Programming-Based Analytical Tools for Forecasting Economic Indicators

Zherlitsyn Dmytro, Doctor of Science (Economics), prof.
ORCID ID 0000-0002-2331-8690
University of National and World Economy, Sofia, Bulgaria

In the rapidly evolving landscape of economic analysis, applying programming-based analytical tools has become paramount. This presentation delves into the intricate world of forecasting economic indicators, highlighting the significance of advanced tools like Prophet, Keras & Tensorflow, PyPorch, etc. We explore the efficacy of various methodologies, including ARIMA, Seasonal ARIMA, and machine learning-based approaches, in predicting trends in economic data. By analyzing different scenarios like energy consumption, financial market trends, and social media traffic, we aim to demonstrate the versatility and precision of these tools. Our comparative analysis sheds light on the practicalities and technicalities of using these programming-based methods, offering insights into their application in real-world forecasting scenarios.

There are a lot of references that cover various methods of economic forecasting. Some of the last of them are Chapman and Desai (2023), who explore macroeconomic predictions using payment data alongside machine learning techniques, suggesting a novel approach to understanding economic trends. Corradin, Billio, and Casarin (2022) discuss forecasting economic indicators through robust factor models, offering a statistical approach to anticipate economic changes. The Frontiers in Energy Research (2023) publication presents a quarterly GDP forecast utilizing a coupled economic and energy feature WA-LSTM model, indicating a link between energy metrics and financial performance. Lastly, Ong, Qiao, and Jadav (2020) introduce a temporal tensor transformation network for multivariate time series prediction, which could have significant implications for predicting complex economic phenomena. Each study contributes to the evolving field of economic forecasting, leveraging data analytics and advanced statistical programming to enhance prediction accuracy.

In time series analysis, a time series is defined as a collection of data points gathered or measured at successive time intervals, usually sorted in chronological order. The characteristics of time series analysis tasks include trends and seasonality, autocorrelation, stationarity, unpredictability, and variability. Practical forecasting tasks encompass macroeconomic indicators, prices of financial assets represented by a large volume of relatively independent data, prices and sales of products, raw materials, web page traffic data, social media data, the development of diseases, pest migration, etc.

The main statistical methods and software tools discussed encompass traditional methods such as ARIMA (AutoRegressive Integrated Moving Average), Seasonal ARIMA (SARIMA), Exponential Smoothing (ES), the Holt-Winters method (HW), machine learning-based tools and other methods. The programming tools for implementation are Python, ranked first in The Top Programming Languages of 2023 by IEEE Spectrum, and R, ranked eleventh.

The forecasting task involves predicting daily volumes of heat, electricity, gas consumption, etc., with a forecasting horizon of 3, 7, and 30 days. The «dirty» data are skewed and require prior cleaning to achieve a maximum of 5 % relative forecasting error for three days. The methods compared include ARIMA and SARIMA (implemented in Python), Facebook Prophet tools, Keras & TF, PyTorch, and machine learning modelling methods such as LightGBM and XGBoost.

Data cleaning and preparation involve standard templates and procedures for data cleansing, detecting outliers, and determining the probability of intentional data distortion (data «illogicality»). Data grouping is based on known characteristics (defined consumer groups, regions), and classical data clustering is based on general traits (technical features, weather conditions, etc.).

The ARIMA model's daily and weekly consumption for three hypothetical consumers from one group showed a MAPE greater than 100 %. The Facebook Prophet tool is simple for implementing ML methods for forecasting time series, automatically identifying trends, detecting annual, weekly, and daily seasonality, and including holidays or special events that may influence trends. It can manually adjust trend changes, seasonality, and other parameters, is effective with large datasets, and is optimized for fast calculations. It provides flexible cross-validation tools, is easy to learn and use by professionals from various fields, and results from Prophet model implementations show a MAPE ranging from 12 to 40 % and an adjusted MAPE ranging from 8 to 36 %, with well-interpreted forecast results.

Machine learning methods (LightGBM and XGBoost) show good results, work well with outliers, do not have problems with collinear features, and can work with missing data (NaN values). These methods showed a relative error of about 3-5 % for most of the observations, but the forecast accuracy dropped sharply from 20 to 150 % for some consumer groups. The interpretability of factors is an irreplaceable tool. Still, these methods are significantly inferior to other forecasting tools regarding model training time with many factorial variables (if this is a limitation).

The use of artificial neural networks analyzed the following models (COMBINATION): Transformer (3 to 14 %), TemporalFusionTransformer (1 to 6 %), LSTNet Multivariate (4 to 12 %), and LSTM Multivariate (5 to 20 %), concerning the work of Yuya Jeremy Ong, Mu Qiao, and Divyesh Jadav (2020) on the Temporal Tensor Transformation Network for Multivariate Time Series Prediction. The time for training and interpreting dependencies remains a challenge.

The forecasting algorithm involves data collection and cleaning, data preparation for modelling, cluster analysis based on characteristics and time behaviour patterns, training several models, and forecasting based on a combination of models, with predictions for three days ahead.

In conclusion, there is currently no universal method or tool for forecasting time series, and the characteristics of the time series significantly influence the choice of forecasting tools, making the preparation and processing of input data the most crucial stage. Facebook Prophet tools provide significant accuracy with low time expenditure on training, hyperparameter tuning, etc., making it «optimal» for most economic tasks. The most accurate methods and tools are those based on artificial neural networks (ANN), but they require users to have specialized knowledge and programming skills. A certain accuracy can only be achieved by combining different methods and tools for complex time series forecasting tasks. The appropriateness of the elective discipline «Modern tools for forecasting time series» with practical examples is highlighted.

References:

1. Baranovskyi, O., Kuzheliev, M., Zherlitsyn, D., Serdyukov, K., & Sokyrko, O. (2021). Cryptocurrency market trends and fundamental economic indicators: correlation and regression analysis. *Financial and Credit Activity: Problems of Theory and Practic*e, 3(38), 249–261. https://doi.org/10.18371/fcaptp.v3i38.237454
2. Chapman, J. T. E., & Desai, A. (2023). Macroeconomic Predictions Using Payments Data and Machine Learning. *Forecasting*, 5(4), 652-683. https://doi.org/10.3390/forecast5040036
3. Corradin, F., Billio, M., & Casarin, R. (2022). Forecasting Economic Indicators with Robust Factor Models. *National Accounting Review*, 4(2), 167-190. https://doi.org/10.3934/NAR.2022010
4. Frontiers in Energy Research. (2023). Quarterly GDP forecast based on coupled economic and energy feature WA-LSTM model. *Process and Energy Systems Engineering*, 11. https://doi.org/10.3389/fenrg.2023.1329376
5. Ong, Y. J., Qiao, M., & Jadav, D. (2020). Temporal Tensor Transformation Network for Multivariate Time Series Prediction. Retrieved from https://arxiv.org/pdf/2001.01051.pdf
6. Zherlitsyn D.M., Kuzheliev M.O., Nechyporenko A.V. (2019) Bioresources Price Trend and GDP Growth Adjustment. *International Journal of Advanced Biotechnology and Research (IJABR)*. Volume 10. Special Issue-1, 2019, pp.379-383.