Radosław Szostak

e-mail: 187234@ue.wroc.pl

ORCID: 0009-0006-0156-7297

Wojciech Świercz

e-mail: 187223@ue.wroc.pl

ORCID: 0009-0002-1906-842X

Wrocław University of Economics and Business

VAR Modeling in the Automotive Industry – Prediction of Volkswagen Prices

DOI: 10.15611/2024.90.1.04
JEL Classification: C53

Quote as: Szostak, R., Świercz, W. (2024). VAR Modeling in the Automotive Industry – Prediction of Volkswagen Prices. In M. Pauka, T. Słoński (Eds.), *Finanse* (pp. 51-59). Publishing House of Wroclaw University of Economics and Business.

Abstract: The purpose of this study is to develop and verify the effectiveness of a Vector Autoregression (VAR) model for use in an investment trading strategy. The research object is the VAR model, which forecasts the daily opening prices of Volkswagen stocks. To achieve stationarity, the model was estimated on the first differences of stock prices. Initially, the model included price changes of five different rivals of Volkswagen from the automotive industry (along with the AR values of Volkswagen), but ultimately, only the price changes of Volkswagen, General Motors, and Honda were included in the model. The study's key findings indicate that the model exhibits strong statistical performance, characterised by a high R^2 value and low Mean Squared Error (MSE) and Mean Absolute Error (MAE). However, the model's prediction accuracy for the direction of price changes is approximately 49%, meaning the model's directional forecasts are correct only almost half of the time.

Key words: VAR, automotive industry, Volkswagen, stock market forecasting

1. Introduction

The automotive industry, characterized by fierce competition and nonstop technological advancements, creates an intriguing environment for analysts and traders. This study aims to develop a Vector Autoregressive (VAR) model for Volkswagen's stock returns, focusing on open prices at daily intervals. The model

will be checked performance wise to determine whether it is suitable for real-life use and implementation in a trading strategy.

VAR models are commonly used for forecasting due to their capability of capturing relations between different variables (Stock & Watson, 2020). VAR models gave promising results in stock market forecasting for markets influenced by other markets in ASEAN region (Suharsono et al., 2017). VAR models are also used for forecasting prices of other assets, like cattle prices (Fanchon & Wendel, 1992). In this article, the authors will determine stock prices of the competition as other variables potentially influencing the price action of the Volkswagen company.

Volkswagen is one of the largest automotive companies, having its roots in Germany, and currently operating in the whole world. Currently, the biggest market for Volkswagen is China, delivering close to 40% of sales and profits. Volkswagen became the subject of controversy, due to colluding to restrict the development and deployment of emissions-cleaning technology between 2009 and 2014. Since then, Volkswagen stock has not really recovered, with attempts to repair its reputation through EV investments proving ineffective. Recently, Volkswagen has also invested in a joint venture with Rivian – the biggest EV rival to Tesla. The company is suitable for trading due to high liquidity and volume of shares traded, and relatively moderate volatility, where some of the conditions were suggested by Świercz & Szostak (2024).

With the automotive industry constantly undergoing critical changes, Volkswagen competes with several major rivals in terms of market capitalization and global presence. These include Toyota, Honda, Ford etc. The interrelationships between these companies' stock performances form a crucial aspect of the analysis. The research presented is operating on several key assumptions about the automotive industry. First, companies operating within this sector are subject to similar market conditions and macroeconomic factors. Second, the prevalence of price wars and competitive pricing strategies suggests potential correlations in stock behavior. Third, the industry-wide push for technological innovation likely creates parallel trends in stock performance. Fourth, similarities in cost structures, given comparable market capitalizations and global reach, may lead to analogous financial outcomes. Fifth, shared marketing channels and strategies could result in correlated market responses. Lastly, common regulatory environments, particularly within the European Union, may induce similar reactions in stock prices. These assumptions lead to a hypothesis that the stock prices of major automotive companies are interrelated. However, the nature of this relationship, and the strength of its influence, remains a point of investigation. While overall market growth might suggest positive correlations, periods of intense competition could potentially lead to negative relationships. The choice of daily intervals for this study is motivated by observed market behaviors, such as highest and lowest returns observed on Mondays and Wednesdays, and highest and lowest volatility observed on Fridays and Wednesdays, respectively (Berument & Kiymaz, 2001). The

aforementioned conditions suggest the possibility of autocorrelation occurring. Through this research, the authors aim to provide insights into the complex dynamics of automotive stock returns and offer a practical tool for traders navigating this challenging market.

To test the performance of the model, the authors used MSE, MAE, coefficient of determination, and propose an efficiency function, as an alternative to standard statistical measures, which can prove misleading in some instances, when changes of low difference but high importance occur.

2. Methodology

Data

The data has been downloaded from yahoo finance for the following companies: Volkswagen AG (VWAGY), Honda Motor Co., Ltd. (HMC), Toyota Motor Corporation (TM), Ford Motor Company (F), General Motors Company (GM), Hyundai Motor Company (HYMTF).

To conduct the analysis, the time series has been divided into two parts. On the in-sample period, the model will be estimated and on the out-of-sample period, the model will be verified. The in-sample period is from 20.06.2014 to 31.12.2023, the out-of-sample period is from 01.01.2024 to 20.06.2024. The model is estimated on the in-sample period and will be tested on the out-of-sample period.

Tools

All the necessary calculations have been conducted in program STATA and Excel.

Model evaluation

The authors decided to use AIC criterion to choose the best version of the model. AIC is commonly used in model selection, as it is relatively easy to interpret. Lower values of AIC mean that less information is lost, thus the model is of a better quality.

Comparison methods

To compare models, three statistical measures were chosen and a simple logical function. R^2 is a measure commonly used in accuracy testing, which can also be used on normalized data, and is easy to interpret (Dziechciarz, 2003).

$$R^2 = 1 - \frac{SSR}{SST},$$

where: SSR – explained sum of squares; SST – total sum of squares.

MSE, which translates to Mean Squared Error, is a metric that is commonly used in machine learning, as it heavily penalizes large errors. It is simply an average squared difference between actual and predicted values. Low MSE means that model performs well with values predicted being close to actual values. High MSE could mean that there are many outlier errors influencing the score.

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

where: n – number of observations; y_i – actual value; $\hat{y}_i$ – value predicted.

MAE (Mean Absolute Error) is yet another well-known measurement, one which is more resilient to outliers than MSE. MAE does not consider the direction of the errors, hence making it easier to interpret. Interpretation is to be concluded in the same units as the units of data. The lower the overall score, the better predictions model makes. A higher score means that the average absolute value of errors is bigger, thus encouraging to adjust the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right|$$

where: n – number of observations; e_i – errors.

The logical function is crucial for the investor as it evaluates the model's efficiency in predicting price directions, determining whether the model is suitable for trading. We consider a minimum price change threshold of $0.02 during the out-of-sample period. If the forecasted price change is less than $0.02, no trade is initiated. Table 1 shows a logical representation of the decision-making process:

Table 1. Logical function

Actual price	Forecasted price	Price change ≥ $0.02	Buy?	Profitable?
grows	grows	YES	YES	YES
grows	goes down	YES	NO	N/A
goes down	grows	YES	YES	NO
goes down	goes down	YES	NO	N/A
any	any	NO	NO	N/A

Source: own elaboration.

The efficiency of the 1-day prediction prior has been calculated by dividing the number of successful predictions by the number of a sum of the successful and unsuccessful predictions.

$$Efficiency = \sum_{i=1}^{n} s \,/\, \sum_{i=1}^{n}\left(s+u\right)$$

where: n – number of observations; s_t – successful predictions; u_t – unsuccessful predictions.

3. Results

To decide whether the time series appears to be stationary, the ADF test was run in STATA. Honda and Hyundai stock prices were found to be stationary at a 5% significance level. However, when one looks at the chart that displays these variables, it is noticeable that there are trends in data.

Before estimating the VAR(p) model on differenced data, the Johansen test for cointegration on the VAR(3) model (chosen based on AIC criterion) on not differenced data was run. The test selected rank 0, which implies that there are no cointegrating relationships among the variables and the Vector Error Correction (VEC) model should not be estimated. Therefore, to proceed with calculations, the first difference was calculated for all variables and differenced prices were found to be stationary at a 5% significance level.

To choose an appropriate VAR(p) model, the STATA function varsoc was used with maximum lag 6. The decision on how many lags the model should have, was based on AIC criterion, being equal to 11.0686 for the VAR(3) model. It indicates that VAR(3) explains the best the variability in the data while penalizing for model complexity. The next run STATA function was varstable, which presented the graph of the roots of the companion matrix, that provided critical information about the stability of a Vector Autoregression (VAR) model. All the eigenvalues were found to be inside the unit circle, so the VAR model is stable. This indicates that the shocks to the system will dissipate over time, and the system will return to equilibrium. Subsequently, the VAR(3) model was estimated using the var function with 3 lags. The outcome of the function was not satisfying because the p-value of most of the coefficients was much higher than 0.1. Thus it was decided to rerun the estimation function several times to achieve *p*-values for coefficients at least at the 10% level, which is a commonly used significance level in financial modelling. It was done by every time deleting one coefficient from the model that had the highest p-value.

After multiple iterations, the final structure of the model was achieved:

Table 2. Backward Stepwise Regression

	Coefficient	Std.Err.	*t*	$P > \|t\|$
VW Lag1.	−0.1180445	0.0217796	−5.42	0.000
VW Lag2.	0.0403226	0.0206495	1.95	0.051
VW Lag3.	−0.0866248	0.022893	−3.78	0.000
Honda Lag1.	0.0829491	0.0273904	3.03	0.002
GM Lag1.	0.0475776	0.0139928	3.40	0.001

Source: own elaboration.

The VAR(3) model contains first, second and third lags for Volkswagen stock prices, one lag for Honda stock price and third lag for GM. The model does not contain the constant value.

In our analysis, we employed the varlmar function in STATA to conduct a Lagrange-multiplier test. The results indicated an absence of autocorrelation at the specified lag order, which is a favorable outcome. This finding suggests that the estimation of a Vector Error Correction Model (VECM) is not warranted in this case.

The decision between using a VAR or VECM model is crucial in time series analysis, particularly when dealing with financial data. VECM estimation is typically justified when there is evidence of autocorrelation in the errors of a Vector Autoregression (VAR) model. Autocorrelation in VAR errors can indicate the presence of long-run relationships between variables that are not captured by the standard VAR framework. Suharsono et al. (2017) conducted a comparative study of VAR and VECM models for ASEAN stock price indices, highlighting the importance of considering the presence of cointegration and error correction terms. Their research underscores the necessity of testing for autocorrelation and applying the appropriate model based on the results. In our case, the absence of significant autocorrelation aligns with their findings for scenarios where VAR models are more suitable than VECM.

$$\hat{Y} = -0.1180445 \cdot VW_{t-1} + 0.0403226 \cdot \mathrm{VW}_{t-2} - 0{,}0866248 \cdot VW_{t-3} + \\ + 0{,}0829491 \cdot H_{t-1} + 0{,}0475776 \cdot GM_{t-2}$$

Figure 1 presents a chart that compares the actual and the predicted Volkswagen stock prices for the out-of-sample period with a chart representing the model's errors:

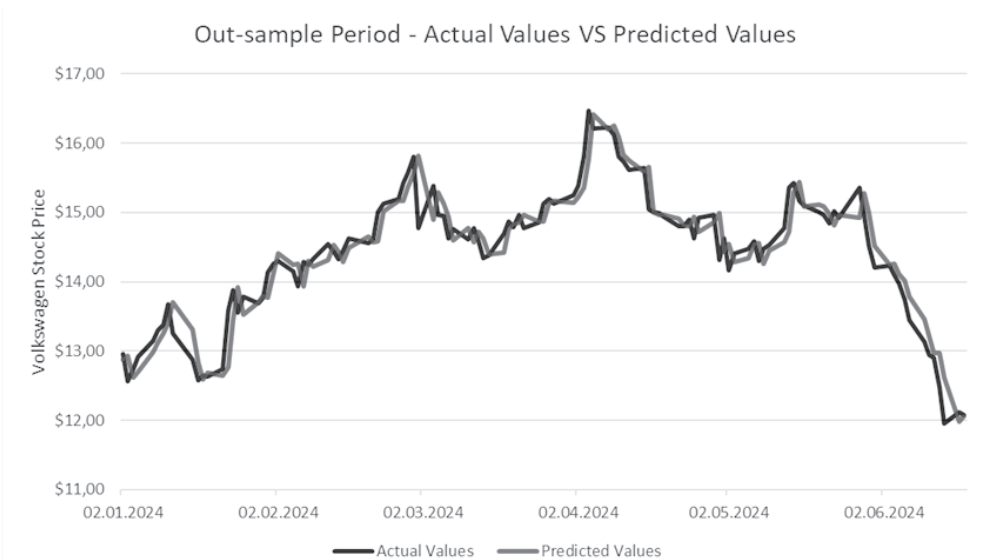

Figure 1. Actual vs. Predicted values

Source: own elaboration.

Figure 2. Errors

Source: own elaboration.

The prediction seems to be quite accurate. However, to surely decide on model's usefulness, statistical measures and logical function were done.

Table 3. Verification measures

Measures	Volkswagen price
R^2	0.912
MAE	0.215
MSE	0.080
Efficiency	49%

Source: own elaboration.

The efficiency of the model turns out to be approximately 49%. This indicates that the model requires adjustments in several areas.

4. Conclusion

In conclusion, it seems that some of companies' stock prices from the same industry are, from a statistical point of view, related to each other. In the presented model, the most influential on Volkswagen's stock price were the stock prices of Honda and General Motors and the past prices of Volkswagen. Statistical verification indicated

that the model is significantly estimated, however the logical function pointed out that this model is not sufficient to be used in a trading strategy, because it did not predict the stock prices directions well enough. To improve the model's efficiency, in the future we will consider including in the model other companies from the automotive industry and other economic variables, which relate to the sector. The model should be tested in different sectors to test the performance across various industries, including relevant variables. We will consider supporting VAR model with different types of models like for instance ARCH or GARCH that are used for error volatility predictions. This is supported by the recent research comparing GARCH models in quantifying VAR risks during stress periods, which found that ARCH and GARCH specifications performed best across both emerging and developed markets during the global financial crisis (Papana et al., 2015). Implementation of the AI part of the model that could interpret daily news related to the company and include it appropriately in the model could significantly improve its performance as well. All models will also be considered in creating a combination model, which will primarily be a machine learning model supported by statistical models. The purpose of the model will be to perform well in a high frequency trading environment.

References

Berument, H. & Kiymaz, H. (2001). Day of the Week Effect on Stock Market Volatility. *Journal of Economics and Finance 25*, 181-193, https://doi.org/10.1007/BF02744521

Dziechciarz, J. (2003). *Ekonometria – metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

Fanchon, P., & Wendel, J. (1992). Estimating VAR Models under Non-stationarity and Cointegration: Alternative Approaches for Forecasting Cattle Prices. *Applied Economics*, *24*(2), 207-217, https://doi.org/10.1080/00036849200000120

Orhan, M., & Köksal, B. (2012). A Comparison of GARCH Models for VaR Estimation. *Expert Systems with Applications*, *39*(3), 3582-3592, https://doi.org/10.1016/j.eswa.2011.09.048

Papana, A., Kyrtsou, C., Kugiumtzis, D., & Diks, C. (2015). Detecting Causality in Non-stationary Time Series Using Partial Symbolic Transfer Entropy: Evidence in Financial Data. *Computational Economics*, 47(3), 341-365, https://doi.org/10.1007/s10614-015-9491-x

Stock, J. H., & Watson, M. W. (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics* (Vol. 2, pp. 415-525). Elsevier, https://doi.org/10.1016/bs.hesmac.2016.04.002

Suharsono, A., Aziza, A., & Pramesti, W. (2017). Comparison of Vector Autoregressive (VAR) and Vector Error Correction Models (VECM) for Index of ASEAN Stock Price. *AIP Conference Proceedings*, *1913*(1), 020032.

Świercz, W., & Szostak, R. (2024). Comparative Analysis of Predictive Models in Stock Market Price Forecasting. In M. Pauka & T. Słoński (Eds.), *Finance* (pp. 2-4). Publishing House of Wroclaw University of Economics and Business.

Sources of data

Yahoo Finance. (n.d.). *Volkswagen AG (VWAGY)*. Retrieved from https://finance.yahoo.com/quote/VWAGY/

Yahoo Finance. (n.d.). *Honda Motor Co., Ltd. (HMC)*. Retrieved from https://finance.yahoo.com/quote/HMC/

Yahoo Finance. (n.d.). *Toyota Motor Corporation (TM)*. Retrieved from https://finance.yahoo.com/quote/TM/

Yahoo Finance. (n.d.). *Ford Motor Company (F)*. Retrieved from https://finance.yahoo.com/quote/F/

Yahoo Finance. (n.d.). *General Motors Company (GM)*. Retrieved from https://finance.yahoo.com/quote/GM/

Yahoo Finance. (n.d.). *Hyundai Motor Co., Ltd. (HYMTF)*. Retrieved from https://finance.yahoo.com/quote/HYMTF/

Model VAR w branży motoryzacyjnej – predykcja cen akcji firmy Volkswagen

Streszczenie: Celem badania jest opracowanie i weryfikacja skuteczności modelu Vector Autoregression (VAR) do wykorzystania w strategii inwestycyjnej opartej na tradingu. Przedmiotem badań jest model VAR, który prognozuje dzienne ceny otwarcia akcji Volkswagena. Aby osiągnąć stacjonarność, model był estymowany na pierwszych różnicach cen akcji. Początkowo model obejmował zmiany cen pięciu różnych rywali Volkswagena z branży motoryzacyjnej (oraz wartości AR Volkswagena), ale ostatecznie uwzględniono w nim tylko zmiany cen Volkswagena, General Motors i Hondy. Kluczowe wyniki badania dowodzą, że model wykazuje silną wydajność statystyczną, charakteryzującą się wysoką wartością R^2 oraz niskimi wartościami Mean Squared Error (MSE) i Mean Absolute Error (MAE). Jednakże dokładność modelu w prognozowaniu kierunku zmian cen wynosi około 49%, co oznacza, że prognozy kierunkowe modelu są poprawne tylko w prawie połowie czasu.

Słowa kluczowe: VAR, branża motoryzacyjna, Volkswagen, predykcja rynku akcji