Time-Series Forecasting Models

Subjects: Computer Science, Artificial Intelligence

Contributor: Yee-Fan Tan

The time-series forecasting method is a suitable pricing solution for Digital Signage Advertising (DSA), as it improves the pricing decision by modeling the changes in the environmental factors and audience attention level toward signage for optimal pricing. However, it is difficult to determine an optimal price forecasting model for DSA with the increasing number of available time-series forecasting models in recent years. Based on the 84 research articles reviewed, the data characteristics analysis in terms of linearity, stationarity, volatility, and dataset size is helpful in determining the optimal model for time-series price forecasting.

time-series forecasting   dynamic pricing   digital signage advertising   data characteristics

model selection

1. Introduction

With recent technology advancements, more dynamic pricing solutions are emerging, applying time-series forecasting methods to predict future activities based on historical data [1]. They provide a strong foundation for time-series forecasting dynamic pricing solutions using statistical, AI, and hybrid models [2][3]. There are several ways to categorize the time-series forecasting models. Wang suggested three main categories of time-series forecasting methods as statistical models, Artificial Intelligence (AI) models, and hybrid models [2]. On the other hand, Neha Sehgal and Krishan classified the time-series forecasting models as stochastic models, AI models, and regression models. However, both the regression and stochastic models are using similar statistical inferences to establish the relationship between the price and other relevant factors. Hence, both stochastic and regression models are classified under the subcategories of statistical models in our review. **Figure 1** elucidates the classification tree of the seven most widely used time-series forecasting models on price estimation.
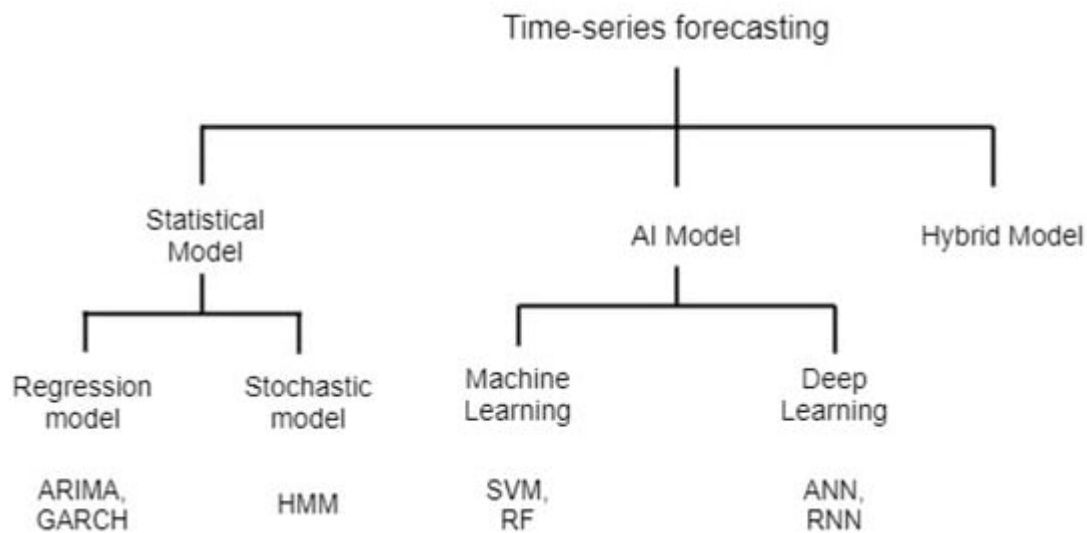
Figure 1. Classification tree of price forecasting models.

The existing price forecasting methods that applied the time-series forecasting models are exemplified in different domains for its applicability. However, as a real-world problem is often complicated, no single model can perform the best in every situation [4]. With an increasing number of models, the model selection decision is also difficult, especially when vital information and knowledge are insufficient. Hence, a model selection framework is crucial for identifying a suitable approach for different data characteristics. From the review, there are four relevant essential data characteristics: namely, data linearity, stationarity, volatility, and the size of the dataset. These data characteristics are essential factors in selecting the optimal forecasting model for pricing decisions, particularly in its usage and interpretability in DSA.

This review aims to structurally analyze the data characteristics of different time-series forecasting models, along with a proposed framework for the optimal model selection of dynamic pricing in DSA. The main contributions of this paper are summarized as follows: - To review the widely used time-series forecasting models, providing an abstract of the different models that can potentially be applied for dynamic pricing. - To discuss the relationship of data characteristics to the time-series forecasting models from each category. - To investigate the applicability of the model selection framework based on the data characteristics of the dataset in DSA.

This paper is divided into four sections. Section 2 presents the review of time-series forecasting models based on the data characteristics. Section 3 shows a summary of the suitable data characteristics for the time-series forecasting models and a proposed framework for optimal model selection based on the data characteristic analysis of DSA. Conclusions are drawn in Section 4 .

2. Review on Time-Series Forecasting Studies

Notably, the stock and electricity markets have always been a focus of time-series forecasting, as shown in **Figure 2** b. Different models have a relatively higher application count in these two fields, which is likely due to the market

complexity that is highly volatile. It is also noted that the application of time-series forecasting is extending to other domains for dynamic pricing recently. The following section introduces the statistical model, AI model, and hybrid model for dynamic pricing solutions.

(a) The distribution of models from 2001 to 2021   (b) The model applications in different domains
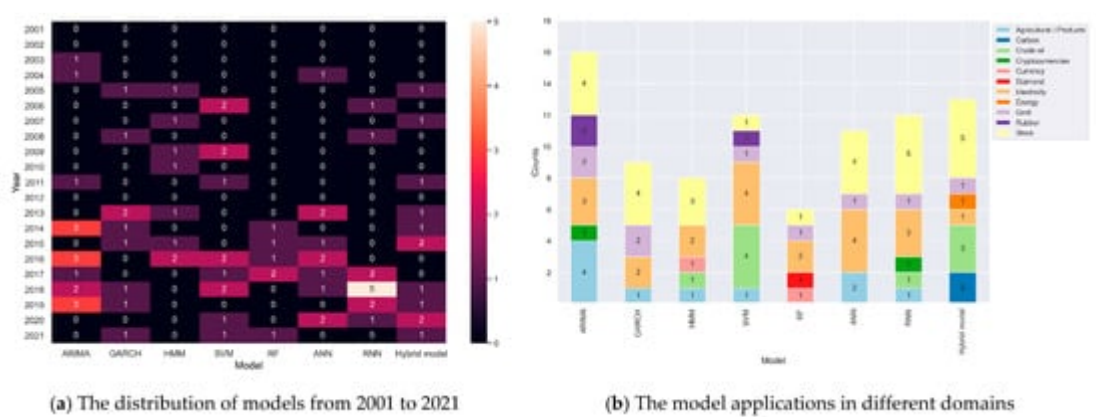
Figure 2. The evolution of time-series forecasting models in different domains over two decades.

A statistical model is a mathematical model that can be used for time-series forecasting [5]. It is classified into two categories of regression and stochastic models.

The stochastic model is widely applied for time-series forecasting [3] to draw inferences of the characteristics of the time-series data [1]. Different from the regression model, the inferences made using the stochastic model are based on the principles of probabilities. On the other hand, the regression model is used to identify the linear relationship between the time-series data and the price outcome. One of the stochastic models, Hidden Markov Model (HMM) is explained in this section.

Although the hybrid model has demonstrated a promising outcome, such a model is usually highly complex and computationally intensive [6]. The determination of the models for combination is often tricky and challenging, as it requires an in-depth understanding of each method. Thus, a strong background of different models should be mastered and well-understood to decide on a suitable hybrid architecture. The hybrid models also have lower flexibility, since it is specifically designed to handle a particular problem. **Table 1** shows the reviewed studies for the hybrid models.

Table 1. Hybrid model time-series forecasting from 2004 to 2021.

Domain	Author	Dataset	Model	Result
	Tang and Diao, 2017 [7]	WIND database (January 2010 to September 2016)	HMM-GARCH	RMSE of 0.0238 and 0.0075
Stock	Pai and Lin, 2005 [8]	Ten stocks dataset (October 202 to December 2002)	ARIMA, SVM, ARIMA-SVM	Hybrid model has the lowest MAE for the included ten stocks

Domain	Author	Dataset	Model	Result
	Raiful Hassan et al., 2007 [9]	Daily stock price of Apple, IBM, and Dell from Yahoo Finance	ANN-GA-HMM	MAPE of 1.9247, 0.84871, and 0.699246 for the stock, respectively
	Wang and Guo, 2020 [10]	Ten stocks dataset (2015 to 2018)	DWT-ARIMA-GSXGB	RMSE of less than 20.3013 for the worst case, the general cases have an RMSE of less than 0.3
	Chen et al., 2020 [11]	Yahoo Finance (September 2008 to July 2019)	MLP-Bi-LSTM with AM	MAE of 0.025393
Crude oil	Shabri and Samsudin, 2014 [12]	Brent crude oil prices and WTI crude oil prices	ANN, WANN	WANN has the best performance with MAPE of 1.31 and 1.39 for Brent and WTI dataset
	Zhang et al., 2015 [13]	WTI and Brent crude oil (January 1986 to 2005) and (May 1987 to June 2005)	EEMD-LSSVM-PSO-GARCH	MAPE of 1.27 and 1.53 for WRI and Brent dataset
	Safari and Davallou, 2018 [14]	OPEC crude oil prices (January 2003 to September 2016)	ESM-ARIMA-NAR	MAPE of 2.44, obtained the lowest error compared to other single models
Energy	Bissing et al., 2019 [15]	Iberian electricity market (February to July 2015)	ARIMA -MLR and ARIMA-Holt winter	ARIMA-Holt Winter has better performance with an MAPE of less than 5.07 for different day forecasting
Carbon	Zhu and Wei, 2013 [16]	European Climate Exchange (ECX) of December 2010 and December 2012	ARIMA-LSSVM	RMSE of 0.0311 and 0.0309 for DEC10 and DEC12
	Huang et al., 2021 [17]	EUA futures from Wind database	VMD-GARCH and LSTM	VMS-GARCH has the best performance with first ranking in terms of RMSE, MAE and MAPE
Electricity	Shafie-khah et al., 2011 [18]	Spanish electricity market of 2002	Wavelet-ARIMA-RBFN	Error variances of less than 0.0049
Gold	Kristjanpoller and Minutolo, 2015 [19]	Gold Spot Price and Gold Future Price from Bloomberg (September 1999 to March 2014)	ANN-GARCH	MAPE of 0.6493 and 0.6621

3. Discussion

After reviewing the different time-series forecasting models presented in Section 2 , it can be concluded that the data characteristics analysis is an important step in deciding the optimal model for forecasting, as each of them performs better on particular data characteristics. Consequently, in order to determine the optimal model selection for DSA, data characteristics analysis is performed. This section presents a summary of the suitable data characteristics for the models introduced in Section 2 , which is followed by the DSA data analysis and a proposed framework for optimal model selection based on the result analysis.

In the review of the seven forecasting models, the importance of the different data characteristics in the optimal model selection for price forecasting is emphasized. **Table 2** shows the suitability of data characteristics for each model. Each model has varying performance for the different data characteristics. The large dataset is generally built up with more than 100,000 observations, and the small dataset has fewer observations. This section presents a summary of the strengths and weaknesses of the models.

Table 2. Optimal model selection with its suitable data characteristics.

Model	Linear	Nonlinear	Stationary	Non-Stationary	Volatile	Non-Volatile	Large Dataset	Small Dataset
ARIMA	✓		✓			✓		✓
GARCH		✓		✓	✓			✓
HMM		✓		✓	✓		✓	
SVM		✓		✓		✓		✓
RF		✓		✓	✓		✓	
ANN		✓		✓		✓	✓	
RNN		✓		✓	✓		✓	

DL models such as ANN and RNN are both neural network models inspired by the human neuro system. They have a few pros and cons in common, such as good generalizability and good performance on nonlinear and non-stationary data modeling. However, these models are also black box in nature, which results in low model transparency. Both DL models have shown better results compared to ML models when the dataset size is sufficient for training. Among ANN and RNN, RNN has demonstrated a better result than ANN in forecasting data with high volatility.

From the interpreted results, SVM is the optimal model for modeling the popularity index, while ARIMA is suitable for modeling both temperature and air humidity. Multiple models can forecast the respective environmental factors instead of using a single model to fit all the factors. As the environmental factors affect the attention level and

consequently the pricing decisions, using multiple models for different data characteristics is better in forecasting the future values of the data for pricing decisions. In our case, where the pricing label is not available, a rule-based system can be applied by setting a threshold value for each environmental factor to generate the price. This can help effectively forecast the important factors that will affect the attention level and pricing decision for DSA. With the optimal models used, a better pricing decision can be made.

4. Conclusions and Future Work

This paper presented a list of widely used time-series forecasting models for dynamic pricing based on an extensive literature review. These models have been proven to have outstanding achievement in different domains. From the review, each model has its strength in dealing with different data characteristics. The advantages and disadvantages of each model are discussed and presented. In addition, the relationships between the models and the different data characteristics are investigated to suggest the optimal model selection for the price forecasting of DSA. From there, an optimal model selection framework is proposed based on the data characteristics analysis of DSA data. This structured review provides a solution of dynamic pricing using time-series forecasting, paving the way for future research on dynamic pricing in DSA.

For future work, an experimental study shall be carried out to examine the applicability of the proposed optimal model selection framework. More time-series data shall be collected to conduct the experimental study for dynamic pricing in DSA to investigate the effectiveness of the optimal model selection framework. Moreover, a customer behavioral targeting approach can be considered in identifying their preferences for better advertising effects [20].

References

1. Athiyarath, S.; Paul, M.; Krishnaswamy, S. A Comparative Study and Analysis of Time Series Forecasting Techniques. SN Comput. Sci. 2020, 1, 175.

2. Wang, D.; Luo, H.; Grunder, O.; Lin, Y.; Guo, H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl. Energy 2017, 190, 390–407.

3. Sehgal, N.; Pandey, K.K. Artificial intelligence methods for oil price forecasting: A review and evaluation. Energy Syst. 2015, 6, 479–506.

4. Gao, W.; Aamir, M.; Shabri, A.B.; Dewan, R.; Aslam, A. Forecasting Crude Oil Price Using Kalman Filter Based on the Reconstruction of Modes of Decomposition Ensemble Model. IEEE Access 2019, 7, 149908–149925.

5. Nisha, K.G.; Sreekumar, K. A Review and Analysis of Machine Learning and Statistical Approaches for Prediction. In Proceedings of the 2017 International Conference on Inventive

Communication and Computational Technologies (ICICCT), Coimbatore, India, 10–11 March 2017.

6. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 2017, 74, 902–924.

7. Tang, L.; Diao, X. Option pricing based on HMM and GARCH model. In Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 3363–3368.

8. Pai, P.-F.; Lin, C.-S. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 2005, 33, 497–505.

9. Hassan, M.R.; Nath, B.; Kirley, M. A fusion model of HMM, ANN and GA for stock market forecasting. Expert Syst. Appl. 2007, 33, 171–180.

10. Wang, Y.; Guo, Y. Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. China Commun. 2020, 17, 205–221.

11. Chen, Q.; Zhang, W.; Lou, Y. Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism, Multi-Layer Perceptron, and Bidirectional Long-Short Term Memory Neural Network. IEEE Access 2020, 8, 117365–117376.

12. Shabri, A.; Samsudin, R. Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model. Math. Probl. Eng. 2014, 11.

13. Zhang, J.-L.; Zhang, Y.-J.; Zhang, L. A novel hybrid method for crude oil price forecasting. Energy Econ. 2015, 49, 649–659.

14. Safari, A.; Davallou, M. Oil price forecasting using a hybrid model. Energy 2018, 148, 49–58.

15. Bissing, D.; Klein, M.T.; Chinnathambi, R.A.; Selvaraj, D.F.; Ranganathan, P. A Hybrid Regression Model for Day-Ahead Energy Price Forecasting. IEEE Access 2019, 7, 36833–36842.

16. Zhu, B.; Wei, Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 2013, 41, 517–524.

17. Huang, Y.; Dai, X.; Wang, Q.; Zhou, D. A hybrid model for carbon price forecasting using GARCH and long short-term memory network. Appl. Energy 2021, 285, 116485.

18. Shafie-khah, M.; Moghaddam, M.P.; Sheikh-El-Eslami, M.K. Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Convers. Manag. 2011, 52, 2165–2169.

19. Kristjanpoller, W.; Minutolo, M.C. Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model. Expert Syst. Appl. 2015, 42, 7245–7251.

20. Saia, R.; Boratto, L.; Carta, S. A Latent Semantic Pattern Recognition Strategy for an Untrivial Targeted Advertising. In Proceedings of the 2015 IEEE International Congress on Big Data, New York, NY, USA, 27 June–2 July 2015; pp. 491–498.