Gerunov, Anton

Article
Automating analytics

Journal of
Economics and Political Economy
www.kspjournals.org
Volume 3 June 2016 Issue 2

Automating Analytics: Forecasting Time Series in Economics and Business

By Anton Antonov GERUNOV [†]

Abstract. With the growing ability of organizations in the public and private sector to collect large volumes of real-time data, the mounting pile of information presents specific challenges for storage, processing, and analysis. Many organizations do need data analysis for the purposes of planning and logistics. Likewise, governments and regulators will need analysis to support policy-making, implementation and controlling. All this leads to the importance of being able to generate large-scale analytics under (sometimes severe) resource constraints. This paper investigates a possible solution – automating analytics with a special focus on forecasting time series. Such approach has the benefit of being able to produce scalable forecasting of thousands of variables with relatively high accuracy for a short period of time and few resources. We first review the literature on time series forecasting with a particular focus on the M, M-2, and M-3 forecasting competition and outline a few major conclusions supported across different empirical studies. The paper then proceeds to explore the typical structure of a time-series variables using Bulgarian GDP growth and show how the ARIMA modeling with a seasonal component can be used to fit economic data of this class. We also review some major approaches to automating forecasting and outline the benefits of selecting the optimal model from a large set of ARIMA alternatives using an information criterion. A possible approach to fit an automated forecasting algorithm on four crucial economic time series from the Bulgarian economy is demonstrated. We use data on GDP growth, inflation, unemployment, and interest rates and fit a large number of possible models. The best ones are selected by taking recourse to the Akaike Information Criterion. The optimal ARIMA models are studied and commented. Forecast accuracy metrics are presented and a few major conclusions and possible model applications are outlined. The paper concludes with directions for further research.
Keywords. Automated analytics, Forecasting, Time series, ARIMA, Business forecasting.
JEL. C22, C53, E37.

1. Introduction

The buzzword of big data seems to dominate the analytics landscape over the past years. It is indeed true that with the growing ability of organizations in the public and private sector to collect large volumes of real-time data, the mounting pile of information presents specific challenges for storage, processing, and analysis. Large organizations could have hundreds or even thousands of metrics tracked across their operations, logistics, finances, sales or services, and resource management. The ability of analytics to add value to operations has long since been recognized but the mounting challenge of analyzing ever increasing

† Sofia University "St. Kliment Ohridski, 125 Tsarigarsko Shosse Blvd. 1113, Sofia, Bulgaria.
☎.
✉. gerunov@uni-sofia.bg

volumes of data can sometimes have prohibitive costs. This is particularly true with sophisticated analytics such as forecasting where trained personnel may be scarce.

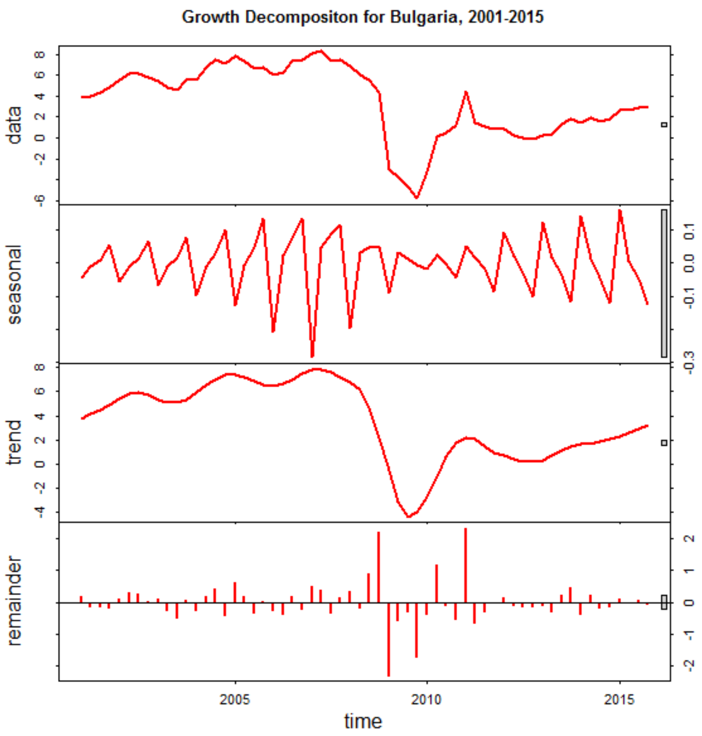
On the other hand many organizations do need data analysis for the purposes of planning and logistics. Likewise, governments and regulators will need analysis to support policy-making, implementation and controlling. All this leads to the importance of being able to generate large scale analytics under (sometimes severe) resource constraints. This paper investigates a possible solution – automating analytics with a special focus on forecasting time series. Such approach has the benefit of being able to produce scalable forecasting of thousands of variables with relatively high accuracy for a short period of time and few resources. We first review the literature on time series forecasting and then proceed to fit an automated algorithm on four crucial economic time series – GDP, inflation, unemployment, and interest rates. A few conclusion and directions for further research are then outlined.

2. Literature Review and Motivation

Due to its importance for planning, forecasting has received significant attention in the statistical, economic, and business literature. May approaches have proposed ranging from relatively simple extrapolative methods through more complicated autoregressive and moving averages methods to very sophisticated machine learning algorithms such as neural networks or random forest models (Hyndman & Athanasopoulos, 2014; Friedman et al., 2001). This rich variety begs the question of which the optimal forecasting model is, and this has been formally tested many times over. One of the first major undertakings in this direction is work by Makridakis et al. (1979) in the late nineteen-seventies which seeks to compare the performance of different forecasting methods on 111 datasets (the M-competition).

This work was continued and later expanded in subsequent research. In the early nineteen-eighties Makrdidakis et al. (1982) presented results (M-2 Competition) comparing 1001 time series forecasted with a number of different methods. Finally, in 2000, Makridakis et al. (2000) published a comparison of the same methods on a sample of 3003 different datasets (the M-3 Competition). The major findings from this study, as well as from other empirical work is that while the most sophisticated methods do not necessarily produce the best forecasting performance, there is still much improvement over naïve estimates to be gleaned from applying formal models. These competitions also show that time series models of the ARIMA class tend to have high accuracy. This is particularly true when it comes to macroeconomic or financial data. In this paper we aim to show precisely how ARIMA models can be automated to produce high quality forecasts at a low price.

To better understand how this class of models works, we note that any given times series is composed of a number of components. Most notably, non-random time series tend to have a trend, and fluctuate around it. They could also have cyclical, seasonal, and random components. Statistical methods can be used to decompose the time series into their constituent parts. Figure 1 displays the decomposition of the Bulgarian GDP growth over the period 2001-2015.



Graph 1. *Decomposition of Growth Time Series for Bulgaria,*
Source: Eurostat

It is the task of the ARIMA models to try and capture the information, contained in the time series and model current variable realization as a function of past ones. In the simplest version of the model, the current realization of a given metric y_t is presented as a weighted function of p previous values y_{t-p} (is an error term). This AR(p) model is defined as follows:

$$y_t = \theta + \sum_{i=1}^{p} \beta_i y_i + \varepsilon_t \tag{1}$$

Additional information can be contained in the error structure of the time series. This can be modeled through a moving average of the error term. Should the analyst use q past values of the error terms to model current variable realization, then we reach a MA(q) of the following form:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \alpha_i \varepsilon_i \tag{2}$$

Combining those two equations one gets a more fuller perspective on the time series, thus reaching the classical ARMA(p, q) model:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_i + \sum_{i=1}^{q} \alpha_i \varepsilon_i + \varepsilon_t \tag{3}$$

Should the time series be integrated of a certain order d, this can also be taken into account, finally reaching the ARIMA(p, d, q) model. A particular strength of this class of models is their ability to accommodate seasonality in the time series, and its mirrors the structure of the deseasoned model. A model with seasonality is thus denoted as ARIMA(p, d, q)(P, D, Q) to account to the autoregressive, integrated, and moving average parts in the seasonal component of the data. Interested readers are directed to Hamilton's (1994) work for further details.

The availability of versatile tools for time series analysis has also spurred the interest in automatic forecasting. Early work on this topic began in the later 1990s, and was further spurred after the M-3 competition. Research by Melard & Pasteels (2000) showed a basic software implementation of automatic ARIMA forecasting which calculates eight models and selects the best one based on lowest residual autocorrelation. While this approach is rather limited and model selection needs significant improvement, this paper showed the possibility of automating time series analytics. Achieving better precision in identifying time series peculiarities was clearly needed and subject to subsequent research such as in the work of Adya et al. (2001). Alternative strands of research have focused on alternative large-scale forecasting (Chakrabarti &Faloutsos, 2002) but they have had only limited impact on theory and practice.

Hyndman & Khandakar (2008) present an automatic forecasting facility implemented in the R statistical language. It is able to handle both exponential smoothing and ARIMA methods, providing a large set of possibilities for the analyst. This is also notable for the customizable selection of models, based on information criteria. Such selection is preferable to the expert or heuristic selection implemented in other settings as it gives an unambiguous preference for a single model among a large set of potentially useful ones. This approach allows for fully automatic forecasting once the analyst or the solution architect has made a few key design decisions. We outline those decisions, demonstrate parameter choices and apply the algorithm to Bulgarian macroeconomic data.

3. Automated Forecasting and Application

Four macroeconomic variables present particular interest for business and policy analytics. Those are the rate of change of GDP (real growth), which proxies disposable income and economic development; the rate of unemployment as an indicator of labor availability and labor costs; interest rates as the cost of capital; and inflation which gives an indication of price dynamics and proxies economic stability. Their accurate and timely forecasting can have crucial implications for planning and intervention at both the firm and the state level. We use those series as examples of how to construct a meaningful architecture for automated analytics.

Data itself are obtained from the Eurostat statistical service, the Bulgarian Ministry of Finance (MF), and the Bulgarian National Bank (BNB) and span the period 2001-2015. GDP growth series are at a quarterly frequency, whereas the unemployment rate, inflation (average HICP of 12 month period), and interest rates on short term business credits (less than 1 year) are at a monthly frequency. Their dynamics can be traced on Figure 1. Up until the crisis which started in 2009, the Bulgarian economy grows robustly, with a decrease in both the rate of unemployment, as well as the interest rates. This period is also accompanied by a burgeoning inflation, which sometimes goes in the double digits. Such developments over this period spurred concern about the overheating of the Bulgarian economy.

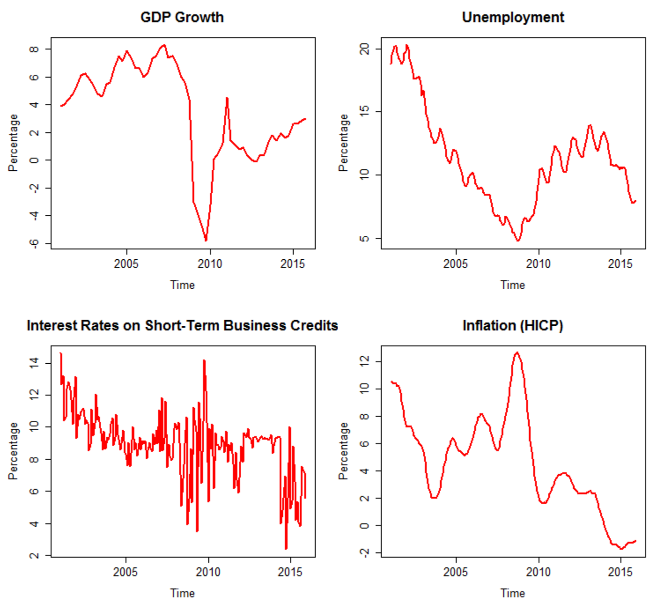
These dynamics changed drastically with the onset of the global economic and financial crisis. In 2009-2010 Bulgaria saw a collapse in growth, and a marked rise in unemployment. The perception of heightened risk and the economic uncertainty also lead to an increase in the volatility of short-term interest rates. In parallel inflationary pressures receded. By the end of the period under study in 2014-2015 the economy mostly recovered and growth has picked up. This led to a decrease in unemployment and a slight decrease in interest for business. Their volatility remains significant, and is also coupled with deflationary pressure.

In short, we observe wild fluctuations of the time series, a regime switch, and pockets of volatility in some of them (interest rates). Unsupervised modeling of such time series presents a particular challenge but also significant opportunity. We use the framework of ARIMA modelling to address this. Initially, we consider the maximum number of lags that would be useful for the series. Since GDP data comes at quarterly frequency, and given its inertia, it would be useful to fit models of up to 8 lags of both the autoregressive and the moving average component (p = q = 8).

The seasonal component can have a much shorter lag structure as it operates across a number of time periods. We select a number of 4 lags for the AR and MA seasonal components. While we do not expect the series to be integrated of order of more than 1, we still provide a buffer by setting a maximum number of possible differences at d = 3, and D = 2, and let the algorithm decide based on a KPSS and OCSB tests. For the growth series these would be maximum number of lags and differences. We thus fit models ranging from ARIMA(0,0,0)(0,0,0) to ARIMA(8,3,8)(4,2,4), or a total of 19,440 different ARIMA models to the growth time series and choose the best among them.

When it comes to the other three time series, we should note that their frequency is much higher and will therefore need a large number of lags to capture effects that operate periodically across time. Thus the AR and MA lags are set at 12 (p = q =12), and the seasonal lags are set at 2. Due to the higher frequency we would expect lower orders of integration and thus set the number of possible differences at d = 2, and D = 1. For the unemployment, interest rates, and inflation time series we thus fit models ranging from ARIMA(0,0,0)(0,0,0) to ARIMA(12,2,12)(2,1,2). For each of those variable we so obtain 9,126 alternative models to pick from.

In order to correctly measure forecast performance, the dataset is split in two sub-samples. The first one ranges from 2001 to the end of 2014 and is used for model training. The data from 2015 comprises the test data set against which we measure the out-of-sample accuracy of the produced forecasts. We thus confront model results will actual realizations upon which the model was not trained in order to gauge real-life model quality.



Graph 2. *Dynamics of Key Macroeconomic Time Series,*
Source: Eurostat, BNB.

Initially we fit all specified alternative models to the training data set. The problem is now straightforward – to select the four best models out of a total of 46,818 alternatives. For this purpose we can use a number of information criteria. Three criteria have become particularly popular in practice – the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), and the corrected Akaike Information Criterion (AICc) (Burnham & Anderson, 2003). To better understand model fit, we define the likelihood function L, equal to the probability p of observing the data x given a model M with a parameter set of θ, or:

$$L = p(x|\theta, M) \tag{4}$$

If we denote the maximized value of this likelihood function as L_{max}, then the BIC of a model with k parameters and a sample size of n is defined as follows:

$$BIC = -2\ln L_{max} + k\ln n \tag{5}$$

Essentially, the information criterion is a measure of model quality, which represents informational loss as data is presented by a given model. Thus it can serve to select the best model among a set of alternatives taking into account the tradeoff between fit and parsimony (or number of parameters). For a given dataset better models have lower values of their information criteria. The BIC is often criticized on the grounds of its difficulty of handling complex collections of model or feature selection, and it is only valid as $n >> k$. This, together with some derivation considerations and performance issues lead many authors to propose using the Akaike Information Criterion instead (Burnham & Anderson, 2004). It is defined as follows:

$$AIC = 2k - 2\ln L_{max} \tag{6}$$

The AIC estimate is valid asymptotically, which means that some corrections needs to be made for finite sample sizes, leading to the corrected version of AIC, or AICc. The formula for univariate series with normally distributed residuals is as follows:

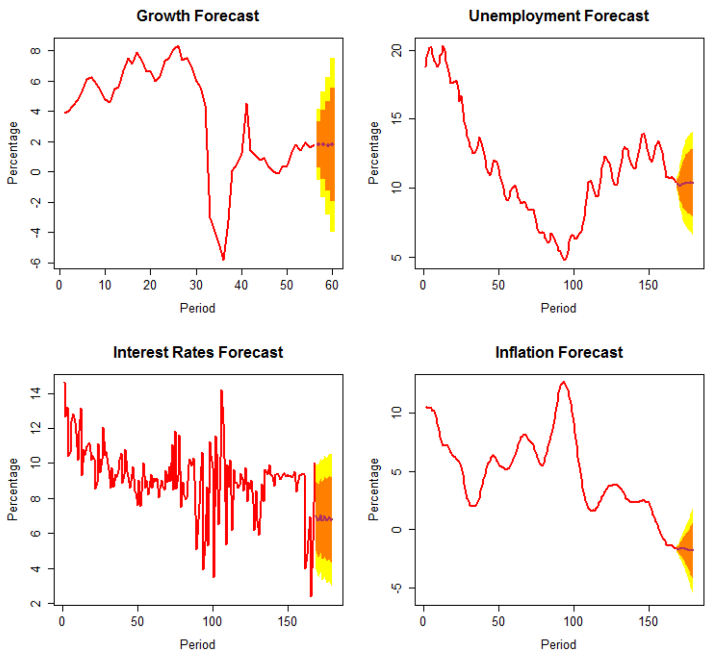
$$AICc = AIC + 2k(k+1)/(n-k-1) \tag{7}$$

The AICc penalizes more heavily models with more parameters than AIC and will thus lead to the selection of more parsimonious ones. In addition to that we should keep in mind that as the sample size grows AICc converges to AIC and this is why many authors recommend it as the primary criterion to use for model selection exercises (Burnham & Anderson, 2003). We follow the literature and use AICc to select the optimal models for our time series but also report the values for alternative ones. From an empirical perspective, the differences between the AIC and AICc for the series under study are very small and any of the criteria will lead to the selection of the same optimal model.

Table 1. *Optimal Models for Specific Time Series*

	Growth	Unemployment	Interest rates	Inflation
Optimal Model	ARIMA(5,1,0)	ARIMA(4,1,3)	ARIMA(3,1,3)	ARIMA(7,1,3)
Akaike IC	189.19	30.88	633.64	-175.99
Akaike IC, corrected	190.94	31.79	634.35	-174.29
Bayesian IC	201.24	55.82	655.47	-141.69

Results for the optimal model for each of the variables are presented in Table 1. The models tend be rather parsimonious and have between 3-7 lags of the variable. All of the series are best fitted by models of their first differences, which is likely due to the clear trends they exhibit. In all cases but growth additional information can be gleaned by taking recourse to the structure of the error terms.



Graph 3. *Automated One-year Ahead Forecasts of Macroeconomic Variables*

We use the optimal models to automatically generate forecasts for one year ahead. In the case of GDP growth this means generating the four-period-ahead forecast, and for the other variable – the 12-period one. Those forecasts are presented in Figure 3 where we clearly see the increasing confidence intervals (shaded areas) as the estimate moves away from observed data, and the typical smoothed form of data. We formally test the generated forecasts against actual realized values and report the accuracy metrics in Table 2. The table contains statistics on the mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE).

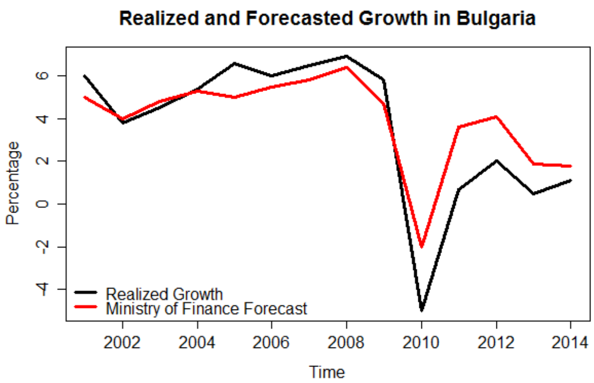
Given the wide fluctuations of data over the period studied, results are rather encouraging. The average MAE of all time series stands at 0.49 for the training set, and at 1.06 in the test set. This means that the automated forecasting procedure produces estimates than are on average off by 1 percentage point from the actual realizations. This varies widely across the variables with interest rates having the largest mean absolute error, and also the largest mean absolute percentage error. Unemployment and inflation are easier to forecast and their MAEs stand at 1.38 and 0.32, respectively. This deviation is relatively small (especially in the inflation case) and the estimates can have very large potential practical use.

Table 2. *Metrics for Forecast Accuracy of Optimal Models*

	ME	RMSE	MAE	MPE	MAPE	MASE
Growth						
Training set	-0.05	1.19	0.64	ND	ND	0.78
Test set	1.01	1.02	1.01	35.91	35.91	1.23
Unemployment						
Training set	-0.03	0.25	0.18	-0.17	1.67	0.53
Test set	-1.17	1.70	1.38	-14.72	16.70	4.12
Interest Rates						
Training set	-0.19	1.53	1.05	-5.92	14.20	0.80
Test set	-0.94	1.80	1.55	-23.97	31.50	1.18
Inflation						
Training set	-0.01	0.13	0.10	-0.14	3.21	0.37
Test set	0.31	0.39	0.32	-25.08	25.66	1.15
Average						
Training set	-0.07	0.77	0.49	-2.08	6.36	0.62
Test set	-0.20	1.23	1.06	-6.97	27.44	1.92

Growth time series are also relatively well fit by the automated model – the MAE in the training set stands at 0.64, and at 1.01 in the test set. The analyst would thus expect such the automated forecast to be an average of 1 percentage point off the mark. Given the complexity of growth determinants, and the large influence of different exogenous shocks, such results are very good. In particular, we should note the large amount of EU-funds invested in the Bulgarian economy as the extension of the 2007-2013 programming period came to an end. The rapid surge of public investment had a sizeable effect on growth, and other key macro aggregates. It is notable that automatic forecasting realized such low error values.

This exercise can be taken as indicative of the potential for automating analytics, and especially forecasting. Accuracy metrics show that such automated models perform well under challenging realistic situations. It is likely that as we move from macroeconomic to microeconomic variables such as sales, which are affected by less exogenous shocks and follow a clearer trend, the forecast accuracy will be even higher. Another point to consider is that those results should be viewed against the benchmark of alternative forecasts, and not against the benchmark of actual realizations. A possible approach is to compare a key forecast used for the purposes of public or business policy to the automated ones. The availability of growth forecasts, done by the Bulgarian Ministry of Finance (MF) as part of the budgeting process allows for that. The MF projections are presented in Figure 4 and compared to actual economic outcomes. While the overall forecast dynamics nicely tracks realizations, there are some significant and relatively consistent deviations, especially in the post-crisis period.



Graph 4. *Ministry of Finance Growth Forecast against Actual Realizations over the Period 2001-2014*

We now proceed to compare the MF projections with the ones, generated by the optimal Auto ARIMA models. Relevant accuracy metrics for the two sets of forecasts are presented in Table 3. All the indicators of accuracy support the conclusion that automated forecasting performs better in terms of accuracy than the estimates, used by the Ministry of Finance in preparation of the state budget. The root mean squared error of the automated forecast stands at 1.02, while the Ministry's forecast has a long-run RMSE of 1.5, which jumps to 2.3 in the crisis and post-crisis period. The mean absolute error of our forecast stands at the modest 1.01, while the Ministry's long-run MAE is at 1.15, jumping to 1.88 in the period 2009-2014. The auto-ARIMA forecasts consistently outperform the long run budget projections on the other accuracy metrics as well. This serves to prove that automated analytics can serve as a viable alternative to more resource-intensive approaches.

Table 3. *Accuracy Comparison between Ministry of Finance and Automatic Forecast for Growth, source: Ministry of Finance and own calculations*

	MF, 2001-2014	MF, 2009-2014	Auto Forecast
Mean Error	-0.4	-1.5	1.01
Mean Squared Error	2.2	5.2	1.04
Root Mean Squared Error	1.5	2.3	1.02
Mean Absolute Error	1.15	1.88	1.01
Mean Percentage Error	-55.59	-138.78	35.91
Mean Absolute Percentage Error	76.34	164.92	35.91

4. Directions for Future Research

Automated analytics present bountiful opportunities in a world with growing data availability and increasing appetite for analysis and insight that sometimes has to operate under significant resource constraints. This paper explores the possibility of fitting a large set of ARIMA models and selecting the most optimal one for the express purpose of generating forecasts. The case study on four macroeconomic time series shows that these unsupervised models achieve excellent forecast accuracy and can even outperform mainstream expert-made projections.

This naturally presents a number of venues for further research. A particular fruitful way forward would be to investigate the performance of alternative algorithms that lend themselves easily to automation. The family of exponential moving average methods or machine learning algorithms such as Random Forests seem to be possible candidates. It would be of natural interest to investigate relative forecast performance depending on the type of the time series in terms of frequency and domain. Finally, the idea of automated analytics can be expanded to other problems of business interest.

In short, automating analytics holds the promise of mainstreaming typical tasks and freeing and empowering people to focus on the complicated problems of strategizing, planning and navigating today's complex economic and business environment, searching for prospects to unlock more value.

References

Adya, M., Collopy, F., Armstrong, J.S., & Kennedy, M. (2001). Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting*, *17*(2), 143-157. doi. 10.1016/S0169-2070(01)00079-6

Burnham, K.P., & Anderson, D.R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media. doi. 10.1007/b97636.

Burnham, K.P., & Anderson, D.R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304. doi. 10.1177/0049124104268644

Chakrabarti, D., & Faloutsos, C. (2002, November). F4: large-scale automated forecasting using fractals. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 2-9). ACM. doi. 10.1145/584792.584797

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics. doi. 10.1007/978-0-387-84858-7

Hamilton, J.D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton university press.

Hyndman, R.J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts. doi. 10.1111/j.1467-856X.2005.00178.x

Hyndman, R.J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software 27*(3) , 1-22.doi. 10.18637/jss.v027.i03

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International journal of forecasting*, *16*(4), 451-476. doi. 10.1016/S0169-2070(00)00057-1

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111-153. doi. 10.1002/for.3980010202

Makridakis, S., Hibon, M., & Moser, C. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society. Series A (General)*, 97-145. doi. 10.2307/2345077

Melard, G., & Pasteels, J.M. (2000). Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting*, *16*(4), 497-508. doi. DOI: 10.1016/S0169-2070(00)00067-4