

MARKO INTIHAR, M.Sc.<sup>1</sup>

E-mail: marko.intihar@um.si

TOMAŽ KRAMBERGER, Ph.D.<sup>1</sup>

E-mail: tomat.kramberger@um.si

DEJAN DRAGAN, Ph.D.<sup>1</sup>

(Corresponding author)

E-mail: dejan.dragan@um.si

<sup>1</sup> Faculty of Logistics, University of Maribor

Mariborska cesta 7, 3000 Celje, Slovenia

Transport Logistics

Preliminary Communication

Submitted: 7 Dec. 2016

Accepted: 3 Mar. 2017

# CONTAINER THROUGHPUT FORECASTING USING DYNAMIC FACTOR ANALYSIS AND ARIMAX MODEL

## ABSTRACT

*The paper examines the impact of integration of macroeconomic indicators on the accuracy of container throughput time series forecasting model. For this purpose, a Dynamic factor analysis and AutoRegressive Integrated Moving-Average model with exogenous inputs (ARIMAX) are used. Both methodologies are integrated into a novel four-stage heuristic procedure. Firstly, dynamic factors are extracted from external macroeconomic indicators influencing the observed throughput. Secondly, the family of ARIMAX models of different orders is generated based on the derived factors. In the third stage, the diagnostic and goodness-of-fit testing is applied, which includes statistical criteria such as fit performance, information criteria, and parsimony. Finally, the best model is heuristically selected and tested on the real data of the Port of Koper. The results show that by applying macroeconomic indicators into the forecasting model, more accurate future throughput forecasts can be achieved. The model is also used to produce future forecasts for the next four years indicating a more oscillatory behaviour in (2018-2020). Hence, care must be taken concerning any bigger investment decisions initiated from the management side. It is believed that the proposed model might be a useful reinforcement of the existing forecasting module in the observed port.*

## KEY WORDS

*container throughput forecasting; ARIMAX model; dynamic factor analysis; exogenous macroeconomic indicators; time series analysis;*

## 1. INTRODUCTION

Nowadays, it is crucial to assure a quality decision support system for generating cargo throughput forecasts in seaports. Namely, high accuracy of the forecasts can essentially influence the appropriate strategy for port development, infrastructure-based investments, and efficient daily management [1]. On the other hand, inaccurate forecasts can lead to negative effects and financial losses that result in inadequate infrastructure investments and poorly chosen port

strategy in the future, as well as improper decisions about the port upgrade or port redesign [2, 3].

After the eruption of the most recent economic crisis, the nature of ports throughput dynamics has significantly changed on the global scale. The throughput time series became much more volatile and complex, and, consequently, the time series modelling procedures evolved into a much more difficult task. Specifically, the simpler forecasting models, which had a leading role in the port economics for many years, are no longer able to capture increased volatility and complexity of the time series. On the contrary, the level of modelling details needs to be more sophisticated, which also requires more advanced modelling mechanisms [4].

As it is known from the scholarly literature, there often exists a certain correlation between the exogenous macroeconomic indicators and the port throughput volume [5, 6]. This interrelation is not surprising since economic indicators, such as GDP (gross domestic product) or import/export, can importantly influence the trade flows and consequently the port throughput volume. The changes in indicators are also visible some time later in the change of throughput dynamics. So, the accessible information carried by these indicators can be, at least in principle, used for making better forecasts of future throughput dynamics shifts.

There are several studies detected, which have more or less successfully engaged external variables into the forecasting modelling design [6]. However, most of these studies have conducted only simpler forecasting models, which would most likely have serious problems to deal with the increased complexity and volatility of the time series in the post-crisis period. Moreover, these studies usually address only a relatively small number of applied external indicators that might influence the throughput. Thus, it could happen that some important indicators, injected as exogenous inputs into a forecasting model, would have been missing.

The primary aim of this research is the construction of an efficient forecasting model, which can accurately forecast the container throughput time series in the Port of Koper, Slovenia. The model also involves a relatively large number of external macroeconomic indicators, which might have a specific impact on the observed throughput. External effects comprise five different exogenous indicators (*GDP, export, import, purchasing power parity (PPP), unemployment rate*) of four different geographical groups (*Slovenia with Hinterland countries, The World, European Union, major Far East Asian exporters*). This gives us a total of 66 external exogenous variables.

The modelling framework combines the Dynamic factor analysis (DFA) [7-10] and the Box-Jenkins ARIMAX (AutoRegressive Integrated Moving-Average model with exogenous inputs) modelling process [11]. They are integrated into the unique four-stage modelling framework, which combines different modelling stages. The DFA is used to extract a lower number of dynamic factors from a much higher number of original external exogenous indicators (stage 1). The extracted factors are injected into the "generator" of ARIMAX models, which creates a whole family of model candidates of different orders regarding the input delay, output delay, and noise delay [12] (stage 2). In stage 3, the diagnostic checking and goodness-of-fit (GOF) testing are applied to each model candidate. In the final stage, stage 4, the best model is automatically selected from the family of all the candidates using a special heuristics.

The forecasting performance of the best selected model is investigated on the test interval for the case of real throughput data. The derived model can efficiently forecast the future dynamics of the throughput time series. Its performance is also compared with the forecasts of a benchmark Holt-Winters model. The prediction results show that our model has outperformed the behaviour of the benchmark model. Since it provides fairly accurate predictions, it might be used to additionally support the existing forecasting module of the Port decision support system (DSS).

The best selected model is also applied to generate forecasts for the forthcoming out-of-sample prediction interval (2017-2020). The forecasts from this interval imply that after the stable rising of container throughput in 2017, more turbulent behaviour might follow in the consecutive years 2018-2020. Accordingly, caution must be taken regarding any bigger actions triggered from the managerial side.

To the best of our knowledge, nearly no similar studies have been applied, which would combine the DFA and ARIMAX methodologies into a unique decision support mechanism. It seems that this gap is not evident in the field of throughput forecasting only, but also in the forecasting research in general.

Moreover, we believe that a specific nature of a specially developed heuristics for the automatic model selection might be another contribution of this paper.

## 2. LITERATURE REVIEW

Since the beginning of the last millenium, many authors have applied various statistical models with the primary aim to get an accurate estimate of the future port throughputs. One of such authors is [13], who proposed an error-correction model for forecasting the demand for Hong Kong's container handling services. The same port was the subject of research carried out by [14], where a multiple linear regression model was used to identify the influence of different economic factors on the forecasted cargo throughput. In a similar way, a linear regression model was conducted in a research of [15], where the estimation of future volumes of Taiwan's import containers was the primary goal. On the other hand, [16] have examined the container transshipment in Germany by applying the Box-Jenkins methodology and exponential smoothing. They have used a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model and Holt-Winters model.

Some papers also report comparative studies about forecasting the models' performance. A typical example of such research are studies conducted by [2] and [17]. For instance, in the work of [2], six univariate forecasting models have been developed at first, and their behaviour was compared later. Within this scope, the following models were applied: the regression model with seasonal dummy variables, classical decomposition model, trigonometric regression model, SARIMA model, the grey model, and finally, the hybrid grey model.

Several authors have conducted models based on statistical learning theory and artificial intelligence. Representative examples of such models are popular Artificial Neural Networks (ANN) models, which have been applied in many scientific fields. The predictive power of this kind of models is demonstrated in the paper [18], where the back propagation neural network was applied to forecast the Guangdong port future throughput. The methodology conducted here has also included the genetic algorithms (GA), as well as the different macroeconomic factors that have had an influence on the port throughput.

In many studies, the authors developed different kinds of hybrid models that generally outperform the individual models. A typical example of successfulness of hybrid models is a study of [4] applied to two Chinese ports, where ARIMA and SARIMA models were combined with the least squares support vector regression. On the other hand, the study conducted by [19] has combined a regression model and GA to forecast the container throughput of Qingdao port, where the benchmark models were significantly outperformed.

The inspection of literature shows that some authors have also used the exogenous macroeconomic indicators in the role of forecasting model inputs. A typical example of such studies are works of [14, 15, 18, 20, 21]. The authors in study [15] have shown the importance of the relationships between the container volumes and different macroeconomic variables. They have applied volumes of export and import containers, industrial production index, population, GDP, agricultural, industrial and service GDPs, GNP (gross national product), GNP per capita, and wholesale for the key factors in the analysis. Gosasang et al. [21] have conducted Thailand's major indicators that affect the number of containers, importing and exporting, such as the GDP, world GDP, inflation, interest, and exchange rates, respectively, population, and the fuel price. In study [14], the indicators such as trade value of import and total exports, GDP, expenditure on buildings, electricity demand, population, and other indicators have been employed. On the other hand, Ping and colleagues [18] conducted the regional indicators of the GDP, overall output value of tertiary industry, and the net export to build a forecasting model, while Wang et al. [20] employed a system dynamics model and regression model involving GDP and international trade-related factors.

In the forecasting field, extendable efforts have been dedicated to the development of different kinds of heuristic-based automatic model selection procedures. For instance, Gomez and Maravall proposed an algorithmic automatic procedure, which has been later also implemented in programs TRAMO and SEATS [22]. More recently, [23] introduced a state space framework for automatic model selection and forecasting based on exponential smoothing methods. On the other hand, authors [24] outlined the so-called RETINA (Relevant Transformation of the Inputs Network Approach) automatic model selection procedure, which wishes to identify a parsimonious set of variables that

are most likely to be significant for the prediction of out-of-sample [25].

Besides the mentioned, there exist many other heuristic approaches for the automatic model selection such as: Genetic algorithms, Evolutionary algorithms, Branch and Bound strategies, etc., just to mention some of them. Conversely to approaches just summarized, our heuristic approach for model selection is different and is based on a plethora of requirements related to information criteria, parsimony, invertibility, stationarity and stability issues.

### 3. PORT THROUGHPUT AND ECONOMY DRIVEN DATA

#### 3.1 Characteristics of the Port of Koper

The Port of Koper, a member of the North-Adriatic Port Association (NAPA), is a modern well-equipped multiuse hub [26, 27]. It is located in the Gulf of Trieste and has 11 terminals in the port area. Besides a container terminal, it also contains the following specialized terminals [26, 28]: car terminal, ro-ro terminal, general cargo terminal (metals & non-metals, coffee, fruits and light-perishable goods, cacao, iron, paper, etc.), timber terminal, livestock terminal, fruit terminal, terminal for minerals, terminal for cereals and fodder, the European energy terminal, alumina terminal, and liquid cargoes terminal.

For the Port of Koper, the Eastern and Central European markets are of a crucial matter, as it performs most of their services for the hinterland countries (i.e. Croatia, Austria, Poland, Italy, Hungary, Southern Germany, Slovakia, the Czech Republic, and others). Since the transit traffic has the major role, the Port of Koper can generally be treated as a transit port [26, 28]. On average, only 29% of the cargo is considered for the Slovenian market, while the remaining 71% is linked to the hinterland countries (see *Figure 1*) [26, 29].

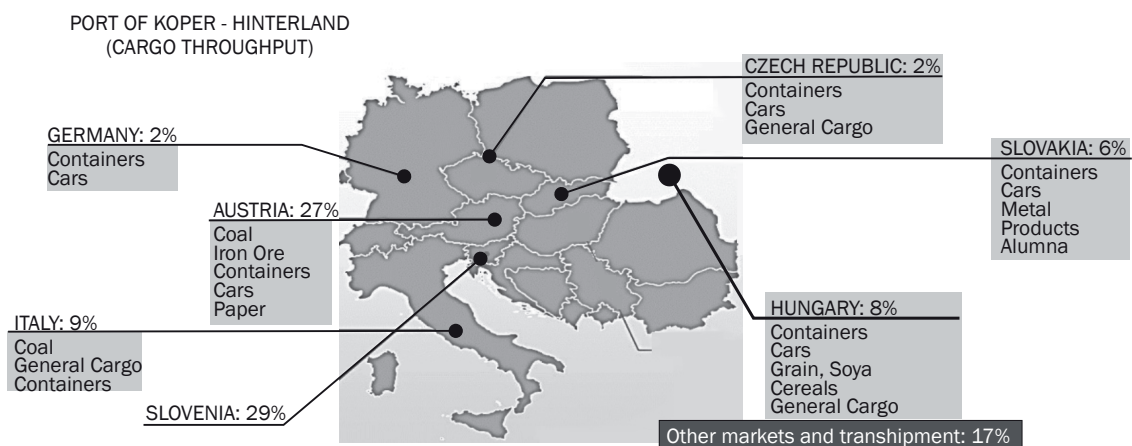


Figure 1 - The distribution of throughput among hinterland countries [29]

Since the NAPA ports (the others are Trieste, Rijeka, and Venice) have a very appropriate geographical position penetrating deeply into the South-East corner of Europe, the waterway from the Far East via Suez to Europe is approximately 2,000 nautical miles shorter than in the case of North-European ports [26, 28]. Thus, these ports represent an interesting gateway to the key European markets.

Regarding the container cargo flows, the ones coming from the Far East (particularly from the ASEAN countries, China, Japan, and S. Korea) are of essential importance for the Port of Koper [30]. The container terminal is linked with the Far East weekly by using two regular direct lines and via feeder services with four important Mediterranean HUB ports [30]. Besides the direct service established by MAERSK and CMA CGM, another direct line was operated in 2010 by four shipping companies (i.e. Hyundai Merchant Marine, Hanjin Shipping, Yang Ming Marine Transport Corporation, and United Arab Shipping Company). This direct line is crucial because of its full implementation in the post-crisis period [30].

### 3.2 Container throughput time series

Our study focuses on modelling and predicting the dynamics of a 20-foot container throughput time series  $y(t), t=1, \dots, 60$  (measured in Kilotonnes (kt)) (see Figure 2). The historical data were observed in the time domain (first quarter of 2002 - last quarter of 2016.) The data are quarterly, so they represent a sequence of 60 measurements, collected by means of [31].

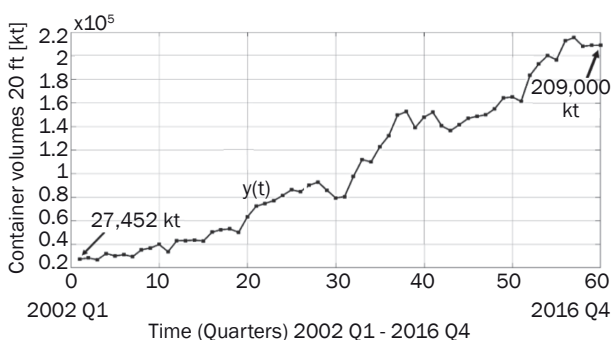


Figure 2 – Quarterly time series of container throughput at the Port of Koper

The observed time series has mostly a steady rising dynamics, with several smaller oscillations and only one significant disturbance, which occurred during the beginning years of the economic crisis (c.f. turbulences in years 2008-2009 (time samples 28-36) in Figure 2). In the crisis eruption period, according to [26], the container terminal handled 343,165 TEUs during the year 2009, which is 3% less in comparison with the year 2008 (353,880 TEUs). Nevertheless, the container data in the post-crisis years confirms that the Port of

Koper is becoming a more and more important regional player regarding the container throughput having a steady long-term rising of dynamics which is expected to persist in the future years [32]. Moreover, the fact is that the container throughput has increased by almost nine times (from 27,452 to 209,000 Kilotonnes) in the last fourteen years, which gives it the leading position among NAPA ports regarding the absolute growth [32].

### 3.3 Influential external macroeconomic indicators

Macroeconomic indicators refer to a statistics about economic activities and enable an analysis of economic performance, as well as predictions of future performance [33]. They are often used in the studies of business cycles and include various economic summaries, indices, and earning reports. Typical indicators are GDP, GNP, unemployment rate, consumer price index, inflation rate, industrial production, consumer leverage ratio, etc. [33].

Many studies (e.g. [18, 20, 21, 34]) have shown that modern ports are flexibly influenced by various economic indicators that exist on the domestic and regional, as well as global scale. Thus, such (endogenous and/or exogenous) variables should be included, particularly for the ports whose container throughput is primarily initiated from the export/import and international trade [20]. These indicators typically comprise the domestic, regional, and/or global GDP, import/export and exchange rate (or PPP), as well as changes in other countries' policy [18, 20, 21, 35, 34]. Since in several econometric studies it has been shown that the exchange rate (ER) volatility negatively influences the trade, sometimes it is better to use PPP to avoid ER misalignment in terms of deviations from PPP [35]. The newest studies also show that there is a direct link between international trade and unemployment rates, while the aggregate unemployment persists even in the long-run equilibrium [36]. Thus, it is recommended to incorporate the unemployment affecting trade flows into the macroeconomic-based models [36].

Following the findings of aforementioned research such as [14, 15, 18, 20, 21], and [35, 36], we have decided to choose five different macroeconomic indicators (see Table 1 [33]). We have used them for those countries or regions on different geographical scales, which are believed to have a strong influence on the observed container throughput dynamics, such as (see Figure 3):

- 1) *Local scale* (port country Slovenia and hinterland countries: Italy, Croatia, Austria, Czech Republic, Germany, Hungary, Slovakia),
- 2) *Regional scale* (EU) and *Global scale* (Entire World),
- 3) *Major Asian Far East exporters* (China, Japan, South Korea, Singapore).

Table 1 – The selected macroeconomic indicators [33]

Macroeconomic indicator	Meaning
GDP per capita (%)	A measure of the total output of a country that takes GDP and divides it by the number of the citizens.
Purchasing power parity (PPP)	An economic theory that compares different countries' currencies via a market "basket of goods" approach.
Import (billions \$)	A good or service brought into one country from another country.
Export (billions \$)	A function of international trade where goods produced in one country are transported to another country for future trade or sale.
Unemployment (%)	A measure of the prevalence of unemployment calculated by dividing the number of unemployed persons by all persons in the labour force.

The reason for choosing the hinterland countries lies in the fact that the major part of the cargo is transferred to them via the Port of Koper (see Figure 1). On the other side, the major Far East exporters have been selected due to their large share of total container cargo amount handled in the Port of Koper. Moreover, we have also considered the fact that according to official statistics [37-40] these Far East countries represent one of the biggest trade partners with the EU [41].

The just mentioned classification gives us 14 different countries/regions with five different macroeconomic indicators. Figure 3 shows that we are dealing with 66 different exogenous variables  $x_i(t), i=1, \dots, 66$ . Since this number of variables is expansively big, we should have applied a data reduction technique to decrease an initial set of variables to a reduced set containing only a few dynamic factors. These factors will retain a mutual cross-sectional dynamics of initial exogenous variables. Since they still carry enough joint macroeconomic information influencing the observed throughput  $y(t)$ , they can be used as inputs to our ARIMAX forecasting model.

The annual time series data for macroeconomic variables  $x_i(t), i=1, \dots, 66$  were collected by means of Eurostat, International Monetary Fund, and the World Bank [42-44]. They were obtained for the same period as it was done for the throughput data (years 2002 to 2016). The measurements were divided into quarters, so that equally long sequences of 60 observations were available for further analyses.

#### 4. METHODOLOGICAL FRAMEWORK

##### 4.1 Conceptual framework

The conceptual framework of our four-stage modelling process, which gives, as a result, the best ARIMAX model, is presented in Figure 4. In the first stage, the DFA is used to extract a lower number of dynamic factors  $u_i(t)$  from a much higher number of standardized original external indicators  $z_k(t)$ . This way, we obtain only a few dynamic factors, while still keeping the crucial information about joint co-movement of original exogenous time series. To save as much

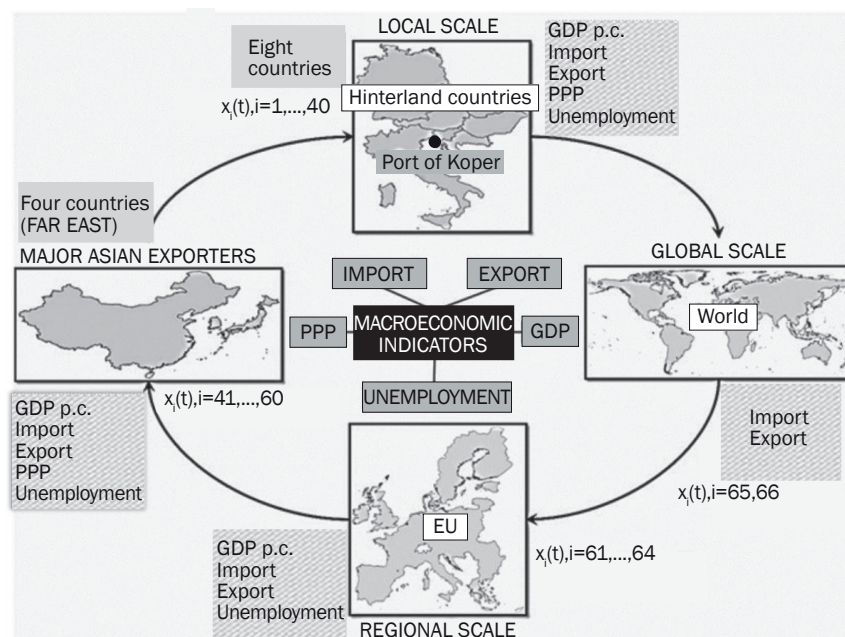


Figure 3 – Framework of selected economic indicators and corresponding geographical scales

information content of original variables as possible, different sizes of factors' space have been checked, which has given us the best combination of dynamic factors.

In the second stage, the extracted factors are injected into the "generator" of ARIMAX models. The latter creates a whole family of model candidates of different orders with respect to the input delay, output delay, and noise delay [12]. At this stage, the model structures and parameter estimates are determined for each model candidate.

In the third stage, the diagnostic checking and goodness-of-fit (GOF) testing are applied to each model candidate. By doing this, each of the model candidates is tested by verifying it via various statistical criteria, such as: the lowest possible residual-based measures (RMSE – Root mean squared error, MAE – Mean absolute error), the Akaike's corrected information criteria AICc, the best possible percentage of model fit to the data, and the fulfilled conditions of parsimony, invertibility, stationarity and stability. Furthermore, additional measures are adopted to examine whether the model residual is close to the white noise without any serious serial correlation or existence of heteroscedasticity.

In the final, fourth stage of the modelling framework, the best model is selected from the family of all

candidates. For this purpose, a special heuristic is applied to determine which model candidate satisfies all the pre-defined criteria most comprehensively.

### 4.2 Dimension reduction based on Dynamic Factor Analysis

The DFA attempts to extract dynamic factors from a given multivariate time series dataset [45]. These factors resemble the addressed time series' common trends and try to capture their interactive effects. When conducting the DFA, the number of dynamic factors  $m$  is selected in advance. In our case, the DFA is applied to reduce the total number of  $N=66$  exogenous macroeconomic variables to a set containing only a few dynamic factors. Since the exogenous variables are measured in different units, it is convenient to standardize them to reach better results when applying the DFA. By doing this, the standardized exogenous time series occur in the following form:

$$z_k(t) = \frac{x_k(t) - \bar{x}_k}{\sigma_k} \quad k = 1, \dots, N \tag{1}$$

where  $\bar{x}_k$  is the time series mean and  $\sigma_k$  is its standard deviation. The relations between the  $z_k(t)$  and  $m$  different dynamic factors  $u_i(t), i=1, \dots, M$  can be expressed as [45, 46]:

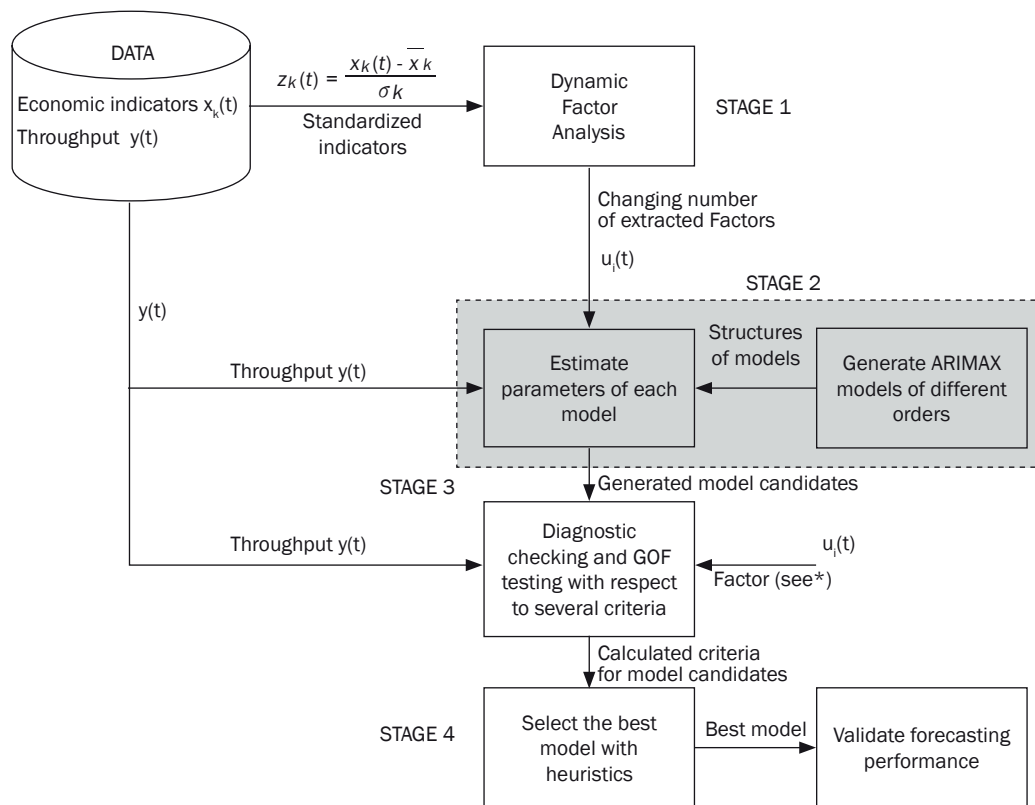


Figure 4 – The conceptual four-stage framework for selecting the best ARIMAX model

$$\begin{aligned}
 z_1(t) &= \lambda_{11} \cdot u_1(t) + \lambda_{12} \cdot u_2(t) + \dots \\
 &\dots + \lambda_{1m} \cdot u_m(t) + \varepsilon_1(t) \\
 &\vdots \\
 z_N(t) &= \lambda_{N1} \cdot u_1(t) + \lambda_{N2} \cdot u_2(t) + \dots \\
 &\dots + \lambda_{Nm} \cdot u_m(t) + \varepsilon_N(t)
 \end{aligned}
 \tag{2}$$

where  $\lambda_{ki}$  is the  $i$ -th factor loading (coefficient) on the  $k$ -th standardized exogenous variable, while  $\varepsilon_k(t) \in N(0, \sigma_{\varepsilon_k})$  is assumed to be the  $k$ -th normally distributed white noise random term. Here, factor loadings describe the strength of interaction between the individual time series and its corresponding dynamic factors. Each of these factors can be modelled as a random walk process, which possesses a form of a slow-moving pattern with sudden increase or decrease that occasionally arises [47]. Following this, we can write:

$$\begin{aligned}
 u_1(t) &= u_1(t-1) + \eta_1(t) \\
 u_2(t) &= u_2(t-1) + \eta_2(t) \\
 &\vdots \\
 u_m(t) &= u_m(t-1) + \eta_m(t)
 \end{aligned}
 \tag{3}$$

where  $\eta_i(t)$  are assumed to be normally distributed white noises terms:  $\eta_i(t) \in N(0, \sigma_{\eta_i})$ . By combining the systems of Equations 2 and 3, we can develop a general state space framework with measured variables  $z_k(t)$  and unmeasured factors (states)  $u_i(t)$ . This way, we can write the following dynamic factor model in vector-matrix form [48]:

$$\begin{aligned}
 \mathbf{z}(t)_{N \cdot 1} &= \mathbf{A}_{N \cdot m} \cdot \mathbf{u}(t)_{m \cdot 1} + \boldsymbol{\varepsilon}(t)_{N \cdot 1} \\
 \mathbf{u}(t)_{m \cdot 1} &= \mathbf{u}(t-1)_{m \cdot 1} + \boldsymbol{\eta}(t)_{m \cdot 1} \\
 \boldsymbol{\varepsilon}(t)_{N \cdot 1} &\in N(\mathbf{0}, \mathbf{R}), \dots, \boldsymbol{\eta}(t)_{m \cdot 1} \in N(\mathbf{0}, \mathbf{Q})
 \end{aligned}
 \tag{4}$$

where  $\mathbf{A} \in (\lambda_{ki})$ ,  $\mathbf{R}$ , is the covariance matrix of measurement noises  $\varepsilon_k(t)$ , while  $\mathbf{Q}$  is the covariance matrix of state noises  $\eta_i(t)$ . One of the general assumptions of DFM modelling process is the one about the mutual independence of vectors  $\boldsymbol{\varepsilon}(t)$ ,  $\boldsymbol{\eta}(t)$  and  $\mathbf{u}_0 = \mathbf{u}(0)$ , which corresponds to the vector of initial states [48].

In the process of estimation of unknown dynamic factors located in  $\mathbf{u}(t)$ , the unknown DFM hyper-parameters hidden in matrices  $\mathbf{A}$ ,  $\mathbf{R}$ , and  $\mathbf{Q}$  should be simultaneously estimated as well. The most suitable way for applying such a kind of estimation is to conduct the maximum-likelihood (ML) estimation method [48]. This method can be proceeded by using the two-step recursive expectation-maximization (EM) algorithm [49].

In general, the EM algorithm alternates between two steps, an expectation (E) step, and maximization (M) step. The E-step generates an expectation of the log-likelihood function which is based on the current parameters estimated by the Kalman filter or smoother. Afterward, the M-step re-computes the parameters by maximizing the expected log-likelihood from the

E-step. These updated parameters are then used to estimate the distribution of unknown states in the next E step. The EM algorithm continues alternating between performing both steps until the convergence to the final solution is achieved.

### 4.3 Presentation of the ARIMAX model

The ARIMAX model represents a composition of the output time series into the following parts: the autoregressive (AR) part, moving average (MA) part, integrated (I) component, and the part that belongs to the exogenous inputs (X). The AR and MA fragments refer to the dependence of the current stationary value of the time series  $y(t)$  from its historical values  $y(t-i)$  and historical values of random errors  $\varepsilon_y(t-k)$ , respectively [50]. On the other side, the integrated component (I) is required in the case when the stationarity of the time series must be achieved. For that purpose, the differencing procedure should be applied to transform the initial non-stationary time series into its stationary counterpart. Finally, the exogenous part (X) reflects the additional incorporation of the present values  $u_i(t)$  and past values  $u_i(t-j)$  of exogenous inputs (dynamic factors in our case) into the ARIMAX model [51].

When dealing with a single input - single output (SISO) ARMAX model (ARIMAX without part I), the following difference equation [12] can be applied:

$$\begin{aligned}
 y(t) + \sum_{i=1}^{n_a} a_i \cdot y(t-i) &= \sum_{j=0}^{n_b} b_j \cdot u(t-j) + \\
 + \varepsilon_y(t) + \sum_{k=1}^{n_c} c_k \cdot \varepsilon_y(t-k)
 \end{aligned}
 \tag{5}$$

where the orders  $n_a$ ,  $n_b$  and  $n_c$  refer to the oldest output delay, input delay, and random noise delay, respectively. Here, the noise term is supposed to have the properties of the white noise. The difference Equation 5 can also be presented in the simpler form of a transfer function-based equation, if the backshift operator  $q$  is applied [52]:

$$\begin{aligned}
 A(q) \cdot y(t) &= B(q) \cdot u(t) + C(q) \cdot \varepsilon_y(t) \\
 \text{where} \\
 A(q) &= 1 + a_1 \cdot q^{-1} + \dots + a_{n_a} \cdot q^{-n_a} \\
 B(q) &= b_0 + b_1 \cdot q^{-1} + \dots + b_{n_b} \cdot q^{-n_b} \\
 C(q) &= 1 + c_1 \cdot q^{-1} + \dots + c_{n_c} \cdot q^{-n_c}
 \end{aligned}
 \tag{6}$$

In the case of multiple inputs, the transfer Function 6 takes the structure:

$$A(q) \cdot y(t) = \sum_{i=1}^m B_i(q) \cdot u_i(t) + C(q) \cdot \varepsilon_y(t)
 \tag{7}$$

where

$$B_i(q) = b_{0i} + b_{1i} \cdot q^{-1} + \dots + b_{n_{bi}} \cdot q^{-n_{bi}}, \quad i = 1, 2, \dots, m$$

The ARIMAX transfer function model has a fairly similar structure as the model in Expression 7, the only

difference is that the output  $y(t)$  must be replaced with its first differentiated equivalent  $\Delta y(t)=y(t)-y(t-1)$ . The calculations in our study have shown that only the first order of differencing was needed to obtain the stationary throughput time series, so the multiple differentiation was not conducted.

#### 4.4 Block scheme of the entire modelling mechanism

On the basis of conceptual framework presented in Figure 4, the entire modelling mechanism is constructed as shown in Figure 5. As can be seen from Figure 5, up to five ( $m=1, \dots, m_{max}=5$ ) different dynamic factors are generated while processing the DFA analysis in the first stage. The VARIMAX rotation procedure is also conducted to improve the results for the obtained values of DFM.

In the next stage, the entire family of different ARIMAX model candidates is created for each possible number (from 1 to 5) of formerly estimated dynamic factors. Furthermore, for each algorithm iteration, at

a given number of the dynamic factors, a different setup of input, output and noise delay orders is also applied. These orders are gradually increasing from value one up to the pre-defined maximum order value ( $n_{a_{max}}=n_{b_{i_{max}}}=n_{c_{max}}=4; i=1,2,\dots,m$ ). By doing this, one model candidate is generated for each single iteration of the algorithm. When all iterations of the algorithm are finished, a family of several hundreds of different ARIMAX models is created. Since we did not only generate the structures of model candidates, but also simultaneously estimated the model parameters as well, this stage requires a significant computational effort.

In the third stage (see Figure 5), the diagnostic checking and GOF testing are applied to each model candidate. Here, several statistical, information-based, and residual-based criteria are computed to investigate the statistical properties, goodness-of-fit, and predictive performance of each model candidate. The following criteria and measures have been considered for every single ARIMAX model [50]:

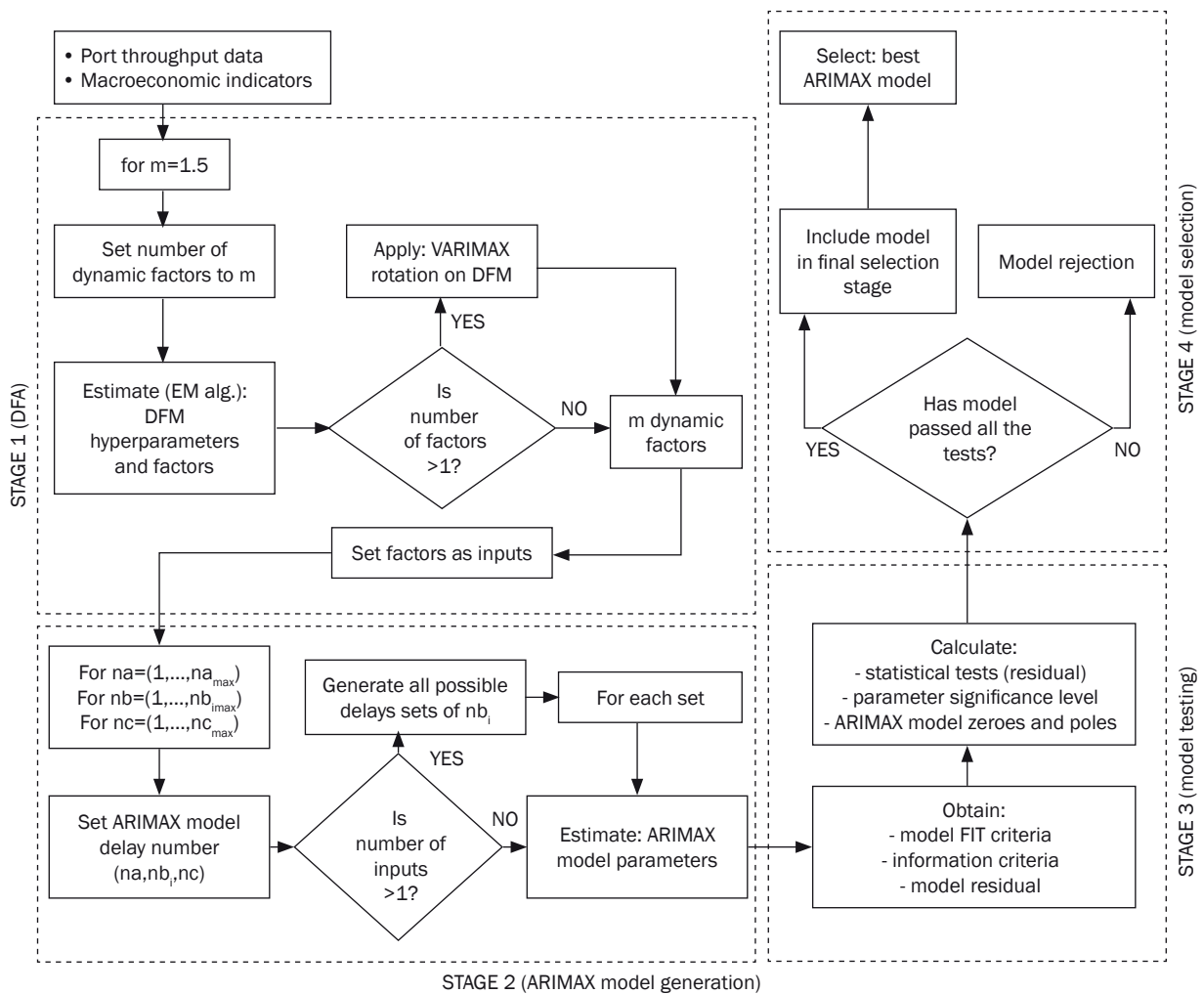


Figure 5 – Block scheme of the entire modelling mechanism



- The root mean square error (RMSE) and the mean absolute estimation (MAE);
- The percentage of model fit (FIT); and
- The information-based Akaike's corrected AICc criterion.

Besides these measures, the Jarque-Bera (JB) test, as well as the Portmanteu (Ljung-Box) LB test [50] have also been applied to investigate whether the model residual holds the properties of the white noise without any serious serial correlation. Moreover, the algorithm at this point calculates the significance level of the model's parameters and examines the model's transfer function zeroes and poles.

The fourth stage in Figure 5 represents the central mechanism of the model selection process. It applies a special heuristics, which tests the quality of all models and gradually leads to the best ARIMAX model. The heuristic reduces the space of initial model candidates at first, and then isolates the best model candidate afterwards. For that purpose, the whole sequence of different tests is conducted for each model candidate with respect to certain logic. To do so, the heuristic uses the calculated criteria from the previous, third stage. While narrowing the range of acceptable models, the heuristics considers the following issues:

- Low values of MAE, RMSE, and AICc measures;
- High value of the FIT measure (high % of FIT);
- Model residual should roughly behave as white noise;
- Fulfilled condition of parsimony (statistically significant parameters);
- Zeroes and poles of the model transfer function must lie within the unit circle (to ensure invertibility, stationarity and stability of the model [11, 12]).

The more detailed working mechanism of an applied heuristic from stage 5 is shown in Figure 6. Firstly, the initial set of model candidates is reduced by excluding those candidates, which do not pass all of the following conditions simultaneously: (1) Fulfilled JB and LB test; (2) All zeroes and poles must lie within the unit circle; (3) More than 90% of estimated parameters are statistically significant; and (4) The candidate has the AICc measure lower than a certain pre-defined threshold level. Since these restrictions are rigorous,

only a group of few model candidates can proceed to the next step of the final model selection process. From this group, the "best" model has yet to be selected. To do so, the candidate with the highest FIT measure is chosen, which represents the final (selected) ARIMAX forecasting model (let us call it the "best" ARIMAX model).

### 5. NUMERICAL RESULTS AND DISCUSSION

The four-stage modelling process was partially implemented in MATLAB technical computing environment, and partially in R statistical environment. The DFA analysis was carried out in R by using the package MARSS [53], while the ARIMAX modelling development was conducted by means of MATLAB System Identification Toolbox [52]. On the other hand, the statistical testing was applied by using the two MATLAB toolboxes, the Statistics and Machine Learning toolbox, and the Econometrics Toolbox.

When the four-stage modelling process was completed, the final ARIMAX model appeared in the following transfer function form (the "best" ARIMAX selected model):

$$\begin{aligned}
 A(q) \cdot \Delta y(t) &= A(q) \cdot [y(t) - y(t-1)] = \\
 &= \sum_{i=1}^5 B_i(q) \cdot u_i(t) + C(q) \cdot \varepsilon_y(t) \\
 A(q) &= 1 - \underbrace{2.376}_{(-16.8)} \cdot q^{-1} + \underbrace{1.815}_{(6.4)} \cdot q^{-2} - \underbrace{0.4658}_{(-3.05)} \cdot q^{-3} \\
 B_1(q) &= \underbrace{-8,838}_{(-3.3)} \cdot q^{-2} + \underbrace{4,064}_{(2.2)} \cdot q^{-3} \\
 B_2(q) &= \underbrace{8,904}_{(2.3)} \cdot q^{-2} - \underbrace{3,376}_{(-2.02)} \cdot q^{-3} \\
 B_3(q) &= \underbrace{-987}_{(-4.8)} \cdot q^{-2} - \underbrace{7,763}_{(-5.7)} \cdot q^{-3} + \underbrace{1,827}_{(2.5)} \cdot q^{-4} \\
 B_4(q) &= \underbrace{10,740}_{(3.07)} \cdot q^{-3} + \underbrace{929}_{(2.44)} \cdot q^{-5} \\
 B_5(q) &= \underbrace{-2,207}_{(-2.6)} \cdot q^{-3} + \underbrace{1,687}_{(1.97)} \cdot q^{-4} \\
 C(q) &= \underbrace{1-2.835}_{(-54.5)} \cdot q^{-1} + \underbrace{2.708}_{(27.8)} \cdot q^{-2} - \underbrace{0.8717}_{(-18.9)} \cdot q^{-3}
 \end{aligned} \tag{8}$$

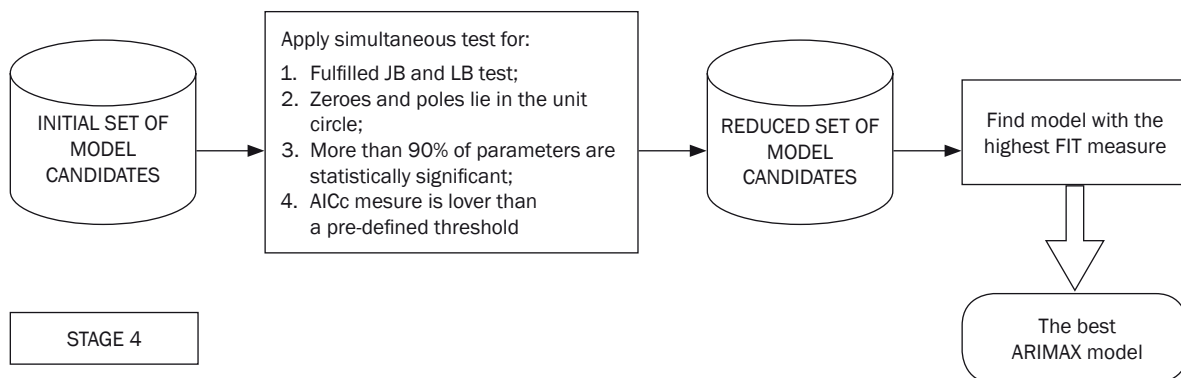


Figure 6 – Working mechanism of the fourth stage heuristics for the final model selection

This model has passed all the heuristic tests from stage 4. The values in parentheses symbolize the  $t$ -values of the corresponding statistics, needed to inspect the significance of parameter estimates. Since some of the calculated  $t$  values have indicated insignificant parameters, they were excluded from the final ARIMAX model.

For the estimation of parameters, the first 44 time samples were used, while in the testing of the model predictive power, the last 16 time samples were applied (see Figure 7). The predictive performance of our final model was also compared with the benchmark Holt-Winters model. Moreover, we have included another ARIMAX model in this comparison (let us call it “the top-fit” or the “best-fit” model), which had achieved even better fit to the data than our Model 8. However, it has been rejected in stage 4 since it had too many insignificant parameters, while some of its transfer function zeroes and poles have dropped outside the unit circle.

### 5.1 Predictive power of forecasting models

Figure 7 shows the real throughput data,  $y(t)$ , the forecasts  $\hat{y}(t)$  of our final ARIMAX selected Model 8, the forecasts  $y(t)$  of the “top fit” (“best fit”) ARIMAX model, and the output  $y_{HW}(t)$  of the benchmark Holt-Winters model. The results in Figure 7 are shown for the estimation interval, test interval, and prediction interval, respectively. The “top fit” model  $y(t)$  follows the real time series  $y(t)$  so close that it is sometimes tough to see the difference. However, the forecasting performance of our final ARIMAX model also provides a relatively good fit to the real throughput data. It is true that there are some more sophisticated details of the real time series dynamics detected, which are not covered by our final model. We believe that there exist two major reasons for this fact as follows:

- Firstly, the observed time series most likely contains a complicated nature of its dynamics, which could not be incorporated into our model. The stochastic mechanism that generates the time series probably has certain nonlinearities, while there might also be some other influential external effects of an unknown source. Although such phenomena remain unmodelled, our model still provides a useful

tool for the prediction of the key variations in the time series dynamics.

- Secondly, the compendium of all required criteria for the selection of our final model was extremely rigorous which means that the best fit to the data was not the only criterion. Nevertheless, despite this, the selected final model ensures a sufficiently good fit, particularly regarding the major trend movements.

The good performance of our final model is also evident from Table 2. Here, all the important criteria (RMSE, MAE, % of FIT, and AICc) are shown for all three compared models. The ARIMAX “top fit” model has achieved the best performance, but from the statistical point of view suffers from some serious deficiencies (non-parsimonious, problems with zeroes and poles). Such deficits can have serious consequences such as unstable future model behaviour and biased or inadequate future forecasts. On the other hand, our final selected model has provided a slightly worse fit performance, but it is better in a statistical sense (all fulfilled criteria) and ensures more stable future behaviour, while keeping the future forecast errors within a reasonably small range. From Table 2 and Figure 7 is also evident that both ARIMAX models outperformed the Holt-Winters model. Since the latter as a triple ex-

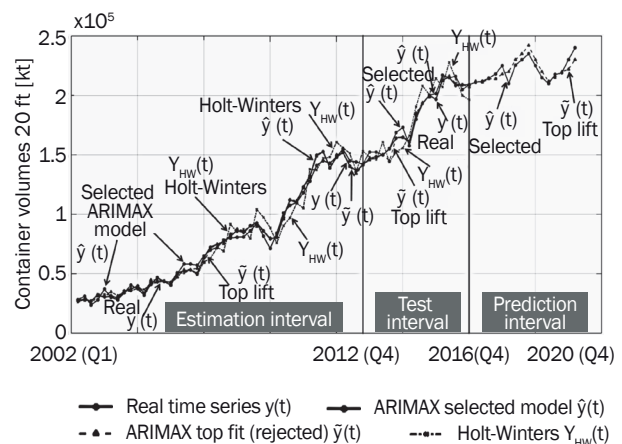


Figure 7 – Actual time series  $y(t)$  and predictions of three models (final selected ARIMAX model- $\hat{y}(t)$ , “top fit” ARIMAX model- $\tilde{y}(t)$ , Holt-Winters- $y_{HW}(t)$ ); Estimation: (2002Q1-2012Q4), Test: (2013Q1-2016Q4), Prediction: (2017Q1-2020Q4)

Table 2 – The performance measures for all three compared models

Model	Fit Criteria			Information Criteria
	RMSE	MAE	%FIT	AICc
ARIMAX final model	4,684.1	3,569.4	89.11	18.09
ARIMAX “top fit” model	2,772.8	1,603.2	94.53	16.05
Holt-Winters model	9,805.1	8,554.0	78.02	/

ponential smoothing filter [54] does not have such sophisticated working mechanism, it suffers from significant overfitting and inadmissible oscillatory behaviour at some specific time points.

## 5.2 Forecasting performance on the out-of-sample prediction interval

Figure 7 also shows forecasts for the forthcoming out-of-sample prediction interval (2017-2020) generated by our final selected model and the “top-fit” model. These forecasts imply that after the stable rising of container throughput in 2017 (time samples 61-65), more turbulent behaviour might follow in the consecutive years 2018-2020 (time samples 66-76). Particularly noticeable is the dropping behaviour of throughput beginning with a period nine quarters after the end of 2016 (see time samples 69-73). This dropping will persist over the entire year 2019, while the previous state of the throughput will be roughly exceeded after the end of year 2020 (see time samples 73-76).

## 5.3 Main findings and discussion of results

An observation of forecasts in Figure 7 convinces us that our final model is capable of giving promising prediction results. It manages to predict and capture all major changes of the container throughput dynamics. However, the forecasts for the forthcoming years till the year 2020 can worry us from the management implication point of view. Namely, if we compare the predicted future scenario in the years 2019-2020 with the throughput dynamics in the years 2008-2009 (time samples 28-36), we can notice certain similarities. At that time period, after the eruption of the crisis in 2008, the throughput dynamics needed almost two years to recover from the negative drop and to capture the previous rising trend. Yet, despite the global downturn, the drop in the container throughput in 2009 was not significantly big, while a noticeable increase followed in 2010 as consequence of the newly applied direct line (Far East Asia-North Adriatic) [30].

Hence, from the mentioned similarity we can conclude that there exists certain likelihood that some significant negative macroeconomic behaviour will happen again in the near future. These fears can also be identified from several other sources (e.g. [55-58]), where researchers and experts express fear of new forthcoming undesired economic events. In several studies, namely, a new economic stagnation or global recession is predicted somewhere in the period 2018-2020 [55-57], while some researchers even warn about an eruption of the new economic crisis [58]. Their assumptions are based on analyses such as “what if” scenario playing analysis [55, 56], where China economic performance represents the biggest concern. On the other side, authors of Malaysian study

[58] conclude that the new recession is expected to occur somewhere in 2018 due to the long cycles of the world recession history. Consequently, this economic slowdown will affect the trade in the observed port in Malaysia.

To summarize, considering the just mentioned warnings, as well as our prediction results, there is a chance that the throughput trend will not persist to keep a steady rising behaviour in the forthcoming years. On the contrary, its dynamics might become more unstable, and there are also possible throughput drops of uncertain magnitude. Since these facts can seriously disturb the big managerial decisions about future investments and other important strategic planning, caution must be taken to prevent bigger decision failures. Contrary-wise, it should be tactically better to wait for a while and see what the future will bring.

## 6. CONCLUSION

In this paper, we have examined the impact of integration of macroeconomic indicators on the accuracy of the container throughput forecasting model. For that purpose, we have introduced the novel modelling approach for the construction of the prediction model, which might represent a useful reinforcement of the existing forecasting module of DSS in the observed Port of Koper.

The modelling process has combined the Dynamic factor analysis and the Box-Jenkins ARIMAX modelling within the unique four-stage modelling framework. The DFA was used to extract the dynamic factors from a much higher number of original external exogenous indicators. The extracted factors were then injected into the “generator” of ARIMAX models, which has created a whole family of model candidates. Those models were afterwards exposed to rigorous diagnostic checking and goodness-of-fit testing. All models that did not fulfil the required pre-defined criteria were rejected. In the final stage of the modelling process, a special heuristic was used to find and select the final ARIMAX model automatically. The applied heuristics has not only ensured that the final model provides a good fit to the data, but also satisfies the requirements related to other strict conditions, such as information criteria, parsimony, invertibility, stationarity and stability issues.

The derived final ARIMAX model was tested on the real container throughput time series data. The forecasting performance has shown promising prediction results regarding the model’s fit to the real throughput data. It managed to capture and predict all important movements of the real time series dynamics, particularly in the case of major changes in the time series’ trend. The model was also employed to generate forecasts for the forthcoming prediction interval

(2017-2020). These forecasts can cause worries since after a presumed stable rising in 2017, more turbulent behaviour might followed in the subsequent period of 2018-2020. Accordingly, caution must be applied about making any bigger investment and planning decisions activated from the managerial side.

One of the goals of our further research is to extend the forecasting module of DSS in such a way, that it will be able to provide accurate forecasts for other major commodities as well. Further, in our future work we also plan to upgrade the applied heuristics for the final model selection in such a way that the isolation of the most adequate model will be achieved in a shorter amount of time. Finally, the input data reduction might perhaps be enhanced via the throughput feedback information, which would help us to investigate whether it is possible to obtain the more suitable configuration of initial exogenous indicators needed for DFA purposes.

asist. mag. **MARKO INTIHAR**, mag. inž. logistike<sup>1</sup>

E-mail: marko.intihar@um.si

izr. prof. dr. **TOMAŽ KRAMBERGER**<sup>1</sup>

E-mail: tomaz.kramberger@um.si

izr. prof. dr. **DEJAN DRAGAN**<sup>1</sup>

E-mail: dejan.dragan@um.si

<sup>1</sup> Fakulteta za logistiko, Univerza v Mariboru

Mariborska cesta 7, 3000 Celje, Slovenija

## **NAPOVEDOVANJE PRETOVORA KONTEJNERJEV S POMOČJO DINAMIČNE FAKTORSKE ANALIZE IN ARIMAX MODELA**

### **POVZETEK**

Glavni namen raziskave je preučevanje vpliva vključevanja makroekonomskih kazalnikov na natančnost napovedovanja modela za predikcijo pretovora kontejnerjev. V ta namen sta uporabljeni dinamična faktorska analiza in Avtoregresivni integracijski model časovnih vrst s premikajočim povprečjem in eksogenimi vhodi (ARIMAX). Obeh metodologiji sta integrirani v okviru inovativne štiristopenjske hevristične procedure. V prvem koraku najprej izluščimo dinamične faktorje iz zunanjih makroekonomskih kazalcev, ki vplivajo na opazovan pretovor. Nato v drugem koraku tvorimo celo družino ARIMAX modelov različnih redov na podlagi prej identificiranih dinamičnih faktorjev. V tretjem koraku izvedemo diagnostično testiranje modela vključno z validacijo kakovosti prileganja napovedi danim realnim podatkom. V sklopu tretjega koraka preverimo številne statistične kriterije, kriterije ocene napake modela, informacijske kriterije, ter preverimo optimalnost strukture modela. V zadnjem koraku s posebno hevrstiko izberemo najboljši model in ga preizkusimo na dejanskih podatkih pretovora kontejnerjev v Luki Koper. Rezultati kažejo, da z uporabo makroekonomskih kazalnikov v predikcijskem modelu lahko dosežemo bolj natančne napovedi pretovora kontejnerjev. Raziskava nas prepriča, da bi predlagani model lahko predstavljal koristno okrepitev obstoječega sistema za napovedovanje v opazovanem pristanišču.

### **KLJUČNE BESEDE**

napovedovanje pretovora kontejnerjev; ARIMAX model; dinamična faktorska analiza; eksogeni makroekonomski indikatorji; analiza časovnih vrst;

### **REFERENCES**

- [1] Zhang C, Huang L, Zhao Z. Research on combination forecast of port cargo throughput based on time series and causality analysis. *Journal of Industrial Engineering and Management*. 2013;6(1): 124-134. doi: 10.3926/jiem.687.
- [2] Peng WY, Chu CW. A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling*. 2009;50(7-8): 1045-1057.
- [3] Button K, Chin A, Kramberger T. Incorporating subjective elements into liners' seaport choice assessments. *Transport Policy*. 2015;44: 125-33.
- [4] Xie G, Wang S, Zhao Y, Lai KK. Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study. *Applied Soft Computing*. 2013;13(5): 2232-2241.
- [5] Dorsser JCMV. A Very Long Term Forecast of the Port Throughput in the Le Havre-Hamburg Range up to 2100. *European Journal of Transport and Infrastructure Research (EJTIR)*. 2012;12(1).
- [6] Wang J, Wang S. *Business Intelligence in Economic Forecasting: Technologies and Techniques*. Business Science Reference; 2010.
- [7] Gupta R, Kabundi A. A large factor model for forecasting macroeconomic variables in South Africa. *International Journal of Forecasting*. 2011;27(4): 1076-88.
- [8] Forni M, Hallin M, Lippi M, Reichlin L. The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics*. 2000;82(4): 540-54.
- [9] Kapetanios G, Marcellino M. A parametric estimation method for dynamic factor models of large dimensions. *Journal of Time Series Analysis*. 2009;30(2): 208-38. doi: 10.1111/j.1467-9892.2009.00607.x.
- [10] Stock JH, Watson MW. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*. 2002;20(2): 147-162. doi: 10.1198/073500102317351921.
- [11] Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*. New Jersey, USA: Wiley; 1994.
- [12] Ljung L. *System identification: theory for the user*. Prentice-Hall; 1987.
- [13] Fung MK. Forecasting Hong Kong's container throughput: an error-correction model. *Journal of Forecasting*. 2002;21(1): 69-80.
- [14] Seabrooke W, Hui ECM, Lam WHK, Wong GKC. Forecasting cargo growth and regional role of the port of Hong Kong. *Cities*. 2003;20(1): 51-64. doi: 10.1016/S0264-2751(02)00097-5.
- [15] Chou C-C, Chu C-W, Liang G-S. A modified regression model for forecasting the volumes of Taiwan's import containers. *Mathematical and Computer Modelling*. 2008;47(9-10): 797-807.
- [16] Schulze PM, Prinz A. Forecasting container transshipment in Germany. *Applied Economics*. 2009;41(22): 2809-15. doi: 10.1080/00036840802260932.
- [17] Chen S-H, Chen J-N. Forecasting container throughputs

- at ports using genetic programming. *Expert Systems with Applications*. 2010;37(3): 2054-8.
- [18] Ping FF, Fei FX. Multivariate forecasting mode of Guangdong province port throughput with genetic algorithms and Back Propagation neural network. *Procedia-Social and Behavioral Sciences*. 2013;96. doi: 10.1016/j.sbspro.2013.08.133.
- [19] Huang A, Lai K, Li Y, Wang S. Forecasting container throughput of Qingdao port with a hybrid model. *Journal of Systems Science and Complexity*. 2015;28(1): 105-21.
- [20] Wang Ya, Chou CCb, Yeo GTa. Application and improvement of a system dynamics model to forecast the volume of containers. *Journal of Applied Science and Engineering*. 2013;16(2): 187-96.
- [21] Gosasang V, Chandraprakaikul W, Kiattisin S. A Comparison of Statistical Technique and Neural Networks Forecasting Techniques for Container Throughput in Thailand. 2011;27(3): 463-82.
- [22] Gómez V, Maravall A. Programs TRAMO and SEATS. Instructions for User. Banco de Espana; 1996.
- [23] Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*. 2002;18(3): 439-54.
- [24] Perez-Amaral T, Gallo GM, White H. A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA). *Oxford Bulletin of Economics and Statistics*. 2003;65(s1): 821-38.
- [25] Castle JL. Evaluating PcGets and RETINA as Automatic Model Selection Algorithms. *Oxford Bulletin of Economics and Statistics*. 2005;67(s1): 837-80.
- [26] Trupac I, Twrdy E. More Competitiveness of the Port of Koper through Supply Chain Integration. *Promet – Traffic & Transportation*. 2010;22(4): 251-257.
- [27] Dragan D, Intihar M, Prah K, Kramberger T. Forecasting of Maritime Cargo Flows to support the Planning Activities in Napa Ports. Report. University of Maribor, Faculty of Logistics; 2013.
- [28] Dragan D, Kramberger T. Forecasting the Container Throughput in the Port of Koper using Time Series ARIMA model. In: *Proceedings of the conference ICLST 14*, 19-21 June 2014, Celje, Slovenia; 2014.
- [29] Brank B, editor. Port of Koper – Gateway for Central and Eastern Europe. *Logistics Forum 'Asia - Engine for Growth'*, Kuala Lumpur, Malaysia; 2013.
- [30] Twrdy E, Trupac I, Kolenc J. Container Boom in the Port of Koper. *Promet – Traffic & Transportation*. 2012;24(2): 169-175.
- [31] EUROSTAT. Port-throughput . Available from: [http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\\_database](http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database) [Accessed 12<sup>th</sup> April 2017].
- [32] Perkovic M, Twrdy E, Batista M, Jankowski S, Gućma L. The Increase in Container Capacity at Slovenia's Port of Koper. *TransNav*. 2013;7(3): 441-8.
- [33] O'Sullivan A, Sheffrin SM. *Economics: Principles in action*. New Jersey, Upper Saddle River: Prentice Hall; 2003.
- [34] Shao AT, Lee H. An exploratory study examining the impact of China's rapid economic growth on the Asian shipping industry. *Journal of International Business and Economics*. 2007;7(1).
- [35] Byrne JP, Darby J, MacDonald R. US Trade and Exchange Rate Volatility: A Real Sectoral Bilateral Analysis *Journal of Macroeconomics*. 2008;30(1): 238-259.
- [36] Belenkiy M, Riker D. Theory and Evidence Linking International Trade to Unemployment Rates. Office of Economics, U.S. International Trade Commission; 2015.
- [37] Commission TE. Singapore - Trade - European Commission 2017. Available from: <http://ec.europa.eu/trade/policy/countries-and-regions/countries/singapore/> [Accessed 17<sup>th</sup> April 2017].
- [38] Commission TE. Japan - Trade - European Commission 2017. Available from: <http://ec.europa.eu/trade/policy/countries-and-regions/countries/japan/> [Accessed 17<sup>th</sup> April 2017].
- [39] Commission EU. China - Trade - European Commission 2017. Available from: <http://ec.europa.eu/trade/policy/countries-and-regions/countries/china/> [Accessed 14<sup>th</sup> April 2017]
- [40] Commission TE. South Korea - Trade - European Commission 2017. Available from: <http://ec.europa.eu/trade/policy/countries-and-regions/countries/south-korea/> [Accessed 17<sup>th</sup> April 2017].
- [41] Trade ECD-Gf. List of the largest trading partners of the European Union 2017. Available from: [https://en.wikipedia.org/wiki/List\\_of\\_the\\_largest\\_trading\\_partners\\_of\\_the\\_European\\_Union](https://en.wikipedia.org/wiki/List_of_the_largest_trading_partners_of_the_European_Union) [Accessed 15<sup>th</sup> April 2017].
- [42] EUROSTAT. Real GDP growth rate. Available from: <http://ec.europa.eu/eurostat/web/products-datasets/-/tsdec100> [Accessed 3<sup>rd</sup> April 2017].
- [43] Imf W. World IMF Economic Outlook Database. Available from: <http://www.imf.org/external/index.htm> [Accessed 3<sup>rd</sup> April 2017].
- [44] Imf W. Economic Outlook Database. Available from: <http://www.imf.org/external/index.htm> [Accessed 4<sup>th</sup> April 2017].
- [45] Zuur AF, Ieno EN, Smith GM. *Analysing Ecological Data*. New York, NY: Springer; 2007.
- [46] Zuur AF, Tuck ID, Bailey N. Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences*. 2003;60(5): 542-52.
- [47] Harvey AC. *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press; 1990.
- [48] Zuur AF, Fryer RJ, Jolliffe IT, Dekker R, Beukema JJ. Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*. 2003;14(7): 665-85.
- [49] McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*, 2<sup>nd</sup> Edition. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008.
- [50] Makridakis S, Wheelwright SC, Hyndman RJ. *Forecasting: Methods and Applications*. 3<sup>rd</sup> Edition. Wiley India Pvt. Limited; 2008.
- [51] Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and Practice*. OTexts (2014). 2014.
- [52] Ljung L. *System Identification Toolbox for use with MATLAB*. 2011; 21.
- [53] Holmes EE, Ward EJ, Wills K. Marss: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal*. 2012;4(1): 11-19.
- [54] Dragan D, Kramberger T, Intihar M. A comparison of Methods for Forecasting the Container Throughput in North Adriatic Ports. Conference IAME 2014, Norfolk, USA; 2014.

- [55] Powell R. Risk scenarios. Asia Regional Strategic Forecast: Preparing For The Paradigm Shift. Globalisation and the Asian business environment: The Economist Corporate Network Risk scenarios. Report. The Economist Corporate Network; 2017.
- [56] Buitter W, Rahbari E, Saunders M, Luna S, Hale J, Menuet G. Is China Leading the World into Recession? Report. Global Economics View; 2016. 30 p.
- [57] Rahman NSFA, Muridan M, Najib AFA. A Maritime Forecasting Method for Analysing the Total Cargo Handling at Johor Port Berhad from 2013 to 2020. International Journal of Business Management and Economic Research IJBMER. 2015;6(3): 187-93.
- [58] Hsieh L. Roots of the next financial crisis: the last one's veterans give views. Reuters. April 19 2016. Available from: <http://blogs.reuters.com/financial-regulatory-forum/2016/04/19/column-roots-of-the-next-financial-crisis-the-last-ones-veterans-give-views/>