Kolegium Analiz Ekonomicznych

**Summary of Accomplishments**

# Demand Forecasting Model in the Steel and Iron Industry

Magdalena Barska

Warszawa 2021

## Dissertation subject and selection rationale

The dissertation "Demand forecasting model in the steel and iron industry" focuses on the demand and sales level analysis for the mentioned industry. We refer to the realized demand, measured by consumption. Demand forecasting for industrial markets receives much less attention than for consumer goods, while it is shaped by different factors, including economic condition, product life cycles, competitors' behaviour, political and environmental decisions. Examining these dependencies sets an interesting research area.

The analysis of future demand and sales volumes is an important aspect of any business activity. It becomes essential for determining the resources and materials needed in the production process and for employment planning. Both the surplus and shortage of production are disadvantageous in a company's perspective. A surplus incurs the costs of product storage and disposal, while a shortage results in the lost opportunity costs or in customer churn.

Many companies favor qualitative methods over quantitative ones and tend to rely on internal expert assessment. They do so to avoid additional implementation and maintenance costs and complexities, which leads to the limited usage of the latter. However, frequently changing business environment of modern companies calls for a suitable forecasting method to examine the dependencies. A good quality forecasting model can be applied to validate the result of a qualitative approach or may become one of the stages of the forecasting process.

The selection of such a model and identification of independent variables is one of the key considerations of the study. Taking into account specific characteristics of the steel and iron industry, i.e. long production process, dependency on economic and environmental policies, it is advisable that the model allows for the extension of the forecast horizon and that it can represent non-observable factors. The steel and iron industry sells its assortment to other industries, rather than to individual customers. Thus, demand fluctuations depend on the general economic conditions. There are many economic prediction use cases found in the literature for industries sharing similar characteristics. An attempt was made to examine if they are applicable to the steel and iron industry.

Many authors refer to the limitations of traditional forecasting methods. The methods are claimed to rely solely on historical data, neglecting the uncertainty factor. To address that, a latent variable model is considered to be applied in the study, as an alternative to the traditional approach.

The business context of the steel and iron industry in Poland needs also to be mentioned. The industry faces many challenges, while new opportunities arise. An appropriate strategy may result in driving a dynamic growth. There is a long-term perspective of domestic demand boost due to infrastructure development, investments in the energy sector, the European Union policies favoring renewable energy, including wind farms. Different growth rates are observed in the industries that purchase metallurgical products. While some of them are dynamically developing, others are in decline. The diversification of product portfolios and sales markets allows to overcome periods of slowdown. This is easier to be achieved by larger plants, having international customers. Entrepreneurs also emphasize the increase in materials and energy costs, challenges in debt recovery, the shortage of qualified employees and strong competition from Asia and Europe.

Enterprises are forced to minimize negative environmental impacts by optimizing their production process. To achieve that, while maintaining competitiveness, Polish steel mills collaborate with universities and research institutes. As it was shown, there are multiple challenges and opportunities ahead of the industry, which increases the importance of an accurate forecast. In addition, steel production may be subject to seasonal fluctuations, which need to be verified prior to building the model.

The above factors, as well as the lack of a similar known analysis for the steel and iron industry, defined the research area.

## Research questions

The aim of the study is to formulate the optimal forecasting method of demand and sales volumes for the heavy industry representative. The cognitive goal is to revisit and systematize quantitative forecasting models and to identify potential impact factors. A practical goal is also distinguished, being the recommendations to implement the model by companies as an alternative or a parallel forecasting method.

In the study it was examined whether the inclusion of economic indicators could improve the quality of a forecast, given the fact that the industry largely depends on economic cycles. An attempt was made to verify whether the models describing stochastic processes are applicable in long-term sales volume forecasting. Determining a suitable model requires eliminating seasonal fluctuations and identifying an adequate economic indicator that explains demand changes.

## Research methods

Answers to the research questions were sought on the basis of a literature review in the field of high-variance time series forecasting and model variables selection. As the purpose of the model is an important selection criterion, the following approaches were of a particular focus: methods allowing for the extension of the forecast horizon, methods allowing to include the economic condition as an independent variable, models suitable to analyze non-measurable factors. Several publications were compared, including demand forecasting methods in the heavy industry (e.g. Briffaut and Lallement 2010), industrial goods forecasting methods (e.g. Rippe 1976), business cycles and macroeconomic indicators forecasting (e.g. Bernardelli 2013), low demand goods forecasting. Time series decomposition methods were described to detect seasonality.

The selection of the impact variables was determined by their applicability in similar research areas and based on the information received from the enterprise about the factor's expected relevance. The decision was also taken based on the analysis of the steel and iron products market where a typical industry representative operates. In particular, the possibilities of including leading indicators were examined.

The selection of good quality and representative data is essential for a reliable model to be built. A company to represent the steel and iron sector was sought for. The market structure is composed of several large plants and several dozen smaller producers. During the examined period, there were three large companies operating on the market, with almost 70% share in the total production. The information was obtained from one of them. Taking into account its size and the market structure, it was assumed that the data were representative for the sector.

The data were checked for quality, especially for missing values and outliers. The critical product lines from the company's perspective were selected. The time series of other product groups showed different patterns and would require another modelling approach. This is related to the varying market conditions of industries being the target product recipients. The data preparation step also requires the identification of seasonal fluctuations.

Out of the two seasonal adjustment procedures, X-12-ARIMA and TRAMO / SEATS, only the latter allowed for a seasonal decomposition of a satisfactory quality. This series was used in further analysis.

In addition to the applicability of models confirmed in similar studies, model selection was driven by their ability to interpret economic indicators as independent variables (ARIMAX) or their ability to incorporate latent variables (hidden Markov model - HMM). The comparison of the models' forecasts seems informative as they rely on different assumptions.

It was noticed that the analyzed time series (dated 2004 to 2015) had a significant drop in 2009, which might indicate a transitory change. As a consequence, some variants of the models were determined for a shortened series in order to ensure stationarity. For the selected methods, the forecast horizon of 6, 12 or 18 months was adopted for the monthly time series. This size of the test set was determined by the number of observations after the potential transitory change.

The comparison of forecasting models was made on the basis of *ex post* forecast measures and information criteria, both sought to be minimized. The hypotheses of statistical tests were verified at a significance level of 0.05, and more than one test was used where it was applicable. For each model, a simulation was carried out to generate a large number of random time series and error values and information criteria compared between the models. The lack of a satisfactory result would be a rationale to revise the stages of model selection process or to refine its parameters. In the literature there are a few universal guidelines for selecting a sample size. Ultimately, it was decided to generate 3000 random series, having previously run a simulation for 1000 samples and concluding that increasing the sample size did not significantly change the result. This is one of the recommendations given by other authors. A satisfactory result of a simulation for the sufficient number of experiments justifies result's generalization.

The achieved results were compared to other authors' findings. Deriving the forecasts of a significantly lower quality or less accurate would be another rationale for revising the model, which was one of the key purposes of this procedure step.

The validation of the model at this stage sets the basis for presenting the outcome to the company and for requesting feedback. This feedback should take into account the usefulness of the forecast, ease of interpretation and acceptable error levels. Failure to meet these conditions should result in model rejection or returning to the model selection phase. The last step of the research procedure is to formulate the results and assess the degree of addressing research questions.

## Source selection

Literature review was conducted on the basis of the similarity criterion, whether related to the industry or the field of forecasting and comparable impact factors. A wide range of classic publications was reviewed, including those describing the theory of supply and demand (Begg et al. 2003), time series forecasting methods (e.g. Zeliaś et al. 2003, Cieślak et al. 2011), latent variable models (e.g. Baum and Eagon 1967, Schrodt 2000, Zwiernik 2005, Cappé et al. 2005), seasonality detection (e.g. Gomez and Maravall 1996, Grudkowska and Paśnicka 2007), establishing forecasting processes in a company (Waters 2002, Rutkowski et al. 2005), the application of leading indicators (Szeplewicz 2011). The analysis of the business context was performed based on industry magazines and publications (e.g. issued by the Polish Steel Association -Hutnicza Izba Przemysłowo-Handlowa and Polish Forging Association – Związek Kuźni Polskich), reports of the Ministry of Economy, press articles, Central Statistical Office data and OECD data.

Data analysis was performed with Gretl, R and JDemetra + software. JDemetra + was introduced by Eurostat in cooperation with the National Bank of Belgium and is commonly applied to the analysis and elimination of seasonality in time series. The software was used to estimate the X-12-ARIMA and TRAMO/SEATS models and to perform a comparative analysis between these two.

## Research outcomes

Demand forecasts were derived based on the ARIMAX model with an economic index and based on the hidden Markow model. The selected methods proved their usefulness to forecast various economic phenomena. Forecast quality was examined and the limitations of the methods were listed. Taking into account a number of model validation criteria (statistical tests, random time series simulation, error values, usability), selected variants of both models were proposed to be included it in the current sales and production planning.

The choice of the ARIMA method with an explanatory variable is justified by the need for measuring the impact of general economic conditions. A number of available economic indicators were examined with the goal to identify the optimal one. Based on the *ex post* forecast measures, the leading indicator (WWK) published by the Bureau of Investment and Economic Cycles and the OECD index for the European Union countries were selected. They provide information on the current status and potential shifts due to their leading properties, when combined in one model. In addition, the selection of the OECD indicator is consistent with the existing target markets and client segments. The model should be further monitored and examined in a longer time horizon. The improvement compared to ARIMA is minor, however, the use of indicators is justified by their leading properties and by the need to estimate a forecast relying not solely on historical values. Based on a series of simulations, the advantage of applying ARIMAX over a long forecasting horizon was demonstrated and random correlation between the variables discarded.

Hidden Markov model determined the potential future states, showing the trend of a demand level over the time. Its application is justified in a long-term planning, where setting directions is required rather than concrete values. The simulation of random time series confirmed the satisfactory result of the first-order two-state model, showing similarities to the actual changes in GDP in Poland in the analyzed period.

The optimal ARIMAX model and the two-state Markov model both determined moderate monthly demand level forecasts for the analyzed period. Due to the differences in model construction, ARIMAX estimated exact values and HMM rounded values. Therefore, the comparison of forecast error measures was omitted, and the comparative criterion is based on the company's own assessment. For HMM, a separate model evaluation system was established, which compares the forecast results with randomly generated series. The argument for the parallel use of the two methods is their independence, achieved by setting different assumptions.

## Theoretical and practical outcomes

The study contributes to the analysis of business environment of the chosen sector, describing existing threats and opportunities. The condition of related industries was also elaborated on. The study extends the existing research outcomes to applying demand and sales forecasting models and seasonality detection methods to the data of a steel and iron industry representative. The results obtained for the ARIMAX model show minor mean error improvements compared to ARIMA, and both forecast have high accuracy in a short time horizon. It was found that several other researchers identified a better indicator for their respective time series (e.g. unemployment rate for GDP measurement, economic sentiment indicator for unemployment measurement – for instance Bielak 2010, Ďurka and Pastoreková 2012). Such a comparison is more difficult in the case of Markov models, because the assessment largely depends on the adopted assessment approach, which may for instance penalize false negative or false positive values or weight the impact of the two. For one of the Markov models similarities were observed in relation to the actual changes in GDP in Poland, although the series is too short to confirm a long-term relationship. Similarly, other authors (including Hamilton 1990) were able to accurately determine the turning points of economic cycles in the United States, corresponding with the data published by the National Bureau of Economic Research.

The practical outcomes refer to determining a model for a representative of the heavy industry and designing recommendations for its application. Augmenting the company's current approach with a forecast estimated through ARIMAX could set the basis for strategic decisions, implemented through short-, medium- and long-term activities. This could further strengthen the position of a specific company on the domestic market, help adjust the offer to customer requirements, support the search for new sales channels and clients. In addition to the activities in the field of employment rationalization, wage system revision or technology modernization, the models could facilitate the adaptation to the new market conditions.

The proposed methods can be included in the existing forecasting process and contribute to the assessment of the market condition. It is recommended to involve experts in the selection of explanatory variables or in estimation of the quantitative forecast parameters.

## Limitations to the research procedure

The research procedure has limitations and requires certain assumptions. The study poses the question of extending the forecast horizon in the steel and iron industry. The conducted analysis, based on including selected economic indicators in the model and examining the usage of stochastic models, only answers the question for a sector representative. The series length, although sufficient to derive a forecast, does not justify drawing conclusions about long-term relationships. Limitations in obtaining data beyond 2015 result in the outcome being relevant to past circumstances. The study was conducted on the time series reflecting the sales levels of one of the largest representatives of the iron and steel industry in Poland. During the focus period, the company market share in total production volumes was above 20%. Additionally, all large companies had to undergo a similar, challenging transition phase when adapting to the market economy. On this basis an assumption was made that the sales volumes are representative for the entire sector. There are however some differences observed in the organizational structure and capital share among those companies.

In this study, the volume of demand is measured by the volume of sales, and therefore represents the realized demand. It should also be noted that only selected product groups were examined, due to the different target recipients of the entire assortment, resulting in different demand curves. For the ARIMAX model, there is no long-term cointegration between the series and the examined indices observed. The selection of a more fitted index could increase forecast quality. The weakness of the proposed Markov models is the generalization of demand values. In the two-state model, a relatively short time series was analyzed, which does not prove the model's long-term ability to identify turning points.

The presented limitations arise mainly from the unavailability of the latest data, especially for other industry representatives.

## Further research directions

The information about customers and geographic regions was not directly used in the study. On that basis the analysis could be further expanded to identify trends for specific regions and customer groups. Moreover, a model could be derived for other product lines. The research can also be conducted based on the demand and sales data for other industry representatives, to verify the impact of localization, capital share and organizational structure factors. Commonly accessible indicators were taken into account in the analysis. Thus, there might be one that better represents fluctuations in demand and sales levels for the steel and industry.

## Literature (selection)


1. Baum, L.E., Eagon, J.A. (1967) *An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model of ecology.* Bulletin of the American Mathematicians Society, nr 73, str. 360-363.
2. Begg, D., Fisher, S., Dornbusch, R. (2003) *Mikroekonomia*. Polskie Wydawnictwo Ekonomiczne, Warszawa
3. Bernardelli, M. (2013) *Nieklasyczne modele Markowa w analizie cykli koniunktury gospodarczej w Polsce.* Roczniki Kolegium Analiz Ekonomicznych, nr 30.
4. Bielak, J. (2010) *Prognozowanie ryku pracy woj. lubelskiego z wykorzystaniem modeli ARIMA i ARIMAX*. Barometr Regionalny, nr 1 (19).

5. Briffaut, J.P., Lallement, P. (2010) *Volatility Forecasting of Market Demand as Aids for Planning Manufacturing Activities*. Service Science & Management, nr 3, str. 383-389.
6. Cappé O., Moulines E., Rydén T. (2005) *Inference in hidden Markov models*. Springer Series in Statistics.
7. Cieślak, M. i in. (2011) *Prognozowanie gospodarcze. Metody i zastosowania.* PWN, Warszawa.
8. Ďurka, P., Pastoreková, S. (2012) *ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series?* Proceedings of 30$^{th}$ International Conference Mathematical Methods in Economics.
9. Gómez, V., Maravall, A. (1996) *Programs TRAMO (Time Series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instruction for the User*. Bank of Spain, nr 9628.
10. Grudkowska, S., Paśnicka, E. (2007) *X-12-ARIMA I TRAMO/SEATS – empiryczne porównanie metod wyrównania sezonowego w kontekście długości próby*. Narodowy Bank Polski, Materiały i studia, nr 220.
11. Hamilton, J.D. (1990) *Analysis of time series subject to changes in regime.* Journal of econometrics, nr 45, str. 39-70.
12. Rutkowski, K. i inni. (2005) *Logistyka dystrybucji*. Szkoła Główna Handlowa, Warszawa.
13. Schrodt, P.A. (2000) *Forecasting conflict in the Balkans using hidden Markov models.* Paper presented at the American Political Science Association meetings, Washington.
14. Szeplewicz, K. (2011) *Wskaźniki wyprzedzające koniunktury - analiza ekonometryczna.* Prace i materiały Instytutu Rozwoju Gospodarczego, IRG SGH, Warszawa, nr 87.
15. Szulc, W. (2014) *Transformacja polskiego hutnictwa żelaza do gospodarki wolnorynkowej (z uzupełnieniami)*. Prace Instytutu Metalurgii Żelaza. Monografie, nr 6.
16. Waters, D. (2002) *Operations Management.* Pearson Education Limited, Essex
17. Zeliaś, A., Pawełek, B., Wanat, S. (2003) *Prognozowanie ekonomiczne: teoria, przykłady, zadania.* Wydawnictwo Naukowe PWN, Warszawa.
18. Zwiernik, P.W. (2005) *Wstęp do ukrytych modeli Markowa i metody Bauma-Welcha.* http://www.mimuw.edu.pl/~pzwiernik/docs/hmm.pdf.

Bovska