STATISTICS IN TRANSITION new series, December 2025
Vol. 26, No. 4, pp. 69–87, https://doi.org/10.59139/stattrans-2025-040
Received – 13.07.2023; accepted – 21.03.2025

Application of the nonlinear splines model to forecast changes in the construction costs index

Mariusz Kubus[1], Łukasz Mach[2], Przemysław Misiurski[3]

Abstract

The research presents the application of the nonlinear splines model to forecast the construction costs index (CCI), which is an important macroeconomic indicator. Due to the long-term nature of the investments in the construction market, we tested our model in a ten-month ahead period. Except minor disruptions, which were likely related to COVID-19, we obtained promising results, which definitely outperformed the classical ARIMA and its variant with nonlinear autocorrelation functions modeled with neural network. The achieved forecast results will enable both the demand and supply in the construction market to be in market equilibrium and minimize the formation of speculative bubbles in the market.

Key words: construction market, nonlinear splines model, forecast, non-stationary time series.

1. Introduction

The market economy of each developed country is based on three factors of production – labor, land and capital. Each of these factors should be developed evenly so that sustainable economic development can occur. A comprehensive study of a country's economic development can be carried out by monitoring an important sector for the economy – the construction sector. By its very nature, this sector reflects the condition and state of development of the three aforementioned factors of production. It is natural that in considering investment projects in the area of the construction sector we must consider the stock of land, labor and capital. The integration of production factors in official statistics is expressed through the construction costs index, since it is

[1] Centre of Education and Mathematics Applications, Opole University of Technology, Opole, Poland. E-mail: M.Kubus@po.edu.pl. ORCID: https://orcid.org/0000-0002-6602-2742.

[2] Institute of Economics and Finance, Faculty of Economics, University of Opole, Opole, Poland. E-mail: L.Mach@uni.opole.pl. ORCID: https://orcid.org/0000-0002-8200-4261.

[3] Department of Enterprise Management, E-business and Electronic Economy, Opole University of Technology, Opole, Poland. E-mail: P.Misiurski@po.edu.pl. ORCID: https://orcid.org/0000-0002-7052-8535.

this index that will reflect directly the capital and indirectly the labor and land required to use this capital to implement the project. In order to plan the investment projects successfully, it is necessary to use data analysis methods capable of forecasting this index. Therefore, the research carried out in this article has tested and demonstrated the usefulness of the nonlinear splines method as a utilitarian forecasting tool, both for stationary and non-stationary time series.

The literature on the construction market and its importance for the economy is not as rich as it might seem. It is dominated mainly by country-specific reports and analyses. On the other hand, Asian researchers seem to dominate theoretical considerations, which seems justified due to the intensive development of construction in these countries. Three-level definition of construction as an economic activity was presented by researchers from Singapore (Pheng & Hou 2019). The issue of the role of construction in the economy was described by Jorge Lopes in the chapter: *Construction in the economy and its role in socio-economic development: role of construction in socio-economic development* (Lopes 2011). An attempt to define the scope of the construction sector was made by Andrew Foulkes and Les Ruddock of the British University of Salford. The latter edited a book on the economic aspects of modern construction (Ruddock & Ruddock 2008). There have also been studies on the importance of the construction market in the economy of specific countries. One can mention here, for example, the publication of scientists from Malaysia regarding Turkmenistan: *Role of the construction industry in economic development of Turkmenistan* (Durdyey & Ismail 2012) or the publication on the Polish economy: *The view of construction companies' managers on the impact of economic, environmental and legal policies on investment process management* (Sobieraj *et al.* 2021). On the other hand, a broader view of the construction market was presented by Chinese researchers: Ye, Lu and Jiang in the article *Concentration in the international construction market* (Ye *et al.* 2009).

A literature analysis of the construction market showed that the topic is important and interesting, but there are not many research works that comprehensively deal with construction markets. This shows that research in this area is needed.

2. Forecasting Construction Cost Index: A Literature Review

Accurate forecasting of the Construction Cost Index (CCI) is essential for effective project planning, cost control, and risk management in the construction industry. Over the past five decades, various forecasting methodologies have been explored, evolving from traditional econometric models to advanced machine learning techniques. This section reviews the key studies on CCI forecasting, with a particular focus on applied methodologies and their effectiveness.

Early research in this field primarily relied on econometric and regression-based models to analyze historical cost trends. One of the pioneering studies was conducted by Williams (1994), who applied neural networks to predict changes in construction cost indices. Wang and Mei (1998) further explored forecasting models for Taiwan's CCI using time series techniques. In subsequent years, autoregressive (AR) and autoregressive integrated moving average (ARIMA) models gained prominence due to their ability to model cost index trends based on past values (Xu & Moon, 2013; Moon, Chi & Kim, 2018). A notable advancement came from Moon and Shin (2018), who introduced the vector error correction model (VECM), capturing long-term equilibrium relationships among cost-related variables. However, these statistical models were constrained by their linear nature, limiting their ability to capture complex, nonlinear patterns in cost fluctuations.

With the advancement of computational techniques, machine learning-based models have emerged as effective tools for CCI forecasting. Elfahham (2019) demonstrated that artificial neural networks (ANNs) could outperform traditional regression and time series methods. Wang and Ashuri (2017) explored machine learning algorithms to enhance CCI predictions, highlighting their adaptability to dynamic construction markets. To improve accuracy, researchers have increasingly integrated classical time series models with machine learning techniques. For example, Kim et al. (2022) combined ARIMA and ANN models to enhance prediction precision. Similarly, Cao and Ashuri (2020) applied a long short-term memory (LSTM) network to predict the volatility of highway construction costs, outperforming traditional forecasting models.

Recent studies have also explored stochastic and network-based forecasting techniques to address the dynamic nature of construction costs. Xu and Moon (2013) utilized a cointegrated vector autoregression (VAR) model for stochastic CCI forecasting, while Joukar and Nahmens (2016) employed a generalized autoregressive conditional heteroskedasticity (GARCH) model to analyze cost index volatility patterns. Additionally, network-based approaches have gained traction as an alternative forecasting method. Zhang et al. (2018) introduced a visibility graph-based forecasting model, while Mao and Xiao (2019) developed a complex network-based approach to improve predictive performance.

The latest research in CCI forecasting has increasingly focused on hybrid methodologies and deep learning techniques. Al Kailani et al. (2024) utilized fuzzy logic and machine learning to enhance prediction accuracy in Jordan's construction sector. Similarly, Altalhoni, Liu, and Abudayyeh (2024) conducted a comprehensive review of existing forecasting techniques, identifying key influential factors and emerging trends.

These studies indicate that artificial intelligence and nonlinear models will continue to play a pivotal role in improving predictive accuracy.

Given the existing literature, it is evident that forecasting methodologies have evolved significantly from traditional statistical models to advanced AI-driven approaches. However, nonlinear modeling techniques, such as splines, remain underexplored in CCI prediction. This study aims to fill this research gap by introducing a nonlinear splines-based model to forecast changes in the construction cost index. By leveraging the flexibility of splines in capturing complex, nonlinear relationships, this approach seeks to enhance forecasting accuracy and provide valuable insights for cost management in the construction industry. This literature review highlights the importance of innovative forecasting methods, reinforcing the relevance of the proposed nonlinear splines model in the broader research landscape.

3. The use of spline methods in research – examples of application

In the field of economics and management, various methods of data analysis are used in the decision-making process. These methods are useful for creating models on the basis of which events can be simulated with the possibility of creating different scenarios of the event and are helpful in the decision-making process. In economics, methods included in the widely understood econometric modelling, as well as the use of artificial intelligence tools, are generally used for this purpose. Commonly used methods include, for example, correlation, regression, econometric modelling, MA, VAR, or ARIMA methods, cf. (Findley et al. 2016; Nieto & Carmona-Benítez 2018; Osiewalski & Osiewalski 2013; Paci & Consonni 2020). In the group of artificial intelligence tools, methods based on neural networks or genetic algorithms are commonly used, cf. (Butler et al. 2021; Cheung et al. 2006; Lin et al. 2021). In the area of modelling economic phenomena, mathematical or simulation methods reserved for technical sciences, i.e. wavelet transform or analysis of the coherence of the wavelet, are relatively rarely used, cf. (Dash & Maitra 2019; Naccache 2011).

In the literature, it is also possible to find few studies showing the usefulness of such a tool as a spline. Studies using the methods of spline can be found in the work of Humphrey & Vale (2004), which demonstrates the usefulness of an elastic cost function when analyzing economies of scale and estimating the cost-effectiveness of bank mergers. The inflexibility of the translogarithmic cost function was pointed out, and the results obtained were compared with more flexible cost functions of the folded curve and Fourier curve types. Using these different approaches, an ex-ante effect on the costs resulting from mergers between 1987 and 1998 was projected using a panel study on a sample of 130 Norwegian banks. Mergers are predicted to result in an average reduction in costs. Predictions using the Fourier approach or compound curves are consistent with calculated actual changes in ex-post-merger costs. Other studies, described in (Monteiro et al. 2008), present a new approach to estimating the risk-neutral probability density function of the future prices of an underlying asset from the prices of options written on the asset. The estimation is carried out in the space of cubic spline functions, yielding appropriate smoothness. The resulting optimization problem, used to invert the data and determine the corresponding density function, is a convex quadratic or semi-definite programming problem, depending on the formulation. Both of these problems can be efficiently solved by numerical optimization software. Monteiro et al. (2008) tested their approach using data simulated from Black–Scholes option prices and using market data for options on the S&P 500 Index. Their results show the effectiveness of the proposed methodology for estimating the risk-neutral probability density function. The application of spline-based phase analysis to macroeconomic dynamics can be found in the work of Gadasina & Vyunenko (2022), in which they use spline-based phase analysis to study the dynamics of a time series of low-frequency data on the values of a certain economic indicator. The approach includes two stages. In the first stage, the original series is approximated by a smooth twice-differentiable function, which is obtained from natural cubic splines. Such splines have the smallest curvature over the observation interval compared to other possible functions that satisfy the choice criterion. In the second stage, a phase trajectory is constructed in the space, which corresponds to the original time series, and a phase shadow as a projection of the phase trajectory onto the plane. The approach is applied to the values of GDP indicators for the G7 countries. The interrelation between phase shadow loops and cycles of economic indicators evolution is shown. The study also discusses the features, limitations, and prospects for the use of spline-based phase analysis (Gadasina & Vyunenko 2022).

On the other hand, the effectiveness of steering processes of complex socio-economic systems using the method of spline analysis has been demonstrated by Ilyasov & Yakovenko (2021), and the utility of tensor spline approximation in economic dynamics with uncertainties was presented in the paper (Chu et al. 2011). Of course, one can also encounter applications of compound curves in fields other than economic science. Examples include studies describing spline functions as an alternative to estimating income-expenditure relationships for beef (Huang & Raunikar 1981), describing evaluating the water sector in Italy through a two-stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines (Vidoli 2011) or forecasting energy demand in transportation by using the multivariate adaptive regression splines (Sahraei *et al.* 2021).

4. Research methodology

4.1. Data and general outline of the research

The data used for the study comes from Eurostat[4] databases and include monthly construction costs (or producer prices) index for Poland. This index is related to new residential buildings, except residences for communities. The data covers the period from January 2000 to June 2022 and the unit of measure is a percentage change on previous period. Since investments in construction are long-term in nature, our model is going to be assessed in ten-month ahead forecast. Therefore, we divided the data into 260 months for training and the rest 10 months for testing purposes. Thus, the last 10 observations were compared to the forecast of the model. This approach somehow simulates the usefulness of the model in the future. Figure 1 depicts monthly construction costs index in the training set.

Having a wide range of time series modelling techniques, from the popular ARIMA approach to artificial intelligence tools, we chose the spline model, which is one of the non-linear and local regression methods. The variant we used, namely smoothing splines, does not require setting the hyper-parameters of the method. This would demand some background knowledge or the use of a validation set, which does not always lead to a stable solution. All of that make this method a convenient tool for practitioners.

Before we fitted the model, the data was pre-processed. At this stage, we dealt with outliers and transformed the data to stabilize the variance. After fitting the model, we performed a residual analysis and then assessed the accuracy of the forecast on the test set. The result was compared to ARIMA, which is one of the most widely used forecasting methods for univariate time series analysis, as well as to neural network variant of ARIMA (Figure 2).

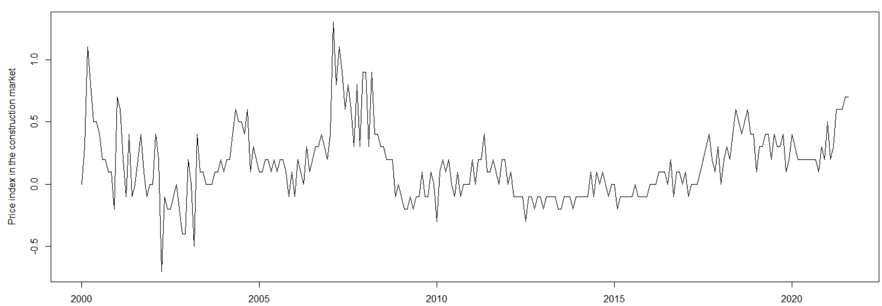

Figure 1. Price index in the construction market up to August 2021 (training set)

[4] https://ec.europa.eu/eurostat/data/database (04.01.2023).

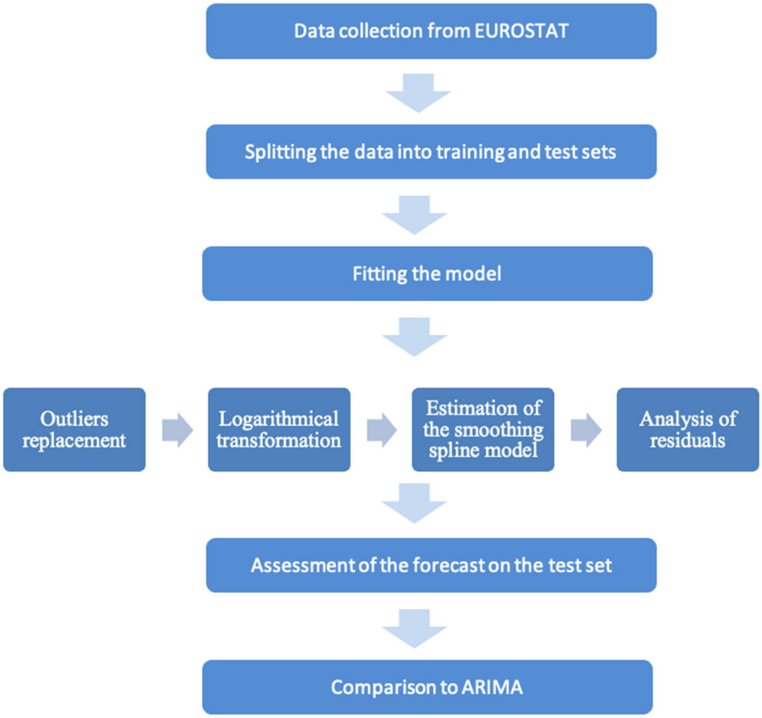

Figure 2. Scheme of the research

All computations were carried out using R packages {tseries} and {forecast}. R is an open source environment freely available on the web site: www.r-project.org. This software is a collaborative project with many contributors and supported by world research community.

4.2. Nonlinear splines model

Nonlinear regression methods have a key meaning in econometric modelling. They found numerous applications for both data independent of time and time series. The popularity they have achieved is because they go beyond simple linear dependencies, which are present in many real domains. The nonlinear approach serves a possibility of extremely accurate data fitting but it is a well-known fact that these kinds of models can be overfitted and, as a result, they usually perform poorly on unseen data. In terms of time series' forecasting, we say that *ex-ante* prognoses are inaccurate. This is unacceptable when prediction is the goal, and that is the case in most economic applications. To deal with the problem of overfitting, the vast majority of regression methods introduce regularization into the model selection stage. Instead of minimizing the quadratic loss function, which consequently results in the best possible fit, the criterion contains a penalty for model complexity. The core idea behind this approach is the trade-off between the accurate fit and the number of model parameters. Simpler models that do not fit the training data perfectly usually perform better on unseen data, i.e. on the forecast horizon in the case of time series. However, an overly simplified model does not reflect a systematic factor in the data and is naturally unlikely to be accurately predicted. The common approach is to use information criteria like AIC, AICc, or BIC to compare candidate models and to choose the final one. In more complex regression methods, i.e. splines, the regularization criterion is embedded in the algorithm of finding the final function formula. This is the case in the regression splines method which we present in this paper.

Generally, regression models can be divided into global or local. In the first case, one curve is fitted to the data in the entire domain, whereas in the second, the domain is partitioned into regions and the regression functions are approximated separately in each of them. Due to the context of the time series we focus on regression with one independent variable, which is time.

The local approach to modelling the regression curve requires choosing K points $\{\xi_k\}$ that divide the domain of X into $K + 1$ contiguous intervals. The points $\{\xi_k\}$ are called knots and it is a set that ξ_1 and ξ_K are equal to the minimal and maximal observation of variable X respectively. In each interval, function f_k is fitted to the data separately. In general, this function can take any form, but a frequent choice is polynomial. Since the high-order polynomial not only tends to overfit but causes estimation problems due to a large number of parameters, a usual choice is degree 3.

The basic problem of a piecewise polynomial function is continuity, thus the constraints:

$$f_k(\xi_k) = f_{k+1}(\xi_k), \tag{1}$$

for each $k \in \{1, \dots, K\}$ are imposed on the knots. Additionally, imposing the equality of the first and second derivatives in the knots:

$$f'_k(\xi_k) = f'_{k+1}(\xi_k);\ \ f''_k(\xi_k) = f''_{k+1}(\xi_k), \tag{2}$$

one can obtain a smooth regression line. That sort of regression models is known as cubic splines. The model has a form of the linear combination of the basis functions:

$$f(X) = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \sum_{k=1}^{K} \beta_k (X - \xi_k)_+^3, \tag{3}$$

where

$$(X - \xi)_+^3 = \begin{cases} (X - \xi)^3 & for\ X > \xi \\ 0 & otherwise \end{cases} \tag{4}$$

and its parameters can be estimated by the least square method. Having given the number of the knots, they are usually set uniformly in quantiles of variable X. As the confidence bands can be relatively wide in the boundary regions, an additional constraint can be imposed. The function is to be linear for $X < \xi_1$ and for $X > \xi_K$. This model is known as a natural cubic spline.

The problem of determining the number and position of knots is naturally solved in the smoothing splines. They simply use all realizations of variable X as the knots, and parameters of the model (3) are estimated by the minimization of the criterion:

$$\sum_{i=1}^{N}(y_i - g(x_i))^2 + \lambda \cdot \int_D (g''(t))^2 dt, \tag{5}$$

where D is a range of t, and λ is a regularization parameter. The first term is a quadratic loss function, which reflects how well the model fits the observed data. The second one is a penalty for a large variability in function g. Interestingly, Hastie *et al.* (2009) showed that the function that minimizes the criterion (5) is a natural cubic spline with knots in the observations of variable X. The absolute value of the second derivative is large when function g is very wiggly near t, and otherwise, for a straight line that is perfectly smooth, the second derivative is equal zero. The amount of the penalty is regulated by the λ parameter, which takes values from zero to infinity. When $\lambda = 0$, there is no penalty and the function g accurately approximates data points. On the other hand, when $\lambda \to \infty$, the regression curve obtains perfect smoothness, and thus it becomes a straight line. Having taken the intermediate values of λ, one has a model that is a compromise between exact fit and amount of smoothness. Setting this parameter has a crucial role in modelling and yielding accurate predictions (or prognosis in the case of time series). The common solution for performing this task is cross-validation. For small samples or one period ahead forecasting, the LOOCV version (leave-one-out cross-validation) is a natural choice:

$$CV(\hat{g}_\lambda) = \frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{g}_\lambda^{-i}(x_i)\right]^2. \tag{6}$$

In this case, the computations demand N models to be built, therefore in practice reformulation of the CV estimate of the mean square error is used:

$$CV(\hat{g}_\lambda) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \hat{g}_\lambda(x_i)}{1 - S_\lambda(i,i)}\right]^2, \tag{7}$$

where $S_\lambda(i,i)$ are diagonal elements of $\boldsymbol{S}_\lambda$, which is the projection matrix in linear fit $\hat{\boldsymbol{y}} = \boldsymbol{S}_\lambda \boldsymbol{y}$. The last formula allows computing LOO criterion with only one fit. The construction of $\boldsymbol{S}_\lambda$ matrix is somewhat technical. As mentioned before, the solution of (5) has a natural spline representation. Given N basis functions, the criterion (5) can be expressed in matrix form and transformed to projection form (for more details see

(Hastie *et al.* 2009)). Replacing the diagonal elements with their mean in (7) one can obtain an approximation that is known as generalized cross-validation:

$$GCV(\hat{g}_\lambda) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \hat{g}_\lambda(x_i)}{1-\frac{trace(S_\lambda)}{N}}\right]^2. \quad (8)$$

The trace of matrix $\boldsymbol{S}_\lambda$ is called the effective degrees-of-freedom and it controls the complexity of the model. Note that the GCV criterion is widely used in multivariate adaptive regression splines (Friedman 2009).

5. Analysis of the results

In this section, we intend to model the costs index on the construction market. Our purpose is to build a model that could be useful in the longer term. Looking at Figure 1 one can observe that the variance in the period before 2009 is clearly higher than afterwards. Before modelling this time series we pre-processed the data. Firstly, the data was cleaned. For this purpose, we used the procedure implemented in {forecast} package for the R program, where time series is decomposed and then the trend (and optionally seasonal component) is removed. For the remainder series, outliers are identified as values greater than Q3+3Q or less than Q1-3Q, and then they are replaced using linear interpolation.[5] In this way observations from (03/2000; 04/2002; 02/2007; 04/2007) appeared to be outliers and their actual values (1.1, -0.7, 1.3, 1.1) were replaced by (0.62293325, -0.04283356, 0.61761271, 0.80996611) respectively. Then the data was logarithmically transformed according to the formula: $y = \ln(1+x)$. The addition of one in the argument of the logarithm was necessary due to the requirement of a domain that must be positive. Even after these actions, the time series remained non-stationary, which was confirmed by augmented Dickey-Fuller test (p-value=0.3281). Although one might expect that there are seasonal fluctuations in the construction market, this is not confirmed by the graph in Figure 3, where indexes are shown in particular months. In addition, we checked this by performing two seasonality tests available in the R package {forecast} (Hyndman & Khandakar 2008). Both tests ruled out the significance of seasonal fluctuations. The Osborn, Chui, Smith and Birchenhall (OCSB) test yielded a test statistic value of -9.9689, while the 5% critical value is -1.803. The Canova and Hansen test, on the other hand, did not yield a basis for rejecting the null hypothesis of no unit roots at seasonal frequencies. The p-values for individual months ranged from 0.1539 to 0.9644.

[5] For more details which are not supplied in {forecast} package see https://robjhyndman.com/hyndsight/tsoutliers/.

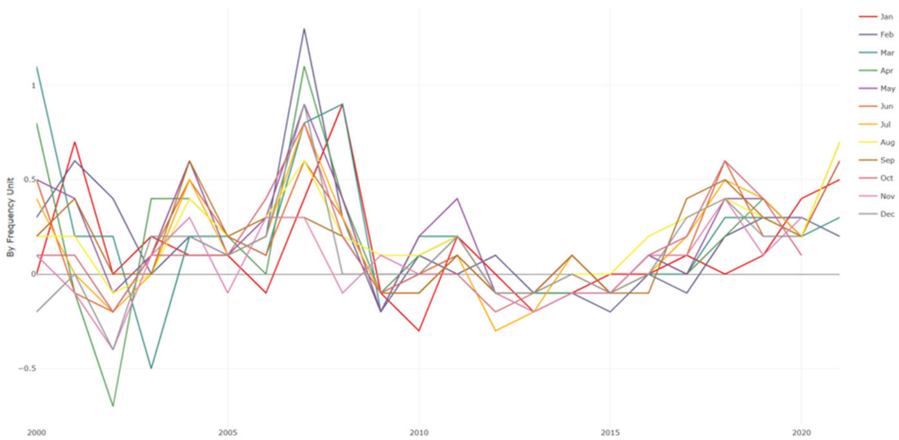

Figure 3. Construction costs index from seasonal perspective

All these preliminary studies were the motivation for choosing the local regression method to model this time series. The spline methodology automatically divides the time axis into periods (intervals) and fits the regression curves to them separately. Due to constraint (1), a continuous line is obtained over the entire time axis. The price index in the construction market will be approximated by a cubic smoothing spline model (3). In order to control the complexity of the model we use the GCV criterion (8), which geometrically results in a smooth regression line. The result is an increasing linear forecast over a 10-month horizon (Figure 4).

Assuming a significance level of 0.05, the independence of the residuals was verified. We used the Ljung-Box test (Ljung & Box 1978) with the number of lags equal to 10. The obtained p-value = 0.26 supports not rejecting the null hypothesis, according to which there is no autocorrelation between residuals. Although the p-value in the Shapiro-Wilk test for residuals is much less than 0.05, outliers appear to be the reason of this result. The isolated bars in the tails of the distribution are clearly visible on the histogram of the residuals (Figure 5). We identified them by assuming that outliers are observations meeting one of the conditions of $y_t < Q_1 - 3Q$ or $y_t > Q_3 + 3Q$, where Q_i are quartiles and Q is the quartile deviation. There were nine such observations from: 03/2000, 01/2001, 04/2001, 02/2002, 01/2003, 03/2003, 04/2003, 10/2004, 01/2010. It should be noted that all but one date from the very early period of the data under consideration. The most recent is from 01/2010, which is halfway through the time series. After removing these atypical residuals, we obtained p-value of 0.1467 in the Shapiro-Wilk test, thus there is no statistical evidence to reject the hypothesis of a normal distribution of residuals in this case.

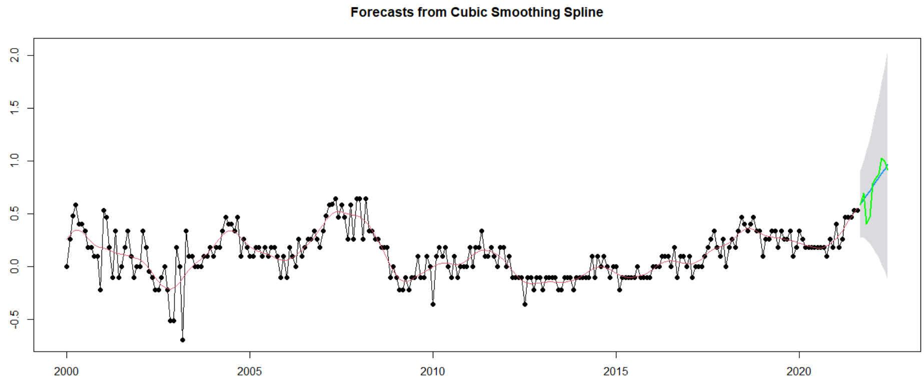

Figure 4. Fitted regression spline curve to the construction costs index (in red). Blue line depicts a 10-month ahead forecast whereas the green one shows observations from the test set

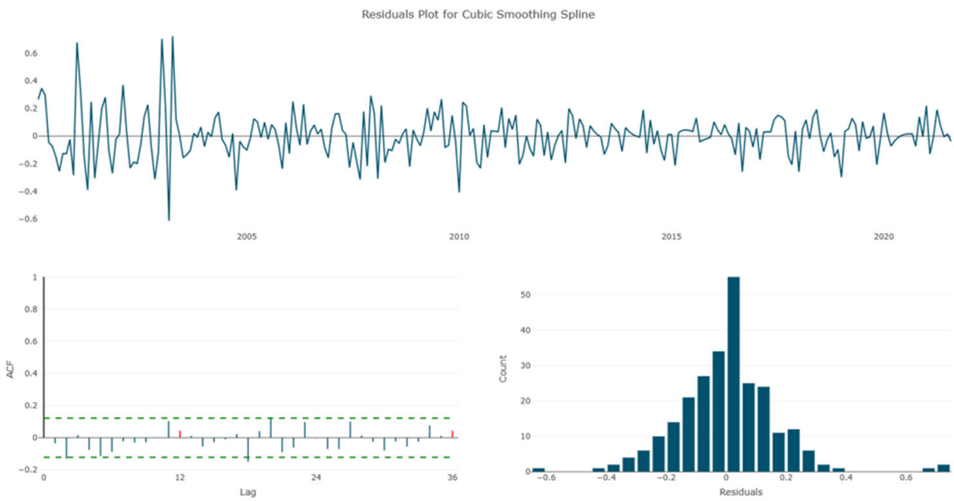

Figure 5. Residuals of the spline model – their autocorrelations and histogram

Table 1 presents the forecast in a 10-month horizon, together with confidence intervals. We obtained the test errors of the forecasts as follows: RMSE = 0.2534, MAE = 0.1943, MAPE = 23.77%. Although the errors are slightly high, note that the biggest differences between the actual values and the forecast are observed in the third and fourth steps of the forecast, which are November 2021 and December 2021 (Figure 6). They mainly influenced the overall values of the considered error measures. At that time, the Polish government introduced a second lockdown due to the COVID-19 pandemic, which could have caused disturbances in the economy, and in particular in the construction market. Note that the cubic smoothing spline model applied to the original data, which was neither cleaned nor transformed, turned out to be definitely inferior: RMSE = 0.3510, MAE = 0.3083, MAPE = 29.38%.

Table 1. The forecast 10 month ahead.

Month	Test observations	Point forecast	Lower limit .95	Upper limit .95
Sep 2021	0.8	0.8104521	0.31633653	1.490045
Oct 2021	1.0	0.8870110	0.31061334	1.716904
Nov 2021	0.5	0.9668073	0.28780969	2.003806
Dec 2021	0.6	1.0499780	0.25092565	2.359440
Jan 2022	1.2	1.1366658	0.20315095	2.794487
Feb 2022	1.3	1.2270194	0.14734652	3.322683
Mar 2022	1.4	1.3211937	0.08593149	3.961584
Apr 2022	1.8	1.4193505	0.02090157	4.733419
May 2022	1.7	1.5216579	-0.04610327	5.666087
Jun 2022	1.5	1.6282917	-0.11374306	6.794486

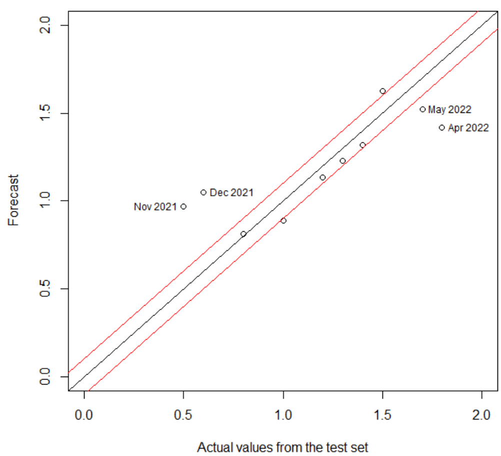

Figure 6. The differences between the actual values of the construction costs index in the test set and the spline model forecast. The red lines show +/- 0.1 outstanding from the exact forecast

One of the most popular method of time series analysis and forecasting is ARIMA, which stands for Auto-Regressive Integrated Moving Average (Box et al. 2008). Thus, it is justified to compare our results with this econometric approach. Instead of analyzing autocorrelation plots to determine the number of lags, degree of differencing, and order of the moving average model, we used the auto.arima function in R (Hyndman & Khandakar 2008). It enables searching of the parameter space guided by information criterion. Since the considered time series is of moderate size, we used an exhaustive search. The final model took the form ARIMA(4,0,0):

$$\begin{array}{l}\hat{Y} = 0.1364 + 0.3593 \cdot Y_{t-1} + 0.1645 \cdot Y_{t-2} + 0.2595 \cdot Y_{t-3} + 0.0759 \cdot Y_{t-4} \\ \quad (0.0632) \quad (0.0625) \qquad (0.0641) \qquad (0.0641) \qquad (0.0634)\end{array} \tag{9}$$

and forecasted decreasing trend 10 month ahead (Figure 7). The standard errors of the coefficients are given in brackets below the model (9). Note that the model is expressed for transformed index of costs (in fact it has a multiplicative form for original data). The errors of the forecast are dramatically high: RMSE= 0.8632, MAE= 0.7216, MAPE= 52.38. Since this model has no predictive abilities we omit the analysis of the residuals.

Then we compared our model with non-linear variant of ARIMA, where autocorrelation function in AR(p) part of the model was estimated with a use of feed-forward neural network with a single hidden layer (Lippi & Zaniolo 2012). As previously, the parameters of the model were exhaustively examined to minimize the information criterion. The model took the form NNet-ARIMA(3,1,2) and also predicted a downward trend in the next 10 months (Figure 8). Similar to classical ARIMA, all error measures considered are high: RMSE= 0.8579, MAE= 0.7138, MAPE= 51.57.

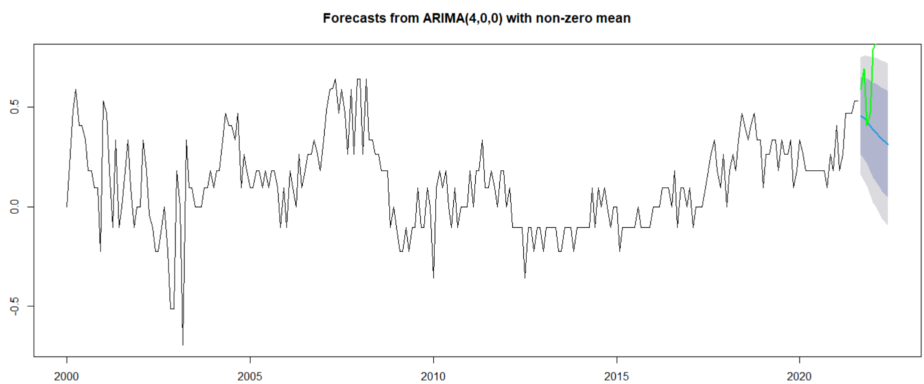

Figure 7. Decreasing forecast (blue line) pointed out by ARIMA model. Actual observations from the test set are in green

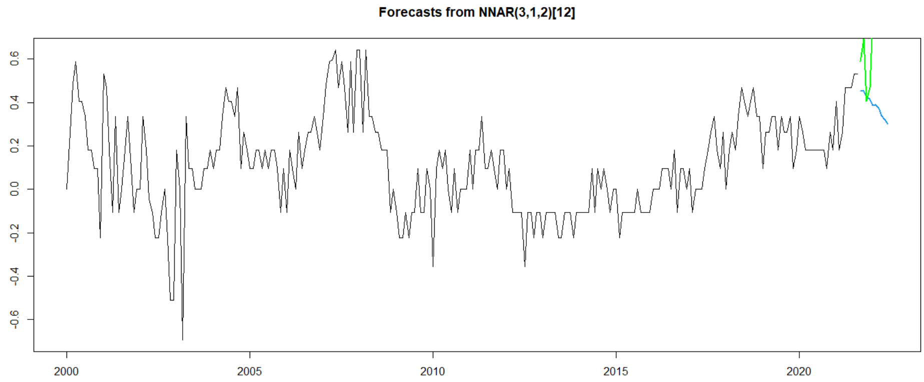

Figure 8. Decreasing forecast (blue line) pointed out by neural network version of ARIMA. Actual observations from the test set are in green

6. Conclusions

The construction costs index is an important component in the macroeconomic system. In the economy, it belongs to at least two of the three factors of production, namely capital and land. The value and dynamics of changes in construction production have a significant impact on the condition of the economic system, and in particular on the condition of the real estate market. The importance of built production in the economic system makes knowledge of its future values a key decision-making resource in modern management. Taking this into account, the authors of this research tested the usefulness of the nonlinear spline method for forecasting changes in the construction costs index in the long term.

Our work showed that nonparametric local regression is capable of being a valuable forecasting tool. In the case of the considered data, the commonly used ARIMA econometric model as well as its extension based on a neural network failed completely. Both ARIMA approaches indicated a decreasing forecasting when the test set data had in fact an increasing trend. The error measures on the test set for the spline model turned out to be not entirely satisfactory, but let us emphasize again that it could have been influenced by the lockdown decisions in Poland during the second wave of the pandemic. These error measures were overstated by the two-month observations. Nevertheless, the spline model detected a sharp increase in the costs index in the coming months, which provides valuable information in making investment decisions. It is noteworthy that smoothing splines can be recommended also due to a lack of tuning parameters, which is convenient for practitioners who are not advanced in data analysis.

References

Al Kailani, H., Sweis, G. J., Sammour, F., Maaitah, W. O., Sweis, R. J. and Alkailani, M., (2024). Predicting construction cost index using fuzzy logic and machine learning in Jordan. *Construction Innovation*. https://doi.org/10.1108/CI-08-2023-0182.

Altalhoni, A., Liu, H., Abudayyeh, O., (2024). Forecasting Construction Cost Indices: Methods, Trends, and Influential Factors. *Buildings*, 14(10), art. no. 3272.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C., (2008). Time Series Analysis: Forecasting and Control (4th ed.). *Wiley*.

Butler, S., Kokoszka, P., Miao, H. and Shang, H. L., (2021). Neural network prediction of crude oil futures using B-splines. *Energy Economics*, 94. https://doi.org/10.1016/j.eneco.2020.105080.

Cao, Y., Ashuri, B., (2020). Predicting the Volatility of Highway Construction Cost Index Using Long Short-Term Memory. *Journal of Management in Engineering*, 36(4). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000784.

Cheung, S. O., Wong, P. S. P., Fung, A. S. Y. and Coffey, W. V., (2006). Predicting project performance through neural networks. *International Journal of Project Management*, 24(3), pp. 207–215. https://doi.org/10.1016/j.ijproman.2005.08.001.

Chu, M. T., Kuo, Ch and Lin, M. M., (2013). Tensor Spline Approximation in Economic Dynamics with Uncertainties. *Computational Economics,* 42, pp. 175–198. https://doi.org/10.1007/s10614-012-9331-1.

Dash, S. R., Maitra, D., (2019). The relationship between emerging and developed market sentiment: A wavelet-based time-frequency analysis. *Journal of Behavioral and Experimental Finance*, 22, pp. 135–150. https://doi.org/10.1016/j.jbef.2019.02.006.

Durdyey, S., Ismail, S., (2012). Role of the construction industry in economic development of Turkmenistan. *Energy Science and Research*, Vol. 29, No. 2, pp. 883–890.

ELFAHHAM, Y., (2019). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal*, 58(2), pp. 499–506.

Findley, D. F., Lytras, D. P. and Maravall, A., (2016). Illuminating ARIMA model-based seasonal adjustment with three fundamental seasonal models. *SERIEs*, 7, pp. 11–52. doi:10.1007/s13209-016-0139-4.

Friedman, J. H., (2009). Multivariate adaptive regression splines. *The Annals of Statistics*, Vol. 19, No. 1, pp. 1–141.

Gadasina, L., Vyunenko, L., (2022). Applying spline-based phase analysis to macroeconomic dynamics. *Dependence Modelling*, 10(1), pp. 207–214. https://doi.org/10.1515/demo-2022-0113.

Hastie, T., Tibshirani, R., Friedman, J., (2009). The Elements of Statistical Learning: Data Mining. *Inference and Prediction*, *Springer*, New York.

Huang, C., Raunikar, R., (1981). Spline Functions: An Alternative to Estimating Income-Expenditure Relationships for Beef. *Journal of Agricultural and Applied Economics*, 13(1), pp. 105–110. doi:10.1017/S0081305200024626.

Humphrey, D. B., Vale, B., (2004). Scale economies, bank mergers, and electronic payments: A spline function approach. *Journal of Banking and Finance*, 28(7), pp. 1671–1696. https://doi.org/10.1016/j.jbankfin.2003.05.003.

Hyndman, R. J., Khandakar, Y., (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 26(3).

Ilyasov, R. H., Yakovenko, V. S., (2021). Spline Analysis of Flow Correlation in Economic Systems, in The Challenge of Sustainability in Agricultural Systems, BOGOVIZ, A. V., eds., *Springer International Publishing, Springer*.

Joukar, A., Nahmens, I., (2016). Volatility forecast of construction cost index using general autoregressive conditional heteroskedastic method. *Journal of Construction Engineering and Management*, 142(1), art. no. 04015051.

Kim, S., Choi, C.-Y., Shahandashti, M. and Ryu, K. R., (2022). Improving Accuracy in Predicting City-Level Construction Cost Indices by Combining Linear ARIMA and Nonlinear ANNs. *Journal of Management in Engineering*, 38(2), art. no. 04021093.

Lin, R. F. Y., Ou, C., Tseng, K. K., Bowen, D., Yung, K. L. and Ip, W. H., (2021). The spatial neural network model with disruptive technology for property appraisal in real estate industry. *Technological Forecasting and Social Change*, 173. https://doi.org/10.1016/j.techfore.2021.121067.

Lippi, M., Zaniolo, F., (2012). A feedforward neural network model for time series forecasting with applications in finance and economics. *Artificial Intelligence in Finance and Economics*, pp. 1–20.

Ljung, G. M., Box, G. E., (1978). On a measure of lack of fit in time series models. *Biometrika*, Vol. 65, No. 2, pp. 297–303. https://doi.org/10.1093/biomet/65.2.297.

Lopes, J., (2011). Construction in the economy and its role in socio-economic development, in New Perspectives on Construction in Developing Countries, OFORI, G., eds., pp. 40–71, London, *Routledge*.

Mao, S., Xiao, F., (2019). A novel method for forecasting Construction Cost Index based on complex network. *Physica A: Statistical Mechanics and its Applications*, 527, art. no. 121306.

Monteiro, A. M., Tütüncü, R. H. and Vicente, L. N., (2008). Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity. *European Journal of Operational Research*, 187(2), pp. 525–542. https://doi.org/10.1016/j.ejor.2007.02.041.

Moon, S., Chi, S. and Kim, D. Y., (2018). Predicting Construction Cost Index Using the Autoregressive Fractionally Integrated Moving Average Model. *Journal of Management in Engineerin*g, 34(2), art. no. 04017063.

Moon, T., Shin, D. H., (2018). Forecasting Model of Construction Cost Index Based on VECM with Search Query. *KSCE Journal of Civil Engineering*, 22(8), pp. 2726-2734.

Naccache, T., (2011). Oil price cycles and wavelets. *Energy Economics*, 33(2), pp. 338–352. https://doi.org/10.1016/j.eneco.2010.12.001.

Nieto, M. R., Carmona-Benítez, R. B., (2018). ARIMA + GARCH + Bootstrap forecasting method applied to the airline industry. *Journal of Air Transport Management*, 71(June), pp. 1–8. https://doi.org/10.1016/j.jairtraman.2018.05.007.

Osiewalski, K., Osiewalski, J., (2013). A Long-Run Relationship between Daily Prices on Two Markets: The Bayesian VAR(2)–MSF-SBEKK Model. *Central European Journal of Economic Modelling and Econometrics*, 5(1), 65–83. https://doi.org/10.24425/cejeme.2013.119253.

Paci, L., Consonni, G., (2020). Structural learning of contemporaneous dependencies in graphical VAR models. *Computational Statistics & Data Analysis*, 144, 106880. https://doi.org/10.1016/j.csda.2019.106880.

Pheng, L. S., Hou, L. S., (2019). The Economy and the Construction Industry, In Construction Quality and the Economy. Management in the Built Environment, PHENG, L.S., HOU, L.S., eds., pp. 21–54, Singapore, *Springer*.

Ruddock, L., Ruddock, S., (2008). The scope of the construction sector: Determining its value, in Economics for the Modern Built Environment. RUDDOCK, L., eds., pp. 99–113, London, *Routledge*.

Sahraei, M.A., Duman, H., Çodur, M. Y. and Eyduran, E., (2021). Prediction of transportation energy demand: Multivariate Adaptive Regression Splines, Energy. *Elsevier Ltd*, Vol. 224. doi:10.1016/j.energy.2021.120090.

Sobieraj, J., Metelski, D. and Nowak, P., (2021). The view of construction companies' managers on the impact of economic, environmental and legal policies on investment process management. *Archives of Civil Engineering*, Vol. LXVII No. 1, pp. 111–129.

Vidoli, F., (2011). Evaluating the water sector in Italy through a two-stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines. *European Journal of Operational Research*, Vol. 212, No. 3, pp. 583–595. doi: https://doi.org/10.1016/j.ejor.2011.02.003.

Wang, C.-H., Mei, Y.-H., (1998). Model for forecasting construction cost indices in Taiwan. *Construction Management and Economics*, 16(2), pp. 147–157.

Wang, J., Ashuri, B., (2017). Predicting ENR Construction Cost Index Using Machine-Learning Algorithms. *International Journal of Construction Education and Research*, 13(1), pp. 47–63.

Williams, T. P., (1994). Predicting changes in construction cost indexes using neural networks. *Journal of Construction Engineering and Management*, 120(2), pp. 306–320.

Xu, J.-W., Moon, S., (2013). Stochastic forecast of construction cost index using a cointegrated vector autoregression model. *Journal of Management in Engineering*, 29(1), pp. 10–18.

Ye, K., Lu, W., Jiang, W., (2009). Concentration in the international construction market. *Construction Management and Economics*, Vol. 27, No. 12, pp. 1197–1207.

Zhang, R., Ashuri, B., Shyr, Y., Deng, Y., (2018). Forecasting Construction Cost Index based on visibility graph: a network approach. *Physica A: Statistical Mechanics and its Applications*, 493, pp. 239-252.