İstanbul Üniversitesi Yayınevi / Istanbul University Press

İstanbul İktisat Dergisi
Istanbul Journal of Economics

İstanbul İktisat Dergisi
Istanbul Journal of Economics,
IJE 2025, 75 (1): 1–16

https://doi.org/10.26650/ISTJECON2023-1366172

Submitted: 27.09.2023
Revision Requested: 13.12.2024
Last Revision Received: 03.01.2025
Accepted: 05.05.2025

Research Article

Open Access

Forecasting the Turkish Manufacturing Industrial Production Index: An Empirical Comparison of Time Series and Machine Learning Models

Ufuk Bingöl [1], Hasan Murat Ertuğrul [2] & Necmettin Alpay Koçak [3]

[1] Bandırma Onyedi Eylül University, Manyas Vocational School, Department of Management and Organization, Balıkesir, Türkiye
[2] Anadolu University, Faculty of Economics, Eskişehir, Türkiye
[3] Hacettepe University, Faculty of Economics and Administrative Sciences, Ankara, Türkiye

Abstract Macroeconomic variables are important in both following cyclical economic developments and answering the questions of decision-makers and investors about the future. In this context, investigating the industrial production index dynamics over time provides rapid and important signals about the general economic prospects. Therefore, the effects of the COVID-19 outbreak on the forecasting performance of economic variables have been increasingly investigated in the literature. This study examines the forecasting performance differences between time series and machine learning models for the Turkish Manufacturing Industrial Production Index) across the pre- and post-COVID-19 periods. Using econometric and machine learning methods, we identified that the time series models performed better before COVID-19, while the machine learning models excelled post-COVID-19. According to the results for the pre-COVID-19 period, the ARDL model, which is a member of the time series model family, produces the best results in terms of forecast performance criteria, however the Principal Component Analysis model, which is a member of the machine learning model family, is found to be the best performing model for the post-COVID-19 period. This finding implies that the forecast performance of the time series and machine learning models is different depending on the COVID-19 outbreak. Time series models produce robust forecast performance before the COVID-19 period, whereas machine learning family member models produce robust results after the COVID-19 period for the Turkish Manufacturing Industrial Production Index variable. These results highlight the shifting utility of model families under economic disruption, offering insights for policymakers and forecasters.

Keywords Forecasting · Industrial production index · Time series models · Machine learning models

Jel Codes C4, C45

Citation: Bingöl, U., Ertuğrul, H. M. & Koçak, N. A. (2025). Forecasting the Turkish manufacturing ındustrial production ındex: an empirical comparison of time series and machine learning models. *İstanbul İktisat Dergisi–Istanbul Journal of Economics*, *75*(1), 1-16. https://doi.org/10.26650/ISTJECON2023-1366172

Corresponding author: Ufuk Bingöl ubingol@bandirma.edu.tr

Forecasting the Turkish Manufacturing Industrial Production Index: An Empirical Comparison of Time Series and Machine Learning Models

Recently, the importance of macroeconomic variables has been increasing, not only for following cyclical economic developments but also for answering the questions of decision-makers and investors about the future. In this context, scrutinising the changes in the industrial production indexes (IPI) over time provides rapid and important signals about the general economic prospects. Because manufacturing is one of the main drivers of the whole business cycle, the IPI index is one of the most important and widely analysed indicators for academics and market professionals (Bulligan et al., 2010). As manufacturing forms the backbone of industrial activity, forecasting the Manufacturing Industrial Production Index (M-IPI) is pivotal for evaluating the economy's trajectory, especially in emerging markets like Türkiye, where manufacturing plays a disproportionately large role in economic output.

The statistical offices publish the IPI on a monthly or quarterly basis to follow changes and developments industrial production. Although the IPI calculation varies depending on the country, it includes the sub-main sectors of mining, energy, and manufacturing. That said, the manufacturing sector's share industrial production in emerging countries is relatively high (Haraguchi et al., 2017). Therefore, forecasting the manufacturing industry production index (M-IPI) is fundamental in terms of future planning for economic development.

In the literature, various econometric models have been employed to predict the IPI or M-IPI. Accurate estimation of production indexes is quite sophisticated due to the scarcity of well-specified and variable-rich economic models. It is not possible to consider all the components of the economy's structural features in a country. Moreover, for variables obtained with short and different collection methods, data stationary, high noise ratios, and non-linear effects, etc. are some of the major obstacles in creating accurate time series models (Moody, 1995).

On the other hand, with recent advances in methodological and technological developments, long- or short-term forecasting methods employed in economic research are evolving and becoming widespread. Today, data science tools are employed together with standard econometric methods, and they produce very successful forecasting performance.

The outbreak of COVID-19 in early 2020 inevitably impacted Türkiye's manufacturing industry, mirroring its effects on the global economy. Imbalances in supply and demand conditions have left both permanent and temporary marks on the manufacturing industry. Therefore, the COVID-19 factor should be taken into account in forecasting the Turkish M-IPI.

This study aims to forecast the Turkish M-IPI using both standard econometric time series models and machine learning techniques, and to identify which kind of models are successful by comparing the forecasting performance of the alternative models. Accurate forecasting of macroeconomic indicators is vital for policymakers and investors. The Industrial Production Index (IPI) is a key economic indicator that reflects manufacturing trends. This study focuses on the Turkish Manufacturing IPI (M-IPI), comparing econometric and machine learning forecasting models. The COVID-19 pandemic underscored the need for robust models that adapt to economic disruptions. Using leading indicators like PMI and RSCI—despite their close timing to IPI—enhances short-term forecasting accuracy. Our study contributes by providing the first comprehensive comparison of these model families in the context of COVID-19. A distinct aspect of this study is the selection of the following predictor variables: the Manufacturing Purchasing Managers Index (PMI) and the Real Sector Confidence Index (RSCI). Both indicators are timely, with the PMI published shortly after the end of the reference month and the RSCI announced a week prior. This makes them effective tools for real-time forecasting of the M-IPI, a perspective not sufficiently explored in the existing literature. The analysis period was divided into two sub-periods, namely the training period and the test period. Using the model parameters determined during the training period, the forecasts for the test period were calculated. The methodology applied in this study is comprehensive, utilising a wide array of models ranging from traditional econometric techniques, such as ARDL and Markov switching regressions, to advanced machine learning approaches like principal component analysis (PCA) and neural networks. The forecasting performance of these models is rigorously compared using multiple metrics, including RMSE, MAE, MAPE, and Theil's U.

This study in addition seeks to fill two critical gaps in the forecasting literature. First, it is an empirical investigation on the time series and machine learning models when predicting the Turkish M-IPI. While a plenty of research has been done in econometric approaches vis-a-vis machine learning algorithms in the forecasting of M-IPI models in the Turkish context, none has looked at the comparative perspective. Second, the analysis includes the effect of COVID-19 pandemic, which marked an unprecedented upheaval in economic systems. By distinguishing between Pre-COVID-19 and Post-COVID-19, we are seeking to analyse the impact of structural economic transformations on the application of these methodologies, thereby testing their stability under extreme situations.

To the best of our knowledge, this is the first paper in the literature comparing the forecasting performance of both time series econometric models and machine learning algorithms for the Turkish M-IPI. We contend that this is the first contribution of this study to the existing literature. This paper also employs two different estimation and forecasting periods in order to analyse the effect of the COVID-19 outbreak on the forecast performance of the alternative models for the Turkish M-IPI. As far as we know, this is the first paper in the literature analysing the effect of the COVID-19 outbreak on the forecasting performance of the Turkish M-IPI, which is the third contribution of this study to the existing literature.

This study consists of five sections. Following the introduction, the second section examines the existing literature. The third section presents the data and methodology employed. Section 4 shares the findings of the empirical analysis and discusses the results. Finally, the conclusion presents an evaluation of the findings and avenues for future research.

Literature Review

The literature on industrial production forecasting started with discussions about the calculation methods. Maher (1957) estimated the turning points in the series by examining linear regression models to compare the cyclical variations and time series in the American Federal Reserve board industrial production with the Diffusion Index developed by Moore (1950). According to his results, false signal formation in the upper and lower limits of fluctuations in the diffusion method and the use of a moving average method with indicators longer than the series negatively affect the forecasting. It has been observed that the linear regression model is more consistent in solving these problems. Stekler (1961) also examined the independent components of the series, as mentioned above, as well as the diffusion index method's prediction performance using leading series regression. The study tested whether leading series regression (LSR) is better at predicting turning points than the diffusion index. Since the independent variables in the LSR are correlated, the impact of the independent components on the estimation was also examined. According to the research results, the diffusion index performed better than the LSR in predicting the turning points in industrial production. On the other hand, the leading basic components of the series have also been tested to be as good as LSR or even better in predicting turning points.

Davies and Scott (1973) examined 24 different sectors that form the industrial production index published by the Central Statistical Office (CSO) of the United Kingdom with naive econometric models vs regressions, and investigated which model successfully predicted turning points and acceleration and deceleration in growth. They divided the forecast performance of industry outputs into two categories, according to the kind of transfer of final expenditure on GDP, whether direct (food, tobacco, etc.) or indirect (cement, petroleum products, etc.). According to the results, regression analysis conducted on both categories predicted sectoral turning points more accurately than naïve methods.

Teräsvirta (1984) studied the Finnish Industrial production monthly time series using an econometric model with linear combinations for short-term forecasting. He compared the prediction success of the principal component analysis and transfer function models through various variables. According to his results, the principal components method predicted more accurately than the autoregressive integrated moving average (ARIMA) models, especially after various turning points in industrial production.

Bodo and Signorini (1987) estimated short-term Italian industrial production using the Holt-Winters and ARIMA models and a business survey and electricity consumption data series. According to their results, the estimation of univariate models is more successful than business surveys, although there is a delay of a few days on the electricity consumption data. However, the researchers stated that an accurate forecast would be provided by combining the models used in their estimation of both data sets with the weighted average method. Bodo et al. (1991) also used the Holt-Winters method to predict Italian industrial production in real-time before the end of the relevant month. In their results, data-based estimations and univariate models with the arithmetic means method provided the most accurate forecast of daily electric use. Bodo et al. (2000) stated that they predicted industrial production in the Euro-zone by using the ARIMA models, and although seasonal factors influenced the ARIMA models, they made a reliable and robust forecast. On the other hand, complex Vector Autoregression (VAR) models applied to increase forecasting success failed due to overparameterization problems. However, when the Eurozone as a single country and the US were included in the VAR models, there was a significant improvement in ARIMA's forecasts.

Moody et al. (1993) were the first to estimate the industrial production data (1950-1970 / 1980-1990) calculated monthly by the American Bureau of Economic Analysis using neural networks models, instead of standard linear autoregressive models. The study compared the predictors using a trivial predictor, a univariate linear autoregression (AR) model, a multivariate linear regression model, and two types of neural network models. Neural network models produced better estimates than the linear models. However, Moody (1995) stated that a single macroeconomic forecasting method would not be sufficient to reduce the estimation risk, and the combination of all models would show optimal performance in the estimation.

Thury and Witt (1998) predicted Austrian and German industrial production using ARIMA time series models. They found that basic structural models for univariate forecasting make more accurate estimations than ARIMA models.

Siliverstovs and van Dijk (2002) estimated industrial production growth rates in G7 countries using linear AR, non-linear and Structural Change models. They concluded that the Markov Switching Model (MSMH) is the best predictor in the short term and the Self-Exciting Threshold Autoregressive (SETAR) models are better in the long term. In this sense, it has been stated that non-linear models give better results than linear models to identify uncertainties.

Bradley and Jansen (2004) state that despite the overfitting problem in non-linear models, it is successful in forecasting except for the Multiple-Regime Smooth Transition Autoregressive (MRSTAR) model.

Heravi et al. (2004) estimated industrial production in Germany, France, and England using neural network models over non-seasonally adjusted time series from 24 different sub-sectors. According to their results, it has been observed that non-linear models are more successful than linear models in determining the direction of change in forecasting. On the other hand, linear models are more successful in estimating the growth rate although neural network models outperform growth direction in the forecasting accuracy criterion.

Zhang and Qi (2005) predicted seasonal and trend time series with Neural Networks (NN), and concluded that NNs made more accurate predictions with deseasonalize data. In this sense, it is stated that well-preprocessed NNs can provide more accurate forecasting than traditional ARIMA models due to their ability to eliminate overfitting problems.

Aminian et al. (2006) predicted the American Real GDP and IP using NN. The study once again showed that NNs are significantly more successful in forecasting than linear regression. Stock and Watson (2006) estimated the US IP growth rate over 130 predictive variables through forecast pooling, dynamic factor, and Bayesian models. They found that multivariate estimation models and non-linear estimation became widespread and made more accurate forecasting in macroeconomic analysis.

In the literature, the Bayesian VAR (Aprigliano, 2020; Barışık & Yayar, 2012; Günay, 2018), VAR-X (Bianchi et al., 2010), ARIMA (Alencar & Rocha, 2016; Öncel Çekim, 2018) and Spectrum analysis (Hassani et al., 2009) methods continued to be employed for estimating industrial production. However, artificial neural network models are becoming quite common, especially in forecasting the parameters of the macroeconomic variables (Babkin et al., 2016; Heravi et al., 2004; Moody, 2012; Polat & Temurlenk, 2011).

Studies that focus on forecasting and predicting in different disciplines, including macroeconomic variables regarding the period before and during the COVID-19 pandemic, are also becoming widespread in this context (Altig et al., 2020; Bildirici et al., 2020; Depren & Kartal, 2020; De Santis & Van der Veken, 2020; Güngör et al., 2021; Jena et al., 2021; Kartal et al., 2021; Larson & Sinclair, 2021; Nikolopoulos et al., 2021; Primiceri & Tambalotti, 2020; H. Zhang et al., 2021). According to the related literature, a common view about the forecast performance of the models during the COVID-19 pandemic is an important challenge. Prior research extensively examined IPI forecasting using econometric models such as ARIMA and machine learning approaches such as Neural Networks. However, these studies rarely consider external shocks like COVID-19, nor do they compare the performance of these models under such conditions. This study fills these gaps by providing empirical evidence on how the pandemic alters the model effectiveness for Turkish M-IPI.

This study aims to forecast the Turkish M-IPI by using both standard econometric time series models and machine learning techniques, which is different from the papers in the existing literature.

Data and Methodology

In the study, we employed the monthly M-IPI (2015=100, seasonally and calendar adjusted) obtained from TURKSTAT as the target (dependent) variable. We employ the Manufacturing Purchasing Manager Index-

PMI (seasonally adjusted) and the Real Sector Confidence Index-RSCI (seasonally adjusted) variables as independent variables in order to explain M-IPI variable. The selection of PMI and RSCI as predictor variables is grounded in their leading indicator properties and strong correlation with IPI. PMI reflects forward-looking measures like new orders and production, which directly precede manufacturing activity. Similarly, RSCI provides insights into managerial expectations and confidence levels, which are pivotal for short-term IPI predictions (referencing Haraguchi et al., 2017; Bodo et al., 1991). Despite their publication shortly before the IPI data, these variables offer significant forecasting advantages due to their early availability.

PMI is an indicator produced for many countries using an internationally comparable method, obtained from surveys conducted on purchasing managers of leading companies in the manufacturing sector, especially on orders, production, and employment. RSCI also aims to reflect the short-term trends of the manufacturing industry by monitoring the evaluations of senior managers about the recent and current situation and their expectations for the future. PMI is disclosed by IHS-MARKIT with a delay of 1 or 2 days after the end of the reference month, while RSCI is announced by the Central Bank of the Republic of Türkiye (CBRT) one week before the end of the reference month. We contend that using RSCI and PMI to predict IPI will also contribute to early estimations of IPI.

Although M-IPI starts from 1986 and RSCI data starts from 2007, PMI data is only available from January 2012. For this reason, January 2012 was chosen as the starting date of the study. Therefore, this study will cover the period from January 2012 to September 2020, which includes 105 observations. Since the M-IPI index is not stationary at its level, annual growth rates, $(IPI_t - IPI_{t-12}) / IPI_{t-12}$, are used.

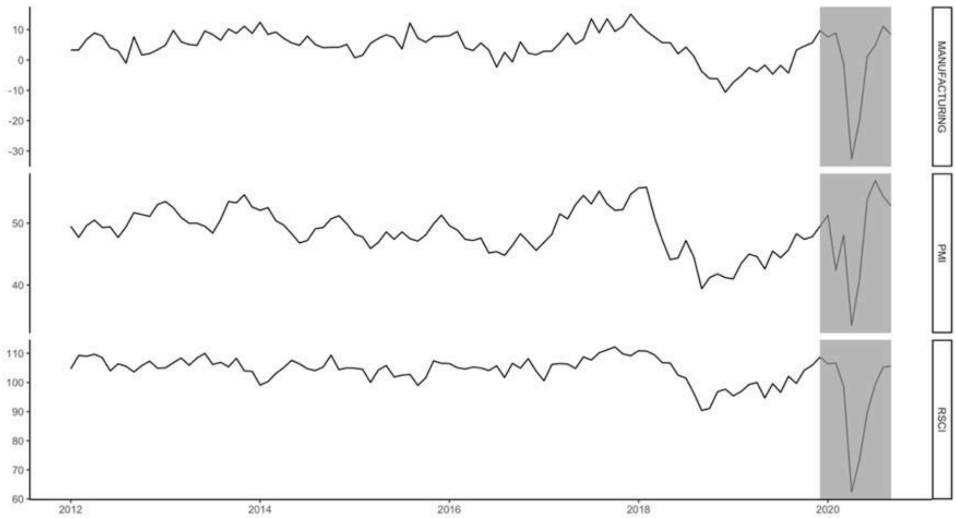
Our sample includes the period covering the COVID-19 (Severe Acute Respiratory Syndrome Coronavirus 2, SARS-CoV-2) outbreak. The COVID-19 outbreak has had a negative effect on the forecasting performance of the models (Altig et al., 2020; Güngör et al., 2021; Jena et al., 2021; Primiceri & Tambalotti, 2020). In order to consider the COVID-19 outbreak's distortion on the forecasting performance of the alternative models, we examined two different estimation and forecasting periods.

The forecasting periods were decided according to the date when the COVID-19 outbreak first appeared in Türkiye. The first COVID-19 case in Türkiye was seen on March 11, 2020. The first estimation and forecasting periods for the alternative models are January 2012-December 2019 and January 2020-March 2020, respectively. The first forecasting period coincided with the pre-COVID-19 outbreak period.

The second estimation and forecasting periods are January 2012-March 2020 and April 2020-September 2020, respectively. The second forecasting period covers the COVID-19 outbreak. By doing so, an attempt at examining the impact of the COVID-19 outbreak on the forecast performance of the alternative models is made. The variables employed in the empirical study are presented in Figure 1.

Figure 1

Turkish Manufacturing IPI, PMI and RSCI data for 2012 –2020



The models employed in the empirical study are presented in Table 1. The investigated models can be grouped under two main headings: time series and machine learning models. Static regression, dynamic regression, and univariate models were used for the time series models[1]. Tree-based, neural-network, advanced tree-based, and decomposition models were used for the machine learning models. The technical details about the investigated machine learning models are presented in the Appendix.

Table 1

Alternative Models

Family	Group	Models
Time series	Univariate	• ARMA
	Static regression	• OLS
		• Fully Modified OLS (FMOLS)
		• Threshold Regression (THRESHOLD)
	Dynamic regression	• Markov Switching Regression (MARKOV)
		• Dynamic OLS (DOLS)
		• ARDL
Machine Learning	Tree-based	• Bagged CART
		• CUBIST
	Neural-network	• Neural Network (NNET)
		• Bayesian Regularized Neural Networks (BRNN)
	Advanced Tree-based	• Random Forest (RF)
		• Quantile Random Forest (QRF)
	Decomposition	• Elasticnet (ENET)
		• Principal Component Regression (PCR)
		• Relevance Vector Machines with a Polynomial Kernel (RVMPOLY)

[1]See (Ertuğrul & Gebeşoğlu, 2020; Ertugrul & Mangir, 2015; Ertuğrul & Seven, 2021) for further technical information regarding the time series models.

After running the alternative models, we compared them according to their forecast performance by employing various performance criteria, namely RMSE, MAE, MAPE and Theil Inequality indicators. The lower the values of these indicators, the closer the forecasts that the model produces are to the observed trends. The forecast performance indicators are described in the Appendix.

Results and Discussion

In order to estimate the best fit model for the M-IPI index, we use alternative time series and machine learning models for the period from January 2012 to September 2020. We employ the PMI and RSCI variables as independent variables. We compare alternative models according to their forecasting performance criteria, including RMSE, MAE, MAPE and Theil Inequality indicators and define the best fit model.

We employed two different estimation and forecasting periods in order to analyse the effects of the COVID-19 outbreak on the forecast performance of the models. The first estimation and forecasting periods for the alternative models are January 2012-December 2019 and January 2020-March 2020, respectively. The second estimation and forecasting periods are January 2012-March 2020 and April 2020-September 2020, respectively. The first forecasting period coincides with the pre-COVID-19 outbreak period, while the second forecasting period covers the COVID-19 outbreak.

The forecast performance comparison of the alternative time series and machine learning models for the pre-COVID-19 period is presented in Table 2.

Table 2

Forecast Performance Comparison of Alternative Models for the Pre-COVID-19 Period

Pre Covid-19 Outbrek Period (Estimation Period: 2012m01-2019m12-Forecast Period: 2020M1-2020M3)					
Models	**Family***	**RMSE**	**MAE**	**MAPE**	**Theil Inequality**
ARDL (1,1,3)	TS	3,028	2,812	84,222	0,235
ARMA (1,1)	TS	4,988	3,364	239,353	0,347
ARMA (3,2)	TS	5,162	3,398	248,008	0,352
DOLS	TS	5,237	4,001	120,688	0,467
FMOLS	TS	6,012	4,565	141,32	0,520
MARKOV	TS	5,517	4,679	185,214	0,450
OLS	TS	5,111	4,371	193,644	0,429
THRESHOLD	TS	5,791	4,803	200,249	0,487
BAGGED CART	ML	7,496	6,385	236,731	1,106
CUBIST	ML	7,050	5,501	190,142	1,040
NEURAL NETWORK	ML	7,283	5,652	190,146	1,075
BAYESIAN REGULARIZED NEURAL NETWORKS	ML	6,497	5,006	166,934	0,959
RANDOM FOREST	ML	6,287	5,63	216,147	0,928
QUANTILE RANDOM FOREST	ML	5,58	5,019	225,508	0,823
ELASTICNET	ML	5,920	4,535	147,967	0,873
PRINCIPAL COMPONENT ANALYSIS	ML	4,241	3,331	103,435	0,326
RELEVANCE VECTOR MACHINES WITH POLYNOMIAL KERNEL	ML	5,924	4,544	147,992	0,877

*TS: Time Series, ML: Machine Learning

According to Table 2, the ARDL model (TS family) was found to be the best fit model for the pre-COVID-19 period according to all forecast performance indicators. All forecast performance indicators imply that the highest forecast performance can be achieved by employing the ARDL model. After the ARDL model, the PCA (principal component analysis. ML family) model was found to be the second best fit model according to all forecast performance indicators. The ARMA(1,1) model (TS family) was found to be the third best model according to three (RMSE, MAE and Theil Inequality Coefficient) out of four forecast performance indicators.

The three least successful forecasting models for the Turkish M-IPI for the pre-COVID-19 period are the Bagge Cart, Cubist, and the Neural Network models. All three unsuccessful models are members of the ML family.

Hence, TS family models performed better than ML family models during the pre-COVID-19 outbreak period.

The forecast performance comparison of alternative time series and machine learning models for the post-Covid-19 periods is presented in Table 3.

Table 3

Forecast Performance Comparison of Alternative Models after the COVID-19 Period

After Covid-19 Outbrek Period (Estimation Period: 2012m01-2019m3-Forecast Period: 2020M4-2020M9)					
Models	**Family***	**RMSE**	**MAE**	**MAPE**	**Theil Inequality**
ARDL (1,1,3)	TS	17,578	10,54	100,718	0,691
ARMA (1,1)	TS	16,449	11,518	66,438	0,862
ARMA (3,2)	TS	19,067	12,733	75,098	0,892
DOLS	TS	5,011	4,056	64,899	0,166
FMOLS	TS	5,717	4,605	77,179	0,193
MARKOV	TS	6,972	5,439	72,771	0,25
OLS	TS	10,695	7,771	73,839	0,459
THRESHOLD	TS	25,208	19,947	429,497	0,772
BAGGED CART	ML	13,543	8,997	106,947	0,806
CUBIST	ML	6,388	4,908	69,138	0,380
NEURAL NETWORK	ML	11,855	8,110	93,001	0,706
BAYESIAN REGULARIZED NEURAL NETWORKS	ML	11,239	7,645	83,878	0,669
RANDOM FOREST	ML	12,797	9,021	121,259	0,762
QUANTILE RANDOM FOREST	ML	13,144	9,438	146,814	0,783
ELASTICNET	ML	5,684	4,412	59,802	0,338
PRINCIPAL COMPONENT ANALYSIS	ML	4,372	3,338	27,875	0,160
RELEVANCE VECTOR MACHINES WITH POLYNOMIAL KERNEL	ML	6,003	4,587	54,336	0,357

According to Table 3, the PCA model (ML family) was found to be the best fit model for the post-COVID-19 outbreak period according to all forecast performance indicators. All forecast performance indicators show that the highest forecast performance can be achieved with the PCA model. After the PCA model, the DOLS model (TS family) was found to be the second best model according to three (RMSE, MAE and Theil Inequality Coefficient) out of four forecast performance indicators. The Elasticnet model (ML family) was found to be the third most successful forecasting model according to three (RMSE, MAE and MAPE) out of four forecast performance indicators.

The three least successful forecasting models for the Turkish M-IPI for the post-COVID-19 period are the Threshold, ARMA and ARDL models. All three unsuccessful models are members of the TS family.

Hence, ML family models demonstrated superior performance compared to TS family member models during the post-COVID-19 outbreak period. The post-COVID-19 period witnessed significant shifts in economic patterns, including non-linear disruptions and higher volatility. Machine learning models such as PCA and Elastic Net excelled due to their ability to uncover underlying data structures and adapt to these non-linear patterns. Traditional time series models, constrained by their reliance on historical linear trends, struggled to adapt to the new economic dynamics introduced by the pandemic (Zou & Hastie, 2005; Moody, 1995).

The differences in the performance of the models may mean that these forecasting tools are effective in certain contexts and therefore are informative in the selection of tools in the event of economic shocks. Given the short- and long-run characteristics of the ARDL model and its ability to capture the co-integration aspect of the DOLS model, they are more ideally suited for periods of economic stability. However, these models had difficulty responding to the new circumstances and times of COVID-19. Restrictions on the use of linear relationships and the inability to employ non-linear patterns were the major constraints. The results indicate that econometric models can be useful for forecasting in normal times but less so in periods of structural breaks and high degree of uncertainty.

With regard to Machine learning models findings, in particular PCA and Elastic Net, these also demonstrated their promise in the time after the COVID-19 pandemic. These models have been designed to conceivably outperform conventional econometric models in moments of economic stress because they are excellent at uncovering underlying architectures and functional non-linearities in data. It is this property of PCA, which allows for reduction in dimensions of data without losing vital components, which enabled predictions to be made during the chaos introduced by the pandemic. Likewise, with the Elastic Net, being able to incorporate regularisation techniques improved the performance of models by resolving multicollinearity and overfitting problems. While machine learning models reach their zenith in complex and crisis situations, they must be provided in input with variables, which are as rich as possible to be able to demonstrate their best.

The forecast performance was evaluated using RMSE, MAE, and MAPE metrics, with additional robustness checks. Empirical results indicate ARDL's robustness in stable conditions, while PCA adapts well during periods of volatilities such as COVID-19 disruptions. These findings indicate that time series models are suitable for predictable environments, whereas machine learning models perform better under structural breaks and economic uncertainty. The results indicate that the forecast performance of the TS and ML models is different depending on the COVID-19 outbreak. The TS family member models produced robust forecast performance before the COVID-19 period, while ML family member models produced robust results after the COVID-19 period for the Turkish M-IPI variable.

Conclusion

Macroeconomic variables are important both for following cyclical developments within economies and for answering the questions that decision-makers and investors have about the future. In this context, investigating IPI over time provides rapid and important cues about the general economic prospects.

Obtaining a robust forecast for M-IPI for emerging countries such as Türkiye could thus contribute to the forecast of the country's overall economy.

COVID-19 disrupted supply chains and altered demand patterns, fundamentally changing PMI dynamics. Due to a lack of literature on this topic, we compared the forecast performance of both econometric time series and machine learning models for the Turkish M-IPI for pre- and post- COVID-19 outbreak. The forecast performance of the models was compared according to various forecast performance criteria that are employed the most in the empirical literature for the period between January 2012 and September 2020. Traditional models struggled to adapt to these rapid shifts, while machine learning approaches leveraged their flexibility to better capture non-linear trends.

According to the model results for the pre-COVID-19 period, the ARDL model, PCA model, and ARMA(1,1) model were found to be the best performing models according to the forecast performance comparison for the Turkish M-IPI variable. Both ARDL and ARMA models are under the time series family, whereas the PCA model is under the machine learning family. The Bagge Cart, Cubist and Neural Network models were found to be the most unsuccessful models according to the forecast performance comparison for the Turkish M-IPI. All three unsuccessful models are within the family of machine models for the pre-COVID-19 period.

The PCA, DOLS and Elasticnet models were found to be the best performing models according to the forecast performance comparison for the Turkish M-IPI variable in that specific order in the post-COVID-19 period. Both the PCA and Elasticnet models fall under the machine learning family and the DOLS model falls under the time series family. Threshold, ARMA and ARDL models were found to be the most unsuccessful models according to the forecast performance comparison for Turkish M-IPI in that order. All three unsuccessful models fall under the time series family for the post-COVID-19 period.

According to the model results, the forecast performance of the time series and machine learning models differed depending on the COVID-19 outbreak. Time series family member models showed robust forecast performance before the COVID-19 period, while machine learning family member models produced robust results after the COVID-19 period for the Turkish M-IPI variable.

The study's limitations are the combinations of variables to explain manufacturing IPI, the selected period coverage, and the selected performance indicators. Expanding the range of predictor variables to include additional macroeconomic and sector-specific indicators could enhance model robustness. While this study does not explicitly incorporate confidence intervals or additional robustness checks, the performance metrics employed (RMSE, MAE, MAPE) provide reliable indicators of model accuracy. Future research could address this by implementing bootstrapped confidence intervals or alternative testing procedures to further validate the findings. The limited number of explanatory variables (PMI and RSCI) may constrain the performance of the machine learning models, which generally benefit from larger feature sets. Expanding the variable set could enable these models to better capture complex relationships and improve the forecasting accuracy. This limitation is acknowledged as a potential avenue for future research. Combination forecasts obtained from the best performance time series and machine learning models can be investigated as alternative models within the framework of further research. Also, forecast performance comparisons for alternative machine learning/time series models using high frequency (daily or intra day) variables are thought to be important for future research. Future research could expand the predictor variable set to include additional macroeconomic indicators, such as energy prices or export data, to enhance the robustness of the machine learning models. This study acknowledges the potential benefits of incorporating a richer variable set, particularly for machine learning approaches, in improving prediction accuracy. Finally, extending the analysis to post-COVID-19 recovery periods could provide insights into the long-term impacts of the pandemic on industrial production forecasting methodologies.

Ethics Committee Approval: This article does not contain any studies with animals or human beings performed by any of the authors. For this type of study, formal consent is not required.

Author Contributions: Conception/Design of Study- U.B., H.M.E., N.A.K.; Data Acquisition- U.B., H.M.E., N.A.K.; Data Analysis/ Interpretation- U.B., H.M.E., N.A.K.; Drafting Manuscript- U.B., H.M.E., N.A.K.; Critical Revision of Manuscript- U.B., H.M.E., N.A.K.; Final Approval and Accountability- U.B., H.M.E., N.A.K.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

Author Details

Ufuk Bingöl

[1] Bandırma Onyedi Eylül University, Manyas Vocational School, Department of Management and Organization, Balıkesir, Türkiye

0000-0003-1834-842X ubingol@bandirma.edu.tr

Hasan Murat Ertuğrul

[2] Anadolu University, Faculty of Economics, Eskişehir, Türkiye

0000-0001-9822-4683

Necmettin Alpay Koçak

[3] Hacettepe University, Faculty of Economics and Administrative Sciences, Ankara, Türkiye

0000-0002-4232-9985

References

Alencar, A. P., & Rocha, F. M. M. (2016). Forecasting the Brazilian Industrial Production Index with Level and Trend Change after the Crisis and SARIMA Models. *International Journal of Statistics & Economics*, *17*(1), 22–29.

Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E., Mizen, P., Parker, N., Renault, T., Smietanka, P., & Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics*, *191*, 104274. https://doi.org/10.1016/j.jpubeco.2020.104274

Aminian, F., Suarez, E. D., Aminian, M., & Walz, D. T. (2006). Forecasting economic data using neural networks. *Computational Economics*, *28*(1), 71–88. https://doi.org/10.1007/s10614-006-9041-7

Aprigliano, V. (2020). A large Bayesian VAR with a block-specific shrinkage: A forecasting application for Italian industrial production. *Journal of Forecasting*, *39*(8), 1291–1304. https://doi.org/10.1002/for.2687

Babkin, A. V., Karlina, E. P., & Epifanova, N. S. (2016). Neural networks as a tool of forecasting of socioeconomic systems strategic development. *Proceedings of the 28th International Business Information Management Association Conference-Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth*, *207*, 11–17. doi:10.1016/j.sbspro.2015.10.096

Bair, E., Hastie, T., Paul, D., & Tibshirani, R. (2006). Prediction by supervised principal components. *Journal of the American Statistical Association*. https://doi.org/10.1198/016214505000000628

Barışık, S., & Yayar, R. (2012). Sanayi üretim endeksini etkileyen faktörlerin ekonometrik analizi. *İktisat İşletme ve Finans*, *27*(316). doi:10.3848/iif.2012.316.3342

Bianchi, C., Carta, A., Fantazzini, D., Elena De Giuli, M., & Maggi, M. A. (2010). A copula-VAR-X approach for industrial production modelling and forecasting. *Applied Economics*, *42*(25), 3267–3277. doi:10.1080/00036840802112349

Bildirici, M., Bayazit, N. G., & Ucan, Y. (2020). Analysing crude oil prices under the impact of COVID-19 by using lstargarchlstm. *Energies*, *13*(11). doi:10.3390/en13112980

Bodo, G., Cividini, A., & Signorini, L. F. (1991). Forecasting the Italian industrial production index in real time. *Journal of Forecasting*, *10*(3), 285–299. https://doi.org/10.1002/for.3980100305

Bodo, Giorgio, Golinelli, R., & Parigi, G. (2000). Forecasting industrial production in the Euro area. *Empirical Economics*, *25*(4), 541–561. https://doi.org/10.1007/s001810000032

Bodo, Giorgio, & Signorini, L. F. (1987). Short-term forecasting of the industrial production index. *International Journal of Forecasting*. https://doi.org/10.1016/0169-2070(87)90006-9

Bradley, M. D., & Jansen, D. W. (2004). Forecasting using a nonlinear dynamic model of stock returns and industrial production. *International Journal of Forecasting*, *20*(2), 321–342. doi:10.1016/j.ijforecast.2004.09.007

Breiman, L. (1996). Bagging predictors. *Machine Learning*. https://doi.org/10.1023/A:1018054314350

Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. In *Classification and Regression Trees*. https://doi.org/10.1201/9781315139470

Bulligan, G., Golinelli, R., & Parigi, G. (2010). Forecasting monthly industrial production in real-time: From single equations to factor-based models. *Empirical Economics*. https://doi.org/10.1007/s00181-009-0305-7

Dan Foresee, F., & Hagan, M. T. (1997). Gauss-Newton approximation to bayesian learning. *IEEE International Conference on Neural Networks-Conference Proceedings*. https://doi.org/10.1109/ICNN.1997.614194

Davies, S. W., & Scott, T. W. K. (1973). Forecasting industrial production. *National Institute Economic Review*, *66*(1), 54–68. https://doi.org/10.1177/002795017306600104

De Santis, R. A., & Van der Veken, W. (2020). *Forecasting macroeconomic risk in real time: Great and COVID-19 Recessions*. https://doi.org/10.2866/019813

Ertuğrul, H. M., & Gebeşoğlu, P. F. (2020). The effect of a private pension scheme on savings: A case study for Turkey. *Borsa Istanbul Review*, *20*(2). https://doi.org/10.1016/j.bir.2019.12.001

Ertugrul, H. M., & Mangir, F. (2015). The tourism-led growth hypothesis: empirical evidence from Turkey. *Current Issues in Tourism*, *18*(7). https://doi.org/10.1080/13683500.2013.868409

Ertuğrul, H. M., & Seven, Ü. (2021). Dynamic spillover analysis of international and Turkish food prices. *International Journal of Finance and Economics*. doi:10.1002/ijfe.2517

Günay, M. (2018). Forecasting industrial production and inflation in Turkey using factor models. *Central Bank Review*, *18*(4), 149–161. https://doi.org/10.1016/j.cbrev.2018.11.003

Güngör, B. O., Ertuğrul, H. M., & Soytaş, U. (2021). Impact of COVID-19 outbreak on Turkish petrol consumption. *Technological Forecasting and Social Change*, *166*, 120637. https://doi.org/10.1016/j.techfore.2021.120637

Haraguchi, N., Cheng, C. F. C., & Smeets, E. (2017). The Importance of Manufacturing in Economic Development: Has It Changed? *World Development*. https://doi.org/10.1016/j.worlddev.2016.12.013

Hassani, H., Heravi, S., & Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, *25*(1), 103–118. doi:10.1016/j.ijforecast.2008.09.007

Heravi, S., Osborn, D. R., & Birchenhall, C. R. (2004). Linear versus neural network forecasts for the European industrial production series. *International Journal of Forecasting*, *20*(3), 435–446. https://doi.org/10.1016/S0169-2070(03)00062-1

Jena, P. R., Majhi, R., Kalli, R., Managi, S., & Majhi, B. (2021). Impact of COVID-19 on the GDP of major economies: Application of the artificial neural network forecaster. *Economic Analysis and Policy*, *69*, 324–339. https://doi.org/10.1016/j.eap.2020.12.013

Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression. *Applied Statistics*. https://doi.org/10.2307/2348005

Larson, W. D., & Sinclair, T. M. (2021). Nowcasting unemployment insurance claims in the time of COVID-19. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2021.01.001

MacKay, D. J. C. (1992). Bayesian Interpolation. *Neural Computation*. doi:10.1162/neco.1992.4.3.415

Maher, J. E. (1957). Forecasting Industrial Production. *Journal of Political Economy*, *65*(2), 158–165. https://doi.org/10.1086/257899

Moody, J. (2012). Forecasting the Economy with Neural Nets: A Survey of Challenges and Solutions. In G. B. Orr & K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade* (pp. 343–367). Springer-Verlag. doi:10.1007/978-3-642-35289-8_22

Moody, J. (1995). Economic forecasting: challenges and neural network solutions Economic Forecasting: Challenges and Neural Network Solutions. *International Symposium on Artificial Neural Networks*, *December*, 1–8. http://digitalcommons.ohsu.edu/csetech%5Cnhttp://digitalcommons.ohsu.edu/csetech/266

Moody, J., Levin, U., & Rehfuss, S. (1993). Predicting the U . S . Index of Industrial Production (Extended Abstract). In M. Novak (Ed.), *PASE '93: Parallel Applications in Statistics and Economics* (Vol. 3, Issue 6, pp. 791–794). VSP International Science Publishers.

Moore, G. H. (1950). *Statistical Indicators of Cyclical Revivals and Recession*.

Nguyen, D., & Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *IJCNN. International Joint Conference on Neural Networks*. https://doi.org/10.1109/ijcnn.1990.137819

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*, *290*(1), 99–115. doi:10.1016/j.ejor.2020.08.001

Öncel Çekim, H. (2018). Türkiye'de sanayi üretim endeksinin zaman serileri yöntemi ile incelenmesi. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *20*(1), 1–8. https://doi.org/10.25092/baunfbed.423143

Polat, Ö., & Temurlenk, S. (2011). Yapay Sinir Ağları Metodolojisi İle Makroekonomik Zaman Serilerinde Öngörü Modellemesi. *Dicle Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *1*(2), 98–106.

Primiceri, G. E., & Tambalotti, A. (2020). Macroeconomic Forecasting in the Time of COVID-19. *Working Paper*, *June*.

Quinlan, J. R. (1992). Learning with continuous classes. *Australian Joint Conference on Artificial Intelligence*.

Quinlan, J. R. (1993). Combining Instance-Based and Model-Based Learning. In *Machine Learning Proceedings 1993*. doi:10.1016/b978-1-55860-307-3.50037-x

Ripley, B. D. (1996). Pattern recognition and neural networks. In *Pattern Recognition and Neural Networks*. https://doi.org/10.1017/CBO9780511812651

Siliverstovs, B., & van Dijk, D. (2002). Forecasting Industrial Production with Linear, Nonlinear, and Structural Change Models. *Econometric Institute Report EI 2003-16*.

Stekler, H. O. (1961). Forecasting Industrial Production. *Journal of the American Statistical Association*, *56*(296), 869–877. https://doi.org/10.1080/01621459.1961.10482131

Stock, J. H., & Watson, M. W. (2006). Forecasting with Many Predictors. *Handbook of Economic Forecasting*, *1*(05), 515–554. https://doi.org/10.1016/S1574-0706(05)01010-4

Teräsvirta, T. (1984). Short-term forecasting of industrial production by means of quick indicators. *Journal of Forecasting*, *3*(4), 409–416. https://doi.org/10.1002/for.3980030404

Thury, G., & Witt, S. F. (1998). Forecasting industrial production using structural time series models. *Omega*, *26*(6), 751–767. https://doi.org/10.1016/S0305-0483(98)00024-3

Tipping, M. E. (2001). Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*. https://doi.org/10.1162/15324430152748236

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S Fourth edition by. In *World* (Vol. 53, Issue March).

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501–514. doi:10.1016/j.ejor.2003.08.037

Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, *87*, 103149. https://doi.org/10.1016/j.annals.2021.103149

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix

Machine Learning Models

The time series and machine learning model types discussed in this study have very detailed explanations in the literature and thus it is not possible to detail it further due to scope limitations. This is particularly the case for the many specifications/estimation options for each model. Therefore, the models used in this study are explained briefly in this section.

Bagging Classification and Regression Tree (Bagging CART)

One of the most well-known decision tree algorithms is the classification and regression tree (CART) algorithm proposed by Breiman et al.(1984). Each decision tree aims to divide the training data set into subgroups. It is essential that these subgroups are homogeneous within themselves; however, intergroup heterogeneity is desirable. A fixed coefficient estimate (e.g. average) is made for each group. Groups essentially contain data corresponding to simple yes-no answers. "Bagging" means that more than one CART model is renewed and combined with the bootstrap method (Breiman, 1996).

Random Forests (RF)

Random forests can be defined as "bagged decision trees". To put it more clearly, it is the merging of unrelated decision trees by analysing different dimensions of the dataset. As applications, they have become simplified algorithms. The algorithm proposed by Breiman (2001) is very popular. The random forests approach can be defined as bringing together unrelated trees or reducing the correlation between trees by adjusting the level of randomness. The difference from the Bagged decision tree is that more trees are brought together in this method.

Quantile Random Forest (QRF)

The QRF is an extension of the random forest approach proposed by Breiman (2001) for percentiles. It uses not only the performances on the output but also the mean and variances in the tree leaves as a target when making predictions. In the QRF method, empirical percentage estimates are used during the estimation.

CUBIST

In this study, the cubist method is a prediction-oriented regression model that combines the studies in Quinlan (1992) and Quinlan (1993). It also creates rules, as seen previously in the decision tree approach, and learns the relationship between the input and output variables as rules. Although similar to decision tree models, its most prominent feature is that it estimates a regression model for each rule.

Neural Network (NNET)

The artificial neural networks method consists of applying an approach used in computer science to statistical problems. The NNET algorithm learns the match between inputs and outputs and applies feedback through a continuous control mechanism. This approach differs in how neurons are modelled. Although it is difficult to represent and explain the complex structure of neurons, it is stated in the literature that it is quite successful in many statistical problems. The algorithm developed by Ripley (1996) and Venables & Ripley (2002) was used in this study.

Bayesian regularized artificial neural networks (BRANN)

Bayesian regularized artificial neural networks (BRANNs) are the more robust version of the classical neural network. Bayesian regularization is a method similar to ridge regression, which is useful in cases of multicollinearity. The algorithm used in this study fits a two-layer neural network as described in MacKay (1992) and Dan Foresee & Hagan (1997). It uses the Nguyen & Widrow algorithm (1990) to assign initial weights and the Gauss-Newton algorithm to perform the optimisation.

Elastic Net (ENET)

The Elastic Net method proposed by Zou & Hastie (2005) was used in this study. This method includes regularisation and feature selection approaches. According to Zou & Hastie (2005), this method, which is claimed to have a superb predictive power, essentially uses the grouping approach. Zou and Hastie (2005) claim that it is superior to other methods when the number of exogenous variables is greater than the number of observations.

Principal component regression (PCR)

Principal component regression (PCR) is a method based on principal component analysis (PCA). In the principal components regression approach, instead of using the exogenous variables directly, the principal components obtained from them are used as exogenous variables in the regression model (Jolliffe, 1982). Often, every possible combination is tried, and the best performing components are used instead of using all of the main components (Bair et al., 2006). The PCR model used in this study uses the singular value decomposition algorithm developed by Martens and Næs (1989).

Relevance Vector Machines with a Polynomial Kernel (RVMPOLY)

The Support Vector Machine (SVM) method performs generalisation operations over the dataset by kernel representation. However, the "Relevance Vector Machine" model used in this study is a Bayesian model and predicts the regression model with an algorithm identical to the support vector machine. It has a stingier parameter specification compared to SVM models (Tipping, 2001).