGARCH models. Overall, as illustrated in Figure *6*, the model highlights that the conditional volatility, the variability of sales at a given time, influenced by past sales and past forecast errors, is highly persistent. The extreme events are more likely than normal. The mean-reversion is also evident in the series.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

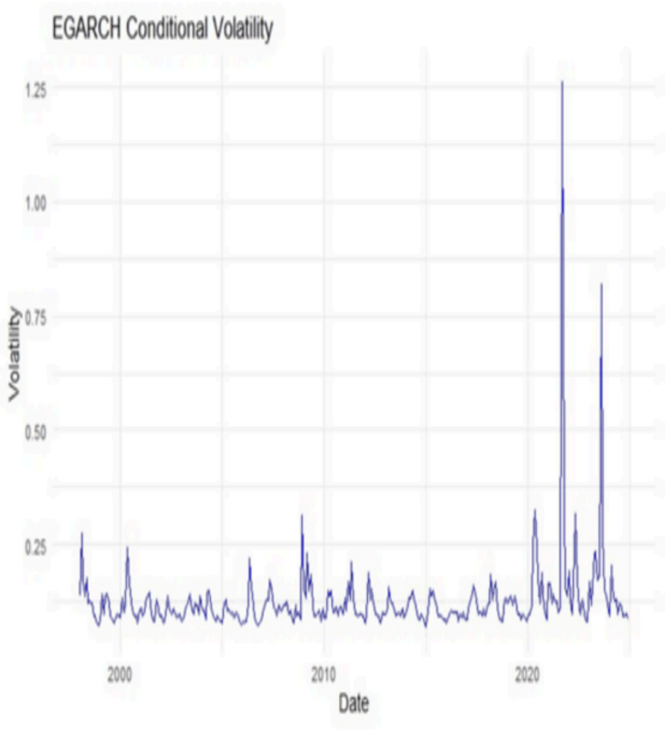


FIGURE 6: Car Sale Volatility Trend

Source: Author

EGARCH-Based Forecast

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

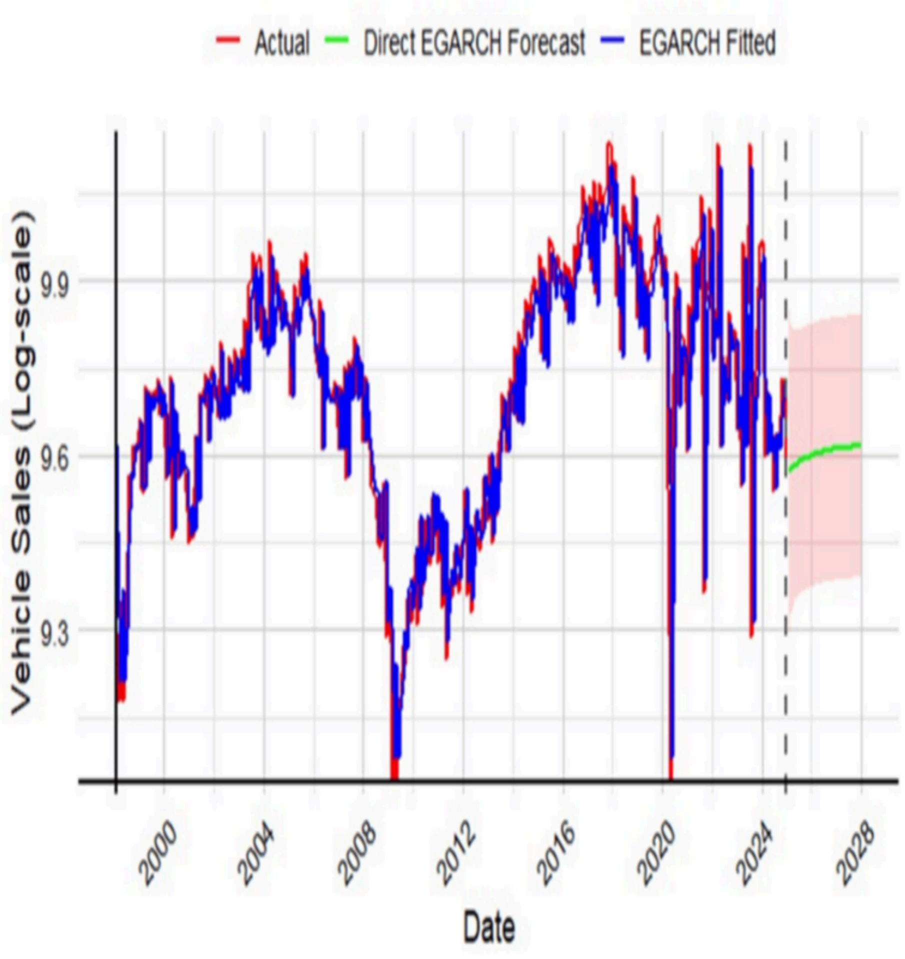


FIGURE 7: Visualization of Actual with Forecast from EGARCH Model

Source: Author. EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity

The EGARCH-based forecasts of the log-transformed car sales in New Zealand for the period January 2025 to December 2027 are illustrated in Figure *7*. The forecast line is much smoother. It has shown a steady upward drift from about 9.56 to 9.61 across the forecast horizon. The trajectory remains broadly consistent with the recent historical pattern, aligned with the long-term trend. In addition, the volatility is slightly higher at the start of the forecast horizon relative to later times. Recent data had high volatility or turbulence, which resulted in the phenomenon. However, the model assumes these shocks decay with time, and volatility gradually returns to its long-run average level. The forecasts of the EGARCH model is illustrated in Figure *8*.

The EGARCH model, by design, produces a more gradual and statistically consistent forecast with explicit volatility

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

estimation. In other words, this is a typical GARCH-family model that forecasts a gradual adjustment toward the long-run mean rather than short-term wiggly patterns. Its steady path, combined with conditional volatility information (σ), helps identify periods of expected uncertainty.

Forecasts Difference Between the LSTM and EGARCH Models

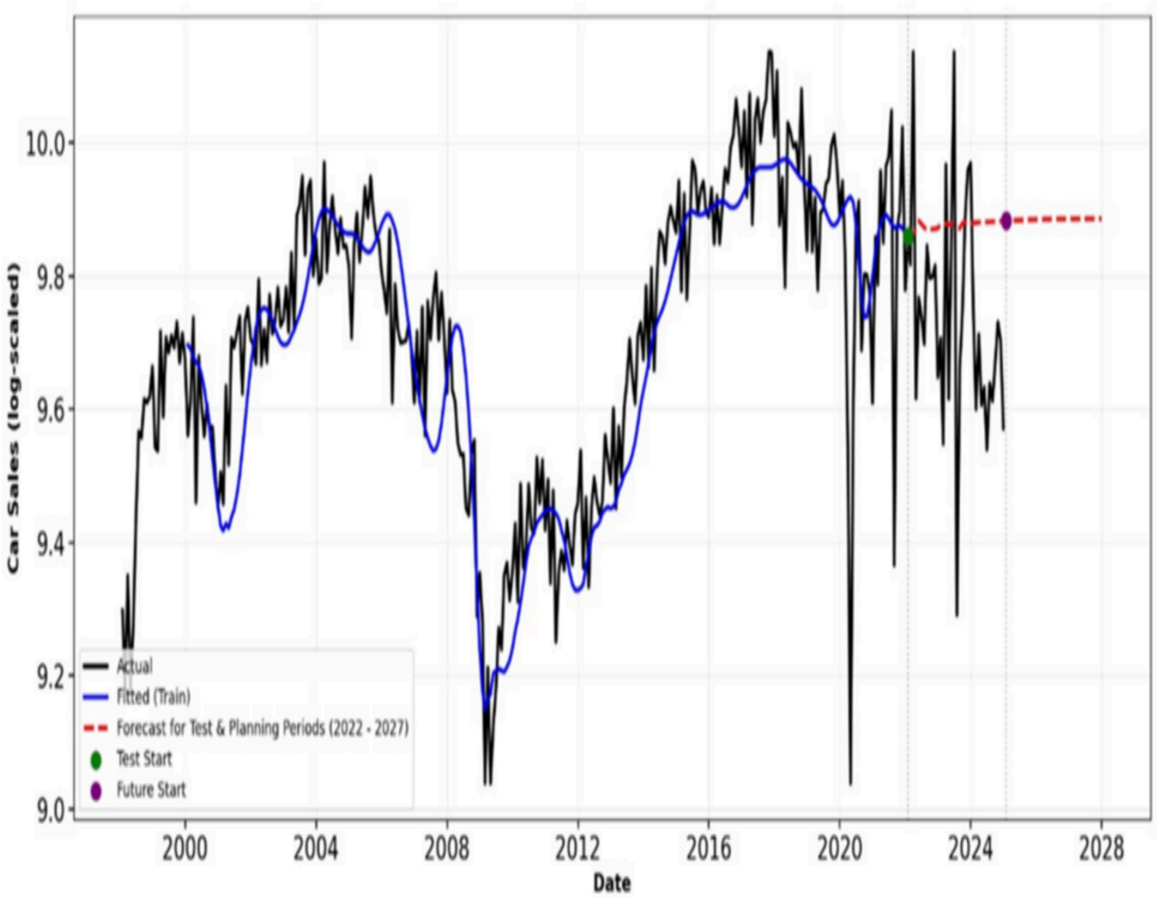


FIGURE 8: Visualization of Actual with Forecast from LSTM Model Trained on Split Dataset

Source: Author. LSTM, Long Short-Term Memory

Figure *8* and Figure *9* show results of trained on the split dataset and full dataset, respectively. They illustrate comparative positions of their forecasted lines relative to their fitted and actual lines. The fitted lines of LSTM models are shifted slightly to the right hand side. In recursive LSTM time series forecasting with a planning horizon of one, the model generates a one-step-ahead prediction using information available up to time t. During both in-sample fitting and out-of-sample forecasting, each predicted value is then fed back as an input to generate the next forecast. As a result, fitted or forecasted series are inherently aligned one time step ahead of the corresponding observed values. This apparent shift does not indicate a misalignment or plotting error; rather, it reflects the causal structure of recursive forecasting and the use of lagged information in the model. When combined with the accumulation of small prediction errors and imperfect learning of trend or seasonality, this mechanism can lead to gradual divergence between the forecasted and actual series over time.

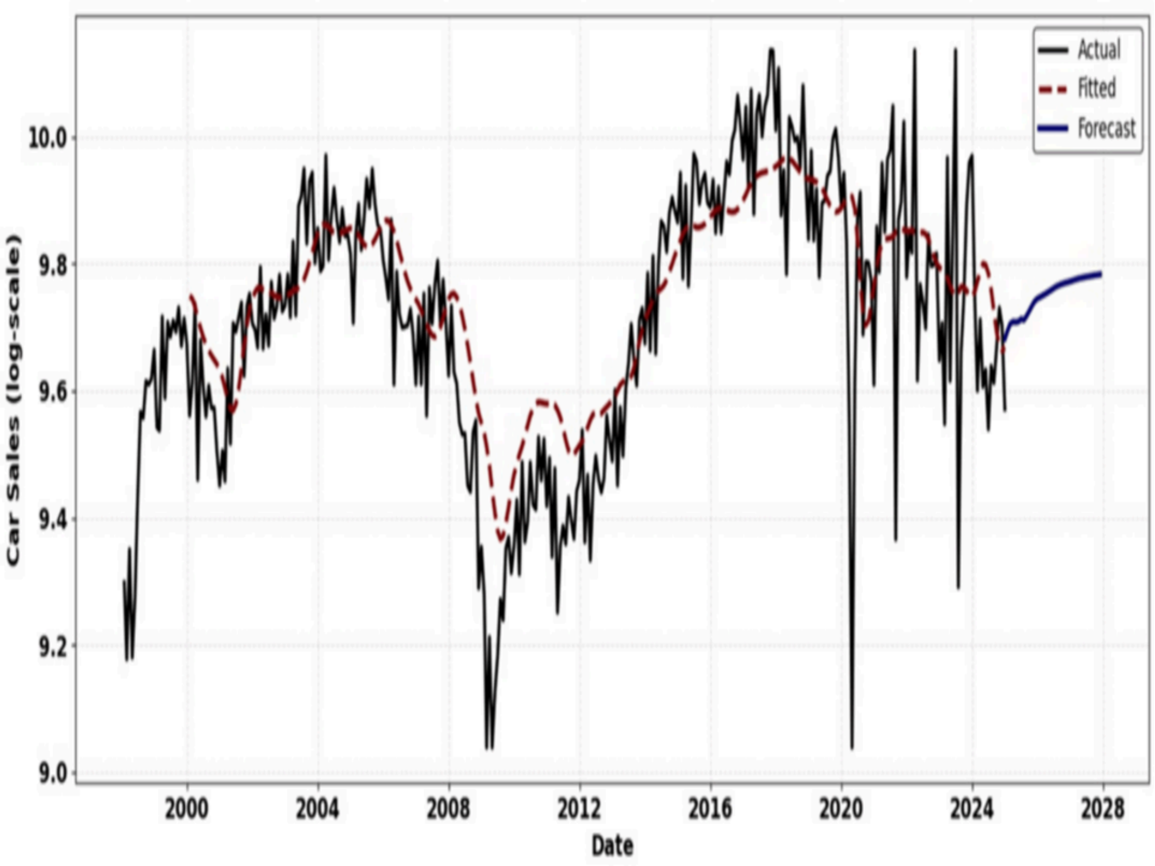


FIGURE 9: Visualization of Actual with Forecast from LSTM Model Trained on Full Dataset

Source: Author. LSTM, Long Short-Term Memory

The forecasts of split data LSTM model are noticeably higher and flatter than full data LSTM model. The trend of fitted values looks slightly lower than actual values in both cases. However, the origins of forecasted values are also closer to the end of fitted values. Nonetheless, the forecasted values at the beginning period are seen inconsistent with the available actual data distribution at the end period. This reflects as source of problems in this split data model.

These forecasts of LSTM models, in general, are approximate to the EGARCH model, but the dynamics of LSTM full data model appear more erratic at early period of planning horizon. The forecasts of the LSTM models are also higher than the EGARCH model. In addition, the LSTM models, particularly full data one, exhibited slightly up-and-down movements across the forecast horizon, while the EGARCH model produced a relatively straight line based on the historical mean. The reason for this behavior of the LSTM models is likely due to their ability to capture short-term patterns and non-linear patterns. The EGARCH models captures conditional volatility rather than directional trends. The EGARCH model does not explicitly show upward or downward trends. Instead, it provides a potential range of volatility based on the sigma value. It reflects uncertainty without a defined direction. This occurs because the EGARCH model models conditional variance rather than a defined trend. This makes EGARCH particularly valuable for risk-sensitive applications, even though its mean forecasts may appear flatter compared to machine learning models such as LSTM.

Comparison Forecast Between Split and Full Data Models

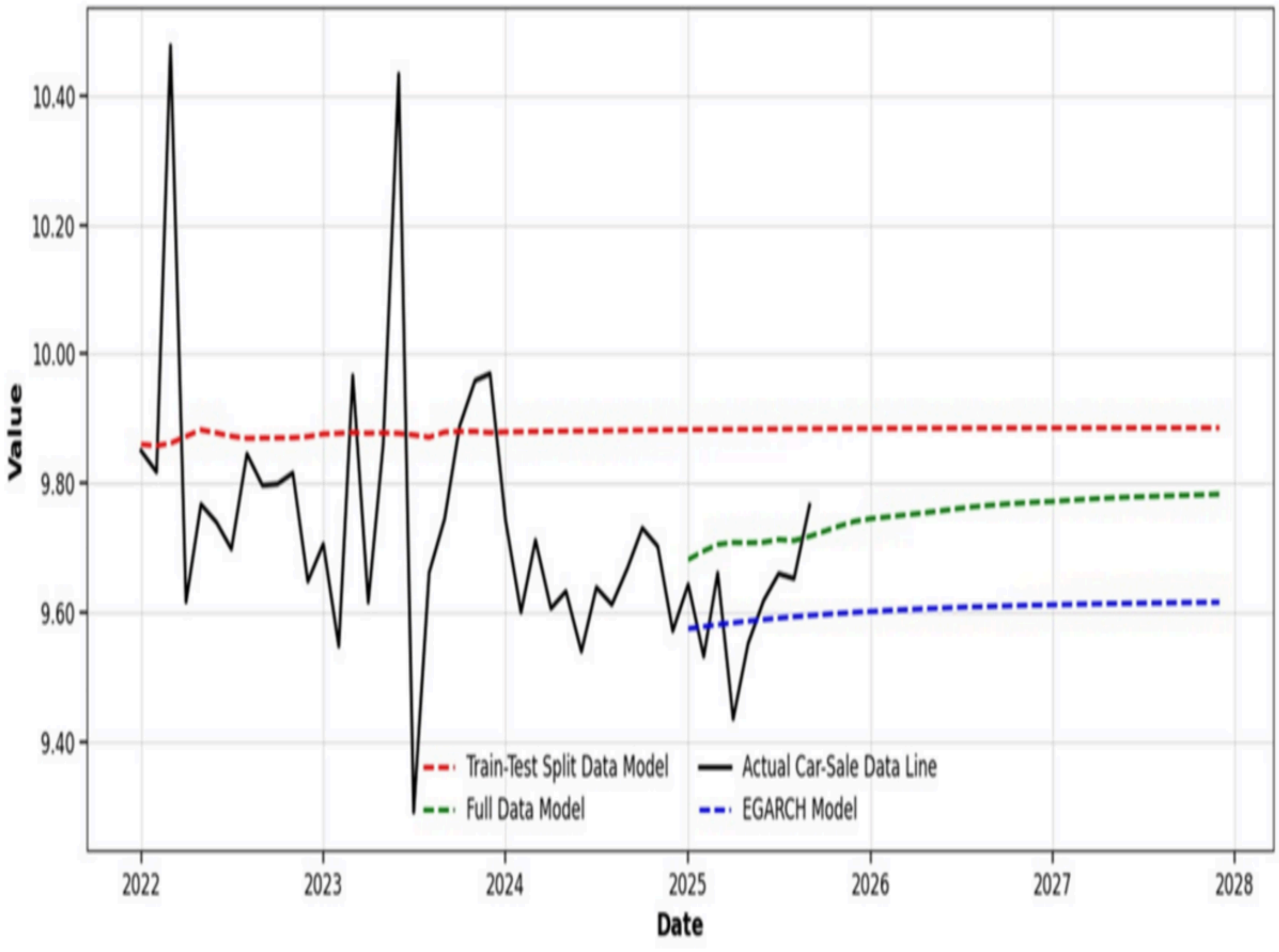


FIGURE 10: Comparative Forecasts of EGARCH, Full-Data LSTM, and Split-Data LSTM Models

Source: Author. LSTM, Long Short-Term Memory; EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity

Figure 10 illustrates the forecasted car sales from three models, the Train-Test Split Data LSTM model (red dashed line), the Full Data LSTM model (green dashed line), and the EGARCH model (blue dashed line), alongside the actual sales (solid black line). Observed sales are extended up to September 2025 to allow a direct comparison between the forecasts and reality. The Train-Test Split Data LSTM forecast originates in early 2022. Its trajectory shows slight monthly fluctuations but remains largely flat across the forecasting horizon. By contrast, the Full Data LSTM forecast, which begins in 2025, demonstrates a steadily increasing trend that gradually levels off toward the end of the planning horizon. However, its line is close to the level and trend of the split model at the end of the planning horizon. The EGARCH forecast, also starting in 2025, exhibits minor monthly variation with a gentle upward slope, though less pronounced than the Full Data LSTM trend.

The divergence between the EGARCH and LSTM forecast trajectories is caused by fundamental differences in how the two models handle long-run dynamics under recursive forecasting. Although both were estimated using a difference-stationary series, the AR-EGARCH framework imposes a mean-reverting structure on the differenced process. As a result, multi-step forecasts gradually stabilize over the planning horizon, producing a relatively smooth and linear projection path.

In contrast, the LSTM model does not impose explicit stationarity or mean-reversion constraints. When forecasts are generated recursively, even small upward drift patterns learned from the data can accumulate over successive steps. This compounding effect leads to progressively higher projections, creating a widening gap relative to the EGARCH trajectory.

Comparing the forecasts, it is apparent that the Split Data LSTM model predicts higher values than the Full Data model even at the start of 2025. Ideally, if the split data were representative of the full dataset, the forecasts in overlapping periods should align closely. The divergence suggests that the recent values in the training period of the split data had higher means than the overall full data series, giving more weight to recent fluctuations and producing elevated forecast levels. The EGARCH model, while also forecasting upward, remains closer to the actual data than the Split Data LSTM forecast, though it underestimates the overall trend relative to the Full Data LSTM model. Overall, the three forecasts differ in timing, slope, and magnitude. The Full Data LSTM captures a clear upward trajectory, the Split Data LSTM overestimates early, and the EGARCH model follows a moderate upward path. These distinctions highlight the sensitivity of LSTM forecasts to the data subset used for training and the tendency of recursive forecasts to amplify differences in recent observations.

A t-test for two samples assuming unequal variances was used to compare the forecasted values of these three models. The comparison results show that the three forecasting approaches produce systematically different outcomes. The Full-Data LSTM and EGARCH models generated very similar average predictions (means of 9.6307 and 9.6037, respectively), whereas the Split-Data LSTM model produced a higher mean of 9.8050, indicating a more optimistic projection of sales. The Split-Data LSTM model also exhibits substantially lower variance than the other two models, which suggests greater stability but potentially reduced sensitivity to fluctuations in the data. In contrast, the EGARCH model displays the highest variance, reflecting greater volatility in its predictions. The t-statistics for the mean comparisons were 10.55 for the full-sample LSTM and EGARCH models, 101.63 for the split-sample LSTM and EGARCH models, and 104.62 for the full-sample LSTM and split-sample LSTM models, with extremely small p-values ($p < 0.001$) for all pairwise tests. The results imply that the differences among the models are statistically meaningful and should be considered when selecting a forecasting approach.

LSTM and EGARCH Hybrid Forecast

The hybrid forecasting framework is a combination of a deterministic LSTM projection with stochastic disturbances generated by an EGARCH process. Figure 11 illustrates the forecast behavior of the integrated method. The LSTM model produced a smooth and steadily increasing forecast path over the period from January 2025 to December 2027. This trajectory indicates the model's ability to capture the underlying long-term dynamics of the log-transformed sales series while filtering out short-term irregularities. The hybrid forecast adds systematic deviations around the LSTM mean path. The EGARCH component generated these deviations. They introduced time-varying volatility and serially correlated shocks. During the initial part of the forecast horizon in 2025, the stochastic disturbances remained relatively small, and the hybrid forecast closely followed the LSTM trajectory. Table 11 shows that the conditional variance during this period remained low, indicating a stable volatility regime. The trajectories show growing departures from the LSTM mean during 2026 and onward. Several positive and negative shocks emerged, causing temporary fluctuations around the underlying trend. These movements coincided with an increase in the estimated conditional variance. These phenomena indicate the presence of volatility clustering. Despite these fluctuations, the long-run upward movement of the series remained intact, as the stochastic component altered short-term behavior gradually.

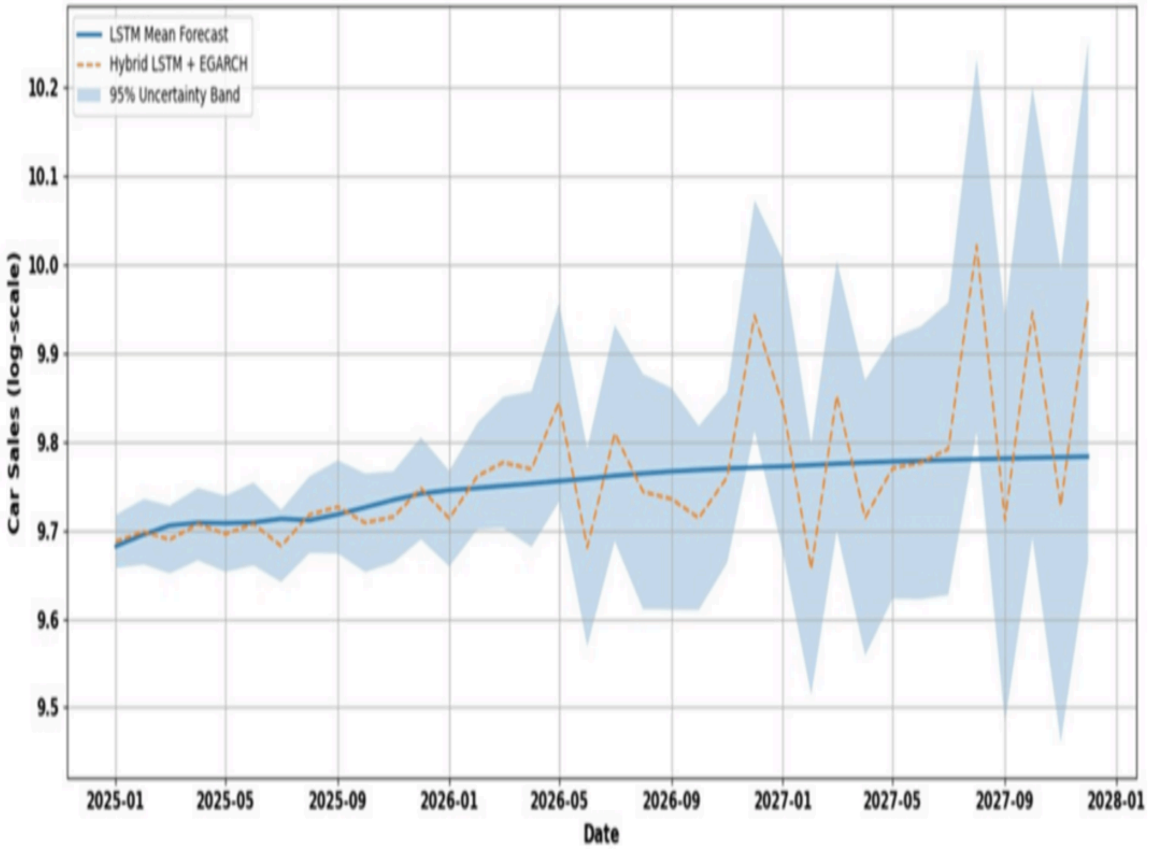


FIGURE 11: LSTM-Curved-Linearity and EGARCH Stochastic Properties Integrated Forecast

Source: Author. EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity; LSTM, Long Short-Term Memory

The hybrid forecast displayed more pronounced deviations from the LSTM mean, particularly toward the end of the horizon. These phenomena are due to substantially increasing conditional variance and corresponding standard deviation which led to a marked widening of the uncertainty bands. This pattern reflects the accumulation of forecast uncertainty over time rather than a deterioration of the underlying trend estimate. The period of 2025, starting from 2024, can be interpreted as a less fluctuating phase in the sales business shock cycle.

The combined results indicate that the LSTM component captures the expected long-term evolution of the series, while the EGARCH component represents realistic short-term uncertainty through heteroskedastic shocks. This integrated structure allows the model to produce forecasts that remain trend-consistent while explicitly accounting for time-varying risk. The widening uncertainty bands imply that forecast reliability declines as the horizon extends, which is a critical consideration for policy analysis and forward-looking decision-making. The hybrid approach therefore provides a more informative representation of future outcomes than a deterministic forecast alone.

SN.	Date	LSTM Forecast	EGARCH Shock	Hybrid Forecast	Conditional Variance	Conditional Std
1	1/01/2025	9.681957	0.005394	9.687351	0.000237	0.015394
2	1/02/2025	9.695276	0.003094	9.69837	0.000361	0.019012
3	1/03/2025	9.705391	-0.01589	9.6895	0.000384	0.019606
4	1/04/2025	9.708503	-0.00157	9.706935	0.000439	0.020951
5	1/05/2025	9.707956	-0.01195	9.696004	0.000484	0.021995
6	1/06/2025	9.708847	-0.00171	9.707138	0.000576	0.024002
7	1/07/2025	9.712968	-0.03062	9.682344	0.000436	0.02088
8	1/08/2025	9.711371	0.006215	9.717586	0.000491	0.022162
9	1/09/2025	9.717937	0.00843	9.726367	0.000722	0.026878
10	1/10/2025	9.725894	-0.01741	9.708484	0.000811	0.028471
11	1/11/2025	9.734279	-0.01915	9.715125	0.000697	0.026398
12	1/12/2025	9.741293	0.005987	9.74728	0.000866	0.029427
13	1/01/2026	9.745313	-0.03202	9.713295	0.000764	0.027636
14	1/02/2026	9.747705	0.01364	9.761346	0.000947	0.030766
15	1/03/2026	9.750166	0.026359	9.776524	0.001427	0.037781
16	1/04/2026	9.752707	0.016277	9.768983	0.002017	0.044914
17	1/05/2026	9.75549	0.08883	9.84432	0.003277	0.057245
18	1/06/2026	9.758419	-0.07786	9.680561	0.003255	0.057053
19	1/07/2026	9.761506	0.048076	9.809582	0.00387	0.062208
20	1/08/2026	9.764185	-0.02062	9.743569	0.004594	0.067779
21	1/09/2026	9.76632	-0.03094	9.735383	0.004087	0.063926
22	1/10/2026	9.76811	-0.05427	9.71384	0.002817	0.053075
23	1/11/2026	9.769707	-0.0098	9.759904	0.002439	0.049384
24	1/12/2026	9.770991	0.171346	9.942337	0.00443	0.066557

25	1/01/2027	9.77205	0.068469	9.840519	0.007099	0.084253
26	1/02/2027	9.773601	-0.11742	9.656182	0.005254	0.072488
27	1/03/2027	9.775082	0.076797	9.851879	0.006083	0.077996
28	1/04/2027	9.776423	-0.06244	9.713982	0.006297	0.079352
29	1/05/2027	9.77758	-0.00759	9.76999	0.005647	0.075149
30	1/06/2027	9.778569	-0.00237	9.776203	0.006186	0.078649
31	1/07/2027	9.779439	0.012482	9.791921	0.007106	0.084297
32	1/08/2027	9.780242	0.241304	10.02155	0.011551	0.107475
33	1/09/2027	9.780982	-0.06967	9.711317	0.013686	0.116986
34	1/10/2027	9.781715	0.164523	9.946238	0.016917	0.130064
35	1/11/2027	9.782442	-0.05518	9.727257	0.018498	0.136007
36	1/12/2027	9.783151	0.175941	9.959092	0.022374	0.149578

TABLE 11: Hybrid LSTM–EGARCH Forecasts and Conditional Volatility Estimates (2025–2027)

Source: Author. EGARCH, Exponential Generalized Autoregressive Conditional Heteroskedasticity; LSTM, Long Short-Term Memory

Discussion

This study explained the phenomenon of private car sales in a fully import-based and tariff-free country. The result showed that the imports increased reasonably after the tariff relaxation. The increase rate was much higher than the population growth rate for the period. The wave of import tariffs represents a kind of reduction in car prices. The tariff relaxes reduced transaction costs, which can be considered a driver of car sales. They increased affordability. However, the sale was substantially influenced by other shocks and recovered after a while from the shocks. Interestingly the COVID-19 had a short-term shock. The car sales boom immediately after COVID-19 was probably driven by low interest rates and better savings of income during the lockdown period. Global passenger car sale statistics shows the car sale recovery is still lower than pre-covid period worldwide including UK (Fastner-World 2022) (Fastner-World 2024) (Hall, 2025). However, the Australia has got good recovery but its growth is not similar to New Zealand. The driving factors are subject to further study. Cars have become essential goods in the daily lives of people in the country, which may be one reason why sales recovered quickly after the COVID-19 downturn. Findings of other studies on the trend of long-term car sales are not adequately comparable with the results of this study. However, Rimašauskas et al. (Rimašauskas et al., 2025) reported a significant decrease in car imports to European countries from 2017 to 2022, even after COVID-19. The trend can be considered a decline in sales. Partsvaniya (Partsvaniya, 2022) reported declining trends of both import and production of passenger cars in Russia from 2011 to 2020. This evidence suggested that the perspective of the New Zealand car sales business is still promising in the current global context. The tariff relaxation can be interpreted as a driver for the continuous growth of car sales.

The findings reveal that vehicle sales in New Zealand's tariff-free context exhibit pronounced volatility clustering, strong mean-reversion, and a heightened likelihood of extreme fluctuations. Such a pattern is common for market mechanism-based sales series (Asteriou and Hall, 2021). The significant negative autoregressive term confirms that sharp increases in imports or sales are typically followed by corrective declines. This suggests that short-term surges do not translate into a stable upward trajectory. At the same time, the large and highly significant persistence parameter indicates that volatility shocks tend to last. The results suggest that car sales will remain unsettled for long periods instead of quickly stabilizing. They also show that extreme sales disturbance events happen more often than normally expected. Although asymmetry effects were not statistically robust. The positive sign hints that both positive and negative shocks could contribute to uncertainty in similar ways.

The LSTM model forecast also showed that the sales remained steadily fluctuating but slightly upward unless other economically favorable conditions arises. A combination of these results suggests that while tariff removal has stimulated higher volumes of vehicle sales, the market continues to experience instability and heightened sensitivity to external shocks. For policymakers and service providers, this means that planning for infrastructure, environmental rules, and credit access should consider not only the steady growth in car ownership but also the chance of sudden and lasting swings in demand. The results of a steady increase by both forecast methods imply that the government requires gradually increasing efforts for future growth, while can use the other efforts to manage current challenges.

This study examined the forecasting implications of the common practice of splitting datasets into training and testing samples, with particular emphasis on time series data characterized by high variability in their underlying distributions. Splitting the dataset is useful for assessing model performance because it allows benchmarking against unseen data and helps guard against overfitting. It can provide a more realistic estimate of prediction error. This study investigated methodological issues arising from the split-data approach, including cross-fold validation in LSTM model evaluation. The results of seven-fold TSCV with both rolling-window and expanding-window approaches showed that forecasting performance can be sensitive to the choice of out-of-sample periods. More importantly, in both approaches, some folds produced very high out-of-sample errors and large negative R^2 values. These results indicated that the model sometimes performed worse than a naïve mean forecast. It was a key subject of investigation whether such poor metrics of model evaluation were caused by problems in the model itself or by limitations of the cross-validation procedure for time series data.

Distributional diagnostics of the training and test datasets for individual folds revealed that, in most folds, the training and test sets differed significantly in mean, variance, and overall distribution. These differences were confirmed statistically using Welch's t-test, Mann-Whitney U test, and Kolmogorov-Smirnov test. Such discrepancies help explain why the model struggled in certain folds: the deterioration in performance often reflected structural instability between training and test periods rather than intrinsic flaws in the model. The results also showed that the structural design of TSCV contributed to the observed variability. By construction, last temporal segments of data cannot be included in training. Recent data, which can carry important information about new trends or volatility changes, were left out during training. This limits the ability of the validation procedure to fully reflect the performance of a model trained on the complete dataset. Comparisons between level and first-differenced data showed that preprocessing had little impact: RMSE, MAE, and R^2 values were nearly identical across folds. The results indicated that the LSTM could learn trends, seasonal patterns, and absolute levels directly without requiring stationarity transformations. The analysis also revealed an underlying cause of negative out-of-sample R-square result.

This study also evaluated what can implications of split-data model in forecasting in real world or beyond experimentation. The split-dataset models exhibited higher explanatory power, as measured by R-square, than models trained on the full dataset. Nevertheless, these higher values should be interpreted with caution. The exclusion of a substantial portion of observations, particularly those with noticeably different distributional characteristics, limits the adequacy of R-square as an indicator of overall model quality. High explanatory power under reduced sample sizes does not necessarily translate into more reliable or meaningful forecasts. Consistent with this limitation, the study found statistically significant differences in forecast outcomes between models trained on full datasets and those trained on split datasets, with full-dataset models producing forecasts that were more consistent and logically aligned with observed dynamics.

Excluding part of the dataset from final model estimation and forecasting is therefore problematic, as it prevents the model from learning from all available information. In this study, the training and test datasets differed both statistically and observationally; for example, the mean level of car sales was higher in the test dataset than in the training dataset. When such distributional differences exist, comparisons of forecast errors across models become less informative. More importantly, the full-dataset model generated forecasts that were closer to observed outcomes in the earlier forecast periods, although the difference narrowed toward the end of the planning horizon. This issue is particularly relevant for machine learning methods such as LSTM models, which require larger sample sizes to support longer look-back windows and extended forecast horizons. Splitting the data further constrains the effective sample size and can weaken the model's ability to capture underlying dynamics. From a statistical perspective, the loss of observations associated with data splitting can reduce true forecasting power by discarding relevant information. By examining this problem in unevenly volatile time series data, this study proposes an approach that alleviates these limitations and improves the robustness and reliability of long-term forecasts. The approach is especially valuable for empirical studies conducted with relatively small datasets.

Previous studies have compared the performance of EGARCH and LSTM models primarily using error-based measures such as MSE and MAE (Pimentel et al., 2025), (Kakade et al., 2022), and (Kim and Won, 2018). For instance, Kim and Won (2018) concluded that a modified LSTM model outperforms GARCH-family models across usable window sizes based on these error metrics. In contrast, the findings of this study suggest that comparing models solely on the basis of forecast errors is not a sufficient criterion for assessing robustness. For business decision-making, forecasts are often valued more for their ability to capture trends and underlying patterns than for precise magnitudes alone. Consequently, a model's capacity to explain and reflect dynamic behavior, such as the LSTM and hybrid models, remains a more critical consideration.

This study showed comparably lower and statistically different forecasts of the EGARCH model than LSTM model. The forecast differences between the EGARCH and LSTM models are not unexpected, as they were estimated using completely different methods and data transformation approaches. The results showed that the LSTM followed the recent upward drift and short-term fluctuations in the data. It captured level shifts (mean behavior), which is likely due to adapting to recent trends, producing higher forecast levels. The EGARCH, by contrast, forecasted lower values because it mean-reverts around the long-run average and models volatility dynamics. More importantly, the LSTM captured patterns in the mean and potentially in volatility, whereas the EGARCH was designed primarily to capture volatility. This study showed differences in forecasted values between the models that were not well explained in previous studies. More importantly the LSTM model captured forecast dynamics over EGARCH model.

This study produced forecasts by combining LSTM and EGARCH models and showed more dynamics than a single model. In this study, the LSTM model captured the mean and trend (first moment) of the data, while the EGARCH model focused on volatility (second moment). The LSTM provided point forecasts of the expected value, whereas the EGARCH produced forecasts of risk and uncertainty in terms of variance and volatility. The dynamics of the results showed that the LSTM was effective in capturing nonlinear patterns and long-term dependencies in the series. LSTM gave a clear view of the expected level and direction of the data, while EGARCH added information about the degree of uncertainty and variability around that path. In other words, one method explained where the series is likely heading, while another explained how stable or risky that path may be. The results of the two models can be said to complement each other. The findings of this study imply that business groups that focus more on expected values and long-term trends can benefit more from the LSTM forecasts, while those concerned with volatility, uncertainty, or risk can gain more from the EGARCH model. The information from the combined approach will be more meaningful for business decision-making.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

Conclusions

This study analyzed private car sales in New Zealand using monthly observations from 1998 to 2024, which is an import duty-free period. New Zealand provides a useful case because it is fully dependent on vehicle imports and operates without tariffs. The results indicate that the removal of tariffs coincided with a sustained increase in car sales over the long run. This expansion, however, did not occur smoothly. Sales were marked by substantial volatility, with clear clustering of high and low fluctuations, frequent extreme movements, and a strong tendency to revert to the long-run mean. These phenomena indicate that the car market reacts sharply to external disturbances, but it does not remain permanently displaced by them.

The ups and downs in car sales reflected the effects of the economic recessions and booms and the COVID-19 pandemic. Sales fell abruptly at the onset of the shock but recovered quickly once restrictions eased. The speed of this recovery suggests that underlying demand conditions remained intact. While tariff liberalization alone cannot explain this outcome, it appears to have worked alongside other supportive factors. In particular, New Zealand's relatively well-developed transport infrastructure, access to credit, insurance coverage, and policy support likely contributed to the market's ability to absorb the shock and rebound. Forecasts for the period 2025-2027 indicate that sales are expected to continue rising, though at a slower pace than in history. Growth appears to be levelling off, with sales approaching a peak rather than entering a new expansion phase. For other tariff-free and import-dependent economies, the evidence suggests that open trade policies are most effective when combined with strong institutional and infrastructural support.

This study demonstrates that cross-fold validation for LSTM models in time-series forecasting has important methodological limitations. The evidence showed that time-series cross-validation can produce very large out-of-sample errors and strongly negative R^2 values when structural differences exist between training and test segments, particularly in upward-trending series, even when the model itself is not mis-specified. Comparisons between level and first-differenced data confirmed that these distortions were not driven by scaling or non-stationarity adjustments. Because the TSCV framework systematically excludes certain temporal segments - especially recent observations - from estimation, it cannot provide a fully robust assessment of a model ultimately trained on the complete dataset. These issues are arising mainly due to data splitting practice. These findings suggest the need for alternative evaluation approaches for time-series-based LSTM models.

This study showed that forecasts based on split datasets differ systematically from those obtained using the full sample. They can often be less consistent, particularly at the beginning of the forecast horizon. Despite the marginal difference of in-sample errors and R-square value, the forecasted values can be statistically and physically different. Models estimated on the full dataset yielded justifiable forecasts, which suggest that they can better capture the underlying data-generating process. These findings imply that, in short and highly variable series, full-sample estimation may be more appropriate and robust for medium- to long-term planning than strict adherence to data-splitting rules borrowed from other machine-learning contexts.

Although the LSTM and EGARCH models produced close results regarding future sales movements, they conveyed different types of information. The LSTM model projected slightly higher upward future sales levels than EGARCH despite a close forecasting origin, reflecting its capacity to capture nonlinear dynamics and recent patterns in the data. In contrast, the EGARCH-based forecasts remained more conservative, consistent with their reliance on parametric structures and mean-reverting behavior. They produced volatility persistence and mean reversion, making them more informative about uncertainty and downside risk. When their results were integrated, the two approaches provided a more complete picture of market behaviour, capturing both nonlinear dynamics and stochastic volatility. This combined perspective is likely to be more useful for decision-makers than relying on a single model alone, particularly in markets exposed to frequent shocks and non-linear phenomena.

In essence, the results suggest that New Zealand's car market has combined long-run growth with considerable short-run instability. Car sales are undoubtedly driven by tariff-free access, but quick recovery from shocks and resilience are more about institutional capacity and market infrastructure. These conclusions are likely to be relevant for other small, open economies where vehicle markets are exposed to external shocks and policy uncertainty.

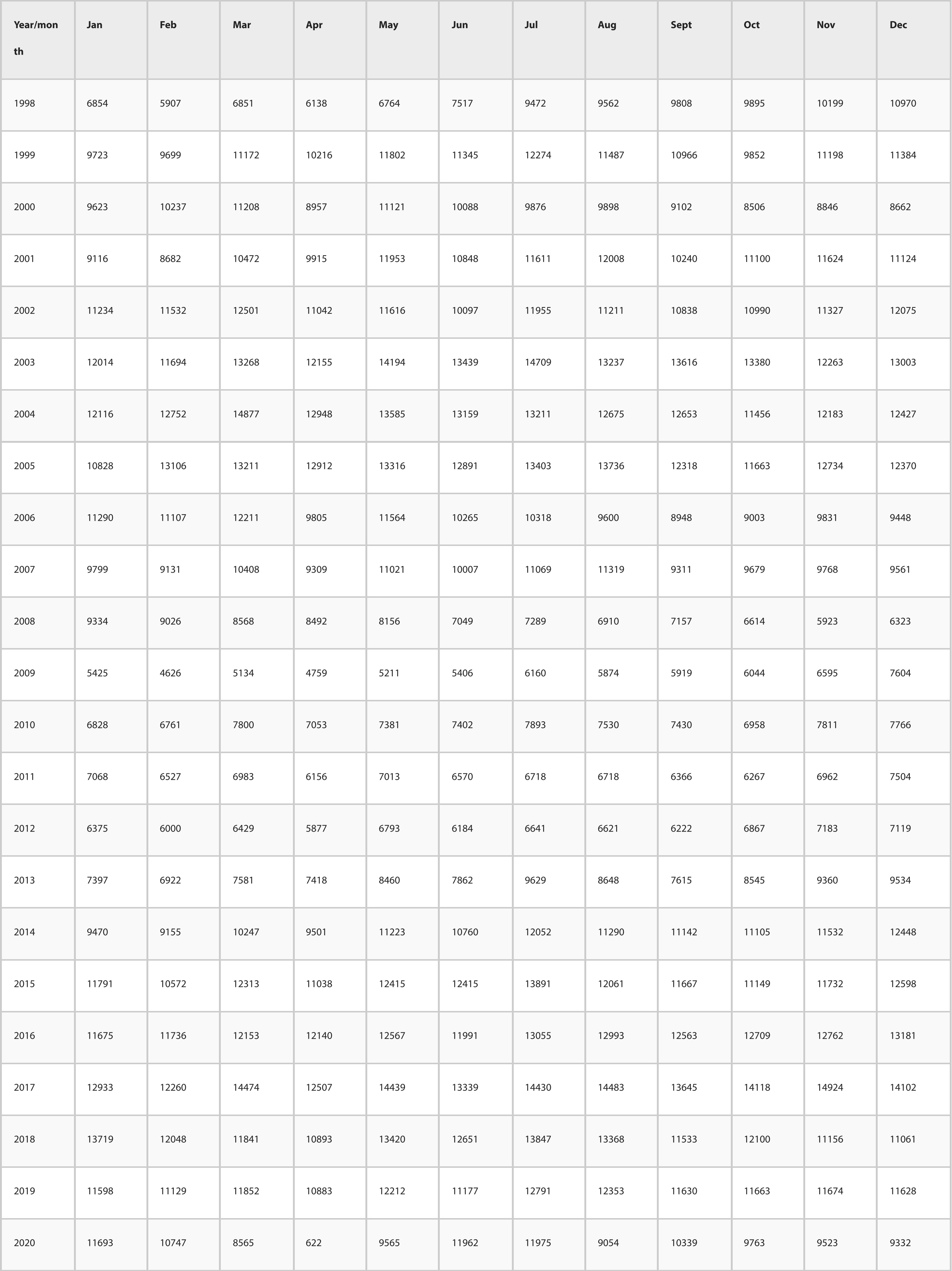
Limitations and future research

This study has substantial limitations. First, it focuses on a single country, New Zealand, where infrastructure, road networks, and institutional facilities supporting car sales are reasonably well developed. Many institutional conditions have also made private vehicles an essential part of daily life for households and businesses. However, the situation in other countries may differ substantially. Secondly, the dataset covering 27 years (324 months) is still relatively modest for adequately train deep learning models such as LSTM. Thirdly, the analysis relied primarily on sales series data. It did not include the sales influencing factors such as interest rates, fuel prices, household income, or exchange rate fluctuations. Similarly, the link between tariff relaxation and sales growth was inferred. It was not rigorously tested with counterfactual or comparative evidence. Fourthly, the modelling approaches themselves (LSTM and EGARCH) are technically different in terms of design and data processing, which affects the results. Finally, this study evaluated model performance between full-dataset and split-dataset training on a modest dataset and relatively simple test methods. The model was evaluated only on data of light-vehicle sales, which were highly influenced by external and internal shocks. These issues may have led to results that are somewhat different from or incomparable to other conditions. It is suggested that future studies work with these limitations or address these issues to rectify the findings.

Appendices

The following appendices present the detailed data and results used in this study, including the raw monthly car sales (Table *12*), cross-fold validation results for rolling and expanding windows (Table *13*), tests for the distribution of training and test datasets for 5-fold rolling and expanding window cross-validation methods (Tables *14* and 15), and distribution properties of training and test datasets for 7-fold expanding window cross-validation (Table *16*)

1. Appendix: Data



Year/month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1998	6854	5907	6851	6138	6764	7517	9472	9562	9808	9895	10199	10970
1999	9723	9699	11172	10216	11802	11345	12274	11487	10966	9852	11198	11384
2000	9623	10237	11208	8957	11121	10088	9876	9898	9102	8506	8846	8662
2001	9116	8682	10472	9915	11953	10848	11611	12008	10240	11100	11624	11124
2002	11234	11532	12501	11042	11616	10097	11955	11211	10838	10990	11327	12075
2003	12014	11694	13268	12155	14194	13439	14709	13237	13616	13380	12263	13003
2004	12116	12752	14877	12948	13585	13159	13211	12675	12653	11456	12183	12427
2005	10828	13106	13211	12912	13316	12891	13403	13736	12318	11663	12734	12370
2006	11290	11107	12211	9805	11564	10265	10318	9600	8948	9003	9831	9448
2007	9799	9131	10408	9309	11021	10007	11069	11319	9311	9679	9768	9561
2008	9334	9026	8568	8492	8156	7049	7289	6910	7157	6614	5923	6323
2009	5425	4626	5134	4759	5211	5406	6160	5874	5919	6044	6595	7604
2010	6828	6761	7800	7053	7381	7402	7893	7530	7430	6958	7811	7766
2011	7068	6527	6983	6156	7013	6570	6718	6718	6366	6267	6962	7504
2012	6375	6000	6429	5877	6793	6184	6641	6621	6222	6867	7183	7119
2013	7397	6922	7581	7418	8460	7862	9629	8648	7615	8545	9360	9534
2014	9470	9155	10247	9501	11223	10760	12052	11290	11142	11105	11532	12448
2015	11791	10572	12313	11038	12415	12415	13891	12061	11667	11149	11732	12598
2016	11675	11736	12153	12140	12567	11991	13055	12993	12563	12709	12762	13181
2017	12933	12260	14474	12507	14439	13339	14430	14483	13645	14118	14924	14102
2018	13719	12048	11841	10893	13420	12651	13847	13368	11533	12100	11156	11061
2019	11598	11129	11852	10883	12212	11177	12791	12353	11630	11663	11674	11628
2020	11693	10747	8565	622	9565	11962	11975	9054	10339	9763	9523	9332

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

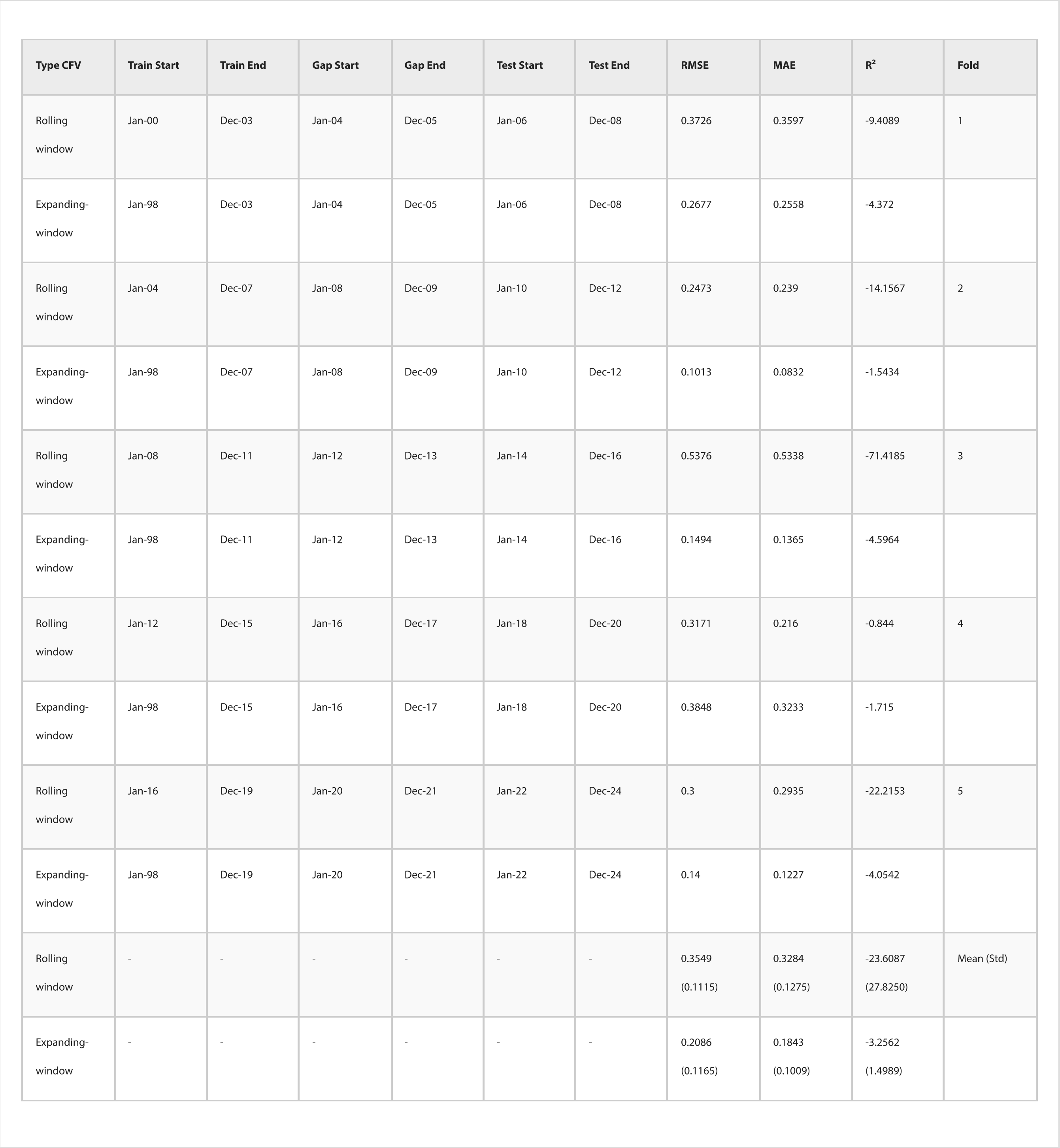
2021	9400	9091	11008	10134	11252	11256	13127	7099	7440	10635	11333	10109
2022	10112	10734	24346	6515	6886	7417	8243	7939	7375	7240	7156	6993
2023	7145	7762	9683	7945	9589	16680	6495	8730	8921	9655	9651	12812
2024	8920	8826	8947	8735	8890	7834	9008	8185	7220	7319	6923	7139

TABLE 12: Summary of the Raw Data Used in this Study

Source: MIA, 2025.

2. Appendix: Results

Cross-Fold-Validation Results

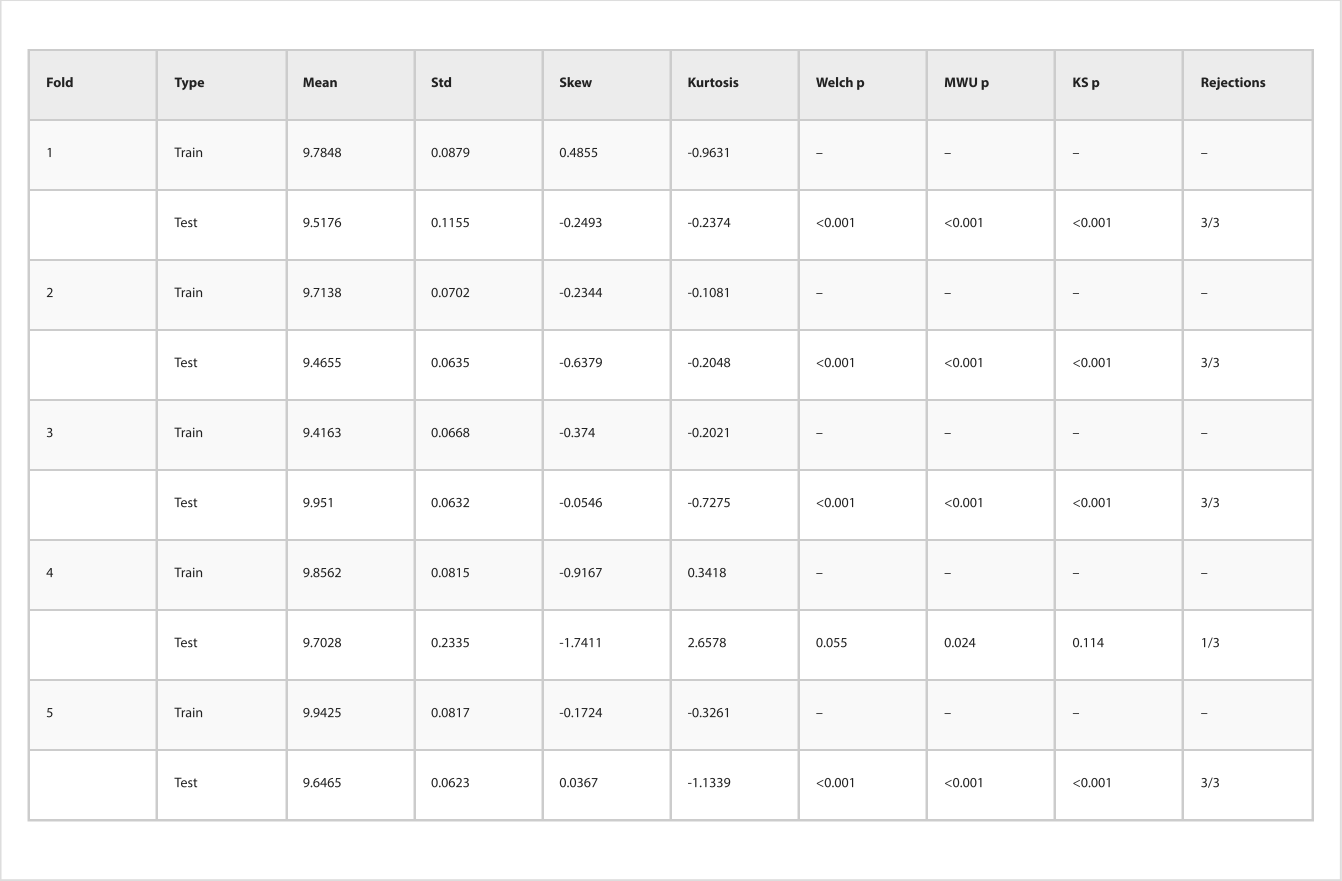


Type CFV	Train Start	Train End	Gap Start	Gap End	Test Start	Test End	RMSE	MAE	R²	Fold
Rolling window	Jan-00	Dec-03	Jan-04	Dec-05	Jan-06	Dec-08	0.3726	0.3597	-9.4089	1
Expanding-window	Jan-98	Dec-03	Jan-04	Dec-05	Jan-06	Dec-08	0.2677	0.2558	-4.372	
Rolling window	Jan-04	Dec-07	Jan-08	Dec-09	Jan-10	Dec-12	0.2473	0.239	-14.1567	2
Expanding-window	Jan-98	Dec-07	Jan-08	Dec-09	Jan-10	Dec-12	0.1013	0.0832	-1.5434	
Rolling window	Jan-08	Dec-11	Jan-12	Dec-13	Jan-14	Dec-16	0.5376	0.5338	-71.4185	3
Expanding-window	Jan-98	Dec-11	Jan-12	Dec-13	Jan-14	Dec-16	0.1494	0.1365	-4.5964	
Rolling window	Jan-12	Dec-15	Jan-16	Dec-17	Jan-18	Dec-20	0.3171	0.216	-0.844	4
Expanding-window	Jan-98	Dec-15	Jan-16	Dec-17	Jan-18	Dec-20	0.3848	0.3233	-1.715	
Rolling window	Jan-16	Dec-19	Jan-20	Dec-21	Jan-22	Dec-24	0.3	0.2935	-22.2153	5
Expanding-window	Jan-98	Dec-19	Jan-20	Dec-21	Jan-22	Dec-24	0.14	0.1227	-4.0542	
Rolling window	-	-	-	-	-	-	0.3549 (0.1115)	0.3284 (0.1275)	-23.6087 (27.8250)	Mean (Std)
Expanding-window	-	-	-	-	-	-	0.2086 (0.1165)	0.1843 (0.1009)	-3.2562 (1.4989)	

TABLE 13: Results of Five-Folds Rolling Window and Expanding Window Cross-Validation Methods

Source: Author

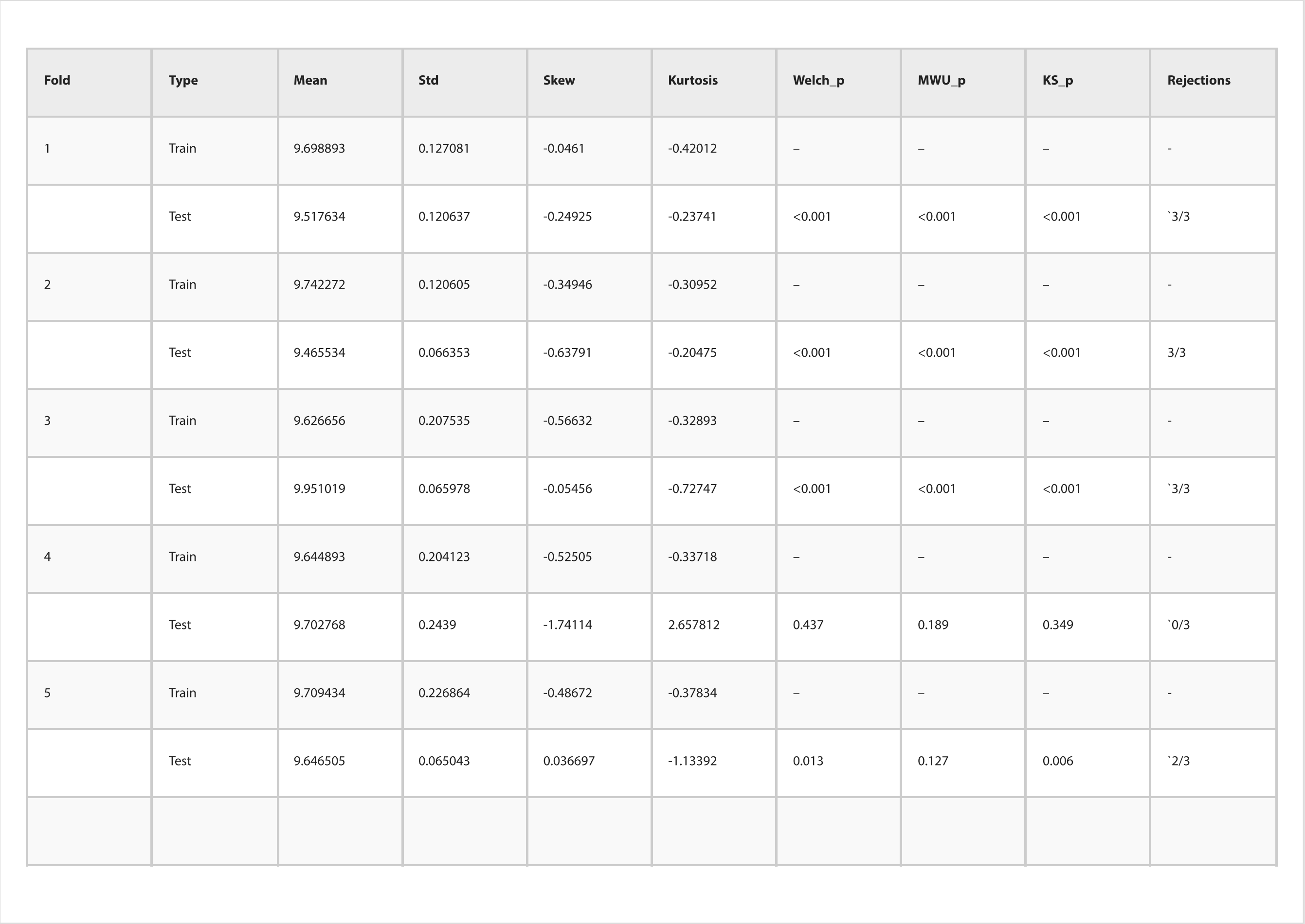
How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6



Fold	Type	Mean	Std	Skew	Kurtosis	Welch p	MWU p	KS p	Rejections
1	Train	9.7848	0.0879	0.4855	-0.9631	–	–	–	–
	Test	9.5176	0.1155	-0.2493	-0.2374	<0.001	<0.001	<0.001	3/3
2	Train	9.7138	0.0702	-0.2344	-0.1081	–	–	–	–
	Test	9.4655	0.0635	-0.6379	-0.2048	<0.001	<0.001	<0.001	3/3
3	Train	9.4163	0.0668	-0.374	-0.2021	–	–	–	–
	Test	9.951	0.0632	-0.0546	-0.7275	<0.001	<0.001	<0.001	3/3
4	Train	9.8562	0.0815	-0.9167	0.3418	–	–	–	–
	Test	9.7028	0.2335	-1.7411	2.6578	0.055	0.024	0.114	1/3
5	Train	9.9425	0.0817	-0.1724	-0.3261	–	–	–	–
	Test	9.6465	0.0623	0.0367	-1.1339	<0.001	<0.001	<0.001	3/3

TABLE 14: Test for Distribution of Training and Test Datasets of 5 Fold Rolling Window Cross-Validation

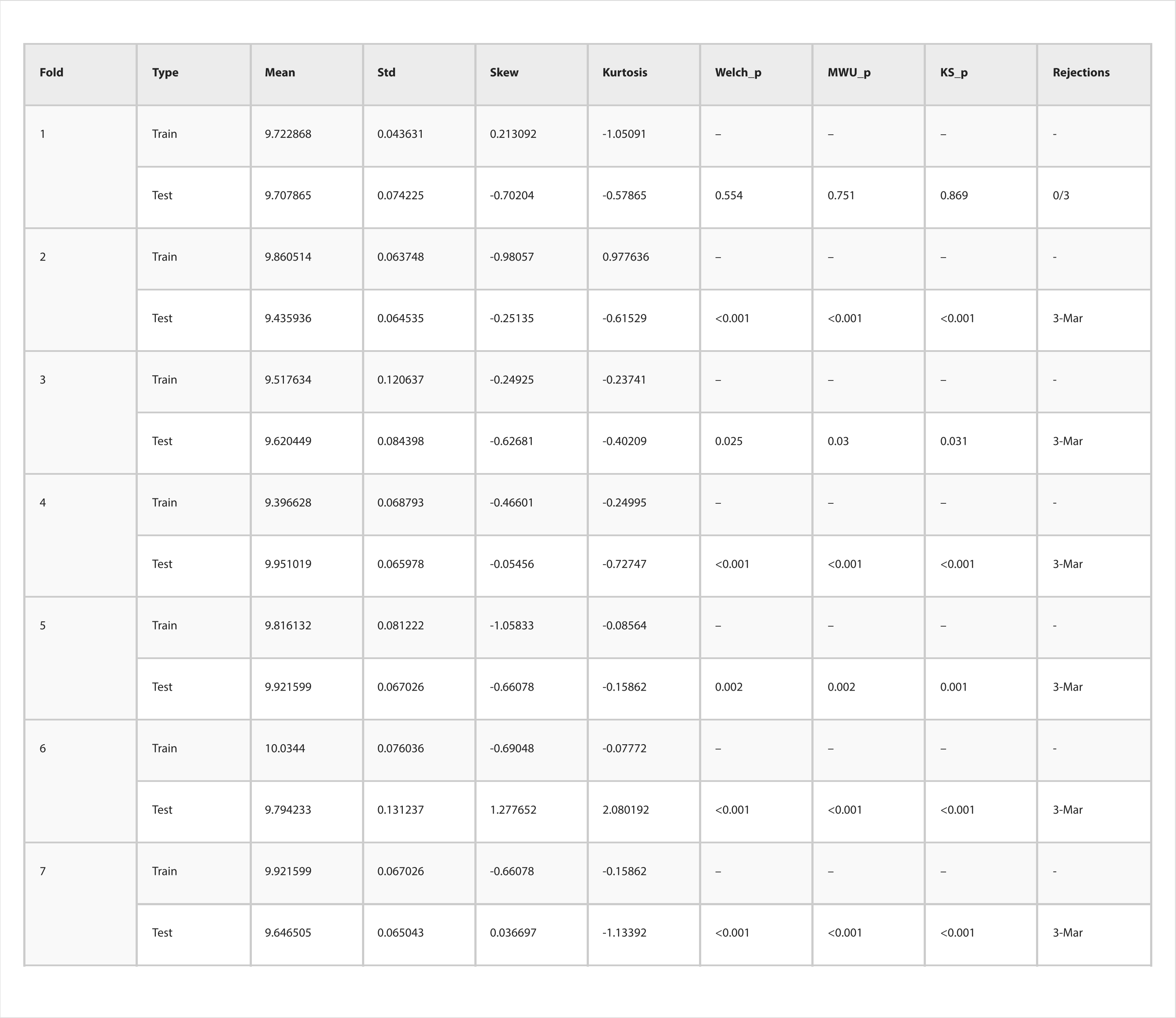
Source: Author



Fold	Type	Mean	Std	Skew	Kurtosis	Welch_p	MWU_p	KS_p	Rejections
1	Train	9.698893	0.127081	-0.0461	-0.42012	–	–	–	-
	Test	9.517634	0.120637	-0.24925	-0.23741	<0.001	<0.001	<0.001	`3/3
2	Train	9.742272	0.120605	-0.34946	-0.30952	–	–	–	-
	Test	9.465534	0.066353	-0.63791	-0.20475	<0.001	<0.001	<0.001	3/3
3	Train	9.626656	0.207535	-0.56632	-0.32893	–	–	–	-
	Test	9.951019	0.065978	-0.05456	-0.72747	<0.001	<0.001	<0.001	`3/3
4	Train	9.644893	0.204123	-0.52505	-0.33718	–	–	–	-
	Test	9.702768	0.2439	-1.74114	2.657812	0.437	0.189	0.349	`0/3
5	Train	9.709434	0.226864	-0.48672	-0.37834	–	–	–	-
	Test	9.646505	0.065043	0.036697	-1.13392	0.013	0.127	0.006	`2/3

TABLE 15: Test Results for Distribution Properties of Training and Test Datasets of 5 Fold Expanding Window Cross-Validation Method

Source: Author



Fold	Type	Mean	Std	Skew	Kurtosis	Welch_p	MWU_p	KS_p	Rejections
1	Train	9.722868	0.043631	0.213092	-1.05091	–	–	–	-
	Test	9.707865	0.074225	-0.70204	-0.57865	0.554	0.751	0.869	0/3
2	Train	9.860514	0.063748	-0.98057	0.977636	–	–	–	-
	Test	9.435936	0.064535	-0.25135	-0.61529	<0.001	<0.001	<0.001	3-Mar
3	Train	9.517634	0.120637	-0.24925	-0.23741	–	–	–	-
	Test	9.620449	0.084398	-0.62681	-0.40209	0.025	0.03	0.031	3-Mar
4	Train	9.396628	0.068793	-0.46601	-0.24995	–	–	–	-
	Test	9.951019	0.065978	-0.05456	-0.72747	<0.001	<0.001	<0.001	3-Mar
5	Train	9.816132	0.081222	-1.05833	-0.08564	–	–	–	-
	Test	9.921599	0.067026	-0.66078	-0.15862	0.002	0.002	0.001	3-Mar
6	Train	10.0344	0.076036	-0.69048	-0.07772	–	–	–	-
	Test	9.794233	0.131237	1.277652	2.080192	<0.001	<0.001	<0.001	3-Mar
7	Train	9.921599	0.067026	-0.66078	-0.15862	–	–	–	-
	Test	9.646505	0.065043	0.036697	-1.13392	<0.001	<0.001	<0.001	3-Mar

TABLE 16: Test Results of Distribution Properties of Training and Test Datasets of 7-Fold Expanding Window Cross-Validation

Source: Author

Methods: hybrid LSTM + EGARCH forecast

The hybrid forecasting approach combines predictions from a recursive LSTM model with stochastic volatility simulated using an EGARCH model. The goal is to produce forecasts that capture both the nonlinear trends (from LSTM) and the time-varying uncertainty or shocks in car sales (from EGARCH).

LSTM Forecasts

The LSTM model is first trained on historical car sales data and used to produce recursive forecasts for the target horizon (e.g., 2025-2027). These forecasts represent the expected sales levels based on past patterns.

EGARCH Parameterization

EGARCH parameters (ω, α, β, γ, φ, μ, ν) are specified to capture volatility dynamics:

α controls the response to past shocks' magnitude.

β determines the persistence of past variance.

γ captures asymmetry, i.e., whether positive or negative shocks have different effects.

φ incorporates autoregressive effects on residuals.

ν is the degrees-of-freedom parameter for the Student-t distribution of shocks.

Initialization

The simulation starts by computing the initial residual (difference between consecutive LSTM forecasts) and initial conditional variance from early LSTM forecast values. This sets the starting point for the stochastic volatility process.

EGARCH Simulation

For each forecast step:

A random shock is drawn from a Student-t distribution with ν degrees of freedom.

The logarithm of the conditional variance is updated recursively based on the previous variance, the new shock, and EGARCH parameters.

The residual for that step is computed as the autoregressive effect of the previous residual plus the scaled stochastic shock.

Conditional variance and standard deviation are calculated for each step.

Hybrid Forecast Construction

The stochastic residuals generated by the EGARCH process are added to the LSTM forecasts at each time step. This produces the hybrid forecast, which incorporates both the trend predicted by the LSTM and the time-varying shocks modeled by EGARCH.

Uncertainty Estimation

Conditional standard deviations from EGARCH are used to construct 95% confidence bands around the hybrid forecast, reflecting the expected range of variation due to stochastic volatility.

Output

The final dataset contains the LSTM forecasts, EGARCH shocks, hybrid forecasts, and associated conditional variances and standard deviations. Visualizations compare the LSTM-only forecast with the hybrid forecast and include uncertainty bands to highlight the effects of stochastic shocks.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Bhubaneswor Dhakal

Acquisition, analysis, or interpretation of data: Bhubaneswor Dhakal

Drafting of the manuscript: Bhubaneswor Dhakal

Critical review of the manuscript for important intellectual content: Bhubaneswor Dhakal

Supervision: Bhubaneswor Dhakal

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The initial work of this study was carried out as part of a project for the Postgraduate Diploma in Data Science, at the University of Canterbury, New Zealand. The data of the study are from open sources. They are available from the Motor Industry Association, 2025. Assistance from a large language model was used to develop plotting code, debug data analysis scripts, and edit the language of the written report. This study used the TensorFlow platform to build, train, and deploy the LSTM model. The reproduced results of the LSTM model may not be precise due to the precision limitations of TensorFlow. There may be some possibility of getting the optimum combination of hyperparameter tuning, which might make small differences in the results. Lastly, I would like to thank Mr Subodh Dhakal of NZTA for assisting me in obtaining the online data used at the start of this project.

References

1. Abotaleb M, Dutta PK: Optimizing convolutional neural networks for univariate time series forecasting: a comprehensive guide. Hybrid Information Systems: Non-Linear Optimization Strategies with Artificial Intelligence. 2024, 459-472. 10.1515/9783111331133
2. An C, Park YW, Ahn SS, Han K, Kim H, Lee SK: Radiomics machine learning study with a small sample size: Single random training-test set split may lead to unreliable results. PLoS One. 2021, 16:e0256152. 10.1371/journal.pone.0256152
3. Asteriou D, Hall SG: Applied Econometrics. Bloomsbury Publishing, London; 2021.
4. Athey S, Imbens GW: Machine learning methods that economists should know about. Annual Review of Economics. 2019, 11:685-725. 10.1146/annurev-economics-080217-053433
5. Ayetor GK, Mbonigaba I, Sackey MN, Andoh PY: Vehicle regulations in Africa: Impact on used vehicle import and new vehicle sales. Transportation Research Interdisciplinary Perspectives. 2021, 10:100384. 10.1016/j.trip.2021.100384
6. Bates S, Hastie T, Tibshirani R: Cross-validation: what does it estimate and how well does it do it?. Journal of the American Statistical Association. 2024, 119:1434-1445. 10.1080/01621459.2023.2197686
7. Bergmeir C, Hyndman RJ, Koo B: A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis. 2018, 120 :70-83. 10.1016/j.csda.2017.11.003
8. Bin-Inqiad W, Dev J, Mustafa A: Machine learning based determination of lateral displacement of fully grouted reinforced shear masonry walls under in-plane axial loading. Journal of Structural Integrity and Maintenance. 2026, 11:2609327.
9. Carnero MA, Peña D, Ruiz E: Standards vs. labels with imperfect competition and asymmetric information. Economics Letters. 2012, 114:61-63. 10.1016/j.econlet.2011.09.032
10. Citroen CL: The role of information in strategic decision-making. International Journal of Information Management. 2011, 31:493-501. 10.1016/j.ijinfomgt.2011.02.005
11. Fastener-World. (2022). Accessed: January 9, 2025: https://www.fastener-world.com/.
12. Fastener-World: Statistics and analysis of global automobile sales. Fastener World International Magazine, Special Industry Report. Fastener-World, Taiwan; 2024.
13. Granger CWJ: Forecasting in Business and Economics. Academic Press, New York; 2014.
14. Hall P: Car Sales Statistics 2025 - UK & Worldwide. Heycar, London; 2025 .

15. Hamzaçebi C, Akay D, Kutay F: Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert Systems with Applications. 2009, 36:3839-3844. 10.1016/j.eswa.2008.02.042
16. Hawinkel S, Waegeman W, Maere S: Out-of-sample R2: estimation and inference. The American Statistician. 2024, 78:15-25. 10.1080/00031305.2023.2216252
17. Hyndman RJ, Athanasopoulos G: Forecasting: Principles and Practice. OTexts, Melbourne; 2018.
18. Irfan M: Understanding vehicle import trends in Pakistan: Insights from economic and regulatory factors. Journal of Energy and Environmental Policy Options. 2020, 3:20-24.
19. Iskhakov F, Rust J, Schjerning B: Machine learning and structural econometrics: contrasts and synergies. The Econometrics Journal. 2020, 23:S81-S124. 10.1093/ectj/utaa019
20. Ismael K, Duleba S: Understanding the motivation and satisfaction of private vehicle users in an Eastern European country using heterogeneity analysis. Vehicles. 2022, 4:409-419. 10.3390/vehicles4020024
21. Kakade K, Mishra AK, Ghate K, Gupta S: Forecasting commodity market returns volatility: a hybrid ensemble learning GARCH-LSTM based approach. Intelligent Systems in Accounting, Finance and Management. 2022, 29:103-117. 10.1002/isaf.1515
22. Kim HY, Won CH: Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications. 2018, 103:25-37. 10.1016/j.eswa.2018.03.002
23. Kisiała W, Kudłak R: Local car markets in an emerging economy: exploring the dichotomous nature of car ownership growth. European Transport Research Review. 2024, 16:1-12. 10.1186/s12544-024-00645-1
24. Konovalov DA, Llewellyn LE, Vander Heyden Y, Coomans D: Robust cross-validation of linear regression QSAR models. Journal of Chemical Information and Modeling. 2008, 48:2081-2094. 10.1021/ci800209k
25. Li J, Xu L, Tang F, Yao D, Zhang C: The impact of public transport priority policy on private car own and use: A study on the moderating effects of bus satisfaction. Transportation Research Part F: Traffic Psychology and Behaviour. 2024, 106:112-127. 10.1016/j.trf.2024.08.010
26. Liu C, Wang C, Tran MN, Kohn R: A long short-term memory enhanced realized conditional heteroskedasticity model. Economic Modelling. 2025, 142:106922. 10.1016/j.econmod.2024.106922
27. Livieris IE, Pintelas P: A novel multi-step forecasting strategy for enhancing deep learning models' performance. Neural Computing and Applications. 2022, 34:19453-19470. 10.1007/s00521-022-07158-9
28. Moles AT, Warton DI, Warman L, et al.: Global patterns in plant height. Journal of Ecology. 2009, 97:923-932.
29. Mouratidis K: Transport and quality of life: The car and its link to subjective well-being, health, and life domains. Transport Policy. 2025, 168:101-111. 10.1016/j.tranpol.2025.04.014
30. Mullainathan S, Spiess J: Machine learning: an applied econometric approach. The Journal of Economic Perspectives: A Journal of the American Economic Association. 2017, 31:87-106. 10.1257/jep.31.2.87
31. MIA. (2025). Accessed: January 9, 2025: https://www.mia.org.nz/Sales-Data/Vehicle-Sales.
32. Partsvaniya VR: The import substitution trap in the realities of the automotive industry. Studies on Russian Economic Development. 2022, 33:203-210. 10.1134/s1075700722020095
33. Pawson E: Cars and the Motor Industry. Te Ara - the Encyclopedia of New Zealand, Wellington; 2014.
34. Penido REK, da Paixão RCF, Costa LCB, Peixoto RAF, Cury AA, Mendes JC: Predicting the compressive strength of steelmaking slag concrete with machine learning: considerations on developing a mix design tool. Construction and Building Materials. 2022, 341:127896. 10.1016/j.conbuildmat.2022.127896
35. Pimentel R, Risstad M, Rogde S, Rygg ES, Vinje J, Westgaard S, Wu C: Option pricing with deep learning: a long short-term memory approach. Decisions in Economics and Finance. 2025, 1-32. 10.1007/s10203-025-00518-9
36. Poldrack RA, Huckins G, Varoquaux G: Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry. 2020, 77:534-540. 10.1001/jamapsychiatry.2019.3671
37. Riker DA: Imports and the Location of Vehicle Production within the United States. US International Trade Commission, Washington, DC; 2024.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6

38. Rimašauskas M, Butrimas D, Besusparienė E: Customs duties for imported motor vehicles in the European Union. European Journal of Economics and Management. 2025, 11:5-12.

39. Ruxton GD: The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. Behavioral Ecology. 2006, 17:688-690. 10.1093/beheco/ark016

40. Shi Y, Arthanari T, Venkatesh VG, Islam S, Mani V: Used vehicle global supply chains: perspectives on a direct-import model. Supply Chain Management: An International Journal. 2022, 27:333-347. 10.1108/scm-06-2020-0238

41. Szarek D, Jabłoński I, Zimroz R, Wyłomańska A: Non-Gaussian feature distribution forecasting based on ConvLSTM neural network and its application to robust machine condition prognosis. Expert Systems with Applications. 2023, 230:120588. 10.1016/j.eswa.2023.120588

42. Stats NZ: New Zealand's greenhouse gas emissions. Statistics New Zealand, Wellington; 2022.

43. Stats NZ: Census - household access to telecommunication systems, tenure, and number of vehicles 2013, 2018, 2023. Statistics New Zealand, Wellington; 2024.

44. Stats NZ: All topics. Statistics New Zealand, Wellington; 2025.

How to cite this article:
Dhakal B (March 24, 2026) Car Sales in a Fully Import-Dependent, High-Ownership Country, New Zealand: Evaluating Forecasting with Deep Learning and Econometric Methods. Cureus J Bus Econ 3 : es44404-025-00040-6. DOI https://doi.org/10.7759/s44404-025-00040-6