QUANTITATIVE METHODS IN ECONOMICS
Volume XXV, No. 3, 2024, pp. 124 – 133

APPLICATION OF ITRANSFORMERS TO PREDICTING PRETERM BIRTH RATE. COMPARISON WITH THE ARIMA MODEL

Marek Karwański https://orcid.org/0000-0001-5192-7920
Urszula Grzybowska https://orcid.org/0000-0001-7342-5382
Institute of Information Technology
Warsaw University of Life Sciences, Poland
e-mail: urszula_grzybowska@sggw.edu.pl; marek_karwanski@sggw.edu.pl
Vassilis Kostoglou
Department of Informatics and Electronic Engineering
International Hellenic University, Greece
e-mail: vkostogl@ihu.gr
Ewa Mierzejewska https://orcid.org/0000-0003-0822-4781
Katarzyna Szamotulska https://orcid.org/0000-0003-1045-7584
Department of Epidemiology and Biostatistics
Institute of Mother and Child, Warsaw, Poland
e-mail: katarzyna.szamotulska@imid.med.pl; ewa.mierzejewska@imid.med.pl

Abstract: In this paper, we study the differences between forecasts obtained with the classical seasonal ARIMA model and forecasts obtained with the neural network model called iTransformers. The analysis is done on Polish data concerning preterm birth from 2015 to 2020. We compare the results and calculate the effect size to conclude the impact of the obtained differences.

Keywords: time series forecasting, seasonal ARIMA, iTransformers, preterm birth

JEL classification: C530, I0, C630

INTRODUCTION

Time series forecasting is important in various domains, including economics or environmental and health policy. Classical methods, with their foundations laid in

https://doi.org/10.22630/MIBE.2024.25.3.11

the 70th by Box and Jenkins[1] [Box, Jenkins 1970], are still widely applied. However, they face increased competition from neural network models, especially those based on attention mechanisms [Vaswani 2017; Ahmed 2023].

We aim to compare the traditional seasonal ARIMA model used for time series forecasting with a recent approach based on Transformers. Namely, we apply a neural network architecture called iTransformers which is suitable for long-term predictions [Liu et al. 2023]. To compare the models, we calculate the differences between the predictions obtained in the statistical and machine learning approaches, test the hypothesis about the predictions, and compare the results by calculating the effect size. The analysis is done on Polish preterm birth data. The problem is important as preterm birth is responsible for a high rate of infant deaths [Brandon, McGrath 2016]. We have built forecasts for 4, 6, and 12 months. Calculations were performed using the GluonTS package with torch ver. 2.3.0+cu114.

LITERATURE REVIEW

Neural networks have been successfully used for decades to classify objects or recognize images, text, and speech. Nowadays, their applications cover virtually all fields, including time series modeling and forecasting. A relatively new survey of deep learning methods for time series forecasting, including Transforme-based architectures, can be found in, e.g., [Lim, Zohren 2021]. A list of the most important statistical and machine learning models with references can be found in a survey by [Miller et al. 2024]. This survey also contains a short description of the newest Transformer-based models and other architectures and forecast quality metrics. Traditional regression or ARIMA models have not lost their relevance for time series forecasting. Also hybrid models are applied [Aijaz, Agarwal 2019]. The data used in our analysis concerns the preterm birth rate. There has not been much research on predicting preterm birth in the last decade. Some examples of statistical modeling for preterm birth rates can be found in [Priya et al. 2024; Gemmill et al. 2021; Sefidkar et al. 2021]. On the other hand, the problem of predicting preterm birth with the help of machine learning methods can be found in [Dench, Joyce 2022; Tzitiridou-Chatzopoulou, Zournatzidou Kourakos 2024; Yu et al. 2024; Zhang et al. 2024; Zhang et al. 2023; Borboa-Olivares et al. 2023].

DATA DESCRIPTION

According to the WHO (World Health Organization), premature birth is a condition in which birth occurs before the 37th week or 259th day of its duration. It is known that the shorter the duration of pregnancy, the lower the infant’s chances of healthy development and survival. In Poland, nearly 7% of pregnancies end in

[1] The famous book with many reissues summarises previous research. Historical details can be found in e. g., [Miller 2024].

premature birth in one calendar year. The data set includes information about the proportion of preterm births for the duration of pregnancy less or equal to 36 weeks in Poland for single live births from 22 weeks of gestation. We have used monthly records from 2015-2020 for the analysis[2]. Cases of stillbirth and multiple pregnancies were excluded from the study. In individual months, births ranged from 20,000 to 40,000, including 1,000 to 2,000 premature births. We have used the percentage rate of premature births in the analysis. Data exhibits seasonal effects with the highest rate in December. There is no visible trend.

METHODS

We compare the traditional ARIMA model applied to seasonally adjusted data with a Transformer-based neural network model called iTransformer. Conventional statistical models, such as regression or ARIMA models, require the data to fulfill certain assumptions. The ARIMA model requires the data to be stationary (constant mean and variance, no trend, no seasonality). Therefore, the data must be prepared before modeling, e.g., by removing stochastic and deterministic trends and seasonality. Neural network models do not require the data to meet similar assumptions. However, the data must be transformed before being fed into the network. It requires positional encoding and model architecture design [Ahmed et al. 2023]. We calculate 4, 6, and 12-month forecasts using ARIMA for X13 adjusted data and ITransformers applied to raw data. To compare the results, we calculate differences in the obtained forecast. The differences were obtained based on MIXED models with autocorrelated errors. We use the effect size measure [Lankens 2013] to better understand the differences' significance.

The classical tool for time series forecasting is the autoregressive–moving-average (ARMA) model and its generalizations which were first published in [Box, Jenkings 1970]. ARMA(p,q) is expressed as the sum of two polynomials, autoregression AR and moving average (MA).

Model $ARMA(p, q)$ can be formulated in B-operator form as:

$$W(B)X_t = V(B)\varepsilon_t$$

or

$$X_t = \frac{V(B)}{W(B)}\varepsilon_t$$

where W() and V() are polynomials of degree p and q, respectively.

In the case of series that do not satisfy the assumption of stationarity but can be reduced to stationary, the integrated moving average ARIMA(p,d,q) model is used.

$$W(B)(1-B)^d X_t = V(B)\varepsilon_t$$

[2] Data is the property of the Institute of Mother and Child. It can be obtained on request.

The parameters that refer to the non-seasonal form of the ARIMA model are: p-the number of autoregressive components (lags), d-the number of differencing applied, q-the number of lagged error components. A seasonal ARIMA model is built by including additional seasonal terms in the ARIMA model with parameters $(P,D,Q)_m$ describing the seasonal part of the model. The terms of the seasonal part of ARIMA are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period. The parameters are: m-length of season, Q-the number of seasonal moving average components, P-the number of seasonal autoregressors. As a result the ARIMA requires adjusting 7 parameters. The combinations are selected based on AIC criterion.

X13 seasonal adjustment

X-13ARIMA-SEATS, is a successor to X-12-ARIMA and X-11. It is a set of statistical methods for seasonal adjustment and other descriptive analysis of time series data that are implemented in the U.S. Census Bureau's software package and are widely used to meet the requirements of forecasting models.

The X13 model assumes that a time series is a sum of components with different rates of change. In particular, it is described by four components:

- Seasonal component
- Trend component
- Cyclical component
- Random component (error)

The general form of the multiplicative model (used in this article) to describe time series looks like:

$$Y_t = Tr_t \cdot Sn_t \cdot \varepsilon_t,$$

where

Y_t - forecasted variable at time t

Tr_t - tendency (trend)

Sn_t - seasonal fluctuations at time t

ε_t - random component at time t (in the analysis of time series the structure of the random component is not considered)

In the calculations we use ARIMA(3,0,0)(0,0,1)[12] model.

Transformers based time series forecasting

Since the famous paper "Attention is All You Need" was published [Vaswani et al. 2017], the focus of the neural network community was paid solely to the new mechanism. Transformers have become the key element in almost all LLM. Numerous models using attention mechanisms have been proposed in time series forecasting, among them Informer, Autoformer, and Pyraformer. A survey on transformers for time series can be found in [Wen et al. 2023; Miller et al. 2024]. GluonTS provides various neural network models used for time series[3].

Much attention was given to the work “Are transformers effective for time series forecasting?” [Zeng et al. 2022], in which transformers application for long-term forecasting has been questioned. The recent model iTransformers is considered the latest breakthrough in long-run time series forecasting. [Liu et al. 2023]. It is regarded to be the best for long-term forecasting[4]. The main difference between this model and the common transformer approach lies in the encoding process, which is quite different; see G3 part of [Liu et al. 2023]. General Transformers’ architecture has been explained in many papers. Therefore, we consider its description here unnecessary.

Effect size

Effect sizes are measures that allow the determination of sample size or to examine how big the investigated effect or difference is [Lakens 2013]. To determine if the calculated RMSE indicates the difference in models’ predictions, we use Cohen’s eta squared comparison given by:

$$\eta^2 = \frac{SS_{effect}}{SS_{total}},$$

where SS is the total sample (corrected) sum of squares, and SS_{effect} is the observed sum of squares due to the effect being tested.

In practice, one often adjusts the model on the training data and tests it on the test data. In this case, the predicted values can be compared with the actual data and the model performance can be assessed using goodness of fit measures (MAE) [Miller et al. 2024]. We aim not to test the model’s performance but to compare the forecasts. Therefore, we calculate the differences between the obtained forecasts and infer the significance of means of differences. We test the hypothesis and estimate the effect size to draw conclusions.

RESULTS

Both ARIMA() and iTransformer models were used to build the forecasts. The ARIMA model was built on a deseasonalized series with the X13 procedure. A forecast of the seasonal component was generated using this procedure. Finally, both forecasts, ARIMA and X13, were combined to obtain the final forecast. The ARIMA hyperparameters were selected using an automatic search strategy for acceptable models. The Akaike Information Criterion (AIC) statistic was used to select the optimal model. Ultimately, the ARIMA(3,0,0)(0,0,1)$_{12}$ model proved the best. Admittedly, the seasonal component appeared here despite the earlier removal of

[3] https://ts.gluon.ai/stable/getting_started/models.html.

[3] https://github.com/thuml/Time-Series-Library?tab=readme-ov-file

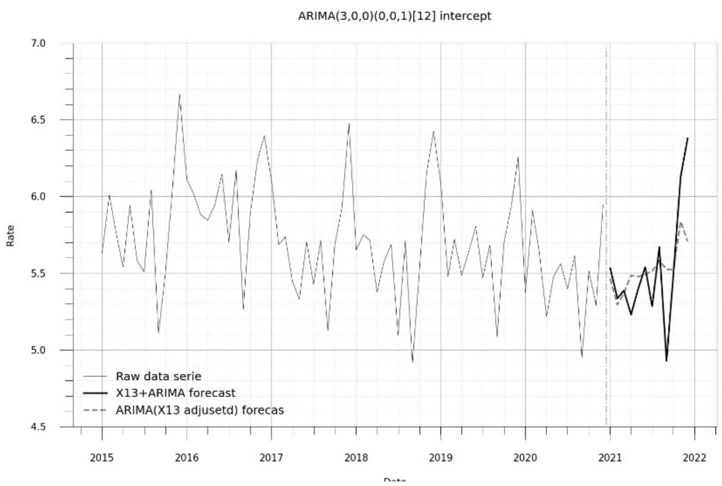
seasonality by the X13 method. Still, the authors left it out because the often-used automatic model selection procedure is not corrected in practice. The paper aimed to test the hypothesis of model equivalence, including the methodologies for using them.

In the data used, an effect is very often encountered in practice: the seasonal component is pronounced, and it determines the short-term forecasts. If this component is removed, we will have a slow-moving time series that ARIMA models can forecast.

In the case of the iTransformers model, the series was divided into a training and a test set, which were used to build the model. The testing set was 12 months, and the training set was 24 months. The training/testing window was shifted by 6 months to obtain further scenarios. The choice of size of the training and testing series was dictated by the expectation of a seasonal period. This choice made it unnecessary to pre-season the data.

Notably, the pattern obtained in this approach did not differ from that obtained in the X13 procedure - the forecasts did not differ within error limits.

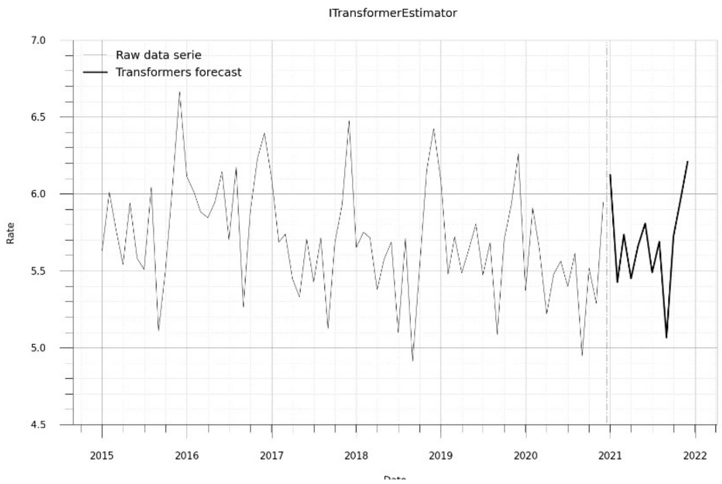
Figure 1. Preterm birth rate. iTransformers forecast for 12 months.



Source: own calculations

Fig. 1 presents the graph of the original time series (dashed line) and the forecast for the sequential 12 months (solid line) obtained for iTransforemrs. The forecast exhibits seasonal behavior with a similar pattern to the original data.

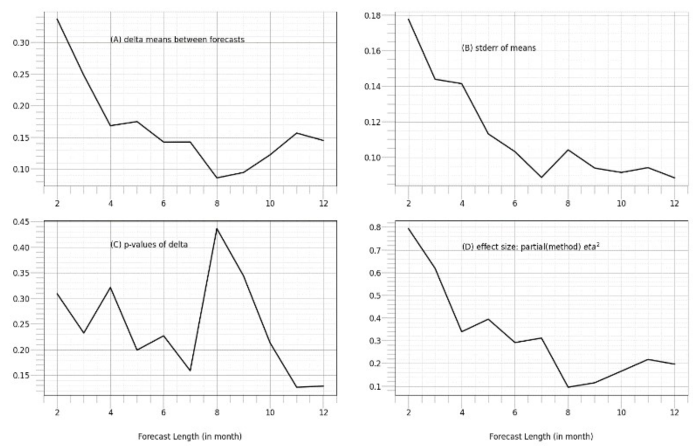
Figure 2. Seasonal ARIMA forecast for 12 months



Source: own calculations

Figure 2 shows the graph of the original time series (dashed line) and the forecast for the sequential 12 months (solid line) obtained for ARIMA X13 adjusted model. The forecast exhibits seasonal behavior with a similar pattern to the original data.

Figure 3. Comparison of forecasts for forecast horizon from 2 to 12 months. (A) differences of means calculated for different horizon values, (B) standard errors of means, (C) p values for hypothesis, (D) effect size for different horizon values.



Source: own calculations

Fig. 3 shows the results of comparing the forecasts obtained from the two models. The Generalized Linear Mixed Model (GLMM) was used for the comparison, which considers that the observations are not independent - the model uses combined observations for individual values of the forecast horizon length. In addition, each observation has a different variance, which leads to the need to introduce weights for each observation. The weights are inversely proportional to the variance.

As the forecast horizon length increases, the standardized distance between models decreases, forecast error increases, and significance tests have less and less power. Therefore, the last graph (D) shows the so-called effect size partial η^2 for the factor: the difference between forecasts.

Table 1. P-values calculated for testing mean differences

Model	Difference between the means	p-value
4 months	0.17	0.32
6 months	0.14	0.225
12 months	0.145	0.12

Source: own calculations

The values shown in Table 1 indicate that it no longer mandates the rejection of hypothesis H_0 of equality of model results.

Table 2. The effect size for different horizon values

Model	η^2	Effect
4 months	0.32	small/medium
6 months	0.3	small/medium
12 months	0.2	small

Source: own calculations

The values shown in Table 2 indicate that the measured difference between the forecast values of the two models can be considered small.

CONCLUSIONS

Our results indicate that both approaches yield similar results. The model based on iTransformer captures well seasonality without any data preparation.

Seasonal variations often affect time series of sub-annual observations, e.g., monthly, quarterly, and weekly. The presence of such variations introduces an essential component to analytical models. Indeed, seasonality usually accounts for most of the total variation within a year. Seasonality results from the fact that some months or quarters of the year are more important in activity or level. The seasonal pattern measures the relative importance of the months of the year and affects both the interpretation and projections of medium- and short-term results. The constant

100% represents the average month. The peak month is December, in which the prematurity rate is almost 15% higher than in the average month; the trough months are September and October, in which the rate is nearly 10% lower than in the average month. The seasonal amplitude, the difference between the peak and trough months of the seasonal pattern, is almost 40%. Research is ongoing to attempt to link the seasonal pattern to biological factors. Forecasting models that take seasonality into account are essential for macro-level management purposes. A methodology based on the X3-X13 procedures has been developed for several decades. It is widely used in economics, although it consists of steps based on moving averages of different lengths. In this article, we wanted to check whether deep learning methods allow for building time series forecasts in the case of seasonality. Based on the obtained results, it can be stated that iTransformer neural networks allow for obtaining results comparable to the X13 procedure. This means that iTransformer neural networks are suitable for modeling seasonal data. It is worth noting that the results are obtained from an analysis based on a self-attention mechanism that examines the global properties of the series - including seasonality, and not the search for local patterns, as in the case of autocorrelation

Comparing these approaches can help find the causes of strong seasonality in the studied problem of preterm births.

We hope our study will contribute to discussing Transformers' application in time series forecasting.

REFERENCES

Ahmed S., Nielsen I.E., Tripathi A. et al. (2023) Transformers in Time-Series Analysis: A Tutorial. Circuits Syst Signal Process 42, 7433-7466, https://doi.org/10.1007/s00034-023-02454-8.

Aijaz I., Agarwal P. (2019) A Study on Time Series Forecasting using Hybridization of Time Series Models and Neural Networks. Recent Patents on Computer Science, 12(5), 827-832, https://doi.org/10.2174/1573401315666190619112842.

Borboa-Olivares H., Rodríguez-Sibaja M.J., Espejel-Nuñez A., Flores-Pliego A., Mendoza-Ortega J., Camacho-Arroyo I., González-Camarena R., Echeverría-Arjonilla J.C., Estrada-Gutierrez G. (2023) A Novel Predictive Machine Learning Model Integrating Cytokines in Cervical-Vaginal Mucus Increases the Prediction Rate for Preterm Birth. International Journal of Molecular Sciences, 24(18), 13851, https://doi.org/10.3390/ijms241813851.

Box G., Jenkins G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Brandon D., McGrath J. (2016) Infant Mortality Rates and Preterm Birth: A Challenge for Advocacy. Advances in Neonatal Care, 16(5), 323-324, https://doi.org/10.1097/ANC.0000000000000344.

Dench D., Joyce T., Minkoff H. (2022) United States Preterm Birth Rate and COVID-19. Pediatrics, 149(5), https://doi.org/10.1542/peds.2021-055495.

Gemmill A., Casey J. A., Catalano R., Karasek D., Margerison C. E., Bruckner T. (2022) Changes in Preterm Birth and Caesarean Deliveries in the United States during the SARS-CoV-2 Pandemic. Paediatric and Perinatal Epidemiology, 36(4), 485-489, https://doi.org/10.1111/ppe.12811.

Lakens D. (2013) Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: a Practical Primer for T-Tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863.

Lim B., Zohren S. (2021) Time-Series Forecasting with Deep Learning: a Survey. Phil. Trans. R. Soc. A, 379: 20200209, https://doi.org/10.1098/rsta.2020.0209

Liu Y., Hu T., Zhang H., Wu H., Wang S., Ma L., Long M. (2023) Itransformer: Inverted Transformers Are Effective for Time Series Forecasting. arXiv preprint arXiv: 2310.06625.

Miller J. A., Aldosari M., Saeed F., Barna N. H., Rana S., Arpinar I. B., Liu N. (2024) A Survey of Deep Learning and Foundation Models for Time Series Forecasting. arXiv preprint arXiv:2401.13912.

Priya T., Sarkar B. K., Sahana S. K. (2024) Regression based Machine Learning Models for Forecasting Preterm Birth Cases. Procedia Computer Science, 235, 830-839, https://doi.org/10.1016/j.procs.2024.04.079.

Sefidkar R., Zayeri F., Kazemi E., Salehi M., Dehnad A., Hafizi M. (2021) A Trend Study of Preterm Infant Mortality Rate in Developed and Developing Countries over 1990 to 2017. Iranian Journal of Public Health, *50*(2), 369-375, https://doi.org/10.18502/ijph.v50i2.5353.

Tzitiridou-Chatzopoulou M., Zournatzidou G., Kourakos M. (2024) Predicting Future Birth Rates with the Use of an Adaptive Machine Learning Algorithm: A Forecasting Experiment for Scotland. International Journal of Environmental Research and Public Health, 21(7), 84, https://doi.org/10.3390/ijerph21070841.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. (2017) Attention Is All You Need. NeurIPS, 2017.

Wen Q., Zhou T., Zhang C., Chen W., Ma Z., Yan J., Sun L. (2022) Transformers in Time Series: A Survey. arXiv preprint arXiv:2202.07125.

Yu Q. Y., Lin Y., Zhou Y. R., Yang X. J., Hemelaar J. (2024) Predicting Risk of Preterm Birth in Singleton Pregnancies using Machine Learning Algorithms. Frontiers in Big Data, 7, https://doi.org/10.3389/fdata.2024.1291196.

Zeng A., Chen M., Zhang, L., & Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting?. Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121-11128, https://doi.org/10.1609/aaai.v37i9.26317.

Zhang F., Tong L., Shi C., Zuo R., Wang L. Wang Y. (2024) Deep Learning in Predicting Preterm Birth: A Comparative Study of Machine Learning Algorithms. Maternal-Fetal Medicine, 6(3), 141-146, https://doi.org/10.1097/FM9.0000000000000236.

Zhang Y., Du S., Hu T., Xu S., Lu H., Xu C., Li J., Zhu X. (2023) Establishment of a Model for Predicting Preterm Birth based on the Machine Learning Algorithm. BMC Pregnancy and Childbirth, 23(1), 779, https://doi.org/10.1186/s12884-023-06058-7.