Jordan neural network for inflation forecasting

Tea Šestanović *

Faculty of Economics, Business and Tourism, University of Split
Cvite Fiskovića 5, 21000 Split, Croatia
E-mail: ⟨tea.sestanovic@efst.hr⟩

Abstract. In times of pronounced nonlinearity of macroeconomic variables and in situations when variables are not normally distributed, i.e. when the assumption of i.i.d. is not fulfilled, neural networks (NNs) should be used for forecasting. In this paper, Jordan neural network (JNN), a special type of NNs is examined, because of its advantages in time series forecasting suitable for inflation forecasting. The variables used as inputs include labour market variable, financial variable, external factor and lagged inflation, i.e. the most commonly used variables in previous researches. The research is conducted at the aggregate level of euro area countries in the period from January 1999 to January 2017. Based on 250 estimated JNNs, which differ in selected variables, sample breaking point and varying parameters (number of hidden neurons, weight value of the context unit), the model adequacy indicators for each JNN are calculated for two periods: in-the-sample and out-of-sample. Finally, the optimal JNN for inflation forecasting is obtained as the best compromise solution between low mean squared error in-the-sample and out-of-sample and low number of parameters to estimate. This paper contributes to existing literature in using JNN for inflation forecasting since it is rarely used for macroeconomic time series prediction in general. Moreover, this paper defines which set of variables contributes to the best inflation forecast. Additionally, JNN is examined thoroughly by fixing certain parameters of the model and alternating other parameters to contribute to the JNN literature, i.e. finding the optimal JNN.

Keywords: euro zone, forecasting, inflation, Jordan neural network

1. Introduction

Neural networks (NNs) are used for modelling and forecasting in many sciences and in economics as well. The feedforward NNs (FNNs) are most commonly used in time series analysis for prediction of stock prices [37], exchange rates [15], GDP [36] etc., and FNNs have been used for inflation forecasting by only a few researches [1, 10, 17, 36]. The main reason for their popularity lies in the fact that FNNs can approximate any nonlinear functional form with high precision [19]. By using FNNs, it is possible to capture nonlinearity and to provide superior accuracy with respect to linear and nonlinear models without the need to construct a specific parametric model. However, their parameters are difficult to interpret and their superior performance in-the-sample does not mean superior prediction on new data, i.e. out-of-sample. Therefore, the overfitting problem needs to be identified. Namely, by increasing the complexity of the model it is possible to obtain an almost perfect fit within the sample, i.e. FNNs can learn the irregular patterns and consequently be less useful for prediction. FNN structure complexity depends, among others, on the number of parameters estimated which depends on the number of inputs (p), outputs, the number of hidden layers, and the number of neurons in the hidden layer (q).

---

*Corresponding author.

Input and output selection depend on the economic theory, while the choice of hidden layers and hidden neurons is left to the researchers' arbitrary decision. Approximation of any nonlinear functional form is always good with a three-layer NN when q is large enough. However, it is not precisely defined how large q should be. Therefore, one way to solve the overfitting problem is model complexity reduction by including only the most important inputs and selecting the optimal number of hidden neurons. This yields to losing part of the forecasting information. Another approach is to include a feedback connection from output to inputs in FNN, i.e. by estimating a type of recurrent NNs (RNNs), particularly the Jordan NN [22]. It compensates the loss of information with parameters reduction, i.e. it complies with the parsimony principle. JNN includes the moving average components (MA) or the lagged error terms, which yields smaller forecasting error, with lower number of inputs than in the FNN, i.e. a lower number of autoregressive (AR) terms is required.

Previous researches do not reach a consensus on the selection of a suitable NN, but on the contrary, provide opposite results. Credibility of the results from existing applications is quite questionable due to missing argumentation and explanation of used RNN structures, i.e. they use different RNNs by default, without any settings specification as well as neglecting the overfitting problem and the choice of optimal number of hidden neurons. However, choosing a suitable NN depends not only on the choice of variables (inputs and outputs) but on numerous settings that need to be predefined and explained in detail. Therefore, in this paper JNN and its application will be explained in detail with its predefined structure. The number of inputs will be selected based on economic theory and varied with the aim to find the set of variables that best predicts inflation for different periods. The number of hidden neurons will be varied to find the optimal JNN for inflation forecasting and the optimal weight of the context unit will be examined. Although in recent economic time series and especially in inflation forecasting, RNNs are seldom used and although there is literature showing drawbacks of RNNs, they remain simple and powerful models. Namely, they keep information about past inputs along with satisfying the parsimony principle since they require a lower number of inputs than FNNs. In this paper, RNNs will be re-detected, investigated thoroughly and their drawbacks will be addressed properly. Moreover, besides the careful selection of JNN parameters, it will be examined whether the change in certain parameters leads to different performances. This paper contributes to the existing literature firstly by using JNN for inflation forecasting as it is rarely used for macroeconomic time series predictions. Secondly, the best set of inputs for inflation modelling and forecasting in periods of deflationary pressures and inflation smoothing in euro zone is found for the two selected periods. Namely, the hypothesis regarding input selection goes along with [20] that argues that models containing more information are better in inflation forecasting. This paper will prove the opposite, i.e. JNN containing only one lag of inflation yields the best forecasts. This research is not in accordance neither with [23] nor [16] who suggest that inflation forecasting using the AR model can be significantly improved by several indicators. However, it goes along with [7] showing good performance of the AR model, where the model containing some indicators is in most, but not all, cases better than the AR model, where the best indicator changes over time. Thirdly, the optimal number of hidden neurons is obtained using the rules of thumb [31, 33] defined from the starting model containing the largest information set, along with the cascading approach, starting from the smallest JNN and adding neurons until achieving the improvement in information criteria as in [4]. Finally, the optimal JNN is selected as the best compromise solution between low in-the-sample and out-of-sample selection criteria. Since it was not possible to obtain the optimal JNN using standard procedures, the PROMETHEE method is used as [2, 5] for other applications.

The remainder of the paper is organized as follows. Section 2 presents the literature review. Section 3 describes methodology. Section 4 presents empirical research results with discussion. Finally, conclusions and directions for the future research are provided in Section 5.

2. Literature review

The literature review on RNNs shows that there is limited research done using RNNs in time series analysis and that there is a time gap between fundamental research, re-detection of RNNs and their use. RNNs and their architectures are often not explained in detail, key elements of architecture are omitted, and they are compared with incomparable econometric models. Input selection is often not consistent with economic theory and the number of hidden neurons is determined by random selection. Most commonly, Elman's NN (ENN) or combinations with JNN, as well as other RNNs, are used. This is not econometrically justified and such research needs to be taken with caution. Research on simulated data and/or non-economic areas give advantage to RNNs cite13 for predicting time series versus linear, nonlinear model and/or FNNs, while [20] give the advantage to FNNs. RNNs are simple but powerful models. Their complexity makes them difficult to train due to the exploding and vanishing gradient problem [8, 32] which can be solved either by using second order algorithms, weight regularization, not learning the recurrent weights, careful initialization of RNN parameters or using the Long Short-Term Memory NNs. Here, it is addressed by careful weight initialization, where recurrent weights are predefined and not learned. Research on economic and financial time series [3, 6, 10, 12, 37] gives preference to RNNs for predicting time series versus linear, nonlinear models and/or FNNs. [34] give preference to RNNs when predicting in the long term, while predicting in the short term FNNs and linear models give better results. [24] does not give a clear answer to which NN has better predictive ability, while [11] prefer other artificial intelligence methods for predicting time series. [15] show that both FNN and RNN are more efficient than GARCH-type models.

In the inflation analysis, there is limited research which involves forecasting with NNs in general, and more limited is the number of research using RNNs. Namely, [32] compare FNN, RBF and RNN with structural and reduced VAR model and show that in the long run RNN has a higher predictive ability. In the short run, it is preferable to use FNN and linear models. However, this study does not consider the overfitting problem or the selection of hidden neurons and there is no explanation of the type of RNN used. [31] compare FNNs with structural, ARIMA, VAR, and BVAR models. FNNs can predict inflation equally well for 3 and 12 months ahead, and for 1 month ahead better than the traditional models. [29] compare linear models and FNNs with a linear component. NNs are better than the best linear model for predicting services index for the euro zone, and are equal or better for more general consumer and producer price indices for different countries. [12] compare RS-VAR and RNN, with VAR model and show that RS-VAR and RNN perform similarly in the short and long run and are significantly better than the VAR model. [9] compare ARIMA, VAR and FNN. FNN gives better results in-the-sample and out-of-sample. This paper provides a detailed insight into the NN architecture, but compares models based on different inputs. [36] compare AR, STAR, AR-NN and FNN. None of the methods dominates others and the optimal method differs by countries, variables, and prediction horizon. STAR is better than the AR model, while FNNs' results are different with respect to the prediction horizon. AR-NN model is inferior to other methods. They conclude that p and q selection is a key factor, with the largest estimated NN giving the worst results, while the simplest giving poor results over long prediction horizons. [10] compare RNN, as a combination of JNN and ENN, and MS-AR model. Both models perform better than the linear model, while for the long-term predictions MS-AR is more suitable. However, input selection is different in different models. [11] use RNN, as a combination of ENN and JNN, evolution strategies and kernel methods. Inputs, as well as the number of hidden neurons vary significantly and a large number of parameters is often estimated. The best results are obtained with a nonlinear AR model based on kernel model. However, these results cannot be compared due to different methodologies and architectures. [17] use Hybrid and ENN and a combination of two NNs. NNs are better in predicting inflation than the AR(1) model in 45% of countries,

while AR(1) is better in 21% of the countries. However, NN architecture is not explained in detail. [1] compare a modified FNN with Box-Jenkins approach. FNN has better predictive abilities. The literature review shows that there is limited research using NNs in time series analysis in general, and even fewer in inflation prediction. NNs are usually not explained in detail, with the key elements of architecture being omitted, i.e. inputs, outputs, and hidden neurons; data division on training, validation and testing; data transformation; initial parameter values; learning rate; number of iterations; learning algorithm; model selection criteria. In this paper, NNs are defined from the econometric perspective, their architecture is carefully chosen for the particular problem and a suitable NN, i.e. JNN, is used for inflation modelling and forecasting.

Central bank's (CB) decisions are determined by the expected future inflation and hence the inflation prediction is crucial for monetary policy making. Given the time lag of its actions, it is important to timely assess and interpret variables that affect inflation which must take into account a large set of determinants and interactions. Different structural macroeconomic, as well as time series models are used, which differ in size, purpose and degree of fundamentals. Although ECB achieved the highest precision in predicting inflation, there is a tendency of inflation underestimation in pre-crisis period and overestimation during the crisis. A review of the models used by CBs to predict inflation in the euro zone is explained in [18]. The most commonly used are ARIMA, VAR and VECM models [27] and rarely used are the nonlinear models [10, 26]. Linear models are limited in the presence of nonlinearity, causing predictions and other conclusions that can be derived from such models to be incorrect. Nonlinear models require that a nonlinear functional form is defined in advance, which can be infinitely many. NNs are used because of the disadvantages of linear and nonlinear models.

Literature on inflation is vast; however, the following researches give an overview of the most frequently used inflation determinants. [14] define a model that consists of product, labour, money, and an external market. [7] group the best indicators into price indices, labour market, fiscal and real variables. [23] defines financial variables that predict inflation as they reflect current and expected monetary policy responses and measures of real economic activity that predict inflation as they provide information on excess demand. The most commonly used variables are lagged inflation [9, 29, 31, 32]; exchange rates [21, 22, 30, 34]; monetary aggregates [11, 27, 28]; interest rates [20]; labour market variables, i.e. wages or unit labour costs [16, 20]; GDP [9, 20, 25]; industrial production [7] and fiscal policy variables. Researchers cannot agree on the inflation determinants since they differ through countries and periods. In this paper, RNN for inflation forecasting is chosen based on the expressed nonlinearity in the inflation process. It incorporates ARIMA model and the key variables that have emerged in empirical research as the main inflation determinants. It relies on the theoretical model [14, 27] which includes the characteristics of demand-pull and cost-push inflation and begins by establishing a balance between aggregate supply and demand. After calculating the logarithms and differentiation, the final model is:

$$\pi_t = \phi_0 + \phi_1\pi_{t-1} + \phi_2\mu_t + \phi_3\omega_t + \phi_4\varepsilon_t + \nu_t, \tag{1}$$

where π_t is the current rate of inflation, π_{t-1} lagged rate of inflation (inertial factor), μ_t monetary expansion rate (financial variable), ω_t rate of change in labour costs (labour market variable), ε_t depreciation rate (external factor), while ν_t captures supply and demand shocks.

3. Methodology

This paper starts from the theoretical model defined in eq.(1). If the assumptions of linearity and normality are not meet, the initial model is extended by adding hidden layer between input and output layer, i.e. several inputs as independent variables and one output as dependent variable. This yields a three-layered FNN. The parameters of the model, i.e. weights are estimated using the backpropagation (BP) learning algorithm. The estimators obtained by the BP algorithm are consistent and asymptotically normally distributed [19]. The name of the method is derived from the way the weights are adjusted. Namely, during the learning, inputs are presented to NN and each sample on the training set is spread forward through NN layer by layer until the output from the network is calculated. The expected output is then compared with the observed output based on which errors are calculated. Errors are used as inputs in backpropagation process when the weights are adjusted backward layer by layer. The adjustments of the weights of the corresponding layer are carried out iteratively. This process is repeated in the training set until the total error converges to the minimum or until the maximum number of iterations is reached [33].

3.1. Jordan neural networks

If FNN is extended by adding a feedback from the output to the input layer, JNN is obtained:

$$\pi_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{oh} g\left(\phi_{ch} + \phi_1 \pi_{t-1} + \phi_2 \mu_t + \phi_3 \omega_t + \phi_4 \varepsilon_t + \phi_{rh} \nu_{t-1}\right)\right) + \nu_t, \qquad (2)$$

where $f(\cdot)$ and $g(\cdot)$ are activation functions. Parameters ϕ_{co} and ϕ_{ch} are constant terms in output and hidden layer respectively. ϕ_{co} $h = 1, 2, ..., q$ are the weights between the hidden and output layer, ϕ_i $i = 1, 2, ..., p$ are the weights between the input and hidden layer, ν_t is the error term. ϕ_{rh} is the weight between the context unit and hidden layer, and ν_{t-1} is the lagged error term. It uses a recursive algorithm [24] which is similar to BP but requires multiple equations for weight adjustment but provides consistent and asymptotically normally distributed estimators. The starting model is tested against the JNN containing only lagged inflation as an input:

$$\pi_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{oh} g\left(\phi_{ch} + \phi_1 \pi_{t-1} + \phi_{rh} \nu_{t-1}\right)\right) + \nu_t, \qquad (3)$$

Model presented by eq.(3) is actually nonlinear ARMA(1,1) with q neurons in a single hidden layer. This type of JNN is just a special case of eq.(2) and thus it can be considered as nested. Therefore, the contribution of proposed variables to inflation predictions can be tested by the null hypothesis $H_0 : \phi_2 = \phi_3 = \phi_4 = 0$. Nevertheless, the same procedure can be obtained for each variable individually. Therefore, three additional models are estimated separately:

$$\pi_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{oh} g\left(\phi_{ch} + \phi_1 \pi_{t-1} + \phi_2 \mu_t + \phi_{rh} \nu_{t-1}\right)\right) + \nu_t, \qquad (4)$$

$$\pi_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{oh} g\left(\phi_{ch} + \phi_1 \pi_{t-1} + \phi_2 \omega_t + \phi_{rh} \nu_{t-1}\right)\right) + \nu_t, \qquad (5)$$

$$\pi_t = f\left(\phi_{co} + \sum_{h=1}^{q} \phi_{oh} g\left(\phi_{ch} + \phi_1 \pi_{t-1} + \phi_2 \varepsilon_t + \phi_{rh} \nu_{t-1}\right)\right) + \nu_t. \qquad (6)$$

3.2. Data

The sample contains monthly data on aggregated euro zone level from 01.1999 to 01.2017, obtained from the ECBs' Statistical Data Warehouse database. Inflation rate is the dependent variable in all models, measured by the annual rate of change of the Harmonized Index of Consumer Prices - HICP (2015=100). The rate of change of monetary aggregate M3 is used as the financial variable. M3 represents money supply or total liquid assets. Instead of wages, unit labour costs (ULC) are used as an indicator of labour market trends. This variable is available at a quarterly level. To obtain monthly data, its approximations are calculated by a third-order polynomial, i.e. by using spline interpolation. Nominal effective exchange rate (NEER) is used as an external factor indicator. It is a weighted domestic currency exchange rate relative to the currency group of 19 major foreign trade partners, showing the value of the domestic currency relative to other currencies traded in the euro zone. Descriptive statistics is calculated for variables with normality, stationarity and independence tests . The results show that inflation is neither normally distributed, stationary, nor i.i.d., i.e. inflation is nonlinear. The results for all variables indicate that the assumptions of normality, stationarity and linearity are disturbed, leading to the necessity of using NNs.

3.3. Analysis

JNN architecture is defined as follows. Inflation rate is the output. Variables are transformed using normalization function. Activation functions are logistic and linear in hidden and output layer respectively. Initial weights are randomly selected from the interval [-1,1] which allows both positive and negative weights as the influences of inputs can have both the positive and negative sign. Learning rate is 0.001 with a maximum number of iterations of 10000, allowing a relatively slow but stable convergence. Learning algorithm is a recursive BP. Only one repetition of training is performed using the same random seed selection for all JNNs . A change in random seed did not lead to a change in results as investigated on several JNN architectures. Other parameters are not investigated enough in literature and/or are application dependent. Inputs depend on economic theory and are varied as in eq.(2)-(6) to investigate which input set best predicts inflation. One hidden layer is used and the number of hidden neurons is selected with attention since performances of JNN can vary significantly. The rules of thumb [31, 33] are used where q should be $\dfrac{N}{5(m+p)}$ or $\dfrac{N}{10(m+p)}$, i.e. they vary from 1 to 5, defined from the initial model. The weight of the context unit λ, representing the long-term memory, is varied with values 0.1, 0.3, 0.5, 0.7 and 0.9 according to [3]. It is not learned or trained but determined in advance to deal with vanishing gradient problem.

JNNs are estimated for 2 scenarios, corresponding to 2 breaking points for JNN estimation, i.e. deflation and stable inflation. Data in both scenarios is divided on *in-the-sample* part used for JNN training, and *out-of-sample* part kept for forecasting. In Scenario 1, the *in-the-sample* contains 192 observations (01.1999-12.2014), and the *out-of-sample* contains last 25 observations (01.2015-01.2017), where the breaking point is the trough of inflation, i.e. the highest deflation period. In Scenario 2, the *in-the-sample* contains data from 01.1999 to 12.2015 (204 observations), and the *out-of-sample* contains last 13 observations (01.2016-01.2017), where the breaking point refers to a period of stable inflation. *In-the-sample* data are additionally divided using 70% for training and 30% for validation. With 5 different inputs, 5 hidden neurons, 5 context unit weights and 2 scenarios, 250 JNNs are estimated in *R Studio* and compared using mean squared error (MSE) and Akaike information criteria (AIC). Finally, top 10 JNNs are selected and compared for 2 scenarios using the PROMETHEE method in the *Visual PROMETHEE*. JNNs are alternatives and the criteria are MSE and AIC. Criteria are equally important and Gaussian preference function is used.

	in-the-sample					out-of-sample				
q	λ=0.1	λ=0.3	λ=0.5	λ=0.7	λ=0.9	λ=0.1	λ=0.3	λ=0.5	λ=0.7	λ=0.9
1	0.4703	0.4652	0.4662	0.4646	0.4661	4.1844	4.3864	4.4454	4.4055	4.3205
2	0.0589	0.0731	0.0592	0.0542	0.0809	3.3378	3.9383	3.7754	3.0389	2.4757
3	0.0589	0.0651	0.0591	0.0524	0.0915	3.3299	3.8308	3.7697	3.0484	2.2587
4	0.0583	0.0670	0.0595	0.0495	0.0780	3.3701	3.8981	3.8161	3.2093	2.6955
5	0.0705	0.0834	0.0575	0.0524	0.0529	3.7851	4.0544	3.7839	3.4747	3.2216

Table 1: *In-the-sample and out-of-sample MSE calculated for JNNs in Scenario 1*

	ϕ	ϕ^+	ϕ^-	Eq.	MSEin	MSEout	AICin	AICout	q	λ	k
JNN42	0.503	0.549	0.046	eq.(3)	0.0542	3.0389	-541.71	45.79	2	0.7	9
JNN43	0.494	0.518	0.024	eq.(3)	0.0524	3.0484	-540.07	53.87	3	0.7	13
JNN44	0.493	0.507	0.014	eq.(3)	0.0495	3.2093	-542.96	63.15	4	0.7	17
JNN117	0.321	0.432	0.111	eq.(6)	0.0504	3.9117	-551.50	56.10	2	0.7	11
JNN120	0.274	0.383	0.108	eq.(6)	0.0476	3.3922	-532.49	82.54	5	0.7	26
JNN118	0.263	0.373	0.110	eq.(6)	0.0495	3.8653	-544.93	65.80	3	0.7	16
JNN70	0.228	0.340	0.112	eq.(4)	0.0489	3.3358	-527.52	82.12	5	0.7	26
JNN17	0.228	0.349	0.121	eq.(2)	0.0504	3.9117	-543.50	64.10	2	0.7	15
JNN103	0.214	0.289	0.075	eq.(6)	0.0537	3.4295	-529.45	62.81	3	0.1	16
JNN123	0.200	0.389	0.189	eq.(6)	0.0488	4.1570	-547.89	67.62	3	0.9	16

Table 2: *Top 10 JNNs with flows, equation, in-the-sample and out-of-sample MSE and AIC, hidden neurons, weight of the context unit and the number of parameters (k) in Scenario 1*

4. Results and discusion

4.1. First scenario

Based on 125 estimated JNNs in Scenario 1, MSEs *in-the-sample* and *out-of-sample* are calculated. Table 1 contains MSEs calculated for eq.(3) in Scenario 1 (the results for eq.(2), eq.(4)-(6) are omitted due to the lack of space but available upon request). The results indicate that JNNs with only 1 hidden neuron do neither learn nor predict well. Increasing q significantly decreases MSEs. In the majority of JNNs, the best results are obtained with 2 to 4 hidden neurons. Increasing q to 5 results in higher MSEs. λ equal to 0.7 gives stable results for all variables, yielding the lowest MSEs. Since it is not possible to find a JNN that has simultaneously low *in-the-sample* and *out-of-sample* MSEs, next step is the elimination of JNNs that do not learn well, i.e. JNNs with MSEs higher than 0.06 are excluded to ease further ranking. Optimal JNN is selected using the PROMETHEE method. The rankings of the top 10 JNNs with net ϕ, positive ϕ^+ and negative ϕ^+ flows, model equation, *in-the-sample* (in) and *out-of-sample* (out) MSE and AIC, the number of hidden neurons, context unit weight, and the number of parameters k are given in Table 2. The results indicate that top 3 JNNs are estimated from eq.(3), i.e. with only one input (lagged HICP). The parsimony principle is satisfied with a small number of estimated parameters. The highest ranked is JNN42 with the lowest MSE and AIC. It is estimated using 2 hidden neurons and context unit weight of 0.7. In 8 out of 10 JNNs, λ is 0.7. The number of hidden units is indicative, i.e. in 8 out of 10 cases q is 2 or 3.

It was expected that models containing more variables would have better performances, which is not the case, since only 1 out of 10 JNNs have all 4 variables in the model. ULC in eq.(5) is not selected among the top 10, i.e. it is not a good inflation predictor. M3 in eq.(4) is ranked 7th, whereas NEER gives solid performances, i.e. it is ranked 5 times among top 10

JNNs, however not in the top 3. It can be concluded that variable selection plays a key role in modelling and forecasting inflation where satisfying parsimony principle leads to a JNN that has both good *in-the-sample* and *out-of-sample* performances, with no overfitting problem.

4.2. Second scenario

In the second Scenario also 125 JNNs are estimated with the same settings as in the Scenario 1. Table 3 contains MSEs calculated for eq.(3) in Scenario 2. The results are assuredly similar as the best performances are obtained with 2 to 4 hidden neurons. λ equal to 0.7 gives again stable results for all variables, yielding the lowest MSEs, whereas other context unit weights that give low MSEs *in-the-sample*, have high MSEs *out-of-sample* and *vice versa*. After following the same procedure as in Scenario 1, the rankings of the top 10 JNNs are given in Table 4. The results indicate that among top 4 JNNs, three are estimated from eq.(3) and one is estimated from eq.(6), making HICP from a previous period and NEER the best two predictors. NEER is ranked among the top 10 three times, whereas eq.(3) 4 times. ULC is not selected among the top 10 neither in this scenario, whereas M3 is ranked the 10th. The highest ranked is JNN44 with optimal values of MSE and AIC *in-the-sample* and *out-of-sample*. λ is 0.7 in 8 out of 10 JNNs which leads to a conclusion that no matter which variables are used and for which periods, the best results are always obtained with λ of 0.7. Moreover, λ of 0.9 is selected as the best in 2 out of top 10 JNNs, i.e. higher values of λ are to be selected in forecasting inflation due to a long memory of the underlaying process. The number of hidden units is also indicative, i.e. in 7 out of 10 cases q is 2 or 3.

When comparing Scenarios 1 and 2, 8 out of top 10 JNNs have the same architecture setting, i.e. no matter which sample is chosen, the same JNNs, if chosen right, always give the best results. This is why it is important to select properly the NN architecture in general. It can be concluded that variable selection plays a key role in modelling and forecasting inflation where satisfying parsimony principle leads to a JNN that has both good in-the-sample and out-of-sample performances with resolved overfitting problem. The findings regarding the variables used are some in favour of previous research, and some yield opposite findings. [20] concludes that the models that include more information are better in inflation forecasting than models based on a narrower information set, while here JNN containing only one lag of inflation yields the best forecasts. [23] and [16] suggest that inflation forecasting using AR model can be significantly improved by several indicators, which is not the case here. [7] show good performance of the AR model, where the model containing some indicators is in most but not all cases better than the AR model and the best indicator changes over time, which was proven by this paper. NEER and M3 were important in [22] and [25] respectively. Moreover, the breaking point significantly influences MSEs as shown from comparison of 2 Scenarios, as there is a significant difference in out-of-sample MSEs with much better inflation predictions in Scenario 2. The unexpected deflation period influenced higher out-of-sample MSEs in Scenario 1 and higher in-the-sample MSEs in Scenario 2, i.e. none of the JNNs, including different scenarios and variables, could anticipate the deflation period.

	in-the-sample					out-of-sample				
q	λ=0.1	λ=0.3	λ=0.5	λ=0.7	λ=0.9	λ=0.1	λ=0.3	λ=0.5	λ=0.7	λ=0.9
1	0.5844	0.5763	0.5766	0.5767	0.5792	3.2638	3.4238	3.5016	3.4650	3.4040
2	0.0583	0.0578	0.0571	0.0576	0.0912	2.4692	2.8968	2.9498	1.6168	1.6992
3	0.0584	0.0574	0.0571	0.0549	0.0893	2.4398	2.8810	3.0127	2.1617	1.6725
4	0.0579	0.0596	0.0579	0.0528	0.0857	2.4980	3.0428	3.0945	1.6952	1.7955
5	0.0598	0.0681	0.0558	0.0533	0.0815	2.8900	3.2969	2.9599	2.1547	1.8571

Table 3: *In-the-sample and out-of-sample MSE calculated for JNNs in Scenario 2*

	ϕ	ϕ^+	ϕ^-	Eq.	MSEin	MSEout	AICin	AICout	q	λ	k
JNN44	0.502	0.532	0.029	eq.(3)	0.0528	1.6952	-565.86	40.86	4	0.7	17
JNN123	0.454	0.521	0.066	eq.(6)	0.0476	2.7123	-589.29	44.97	3	0.9	16
JNN42	0.446	0.509	0.063	eq.(3)	0.0576	1.6168	-564.39	24.25	2	0.7	9
JNN43	0.369	0.414	0.045	eq.(3)	0.0549	2.1617	-566.19	36.02	3	0.7	13
JNN117	0.360	0.475	0.115	eq.(6)	0.0511	3.0687	-584.81	36.58	2	0.7	11
JNN23	0.313	0.433	0.119	eq.(2)	0.0476	2.7123	-577.29	56.97	3	0.9	22
JNN17	0.268	0.396	0.128	eq.(2)	0.0511	3.0687	-576.81	44.58	2	0.7	15
JNN45	0.264	0.349	0.085	eq.(3)	0.0533	2.1547	-556.05	51.98	5	0.7	21
JNN118	0.245	0.383	0.128	eq.(6)	0.0510	3.0389	-575.27	46.45	3	0.7	16
JNN70	0.222	0.365	0.143	eq.(4)	0.0498	2.4484	-559.87	63.64	5	0.7	26

Table 4: *Top 10 JNNs with flows, equation, in-the-sample and out-of-sample MSE and AIC, hidden neurons, weight of the context unit and the number of parameters (k) in Scenario 2*

5. Conclusion

In this paper, the aim was to investigate JNNs in inflation modelling and forecasting. Namely, the influence of inputs and hidden units' selection, the choice of the context unit weight and different sampling on JNN performances *in-the-sample* and *out-of-sample* is examined. A total of 250 JNNs are estimated and compared using MSE and AIC. The optimal model selection follows two steps. Firstly, JNNs that do not learn well are eliminated from further examination. Secondly, using the PROMETHEE method, JNNs are ranked and the top 10 JNNs are compared within 2 scenarios. The results show that in most cases the simplest JNNs are ranked the highest, i.e. JNNs with lagged inflation or JNNs with additional variable (NEER or M3). ULC is not ranked among the top 10, i.e. it is not a good inflation predictor. It was expected that models containing more variables would perform better. This is not the case since only 1 or 2 out of the top 10 JNNs have all 4 variables. The research is not in accordance neither with [25] nor [16] suggesting that inflation forecasting using AR model can be significantly improved by several indicators. Both variable and sample selection play the key role in modelling and forecasting inflation where satisfying parsimony principle leads to a JNN that has both good *in-the-sample* and *out-of-sample* performances. Moreover, 80% of top 10 JNNs have 2 or 3 hidden neurons and the weight of the context unit of 0.7. The best JNN for inflation modelling and forecasting, no matter which scenario is observed, contains only lagged dependent variable with 2 and 4 hidden neurons in Scenario 1 and 2 respectively, and weight of the context unit of 0.7. An optimal number of hidden neurons was obtained using the rules of thumb, along with the cascading selection approach as in [4]. Moreover, optimal context unit weights have not been thoroughly investigated previously and if mentioned they were usually set in advance. This paper investigated which value is appropriate in this application as in [3]. Since the optimal context unit weight is 0.7 it can be concluded that it is data-driven and a very important parameter in selection of appropriate JNN structure.

Inflation has a key role for monetary and fiscal policy decision makers, where the appropriate inflation forecasting enables them to perceive in advance its direct and indirect effects on the future state of the economy. Therefore, this paper points to NNs in general and especially JNNs for forecasting inflation, which is important not only for researchers but also practitioners and policy makers. However, input selection is based on a theoretical model which limits its number. Selecting different inputs one can investigate the robustness of the results. The research is conducted on aggregated euro zone level. It would be worth verifying the robustness of the selected architecture in other countries. Moreover, there is still a challenge for the researchers to improve the performances of NNs in forecasting time-series in general. The use of different

algorithms or other NN architectures with different fixed and varying parameters opens up space for future work and further studies. Directions for future research include modelling and forecasting time-series using different RNNs since other types have been developed, which are proven to be more efficient than standard NNs, whose results are insufficiently analysed in economics.

Acknowledgements

References

[1] Al-Maqaleh, B. M., Al-Mansoub, A. A. and Al-Badani, F. N. (2016). Forecasting using Artificial Neural Network and Statistics Models. *International Journal of Education and Management Engineering*, 6(3), 20–32. doi: 10.5815/ijeme.2016.03.03

[2] Araz O. U. (2005). A Simulation Based Multi-criteria Scheduling Approach of Dual-Resource Constrained Manufacturing Systems with Neural Networks, 1047–1052. In Zhang, S. and Jarvis, R. (Eds): *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), 3809, Berlin: Springer. doi: 10.1007/11589990_138

[3] Arnerić, J., Poklepović, T. and Aljinović, Z. (2014). GARCH based artificial neural networks in forecasting conditional variance of stock returns. *Croatian Operational Research Review*, 5(2), 329–343. doi: 10.17535/crorr.2014.0017

[4] Arnerić, J., Poklepović, T. and Teai, J. W. (2018). Neural Network Approach in Forecasting Realized Variance Using High-Frequency Data. *Business Systems Research Journal*, 9(2), 18–34. doi: 10.2478/bsrj-2018-0016

[5] Azadnia, A. H., Saman, M. Z. M., Wong, K. Y., Ghadimi, P. and Zakuan, N. (2012). Sustainable Supplier Selection based on Self-organizing Map Neural Network and Multi Criteria Decision Making Approaches. *Social and Behavioral Sciences*, 65, 879–884. doi: 10.1016/j.sbspro.2012.11.214

[6] Balkin, S. (2001). On Forecasting Exchange Rates Using Neural Networks. *International Journal of Forecasting*, 17(1), 139–140. doi: 10.1016/s0169-2070(00)00047-9

[7] Banerjee, A., Marcellino, M. and Masten, I. (2005). Leading Indicators for Euro-area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics*, 67, 786–813. doi: 10.1111/j.1468-0084.2005.00141.x

[8] Bengio, Y., Simard, P. and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi: 10.1109/72.279181

[9] Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M. and Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. *Applied Economics*, 37(6), 665-680. doi: 10.1080/0003684052000343679

[10] Binner, J. M., Elger, C. T., Nilsson, B. and Tepper, J. A. (2006). Predictable non-linearities in U.S. inflation. *Economic Letters*, 93, 323-328. doi: 10.1016/j.econlet.2006.06.001

[11] Binner, J. M., Jones, B., Kendall, G., Tino, P. and Tepper, J. (2007). Evolution, recurrency and kernels in learning to model inflation. *Aston Business School Working Paper*, 6, Birmingham: Aston University. http://publications.aston.ac.uk/id/eprint/24158

[12] Binner, J. M., Elger, T., Nilsson, B., Tepper, J. A. (2004). Tools for non-linear time series forecasting in economics - an empirical comparison of regime switching vector autoregressive models and recurrent neural networks, 19, 71–91. In Binner, J. M., Kendall, G. and Chen, S. H. (Eds): *Applications of Artificial Intelligence in Finance and Economics* (Advances in Econometrics), Emerald Group Publishing Limited. doi: 10.1016/s0731-9053(04)19003-8

[13] Brezak, D., Bacek, T., Majetic, D., Kasac, J. and Novakovic, B. (2012). A comparison of feed-forward and recurrent neural networks in time series forecasting. *IEEE Conference on Computational Intelligence for Financial Engineering and Economics*, 1–6. doi: 10.1109/cifer.2012.6327793

[14] Bruno, M. and Melnick, R. (1995). High Inflation Dynamics: Integrating Short–Run Accommodation and Long–Run Steady–States, *Bank of Israel Working Paper*, 6. https://ideas.repec.org/p/boi/wpaper/1995.06.html

[15] Chaudhuri, T. D and Ghosh, I. (2016). Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework. *Journal of Insurance and Financial Management*, 1(5), 92–123.

[16] Chauvin, V. and Devulder, A. (2008). An inflation forecasting model for the euro area. *Banque de France Working Paper*, 192. doi: 10.2139/ssrn.1680375

[17] Choudhary, M. A. and Haider, A. (2012). Neural network models for inflation forecasting: an appraisal. *Applied Economics*, 44(20), 2631–2635. doi: 10.1080/00036846.2011.566190

[18] Fagan, G. and Morgan, J. (2005). *Econometric Models of the Euro-area Central Banks*. Edward Elgar Publishing. doi: 10.4337/9781845428037

[19] Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. doi: 10.1016/0893-6080(89)90020-8

[20] Jansen, E. S. (2004). Modelling Inflation in the Euro Area. *ECB Working paper*, 322.

[21] Jaumotte, F. and Morsy, H. (2012). Determinants of Inflation in the Euro Area: The Role of Labor and Product Market Institutions. *IMF Working Paper*, 12(37), 1–37. doi: 10.5089/9781463933807.001

[22] Jordan, M. I. (1997). Serial Order: A Parallel Distributed Processing Approach. *Advances in Psychology*, 121, 471–495. doi: 10.1016/s0166-4115(97)80111-2

[23] Kozicki, S. (2001). Why do central banks monitor so many inflation indicators? *Federal Reserve Bank of Kansas City - Economic Review*, 86(3), 5–42.

[24] Kuan, C. M. and Liu, T. (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, 10(4), 347–364. doi: 10.1002/jae.3950100403

[25] Lim, Y. C., and Sek, S. K. (2015). An Examination on the Determinants of Inflation. *Journal of Economics, Business and Management*, 3(7), 678–682. doi: 10.7763/joebm.2015.v3.265

[26] Lolić, I. and Sorić, P. (2015). Nonlinear effects in the Croatian inflation generating process. *International Journal of Monetary Economics and Finance*, 7(3), 175–191. doi: 10.1504/ijmef.2014.066483

[27] Malešević Perović, L. (2009). Kointegracijski pristup analizi inflacije u Hrvatskoj. *Financijska teorija i praksa*, 33(2), 201–218.

[28] Mańdziuk, J. and Mikołajczak, R. (2002). Chaotic time series prediction with feed-forward and recurrent neural nets. *Control and Cybernetics*, 31(2), 1–24

[29] McNelis, P. and McAdam, P. (2005). Forecasting inflation with thick models and neural networks. *Economic Modelling*, 22(5), 848–867. 10.1016/j.econmod.2005.06.002

[30] Moretti, L. (2013). Determinants of inflation differentials in the euro area. *Center for Financial Studies Working Paper*, Goethe University Frankfurt.

[31] Moshiri, S. and Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3), 201–217. doi: 10.1002/(sici)1099-131x(200004)19:3%3C201::aid-for753%3E3.3.co;2-w

[32] Moshiri, S., Cameron, N. and Scuse, D. (1999). Static, Dynamic, and Hybrid Neural Networks in Forecasting Inflation. *Computational Economics*, 14(3), 219–235. doi: 10.1023/A:1008752024721

[33] Patterson, D. W. (1995). *Artificial Neural Networks: Theory and Applications*. Prentice Hall

[34] Stoica, O. and Damian, M. (2013). The Analysis of Inflation Rate Dynamics in Central and South-Eastern Europe's States in the Context of The EU Accession. *International Journal of Economics and Finance*, 5(2), 102–110. doi: 10.5539/ijef.v5n2p102

[35] Sundermeyer, M., Oparin, I., Gauvain, J. L., Freiberg, B., Schluter, R. and Ney, H. (2013). Comparison of feedforward and recurrent neural network language models. *IEEE International Conference on Acoustics, Speech and Signal Processing*. doi: 10.1109/icassp.2013.6639310

[36] Teräsvirta, T., van Dijk, D. and Medeiros, M. C. (2005). Linear models, smooth transition autoregressions and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21(4), 755–774. doi: 10.1016/j.ijforecast.2005.04.010

[37] Wang, J., Wang, J., Fang, W. and Niu, H. (2016). Financial Time Series Prediction Using Elman Recurrent Random Neural Networks. *Computational Intelligence and Neuroscience*, ID 4742515, 1–14. doi: 10.1155/2016/4742515

[38] Zhang, G. P. (2007). Avoiding Pitfalls in Neural Network Research. *IEEE Transactions on Systems, Man and Cybernetics*, Part C (Applications and Reviews), 37(1), 3–16. doi: 10.1109/tsmcc.2006.876059