



Export sales forecasting using artificial intelligence

Vahid Sohrabpour^a, Pejvak Oghazi^{b,*}, Reza Toorajipour^c, Ali Nazarpour^d

^a Department of Operations Management, Copenhagen Business School, Copenhagen, Denmark

^b School of Social Sciences, Södertörn University, SE-141 89 Stockholm, Sweden

^c School of Innovation, Design and Engineering, Mälardalen University, Box 325, 631 05 Eskilstuna, Sweden

^d School of Business, Maynooth University, Maynooth, Co. Kildare, Ireland

ARTICLE INFO

Keywords:

Causal forecasting
Modeling
Export sales forecast
Genetic programming
Artificial intelligence

ABSTRACT

Sales forecasting is important in production and supply chain management. It affects firms' planning, strategy, marketing, logistics, warehousing and resource management. While traditional time series forecasting methods prevail in research and practice, they have several limitations. Causal forecasting methods are capable of predicting future sales behavior based on relationships between variables and not just past behavior and trends. This research proposes a framework for modeling and forecasting export sales using Genetic Programming, which is an artificial intelligence technique derived from the model of biological evolution. Analyzing an empirical case of an export company, an export sales forecasting model is suggested. Moreover, a sales forecast for a period of six weeks is conducted, the output of which is compared with the real sales data. Finally, a variable sensitivity analysis is presented for the causal forecasting model.

1. Introduction

Sales forecasting is essential in production, transportation and decision-making at all levels of a firm's supply chain (Verstraete et al., 2020). The importance of sales forecasting has long been recognized by both academics and practitioners (Makridakis, 1990; Remus and Simkin, 1987). To be more specific, such predictions are fundamentally crucial for firms' planning and strategy (Makridakis, 1996), allocation of corporate resources (Stein, 1997), marketing (Crittenden et al., 1993), logistics and supply chain (Hyndman and Athanasopoulos, 2018). Forecasts have become profoundly important because of shorter lead times, increasing customer expectations and the need to deal with limited resources (Boone et al., 2019). Effective demand forecasting usually lowers inventory levels and improves customer service (Kerkkänen et al., 2009). Accurate forecasting also plays an essential role in retail operations and manufacturing (Mentzer and Bienstock, 1998; Trappey and Wu, 2008). In some industries, forecasting acts as an input to many operations and business decisions that affect the profitability of the firm; in production planning, long-term forecasting is employed to determine an adequate level of manpower and acts as an input for business planning, such as planning for expansion or contraction of production units (Sa-ngasoongsong et al., 2012).

Normally, when a firm intends to implement a forecasting system, there is a tendency to replicate the concepts, targets, and principles from other companies to accelerate the implementation process; and because most forecasting methods have been developed for consumer products, there is a risk that unrealistic targets and unsuitable error metrics are adopted when the environment is different (Kerkkänen et al., 2009). Lack of forecasts or inaccurate forecasting can lead to lost sales or excess inventory, which imposes extra costs on the firm (Lawrence et al., 2000; Lawrence and O'Connor, 2000). For instance, inaccurate demand forecasts have led to suboptimal levels of the production workforce (Monks, 1987), which subsequently entails workload imbalances and increases in the costs of hiring, firing, and overtime labor activities; in many cases, the outsourced supply of materials and inaccurate forecasts could lead to a lack of supplies for production (Sa-ngasoongsong et al., 2012). Therefore, swift responses to sales results with forecasting error distribution contribute to a reduction in demand risks (Tanaka et al., 2012) and the balance of production materials and inventory. Such responses become even more critical in times of disasters, unpredictable incidents and unstable market circumstances. Govindan et al. (2020) argues that long-term, rapidly growing disasters such as COVID-19 (Cronavirus) confuse governments and decision-makers and fundamentally disrupt community activities and supply chains; the novelty and ambiguity of

* Corresponding author.

E-mail addresses: vs.om@cbs.dk (V. Sohrabpour), pejvak.oghazi@sh.se (P. Oghazi), reza.toorajipour@mdh.se (R. Toorajipour), ali.nazarpour@mu.ie (A. Nazarpour).

<https://doi.org/10.1016/j.techfore.2020.120480>

Received 20 June 2020; Received in revised form 10 November 2020; Accepted 16 November 2020

Available online 24 November 2020

0040-1625/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

COVID-19 makes it difficult to deal with, which ultimately entails the shortage of resources and disruptions among supply chains. Moreover, lockdowns and limited movement of goods make distribution centers suddenly inaccessible and interrupt the entire supply chain (Kumar et al., 2020).

While sales forecasting is a widely researched area (Winklhofer and Diamantopoulos, 2002), few empirical studies have focused on export sales forecasting (ESF) in comparison to domestic sales forecasting (see Çabuk (2019); Nie and Oksol (2018); Okorie and Ohakwe (2018); Wanto et al. (2019)). Special attention should be paid to ESF because exporting is an important factor for an extensive number of firms, which is characterized by a greater degree of uncertainty that accordingly entails a unique set of challenges (Winklhofer and Diamantopoulos, 2002). Sales forecasting (especially in the form of exports) plays a critical role in linking internal decision-making and unmanageable external factors that can affect an organization (Davis and Mentzer, 2007). Winklhofer and Diamantopoulos (1996) argue that the distinguished nature of ESF can be traced to three principle factors. The first is its internal organization, which is complex due to the involvement of country/regional managers, export agents and other third parties. Second, required data in the form of input into the forecasting process can be hard to obtain due to obstacles associated with data collection, such as availability issues, accessibility and permissions and quality of the data (Douglas and Craig, 1983). Third, different countries or regions might require different forecasting models or different techniques in terms of forecasting model development, as in some markets, data availability issues and local expertise might restrict forecasting approaches (Rice, 1997). These factors make export sales a distinct and challenging type of forecasting that cannot rely only on time series forecasting methods to be empirically precise and trustworthy. In this study, we intend to answer the following research question: Can GP accurately model and forecast export sales behavior in an unstable market?

Historically, forecasting research intended to find the best model for the existing data set (De Gooijer and Hyndman, 2006). Regardless of the context, there are various methods for sales forecasting. Usually, demand (and sales) forecasting has been addressed by traditional time series methods, including the ARIMA model and the Holt-Winters approach, but new AI-based methods have gained interest due to their ability to enhance predictive performance and model nonlinear patterns (Karmy and Maldonado, 2019). Verstraete et al. (2020) argue that traditional statistical forecasting methods extrapolate historical trends and seasonal fluctuations to predict the future, and as a result, these methods are incapable of predicting environmental macroeconomic changes in the business that usually influence demand significantly. To handle these changes, firms either manually adjusted their statistical forecasts or relied on experts' judgmental forecasts. However, these approaches are biased, since humans are generally inefficient in making such adjustments, and the process is time consuming. Moreover, Abarbanell and Bernard (1992), Ali et al. (1992) and Mendenhall (1991) have shown that judgmental forecasting methods are inefficient and biased. Input-driven or causal models can produce substantially higher accuracy in comparison to univariate models, especially when the input sequences are known (Murray and Ringwood, 1994). Causal forecasting methods use a predictive causal model, which defines the causal relationships between a set of dependent variables (forecast variable) and a set of independent variables (input causal factors or predictors) (Jeong et al., 2002). In ESF, several forecast variables affect the behavior of future sales. Therefore, causal forecasting is an adequate method for modeling and forecasting export sales.

Causal sales forecasting based on Genetic Programming is rarely studied, especially with a focus on real empirical data from an unstable market with multiple fluctuating variables. To answer our research question, this study intends to model and forecast export sales using a framework that employs an artificial intelligence technique - Genetic Programming - to build a causal equation that describes the relationship between the independent and dependent variables. This model not only

clarifies the relationship between variables but is also capable of forecasting future sales with high accuracy.

This study is structured as follows: in the first section, introductory explanations and theoretical frameworks are presented. In section two, Genetic Programming is explained, and the relevant details, such as definitions, steps and fitness function, are discussed. In the third section, the research methodology and framework are presented. In section four, data analysis and modeling are discussed, section five covers the results and discussion, and finally, concluding remarks are provided in section six.

2. Literature review

There are a limited number of studies on ESF (Lehmann, 2015), and the majority of the studies that research ESF focus on the improvement of export forecasts, export sales forecast accuracy and performance. Fiorito and Kollintzas (1994) found for the G7 countries that exports are procyclical and coincide with the business cycle of total output. Baghestani (1994) and Cardoso and Duarte (2006) highlight that survey results obtained from professional forecasters improve export predictions in the US and Portugal consecutively. Diamantopoulos and Winklhofer (1999) examined several linkages between firm and export characteristics in addition to the accuracy of short- and medium-term ESF. Winklhofer and Diamantopoulos (2003) also tested a path model of ESF behavior and performance to find the key determinants of export sales forecasting performance. Keck et al. (2010) demonstrated that the OECD25 trade forecasts can be enhanced by applying standard time series methods in comparison to a naïve prediction. Moreover, Ca'Zorzi and Schnatz (2010) forecasted extra Euro-area exports using various price and cost competitiveness measures and found that for a recursive estimation approach, the effective exchange rate based on the export price index outperforms the other measures. Wang et al. (2011) found that standard autoregressive integrated moving average (ARIMA) models are capable of improving export forecasts in comparison to heuristic methods for Taiwan. Jannsen and Richter (2012) forecasted German capital goods exports using a capacity utilization weighted indicator obtained from major export partners, while Elstner et al. (2013) used both hard data and indicators from the Ifo business survey to enhance forecasts for German exports.

Although the aforementioned studies focus on export forecasting, they provide a general perspective, mainly country/region oriented. Therefore, such studies do not provide a bespoke framework for modeling and forecasting export sales behavior from the perspective of an organization for a specific period of time or a particular product. The current study contributes to the existing body of literature by developing and testing a genetic programming-based causal export sales forecasting framework using novel software that is capable of accurately modeling and forecasting the behavior of export sales.

2.1. Genetic programming: definitions and steps

Genetic programming (GP) is an artificial intelligence technique derived from the model of biological evolution (Banzhaf, 2001). Here, the term "evolution" refers to an artificial procedure similar to the natural evolution of living organisms based on Darwin's theory of evolution by natural selection (Darwin, 1859). GP seeks to replicate the natural evolution in which entities evolve to solve problems (Conrad et al., 2016). The aim of GP is to evolve a program that, when appraised, will create behavior as close as possible to a desired goal (Williams and Molteno, 2008). This technique has demonstrated advantages compared to traditional fixed-length chromosome genetic algorithm (GA) approaches (Koza and Koza, 1992). The evolutionary method of "tree-based" GP was first defined by Cramer (Cramer, 1985). GP is a subset of the GA and uses its operators, such as remix and mutation. In this process, trees are created and used as models that are part of a population. Therefore, GP is capable of modeling and formulating mathematical

equations and simulating different lines and curves (Bardool et al., 2016).

In the GP, an individual is simultaneously presented by phenotype and genotype. The individuals in a population are trees of different shapes and sizes. A tree includes nodes (functions) and leaves (constants or variables). These elements (functions, variables and constants) are determined by the user. Thus, the genetic program runs an evolutionary search in an enormous scope of equations, which can be expressed by the initial values (Bardool et al., 2016). Vanneschi and Poli (2012) argue that GP is capable of solving many problems in a competent manner, and, in a sense, it represents an endeavor to achieve the very ambitious goal of having the computer figure out an implementation for what a program is aimed to do, without specifying how, and just defining what. Ruican et al. (2008) suggested a detailed flowchart for the GP paradigm represented in Fig. 1. The GP process includes:

- Step 1: GP begins with generating a random population of solutions of size N .
- Step 2: A specific fitness value is assigned to each solution of the populations.
- Step 3: Applying genetic operators, a new generation of individuals is produced from the previous generation. Choosing the GP operators is the most important part of the process. GP uses genetic operators to find new solutions with fitness values that are better than those of the parents (previous generation). If the selected GP operator is a reproduction, an individual is selected from the present population and then cloned into the new population. The reproduction operator imitates the principle of natural selection and survival of the fittest. For a crossover operator, two random individuals are selected from the population, and a subtree is chosen from each of the selected individuals. These subtrees are cross-pollinated to produce two new offspring. For a mutation operator, a random node is selected from the parent tree. Then, the selected subtree is deleted or replaced by a newly generated random subtree.
- Step 4: The previous step (Step 3) continues until the new population reaches the N number of members.
- Step 5: Steps 2–4 are repeated to find the desired solution, and if the target solution is not found, after a predefined number of generations, the GP operation terminates.

The choice of forecasting method depends on various factors, such as the time period, the cost of preparing forecasting and benefit resulting from its use, desired level of accuracy, forecasting period, data quality,

availability, and the level of complexity of the relationships (Bintang et al., 2019). Our aim in this study is to model and forecast the behavior of sales in an unstable market with chaotic data that has neither a systematic pattern nor an apparent trend in its structure. Therefore, instead of traditional forecasting methods, we employed GP, an AI technique that is useful in modeling and formulating mathematical equations and is capable of finding meaningful relationships between several variables in a chaotic dataset. To this end, we employed GP-based EUREQA software (Liu et al., 2020) that is capable of analyzing the data by automatically testing many different equation forms (Yang et al., 2020). Unlike previous data-driven methods that prespecify a form and then proceed to find the optimal relationship, EUREQA (Schmidt and Lipson, 2013) has been successfully applied to various studies to find a new predictor without any prespecified forms (Liu et al., 2020).

2.2. Genetic programming: fitness function

A fitness value is assigned to every program in the population, which represents the degree to which the program solves the problem of interest (Vanneschi and Poli, 2012). There are different methods for determining the fitness values. In one method, the fitness value is the difference between the output and the expected value. Clearly, the smaller the value, the better. In this study, we employ the GP process (including the fitness function) of EUREQA software.

3. Methodology

3.1. Design and framework

The current study aims to model sales behavior, forecast export sales for a particular period of time, and conduct a variable sensitivity analysis for an export company as a case study. To this end, an AI technique, namely, GP, is used to model and analyze the empirical data provided by the exporter. The research framework for this study is illustrated in Fig. 2, which encompasses four steps. This framework is adapted from Fan et al. (2017) with major modifications.

Step 1-data collection: In this step, the data for six variables (five independent variables and one dependent variable) are collected. These empirical data are obtained from an export sales company willing to participate in this study. The detailed data collection process is explained in section 3.2.

Step 2-data preprocessing: In this step, the goal is to prepare the raw data for the next step. Therefore, a strategy is employed to address

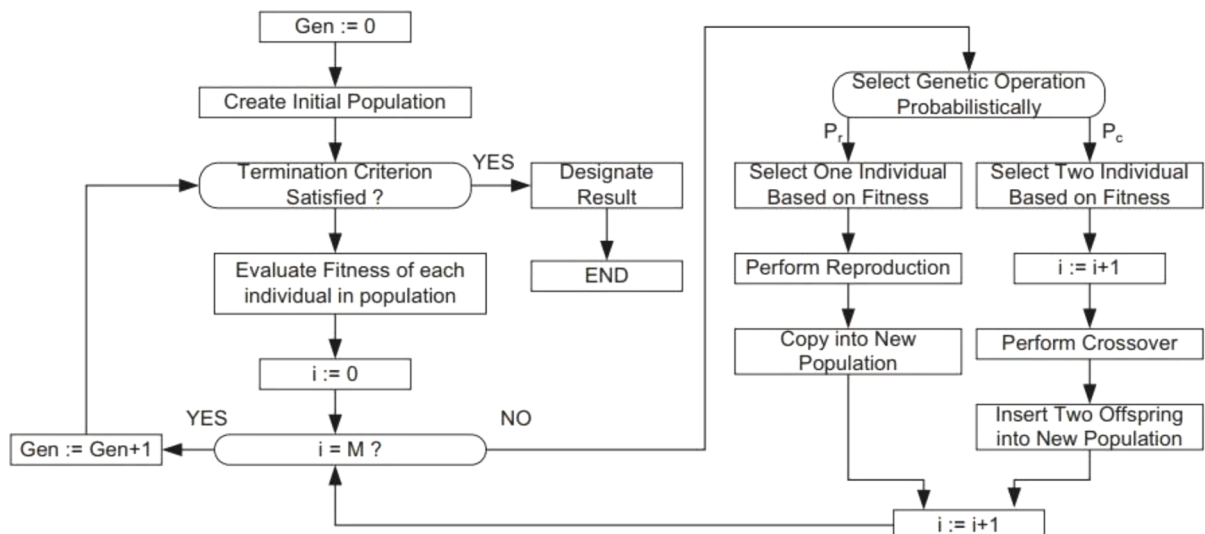


Fig.. 1. Framework for the GP paradigm (Ruican et al., 2008).

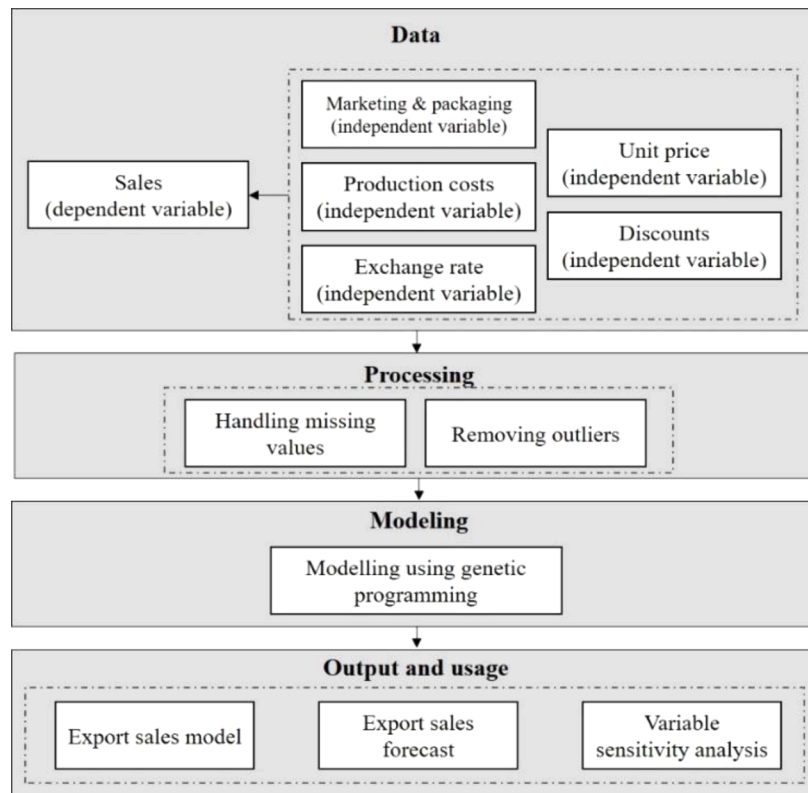


Fig.. 2. Research framework, adapted from Fan et al. (2017).

missing values and outliers.

Step 3-modeling: In this step, the data are applied to the genetic program. Accordingly, a new causal forecasting model is developed. The forecasting accuracy of this model is calculated using various error metrics. This step is presented in section 4, data analysis and modeling.

Step 4-output and usage: This step consists of three parts. The GP-based causal ESF model is presented along with accuracy indices; an export sales forecast is presented for a six-week period of time, and the result is compared with the real sales data; and a variable sensitivity is conducted to measure the impact of each variable on sales. This step is presented in section 5, resulting in a sand discussion.

3.2. Data collection

The fast growth of the Internet and information technology has led to the availability of an enormous amount of external data. Estimates by IBM suggest that 43 trillion GB of data will be made in 2020, approximately 300 times the volume produced in 2008. This expansion in data availability leads to a shift from finding the best model to finding the right data in the form of causal forecasting methods (Verstraete et al., 2020). The abundance of data also simplifies the more accurate causal forecasting for a wide range of firms.

As mentioned earlier, the prevailing forecasting techniques that use time series techniques predict the behavior of future sales (or demand) assuming that level, trend and seasonal fluctuations will follow the same rules in the future and that it will continue without any considerable distortion. These methods ignore the causal relationship between the independent factor (such as export sales) and other influential dependent factors. In this research, we employed a causal forecasting method using GP that strives not only to forecast export sales but also to define the relationship in the form of a mathematical model and to analyze the sensitivity and the impact of each independent variable on the dependent variable, which are other capabilities that time series techniques cannot offer.

To build the causal model, we first identified the variables. We selected variables primarily based on two factors: first, the availability of data, and second, the literature support. This study is originally inspired by a real problem in Middle Eastern companies that faced fluctuations in their sales and other influential factors. The main goal was to model and forecast their sales with high accuracy, which was not attainable using time series techniques and to identify variable sensitivity and the impact of each independent variable on export sales.

Since this is an export sales forecasting, clearly, the dependent variable is the export sales. Various factors affect sales. Baldauf et al. (2000) suggested a model including predictors that impact export sales and performance and argued that costs are one of the most influential factors in export sales. According to Chopra and Meindl (2016), environmental and economic factors play a key role in causal forecasting methods. Therefore, the exchange rate was selected as an independent factor that was also highlighted by the export company. The existing literature emphasizes the impact of marketing mix factors as important influencers on sales forecasts (Chase Jr, 1998; Fildes et al., 2019; Jeong et al., 2002). Therefore, three other variables were added to our database: marketing and packaging, price and discounts. In aggregate, we collected data for one dependent variable and five independent variables. The sales data for 58 weeks were gathered, from which 54 data rows (weeks) were used in the GP, and 4 rows were used for forecasting. The descriptive statistics for the 54 variables are presented in Table 1.

3.3. The case of export company

The aforementioned data were collected from a small-sized Middle Eastern export company that operates in the fresh and dried dairy industry. The production portfolio of this company includes several items, such as whey protein concentrates 80% and 60%, skimmed milk, butter, whipped and normal butter, fresh milk, etc. However, the primary product of this company is powdered milk, which was chosen as the target product of this study. Consequently, all of the data were gathered

Table. 1

Descriptive statistics for the empirical data.

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Sum</i>	<i>Mean</i>	<i>Std. Deviation</i>
Sales	54	5,417,500,000	58,040,000,000	1,684,487,114,026	31,194,205,815.30	11,534,946,544.956
Packaging & advertising	54	121,613,112	1,805,903,814	50,207,862,933	929,775,239.50	414,134,197.835
Exchange rate	54	99,875	149,214	6,731,613	124,659.50	11,496.531
Costs	54	2,688,020,000	53,918,831,600	1,541,207,207,180	28,540,874,207.04	11,267,026,153.120
Discounts	54	5,417,500	290,200,000	5,688,492,047	105,342,445.31	72,244,143.548
Price	54	175,000	206,500	10,065,550	186,399.07	10,699.059

for this product. Its market includes mostly neighboring countries, including but not limited, Persian Gulf countries, western Asian countries, Turkey and Azerbaijan. Data were obtained from the company's main headquarters in the Middle East. The values of all variables are presented in the currency of the country of origin. Table 2 presents the descriptive information of the company. The chosen company is a typical case and not a special one. In particular, our chosen company is neither a leader nor has a monopoly for any of its products, as there are competitors in the region. Such characteristics can help us to enhance the generalizability of our results. In addition, our chosen company operates in an unstable market and in a less studied region. Companies and organizations in developed countries are usually abundant in studies; therefore, we chose a less prevailing region to fill the relevant gap as well (Asian et al., 2019).

3.4. Data access issues

Data access issues have always been an important yet restricting part of any empirical research. These issues are related mainly to the security policies of a firm when providing data sources for users (Abidi et al., 2019). Obtaining access to adequate data can cause problems in many ways, such as delays in access and the type of data (nonstandard formats), and it can place heavy demands on the project's resources (Golder et al., 2016; Schroll et al., 2016; Strom et al., 2016; Wieseler et al., 2013). In some cases, a significant amount of effort is spent on access issues instead of the actual scientific question (Santoro et al., 2018). Therefore, issues regarding accessing data should be considered at an early stage (Golder et al., 2019). For the current study, we contacted several companies with larger volumes of data, but our access to their databases was limited due to their data security policies. Finally, although we identified several options for our study, access issues were determinant factors in finalizing our choice. While larger data sets could provide more details and employ the full potential of the GP, the current source of data is compatible with this research and adequate for our analyses.

4. Data analysis and modeling

4.1. Variables

One of the main capabilities of GP is finding an optimal mathematical model that can predict and describe a system based on the given input-response data (Chen et al., 2018). In other words, GP is capable of modeling the impact and behavior of a set of variables. To start the GP process, a basic model defining the relationship between the dependent and independent variables is presented to the GP. The aim is to define the independent and dependent variables for the genetic program. This

Table. 2

Summary of information of the export company.

<i>Company's field</i>	<i>Number of staff</i>	<i>Size</i>	<i>Main product for ESF</i>	<i>Average turnover</i>	<i>Age</i>	<i>Market size</i>
Dairy	54	Small	Powdered milk	3979 billion	11	3%

basic equation is:

$$y = f(m, x, c, d, p)$$

where y is the dependent variable, weekly sales; m is the costs of packaging and marketing; x is the weekly exchange rate; c is the production cost; d represents the discounts; and p is the unit price for every kilogram of the sold product.

4.2. Data processing and error metrics

To build a GP-based forecasting model, we divided the entire 60-week data period into three separate parts as follows:

- Part 1: The data for six weeks of sales were separated and presented to the program in any way. The aim here was to keep a piece of data that the GP program has never encountered in any way, which we can use for forecasting and comparing with the real export sales data. While the GP conducts such a method for 54 weeks of data (training and validation data), we wanted to further test the model and its reliability.
- Part 2: 75 percent of the data was used for training; this part is used to identify the mathematical model that describes the relationship between variables.
- Part 3: 25 percent of the data was used for validation; this part is used for comparison between the generated model and the real data. To evaluate the quality of a model based on this comparison, error metrics are employed.

4.3. Error metrics

Error metrics are measurement tools for calculating and comparing the accuracy of different predictions (Jackson et al., 2019). Put simply, error metrics measure the quality of a model by comparing the predictions against actual values. Here, we use four different error metrics to assess the quality of the causal forecasting model:

- Mean absolute error (MAE): The mean absolute error is a well-known statistical tool that measures the closeness of the predictions to the outcomes (Adetiloye and Awasthi, 2017). It provides the average magnitude of forecast errors (Kato, 2016). It is a common index in scientific studies for measuring the precision of predictions. The mean absolute error (E_i) of an individual program is evaluated by the equation:

$$MAE : f(E_i) = \frac{1}{N} \sum_{i=1}^B |y_i - f(x_i)|$$

- Mean squared error (MSE): Another popular index for assessing the quality of predictions is MSE. It is used to measure the average of the squares of the errors. The value of MSE depends on the units of the predicted variable and varies from zero to infinity (Gupta et al., 2009); it is never negative. The equation defining MSE is as follows:

$$MAE : f(E_i) = \frac{1}{N} \sum_{i=1}^B (y_i - f(x_i))^2$$

- 3 R squared goodness of fit (R^2): The R^2 is a statistic that represents the percentage of the variance in the dependent variable (here, sales or y), which the independent variables collectively explain. This index is used to measure the strength of a model on a 0–100 percent scale. R^2 is a widely used goodness of fit measurement index, the characteristics of which are known in the applied research (Colin Cameron and Windmeijer, 1997). Its equation is as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where SS_{tot} is proportional to the total variance, and SS_{res} is the residual sum of squares.

- 4 Correlation coefficient: The correlation coefficient is used to measure the direction and strength of the relationship between variables, and it takes on values between -1.0 and $+1.0$ (Mun, 2014). It measures the agreement between paired variables (Shaw et al., 2018). The following expression gives the formula for this index:

$$\frac{1}{n-1} \sum_{i=1}^n \left(\frac{f(x_i) - f(\bar{x})}{S_{f(x)}} \right) \left(\frac{y - \bar{y}}{S_y} \right)$$

where $S_{f(x)}$ and $S_{f(y)}$ are standard deviations.

4.4. Variable sensitivity analysis

Variable sensitivity analysis (VSA) is concerned with how the variations in the outcome of a mathematical model or system can be allocated to different sources of variation in its inputs (Saltelli, 2002; Saltelli et al., 2008). Sensitivity analysis is an underappreciated part of research that determines the amount of variations in an output in response to specific fluctuations of the inputs (Simske, 2019). In other words, using this type of analysis, we can identify the impact of each independent variable (such as marketing) on the dependent variable (export sales).

For every independent variable, we computed five indices as follows:

- 1) Sensitivity: This is the relative impact that an independent variable has on the target variable (y) within this model. Defined as $\left| \frac{dy}{dx} \right| \frac{\sigma(x)}{\sigma(y)}$ evaluated at all input data points.
- 2) Percent (%) Positive: This is the likelihood that an increase in the intended variable will increase the target variable. For example, with an 80% positive in the independent variable, then 80% of the time increases in this variable will lead to increases in the dependent variable (the remaining 20% may decrease it or have no impact). If the percent positive equals zero, then increases in this variable will not increase the dependent variable. Defined as the percent of data points where $\frac{\delta_y}{\delta_x} > 0$.
- 3) Positive Magnitude: This is the amount of positive impact when increases in this variable cause increases in the target variable. Defined as $\left| \frac{dy}{dx} \right| \frac{\sigma(x)}{\sigma(y)}$ at all points where $\frac{\delta_y}{\delta_x} > 0$.
- 4) Percent (%) Negative: This is the likelihood that an increase in this variable entails a decrease in the dependent variable. For instance, if the percent negative equals 55% for an independent variable, then 55% of the time, increases in this variable will cause decreases in the target variable. If the percent negative equals zero, then increases in this variable will not decrease the dependent variable. Defined as the number of data points where $\frac{\delta_y}{\delta_x} < 0$.

- 5) Negative Magnitude: This is the amount of negative impact on a dependent variable when the intended variable increases. Defined as

$$\left| \frac{dy}{dx} \right| \frac{\sigma(x)}{\sigma(y)} \text{ at all points where } \frac{\delta_y}{\delta_x} < 0.$$

where $\left| \frac{dy}{dx} \right|$ is the partial derivative of y with respect to x ; $\sigma(x)$ and $\sigma(y)$ are the standard deviations of x and y , respectively; $|x|$ denotes the absolute value of x , and \bar{x} represents the mean of x .

Before running the GP, a strategy was employed to address probable missing values and outliers. Implementing this strategy, we found that there were very few outliers, and we detected no missing variables.

5. Results and discussion

After preparing the data, we ran the GP and sensitivity analysis to model, forecast, and analyze the sensitivity of each independent variable. Here, we first present the GP-based export sales model and discuss its predictive quality according to the error metrics results.

5.1. Export sales model

The GP conducts millions of evaluations per second to test and produce viable mathematical models describing the relationship between dependent and independent variables. Our GP conducted 5 to 6 million evaluations per second using an intel Core i5 with 2.5 GHz speed. Usually, the more complicated outcomes of a GP are capable of producing much more accurate results (in terms of forecasting, modeling, etc.). The result of running the GP is presented as a mathematical model as follows:

$$y = 9.63m + 1.2c + 7.72e - 7xc + \frac{-4.78e4c}{p} - 2.02e - 5mp - 4.92e7 \cdot \cos(2.02e - 5m) \quad (1)$$

Four error metrics are calculated for this model: R squared goodness of fit, correlation coefficient, mean squared error and mean absolute error. The following table presents the values for each of these metrics.

As Table 3 shows, this model has an R-square goodness of fit value of 98%, which means that 98% of the variance in the dependent variable, which is export sales here, can be explained by the independent variables collectively. It also indicates the strength between the GP-based model and the dependent variable. The most important reason for the high value of R^2 for this model is the causal nature of the model, which is based on the relationships between the dependent and the independent variables. Unlike time series techniques for sales and demand forecasting, which disregard the interconnections between variables and just extrapolate the future trends based on those of the past, causal techniques identify these relationships and forecast the future based on them. Another reason for the high value of R^2 is the capacity of GP to produce complicated mathematical models that analyze the training data set in detail and create a model that is profoundly accurate and close to the validation data set. This leads to the production of accurate but complicated and large mathematical models.

Another index that evaluates the strength of the relationship between variables is the correlation coefficient. As Table 3 presents, our GP-based model has a correlation coefficient of 0.98, which indicates a very strong positive relationship between the predicted and actual sales data. Here, MAE has an approximate value of 76.2 million (currency of the origin country), and MSE has a value of 1.83. To demonstrate the

Table 3
Error metric values.

Error metric	R-square goodness of fit	Correlation Coefficient	Mean Squared Error	Mean Absolute Error
Value	98%	0.98	1.83791e16	76,247,770

closeness of the predicted and actual export sales values better, Fig. 3 is presented.

Real export sales behavior is depicted as a blue line, and the forecasted sales are depicted as orange dots. As the figure shows, the predicted export sales behavior is almost coincident with the real sales behavior, which is due to the causal nature of the model and the high accuracy of the prediction. The causal forecasting technique identifies the relationships between variables and defines a model that is capable of identifying future trends based on the value of independent variables. This not only increases the accuracy of a model but also improves its reliability.

5.2. Empirical export sales forecast

As mentioned earlier, we divided the 54 weeks of data into two parts in the genetic program, training data (75%) and validation data (25%), where it evaluates the quality of the model using four error metrics. However, to assess the predictive quality of the model further and to conduct an empirical export sales forecast, we used six weeks of data that have never been presented to the GP in any way. The data cover the six weeks from 11/15/2019 to 12/26/2019. The GP-based causal forecast model (1) identifies the relationship between the variables (five independent variables and one dependent variable). To conduct the forecast, we calculated the output (export sales forecast) by using the inputs (data of marketing, exchange rate, costs, discounts, unit prices). Then, we compared the outputs with the real export sales that were reported by the company. The results were very close to the reported sales and confirm the predictive quality and accuracy of the model. Table 4 presents the real and forecasted export sales along with the absolute error of forecast for each week, R^2 goodness of fit and MAE. Fig. 4 demonstrates the forecasted and real export sales behavior for the six-week period.

5.3. Variable sensitivity analysis

VSA specifies the way that multiple values of independent variables impact the target dependent variable. In other words, VSA evaluates how different sources of variation in a mathematical model (such as the present GP-based causal model) affect the model's output variation. VSA can be used for models with one or more independent variables. Here, the impact of five variables is analyzed, namely, marketing and packaging, production costs, exchange rate, discounts and unit prices. Table 5 presents the summary of VSA for the GP-based model (1).

Sensitivity analysis results indicate that production costs (c) have the strongest influence on export sales. For production costs, the positive magnitude is approximately 0.96, and the % positive indicates that 100 percent of the time, increases in this variable will lead to increases in sales. Production costs change with time due to different economic circumstances, inflation, supplier prices, etc. The positive impact of

costs on sales can be traced to several reasons. First, since the production costs are directly affected by the quality of the raw materials, the positive impact of costs can be linked to the higher quality of the final product. This applies to the impact of unit prices as well. Moreover, the direct relationship between sales and production quantity and costs is another factor that can justify the positive impact of costs on sales.

The second most sensitive variable is "m", representing packaging and marketing. Marketing and packaging costs are a small fraction of this company's budget, mainly because the company uses simple direct ways of retaining old customers and attracting new ones, and its advertising and digital marketing campaigns are very limited. Therefore, the main advertising efforts consist of direct connections and advertisements on product packages. For marketing and packaging, the value of % positive indicates that 66% of the time, increases in this variable will lead to increases in sales with a magnitude of 0.15, and 34% of the time, increases in this variable will lead to decreases in sales with a magnitude of 1.0.

For unit price (p) and exchange rate (x), again, analysis indicates a 100 percent positive, which means that an increase in these variables will always lead to an increase in sales. Usually, the increase in costs leads to a loss of sales. This happens for a great majority of businesses. For an exporter, however, it depends on various factors. The increase in exchange rates means that the power of foreign currencies increases in comparison to the domestic currency. Therefore, the foreign buyer is capable of buying more products with the same amount of money than before the exchange rate increased. This justifies the positive impact of exchange rates (x) on sales. Accordingly, the exporter can increase the prices to compensate for the effect of the increase in exchange rates. Thus, an increase in prices will also lead to an increase in sales.

VSA indicates that discounts have an ignorable impact on sales. Discounts (d) also did not appear in the GP-based model, which means that this variable does not have a significant effect on sales. Therefore, the export company can fundamentally cut the budget for discounts and improve other elements, such as marketing and packaging.

6. Conclusions

Sales forecasting is a major issue and an important task in supply chain and production management. This importance is the result of its influence on several levels of the organization. Inaccurate or lack of forecasting in a firm can lead to poor inventory and material flow management, loss of sales or excess of products and customer dissatisfaction. If such forecasting problems are coupled with other supply chain deficiencies, such as the bullwhip effect, the results could be catastrophic. While historical sales forecasting techniques prevail in research and in practice, causal methods are proven to be more accurate and precise. The difference between traditional time series and causal methods is highlighted when sales behavior changes in an unstable pattern, and there are apparently unpredictable fluctuations in its trend.

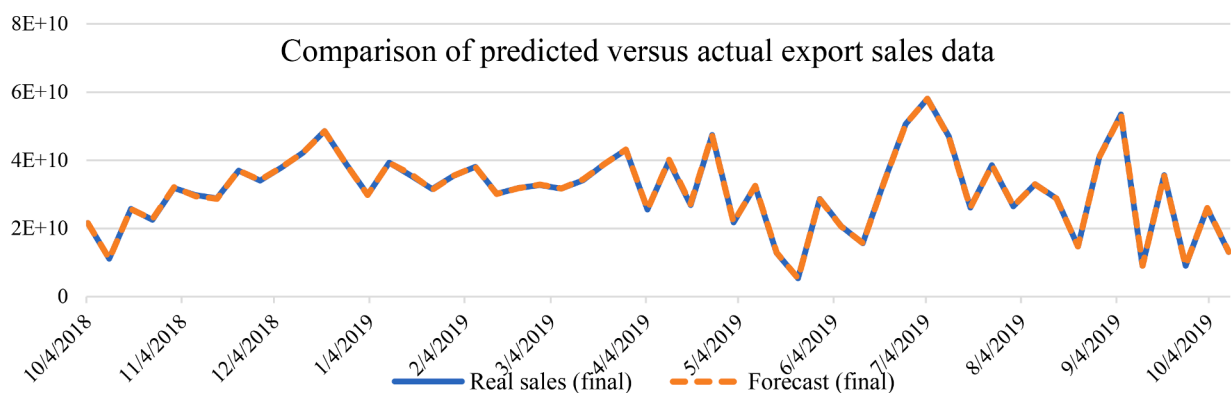
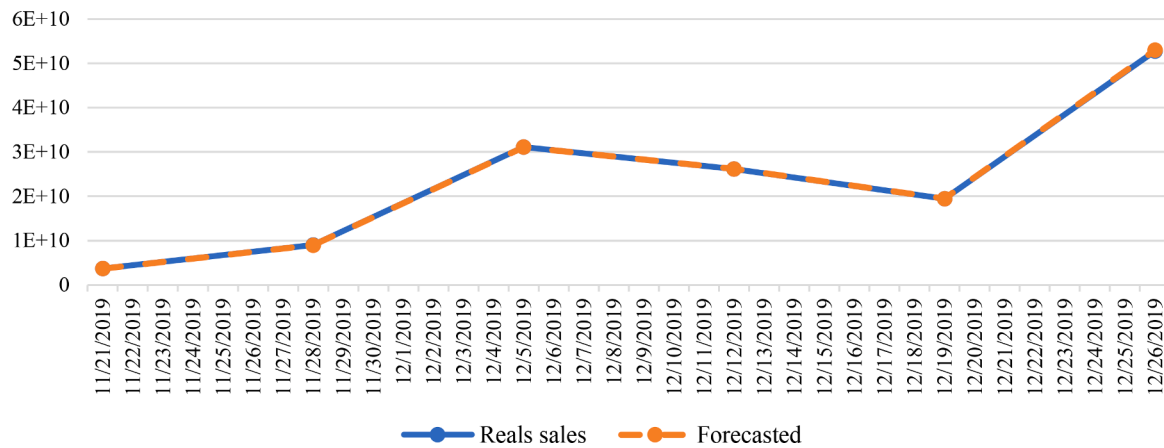


Fig.. 3. Comparison of predicted vs actual export sales data.

Table. 4

Comparison of real and forecasted export sales.

Date(the week end in)	Export sales forecast(y_f)	Real export sales(y_r)	Absolute error $ y_f - y_r $	MAE	R ² goodness of fit
12/26/2019	52,960,900,000	52,755,500,000	205,400,000	59,971,833.33	0.98%
12/19/2019	19,477,000,000	19,460,250,000	16,750,000		
12/12/2019	26,180,000,000	26,151,250,000	28,750,000		
12/05/2019	31,101,200,000	31,115,001,000	13,801,000		
11/28/2019	8,985,780,000	9,052,500,000	66,720,000		
11/21/2019	3,692,090,000	3,720,500,000	28,410,000		

Forecasted and real export sales behavior for the six-week period**Fig.. 4.** Forecast and real export sales behavior.**Table. 5**

Variable sensitivity analysis.

Variable	Sensitivity	% Positive	Positive magnitude	% Negative	Negative magnitude
c	0.9597800	100%	0.9597800	00%	0
m	0.1378100	66%	0.1543600	34%	0.10564
p	0.0081528	100%	0.0081528	00%	0
x	0.0039138	100%	0.0039138	00%	0
d	0	00%	0	00%	0

This happens in unstable economies or in markets in which economic factors affect sales behavior in a complex manner. To be able to forecast future sales in such circumstances, merely relying on historical time series techniques is not useful. Thus, causal forecasting methods capable of identifying the relationships between variables in a sales model are used to determine future sales behavior.

In this study, we employed GP to model the export sales for a Middle Eastern company that was facing fluctuations in sales and other relevant influential factors. The outcomes include the GP-based export sales model, a forecast for a six-week period and variable sensitivity analysis. To assess the predictive quality of the model, four error metrics were calculated: R-square goodness of fit, correlation coefficient, mean squared error, and mean absolute error. The results suggest a high accuracy and predictive precision for the model, which is rooted in its causal nature. Moreover, the predicted export sales for the six-week period also have a high R² goodness of fit, which uses a data set that has never been presented to the GP before. Variable sensitivity analysis was also conducted to evaluate the impact of each variable on the target variable (sales). While the analysis suggests that marketing, exchange rate, price, and costs have a positive magnitude, discounts have a zero magnitude, which means that an increase in this factor (discount) will not lead to any significant positive or negative change in sales.

A great majority of the studies on sales forecasting in production and

supply chain management are based on previously used methods (especially time series methods) and a series of predefined assumptions and formulas that affect the results of these studies at several levels. In some cases, these forecasting studies are not even practically applicable for use by organizations and commercial companies. In contrast, this research suggests a framework for export sales forecasting that has generated a model using GP and by analyzing empirical data from an export company that has experienced fluctuations in sales due to unstable market circumstances. Therefore, the contribution of this research is twofold. First, this study contributes to the literature by addressing the existing gap in causal export sales forecasting and employment of GP for this purpose. Moreover, this study is conducted based on real empirical sales data gathered from an active company in the Middle East, and the forecasting model is generated regardless of any predefined relationships and assumptions between variables. Second, this study contributes to practice by suggesting a novel GP-based framework for export sales forecasting that has high accuracy and is capable of clarifying the interrelationship between sales-related variables. Such a model can help firms reduce costs by providing clear insight into future sales and opportunities and prevent loss of sales. It also leads to accurate control and management of inventory, material flow, workforce and production costs.

The empirical data for this study were collected from an export company in the Middle East. As Fig. 3 shows, sales have severe fluctuations that are the result of unstable market conditions. However, the GP-based causal model is capable of forecasting sales with high accuracy (see Fig. 4 and Table 4). Accordingly, we conclude that this framework is also applicable to other export sales companies in other regions, especially those with more stable markets and economic conditions. Additionally, this method is also applicable to other data sets with more independent variables and a larger amount of data. This means that other organizations that have larger data sets and operate in other regions are also suitable for study using the presented framework.

The current study is primarily inspired by real market problems and

sales fluctuations under unstable economic circumstances, and the collected data were obtained from a company that is currently facing these problems. The collected data were sufficient for our research, but larger data sets comprising longer periods of sales data along with other influential variables are preferred for future studies.

7. Limitations and future study

Access issues to sufficient data have always been a major concern for researchers. For the present study, despite the vast investigations of the authors, only a limited source of data was obtained. These data had the basic requirements for the research, and the quality of the data was acceptable. However, to reach the full potential of Genetic Programming in causal sales forecasting, larger data sets with more relevant independent variables are preferred. Thus, future studies can employ more detailed data sets to achieve deeper insights and provide GP-based models for more complicated situations. Furthermore, other industries, including local retail sales, domestic industrial and governmental sales and online wholesales (such as that of Alibaba.com), can be targeted by GP-based causal sales forecasting methods. One of the limitations of this method is that it only works as a bespoke tool for modeling and forecasting sales based on the available data, meaning that if one organization employs the method, they have to provide sufficient and relevant data for each relevant variable. Therefore, if data for an influential variable are not available, the applicability and quality of the outcome models diminish. Finally, detailed comparisons of traditional sales and demand forecasting techniques (e.g., ARIMA, box-Jenkins, Holt-winter's, etc.) with AI-based methods (e.g., GP and GA-based methods, artificial neural network-based methods, etc.) can provide deep insight into the weaknesses and strengths of each of these methods. This can also lead researchers to devise hybrid forecasting methods that remove the limitations of both the GP and GA-based methods.

8. Author statement

We are excited to submit the manuscript, "Export Sales Forecasting Using Artificial Intelligence" to *Technological Forecasting & Social Change*.

We believe that our topic and empirical approach is a good fit with the aims of the journal. We draw on a data-set, impose a variety of specifications and tests to assess the validity of our inferences.

We thank you, in advance, for the opportunity to have our manuscript reviewed by *Technological Forecasting & Social Change*. We greatly look forward to receiving feedback; please let us know if there is anything else you need as we move forward.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.techfore.2020.120480](https://doi.org/10.1016/j.techfore.2020.120480).

References

- Abarbanell, J.S., Bernard, V.L., 1992. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *J. Finance* 47, 1181–1207.
- Abidi, S., Essafi, M., Guegan, C.G., Fakhri, M., Witt, H., Ghezala, H.H.B., 2019. A web service security governance approach based on dedicated micro-services. *Procedia Comput. Sci., Knowl.-Based Intell. Inform. Eng. Syst.: Proceed. 23rd Int. Confer. KES2019* 159, 372–386. <https://doi.org/10.1016/j.procs.2019.09.192>.
- Adetiloye, T., Awasthi, A., 2017. Chapter 8 - predicting short-term congested traffic flow on urban motorway networks. In: Samui, P., Sekhar, S., Balas, V.E (Eds.), *Handbook of Neural Computation*. Academic Press, pp. 145–165. <https://doi.org/10.1016/B978-0-12-811318-9.00008-9>.
- Ali, A., Klein, A., Rosenfeld, J., 1992. Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS. *Account. Rev* 183–198.
- Asian, S., Pool, J.K., Nazarpour, A., Tabaeian, R.A., 2019. On the importance of service performance and customer satisfaction in third-party logistics selection. *Benchmark. Int. J.*
- Baghestani, H., 1994. Evaluating multiperiod survey forecasts of real net exports. *Econ. Lett.* 44, 267–272.
- Baldauf, A., Cravens, D.W., Wagner, U., 2000. Examining determinants of export performance in small open economies. *J. World Bus.* 35, 61–79. [https://doi.org/10.1016/S1090-9516\(99\)00034-6](https://doi.org/10.1016/S1090-9516(99)00034-6).
- Banzhaf, W., 2001. Artificial Intelligence: genetic Programming. In: Smelser, N.J., Baltes, P.B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*. Oxford, Pergamon, pp. 789–792. <https://doi.org/10.1016/B0-08-043076-7/00557-X>.
- Bardool, R., Javanmardi, J., Roosta, A., Mohammadi, A.H., 2016. Phase stability conditions of clathrate hydrates for methane+aqueous solution of water soluble organic promoter system: modeling using a thermodynamic framework. *J. Mol. Liq.* 224, 1117–1123. <https://doi.org/10.1016/j.molliq.2016.09.084>.
- Bintang, A., Huang, W.-C., Asmara, R., 2019. Forecasting of Indonesia seaweed export: a comparison of fuzzy time series with and without markov chain. *Agric. Soc. Econ. J.* 19, 155–164. <https://doi.org/10.21776/ub.agrise.2019.019.3.4>.
- Boone, T., Ganesan, R., Jain, A., Sanders, N.R., 2019. Forecasting sales in the supply chain: consumer analytics in the big data era. *Int. J. Forecast., Spec. Sect.: Supply Chain Forecast.* 35, 170–180. <https://doi.org/10.1016/j.ijforecast.2018.09.003>.
- Çabuk, Y., 2019. Forecasting the export of wood panel industry in Turkey and determining the best method of forecast. *Bartın Orman Fakültesi Derg* 21, 426–431.
- Cardoso, F., Duarte, C., 2006. The use of qualitative information for forecasting exports. *Banco Port. Econ. Bull.* Winter 67–94.
- Ca'Zorzi, M., Schnatz, B., 2010. Explaining and forecasting Euro area exports which competitiveness indicator performs best? presented at the dimensions of competitiveness. *CESifo Sem. Ser.* September 121–148.
- Chase Jr, C.W., 1998. The role of the demand planner in supply chain management. *J. Bus. Forecast.* 17, 2.
- Chen, C., Luo, C., Jiang, Z., 2018. A multilevel block building algorithm for fast modeling generalized separable systems. *Expert Syst. Appl.* 109, 25–34. <https://doi.org/10.1016/j.eswa.2018.05.021>.
- Chopra, S., Meindl, P., 2016. *Supply Chain Management: Strategy, Planning, and Operation*. Pearson Education.
- Colin Cameron, A., Windmeijer, F.A.G., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* 77, 329–342. [https://doi.org/10.1016/S0304-4076\(96\)01818-0](https://doi.org/10.1016/S0304-4076(96)01818-0).
- Conrad, E., Misenar, S., Feldman, J., 2016. Chapter 9 - Domain 8: software Development Security (Understanding, Applying, and Enforcing Software Security). In: Conrad, E., Misenar, S., Feldman, J. (Eds.), *CISSP Study Guide (Third Edition)*. Syngress, Boston, pp. 429–477. <https://doi.org/10.1016/B978-0-12-802437-9.00009-6>.
- Cramer, N.L., 1985. A representation for the adaptive generation of simple sequential programs. *Present. Proceed. First Int. Confer. Genet. Algorithms* 183–187.
- Crittenden, V.L., Gardiner, L.R., Stam, A., 1993. Reducing conflict between marketing and manufacturing. *Ind. Mark. Manag.* 22, 299–309. [https://doi.org/10.1016/0019-8501\(93\)90026-4](https://doi.org/10.1016/0019-8501(93)90026-4).
- Darwin, C., 1859. *The origin of species; and, the descent of man*. Mod. Libr.
- Davis, D.F., Mentzer, J.T., 2007. Organizational factors in sales forecasting management. *Int. J. Forecast.* 23, 475–495. <https://doi.org/10.1016/j.ijforecast.2007.02.005>.
- De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. *Int. J. Forecast., Twenty five Years Forecast.* 22, 443–473. <https://doi.org/10.1016/j.ijforecast.2006.01.001>.
- Diamantopoulos, A., Winklhofer, H., 1999. The impact of firm and export characteristics on the accuracy of export sales forecasts: evidence from UK exporters. *Int. J. Forecast.* 15, 67–81. [https://doi.org/10.1016/S0169-2070\(98\)00069-7](https://doi.org/10.1016/S0169-2070(98)00069-7).
- Douglas, S.P., Craig, C.S., 1983. *International Marketing Research*. Prentice-Hall, Englewood Cliffs, NJ.
- Elstner, S., Grimme, C., Haskamp, U., 2013. Das ifo Exportklima—ein Frühindikator für die deutsche Exportprognose. *Ifo Schnell.* 66, 36–43.
- Fan, Z.-P., Che, Y.-J., Chen, Z.-Y., 2017. Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis. *J. Bus. Res.* 74, 90–100. <https://doi.org/10.1016/j.jbusres.2017.01.010>.
- Fildes, R., Ma, S., Kolassa, S., 2019. Retail forecasting: research and practice. *Int. J. Forecast.* <https://doi.org/10.1016/j.ijforecast.2019.06.004>.
- Fiorito, R., Kollintzas, T., 1994. Stylized facts of business cycles in the G7 from a real business cycles perspective. *Eur. Econ. Rev.* 38, 235–269.
- Golder, S., Loke, Y.K., Wright, K., Norman, G., 2016. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. *PLoS Med* 13, e1002127.
- Golder, S., Peryer, G., Loke, Y.K., 2019. Overview: comprehensive and carefully constructed strategies are required when conducting searches for adverse effects data. *J. Clin. Epidemiol.* 113, 36–43. <https://doi.org/10.1016/j.jclinepi.2019.05.019>.
- Govindan, K., Mina, H., Alavi, B., 2020. A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: a case study of coronavirus disease 2019 (COVID-19). *Transp. Res. Part E Logist. Transp. Rev.* 138, 101967. <https://doi.org/10.1016/j.tre.2020.101967>.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hyndman, R.J., Athanasopoulos, G., 2018. *Forecasting: principles and practice*. OTexts.
- Jackson, E.K., Roberts, W., Nelsen, B., Williams, G.P., Nelson, E.J., Ames, D.P., 2019. Introductory overview: error metrics for hydrologic modelling - a review of common practices and an open source library to facilitate use and adoption. *Environ. Model. Softw.* 119, 32–48. <https://doi.org/10.1016/j.envsoft.2019.05.001>.
- Jannsen, N., Richter, J., 2012. Foreign capacity utilisation and German exports of capital goods. *Wirtsch.-Hambg.* 92, 833.

- Jeong, B., Jung, H.-S., Park, N.-K., 2002. A computerized causal forecasting system using genetic algorithms in supply chain management. *J. Syst. Softw.* 60, 223–237. [https://doi.org/10.1016/S0164-1212\(01\)00094-2](https://doi.org/10.1016/S0164-1212(01)00094-2).
- Karmy, J.P., Maldonado, S., 2019. Hierarchical time series forecasting via Support Vector Regression in the European Travel Retail Industry. *Expert Syst. Appl.* 137, 59–73. <https://doi.org/10.1016/j.eswa.2019.06.060>.
- Kato, T., 2016. Chapter 4 - Prediction of photovoltaic power generation output and network operation. In: Funabashi, T. (Ed.), *Integration of Distributed Energy Resources in Power Systems*. Academic Press, pp. 77–108. <https://doi.org/10.1016/B978-0-12-803212-1.00004-0>.
- Keck, A., Raubold, A., Trupia, A., 2010. Forecasting international trade: a time series approach. *OECD J. J. Bus. Cycle Meas. Anal* 2009, 157–176.
- Kerckhansen, A., Korpela, J., Huiskonen, J., 2009. Demand forecasting errors in industrial context: measurement and impacts. *Int. J. Prod. Econ.* 118, 43–48. <https://doi.org/10.1016/j.ijpe.2008.08.008>.
- Koza, J.R., Koza, J.R., 1992. *Genetic programming: On the Programming of Computers By Means of Natural Selection*. MIT press.
- Kumar, M.S., Raut, D.R.D., Narwane, D.V.S., Narkhede, D.B.E., 2020. Applications of industry 4.0 to overcome the COVID-19 operational challenges. *Diabetes Metab. Syndr. Clin. Res. Rev.* 14, 1283–1289. <https://doi.org/10.1016/j.dsx.2020.07.010>.
- Lawrence, M., O'Connor, M., 2000. Sales forecasting updates: how good are they in practice? *Int. J. Forecast* 16, 369–382. [https://doi.org/10.1016/S0169-2070\(00\)00059-5](https://doi.org/10.1016/S0169-2070(00)00059-5).
- Lawrence, M., O'Connor, M., Edmundson, B., 2000. A field study of sales forecasting accuracy and processes. *Eur. J. Oper. Res.* 122, 151–160. [https://doi.org/10.1016/S0377-2217\(99\)00085-5](https://doi.org/10.1016/S0377-2217(99)00085-5).
- Lehmann, R., 2015. Survey-based Indicators vs. Hard data: What improves Export Forecasts in Europe? (No. 196). Ifo Working Paper Series, Ifo Working Paper Series. Ifo Institute - Leibniz Institute for Economic Research at the University of Munich.
- Liu, M.-Y., Huai, W.-X., Yang, Z.-H., Zeng, Y.-H., 2020. A genetic programming-based model for drag coefficient of emergent vegetation in open channel flows. *Adv. Water Resour.* 140, 103582. <https://doi.org/10.1016/j.advwatres.2020.103582>.
- Makridakis, S., 1996. Forecasting: its role and value for planning and strategy. *Int. J. Forecast.* 12, 513–537. [https://doi.org/10.1016/S0169-2070\(96\)00677-2](https://doi.org/10.1016/S0169-2070(96)00677-2).
- Makridakis, S.G., 1990. *Forecasting, Planning, and Strategies for the 21st Century*. Free Press, New York/London.
- Mendenhall, R.R., 1991. Evidence on the possible underweighting of earnings-related information. *J. Account. Res.* 29, 170–179.
- Mentzer, J.T., Bientock, C.C., 1998. *Sales Forecasting management: Understanding the techniques, Systems and Management of the Sales Forecasting Process*. SAGE Publications, Incorporated.
- Monks, J.G., 1987. *Operations Management: Theory and Problems*. McGraw-Hill.
- Mun, J., 2014. Chapter 3 - A Primer on Quantitative Risk Analysis. In: Glantz, M., Kissell, R. (Eds.), *Multi-Asset Risk Modeling*. Academic Press, San Diego, pp. 63–118. <https://doi.org/10.1016/B978-0-12-401690-3.00003-2>.
- Murray, F.T., Ringwood, J.V., 1994. Improvement of electricity consumption forecasts using temperature inputs. *Simul. Pract. Theory* 2, 121–139. [https://doi.org/10.1016/0928-4869\(94\)90008-6](https://doi.org/10.1016/0928-4869(94)90008-6).
- Nie, J., Oksol, A., 2018. Forecasting current-quarter US exports using satellite data. *Econ. Rev.-Fed. Reserv. Bank Kans. City* 103, 1–20.
- Okorie, I.E., Ohakwe, J., 2018. Forecasting Nigeria's inflation and the world prices of her major agricultural export commodities with probability distributions via VaR and ES and estimating their dependence via copula. *Commun. Stat. Case Stud. Data Anal. Appl.* 4, 28–45.
- Remus, W., Simkin, M., 1987. *Integrating forecasting and decision making*. Handb. Forecast. Manag. Guide 2nd Ed. Wiley, N. Y. USA.
- Rice, G., 1997. Forecasting in US firms: a role for TQM? *Int. J. Oper. Prod. Manag.* 17, 211–220.
- Ruican, C., Udrescu, M., Prodan, L., 2008. A Genetic Algorithm Framework Applied to Quantum Circuit Synthesis. pp. 419–429. [10.1007/978-3-540-78987-1_38](https://doi.org/10.1007/978-3-540-78987-1_38).
- Saltelli, A., 2002. Sensitivity analysis for importance assessment. *Risk Anal.* 22, 579–590.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity analysis: the Primer*. John Wiley & Sons.
- Sa-ngasongsong, A., Bukkapatnam, S.T.S., Kim, J., Iyer, P.S., Suresh, R.P., 2012. Multi-step sales forecasting in automotive industry based on structural relationship identification. *Int. J. Prod. Econ.* 140, 875–887. <https://doi.org/10.1016/j.ijpe.2012.07.009>.
- Santoro, M., Andres, V., Jirka, S., Koike, T., Looser, U., Nativi, S., Pappenberger, F., Schlummer, M., Strauch, A., Utech, M., Zsoter, E., 2018. Interoperability challenges in river discharge modelling: a cross domain application scenario. *Comput. Geosci.* 115, 66–74. <https://doi.org/10.1016/j.cageo.2018.03.008>.
- Schmidt, M., Lipson, H., 2013. *Eureqa (version 0.98 beta)[software]*. Nutionian Somerv. Mass.
- Schroll, J.B., Penninga, E.I., Göttsche, P.C., 2016. Assessment of adverse events in protocols, clinical study reports, and published papers of trials of orlistat: a document analysis. *PLOS Med* 13, e1002101. <https://doi.org/10.1371/journal.pmed.1002101>.
- Shaw, P.A., Johnson, L.L., Proschan, M.A., 2018. Chapter 27 - Intermediate Topics in Biostatistics. In: Gallin, J.I., Ognibene, F.P., Johnson, L.L. (Eds.), *Principles and Practice of Clinical Research (Fourth Edition)*. Academic Press, Boston, pp. 383–409. <https://doi.org/10.1016/B978-0-12-849905-4.00027-7>.
- Simske, S., 2019. Chapter 5 - Sensitivity analysis and big system engineering. In: Simske, S. (Ed.), *Chapter 5 - Sensitivity analysis and big system engineering*. Meta-Anal. Morgan Kaufmann 187–201. <https://doi.org/10.1016/B978-0-12-814623-1.00005-8>.
- Stein, J.C., 1997. Internal capital markets and the competition for corporate resources. *J. Finance* 52, 111–133.
- Strom, B.L., Buyse, M.E., Hughes, J., Knoppers, B.M., 2016. Data sharing—Is the juice worth the squeeze? *N. Engl. J. Med.* 375, 1608–1609.
- Tanaka, K., Akimoto, H., Inoue, M., 2012. Production risk management system with demand probability distribution. *Adv. Eng. Inform., Netw. Supply Chain Syst. Integr. Mass Custom. Sustain. Behav.* 26, 46–54. <https://doi.org/10.1016/j.aei.2011.07.002>.
- Trappey, C.V., Wu, H.-Y., 2008. An evaluation of the time-varying extended logistic, simple logistic, and Gompertz models for forecasting short product lifecycles. *Adv. Eng. Inform.* 22, 421–430.
- Vanneschi, L., Poli, R., 2012. Genetic Programming — Introduction, Applications, Theory and Open Issues. In: Rozenberg, G., Bäck, T., Kok, J.N. (Eds.), *Handbook of Natural Computing*. Springer, Berlin, Heidelberg, pp. 709–739. https://doi.org/10.1007/978-3-540-92910-9_24.
- Verstraete, G., Aghezzaf, E.-H., Desmet, B., 2020. A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Comput. Ind. Eng.* 139, 106169. <https://doi.org/10.1016/j.cie.2019.106169>.
- Wang, C.-C., Hsu, Y.-S., Liou, C.-H., 2011. A comparison of ARIMA forecasting and heuristic modelling. *Appl. Financ. Econ.* 21, 1095–1102.
- Wanto, A., Hayadi, B.H., Subekti, P., Sudrajat, D., Wikansari, R., Bhawika, G.W., Sumartono, E., Surya, S., 2019. Forecasting the export and import volume of crude oil, oil products and gas using ANN. *J. Phys. Conf. Ser.* 1255, 012016. <https://doi.org/10.1088/1742-6596/1255/1/012016>.
- Wieseler, B., Wolfram, N., McGauran, N., Kerekes, M.F., Vervölgyi, V., Kohlepp, P., Kamphuis, M., Grouven, U., 2013. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished clinical study reports with publicly available data. *PLoS Med* 10, e1001526.
- Williams, P.J., Moltano, T.C.A., 2008. A comparison of genetic programming with genetic algorithms for wire antenna design [WWW Document]. *Int. J. Antennas Propag.* <https://doi.org/10.1155/2008/197849>.
- Winklhofer, H., Diamantopoulos, A., 2003. A model of export sales forecasting behavior and performance: development and testing. *Int. J. Forecast.* 19, 271–285. [https://doi.org/10.1016/S0169-2070\(01\)00146-7](https://doi.org/10.1016/S0169-2070(01)00146-7).
- Winklhofer, H., Diamantopoulos, A., 2002. A comparison of export sales forecasting practices among UK firms. *Ind. Mark. Manag.* 31, 479–490. [https://doi.org/10.1016/S0019-8501\(01\)00206-1](https://doi.org/10.1016/S0019-8501(01)00206-1).
- Winklhofer, H., Diamantopoulos, A., 1996. First insights into export sales forecasting practice: a qualitative study. *Int. Mark. Rev.* 13, 52–81.
- Yang, J., Estefen, S.F., Fu, G., Paz, C.M., Lourenço, M.I., 2020. Collapse pressure of sandwich pipes with strain-hardening cementitious composite - Part 2: a suitable prediction equation. *Thin-Walled Struct.* 148, 106606. <https://doi.org/10.1016/j.tws.2020.106606>.

Dr. Vahid Sohrabpour has a multidisciplinary approach in research. He is passionate about the research on Industry 4.0 and the use of technologies such as Internet of Things (IoT), Cloud Computing, Block Chain, Big Data and Artificial Intelligence in Supply Chain Management. These technologies lead to establish Smart Supply Chain, Smart Manufacturing and Smart Product which are related to his background in Packaging Logistics, Supply Chain Management, Operations Management, Quality Management and Computer Engineering.

Dr. Pejvak Oghazi is Professor in Business Studies and head of Department at Sodertorn University, School of Social Sciences. He holds an MSc in Industrial and Management Engineering in addition to a PhD in Industrial Marketing from Lulea University of Technology in Sweden. Prior to his current position, Professor Oghazi worked as an industrial manager at national and international level. Professor Oghazi's current research interests revolve around topics in Digitalization, Business models, SCM and Marketing. His-research has been published in internationally renowned academic journals. His-work is also regularly presented at international academic conferences.

Reza Toorajipour is PhD candidate at Mälardalen University. After completing his master's program, he worked as a researcher for more than two years at Shiraz University Innovation Center where he conducted several studies on the employment of AI and Blockchain in business. His-current research interests include the application of artificial intelligence and blockchain technology in supply chain management, marketing and business; business model innovation, digitalization and circular business models.

Dr. Ali Nazarpour is Assistant Professor in Management at Maynooth University School of Business. Ali has been awarded PhD in Operations and Supply Chain Management from University College Dublin, Smurfit School of Business in 2016. He earned his BSc degree in Industrial Engineering and Master's degree in Business Administration from Iran University of Science and Technology. Prior to his PhD studies, he worked in the construction sector and in the automotive industry where he served as Sales Supervisor, Marketing and Sales Planning Chief, and Inventory Management Project Manager.