

FORECAST OF CONSUMER BEHAVIOUR BASED ON NEURAL NETWORKS MODELS COMPARISON

M. Štencl, O. Popelka, J. Šťastný

Received: November 30, 2011

Abstract

ŠTENCL, POPELKA, ŠŤASTNÝ: *Forecast of consumer behaviour based on neural networks models comparison.* Acta univ. agric. et silvic. Mendel. Brun., 2012, LX, No. 2, pp. 437–442

The aim of this article is comparison of accuracy level of forecasted values of several artificial neural network models. The comparison is performed on datasets of Czech household consumption values. Several statistical models often resolve this task with more or fewer restrictions. In previous work where models' input conditions were not so strict and model with missing data was used (the time series didn't contain many values) we have obtained comparably good results with artificial neural networks. Two views – practical and theoretical, motivate the purpose of this study. Forecasting models for medium term prognosis of the main trends of Czech household consumption is part of the faculty research design grant MSM 6215648904/03/02 (Sub-task 5.3) which defines the practical purpose. Testing of nonlinear autoregressive artificial neural network model compared with feed-forward neural network and radial basis function neural network defines the theoretical purpose. The performance metrics of the models were evaluated using a combination of common error metrics, namely Correlation Coefficient and Mean Square Error, together with the number of epochs and/or main prediction error.

artificial neural networks, forecasting methods, customer behaviour

Prediction of tendencies in consumer behaviour of Czech (or any other) households is one of key components of determining current economic development. Especially in the period of economic recessions and shocks, household consumption becomes one of the leading elements of GDP growth of both Czech Republic (Singer, 2007) and Slovak Republic (Senaj, 2007). Household consumption is also the most stable and therefore the most predictable part of demand. This is caused by a connection to development of real wages, where another connection can be derived to the trend of running in debts. Knowing the trend of running in debts is also very important as it influences the development of whole household consumption. The influence is particularly important at time of economic oscillations, which is a very current topic.

For constructing the above mentioned prognoses, various statistical methods are heavily used. These have their natural limitations, so other approaches are being evaluated. One of such approaches are algorithms, which take advantage

of statistical learning, but they are built on different fundamentals. During the work on research design grant No. MSM 3215648904/03/02 – *Development of relationships in the business sphere as connected with changes in the life style of purchasing behaviour of the Czech population and in the business environment in the course of processes of integration and globalization* we have created several prediction models based on artificial neural networks. Testing proved that in specific cases the approximation capabilities (including the following prognosis) of these methods were excellent. The approximation tests were performed against a dataset containing Czech household consumption data collected using the COICOP methodology.

The objective of this article is to evaluate results obtained with different artificial neural networks models compared to real development of a given household consumption indicator. We use nonlinear auto-regressive model of neural network, which is further compared with previously tested models based on artificial neural networks (ANN). Specifically – multilayer perceptron neural

network (MLP) and radial basis functions network (RBF). Part of the evaluation is also assessment of applicability of used neural networks for forecasting main tendencies of consumer behaviour of Czech households – mainly the adaptation capabilities of these models.

Actual experiments include testing of different types of neural networks for prediction of trends in indexes of consumer behaviour. For the testing we have used data sets of European household consumption indicators, which are created using the COICOP methodology. The prognosis of specific indicators computed by one neural network type is compared to the real value (not included in the training data set) and also to the value computed by a different type of neural network. In the conclusion of this paper, we formulate possible advantages of using neural networks for constructing medium-term prognosis of development trends. This includes selecting the appropriate neural network model for a given specific household consumption indicator.

Current State of the Art

Models for prognosis construction are usually based on principles of the Box-Jenkins methodology. As stated in (Artl *et al.*, 2007) this methodology is also used as a base for modelling non-stationary time series, time series with long memory and seasonal time series. Class AR, MA and ARMA time series models are also characteristic with a specific form of autocorrelation function and partial autocorrelation function. Models based on Box-Jenkins methodology therefore first identify stationary parts of a given time series, and then seasonal parts with their respective auto-correlating functions. In the second phase, parameters of the ARMA model are determined using a computational model, which best fits the course of cleaned time series. Again, there are a number of methods available for computing what is the best fitting model. Notable methods include MLE (maximum-likelihood estimation) and non-linear least square method (Popelka *et al.*, 2009). When a best fit function is found, validation and testing of the model can be performed. In case either of validation or testing fails, the process restarts from the very beginning – identifying stationary and seasonal parts of the time series. We obtain a fitting model, when both testing and validation are passed. Then the actual prognosis of the time series can be constructed (Artl *et al.*, 2007).

All of the abovementioned methods or time series types (AR, MA and ARMA) are based on linear functions. There are also non-linear GARCH class models (Hastie *et al.*, 2009; McNelis, 2005). An excellent overview of used non-linear models of time series is given in (De Gooijer and Kumar, 1992). Apart from the mentioned linear and non-linear methods, a number of various other methods are used in experiments (Mills, 2009). For example it is a simple moving average method is used

(Makridakis *et al.*, 1998); intuitive (expert) methods such as DELPHI (Lawrence *et al.*, 2006); regression analysis (Montgomery *et al.*, 2008), cell-based analysis (Anderson, 2000); methods of predictive modelling (Senensky, 2008).

Although the Box-Jenkins methods are successful, it is also possible to find a number of successful applications of statistical learning and machine learning. The former group includes for example the Support Vector Machines (SVM) method. The latter group includes models based on artificial neural networks. As we mentioned in the introduction – this article concentrates on artificial neural networks.

In (Zhang *et al.*, 2004) three main reasons for using neural networks are stated as: neural networks are capable of adaptation to given data; neural networks are capable of generalization even when the input data set contains noise or missing values (trained network has the capability of correctly filling the value without affecting the prediction); neural networks act as an universal approximator for an arbitrary continuous function with arbitrary precision (Hornik, 1993; Šíma *et al.*, 1996). It is also worth mentioning that neural networks have strong capabilities of producing compromise, while they introduce uncertainty (Zhang *et al.*, 2004).

It is important to note that neural networks are non-linear models unlike the widely used Box-Jenkins models, which are linear. Since neural network is more of a generic framework than a ready-to-use model, there is high number of concrete models and equally high number of different applications. Among those, which most notably influenced the evolution of neural networks (both in applications and research), there are multi-layer perceptron network, Hopfield network and Kohonen self-organizing map (SOM) (Zhang *et al.*, 2004). Each of these mentioned network types is a representative of a different approach to memory management and different (possibly better) results for different types of applications. For working (especially forecasting) with time-series the multi-layer perceptron network proved to be most successful in experimental tests. For applications of universal approximators the Radial Basis Function networks are also used as well as networks with semi-local units (Šíma *et al.*, 1996).

Multi-layer perceptron neural network is considered to be the most universal neural network. It belongs to the group of *static neural networks*. In a simplified way, this means that this network is based on algebraic expressions (Dreyfus *et al.*, 2005). These networks are also part of *feed-forward neural networks*. MLP network is most widely used network type for predictions of time series and also for analysis and forecasting of deterministic time series with or without noise or time series with missing data. A number of articles have been published in the areas of finance, banking and economics (McNelis, 2005; Wang, 2009; Zhang *et al.*, 2004).

Apart from the group static neural networks, there are *dynamic neural networks*. This group includes the Hopfield network mentioned before and recurrent (cyclic) neural networks. These include Elman models of neural networks (Dreyfus *et al.*, 2005) and non-linear autoregressive neural networks (NAR) (Khashei *et al.*, 2010, Yu *et al.*, 2009). Underlying models of dynamic neural networks are based on differential or partial differential equations, where the independent variable is time (Dreyfus *et al.*, 2005).

Used data sets and metrics

To be able to reliably assess the usability of a given method for a given application, it is necessary to verify the computational method on a representative reliable set of data. In our experiments the data constitute of statistics individual consumption expenditure of households collected using the COICOP methodology. This methodology was defined by United Nations Statistics Division. Together with the data we have defined the comparison metric for evaluating partial results. The following two sections do explain these in detail.

Individual consumption according to COICOP

The COICOP (Classification of Individual Consumption According to Purpose) methodology is defined by United Nations Statistics Division. Together with three other classifications (COFOG, COPNI, COPPI) it falls into census category, which tracks individual consumption according to purpose. Within each of these censuses the primary data was gathered by individual local organizations under the supervision of corresponding national statistical office. The data are divided using national designation – for example the data set with Czech households is marked CZ-COICOP. Each data set is further divided into 14 divisions, 58 groups and 157 classes. Each dataset provides information about the amount of expenses and about structure of expenses of individual households. Within each expenses group, the data can be sorted by various factors and influences (e.g. price movements, market situation) on the structure of individual expenses, which cannot be obtained from other sources.

In Czech Republic the survey was conducted on a set of 3000 households selected by quota sampling (Czech Statistical Office, 2009). The actual statistic set contains two sub-sets – *basic* and *supplementary*. The basic set was selected so that its' division represents the actual division of households in Czech Republic according to social groups and other criteria. Supplementary set includes 400 households and was created in order to obtain data from households with minimum income. Although these are included in the basic set, their amount there is not sufficient for ensuring representativeness of the data. Individual consumption does not include investments, construction expenses etc. The data

was collected during years 1999 to 2008 with quarter of a year period.

In this article, the described dataset was used mainly for verifying validity of used methods. For actual deduction of general conclusions a more reliable dataset should be used. The data set for experiments was chosen within previously published works (Štencl *et al.*, 2010).

Comparison metrics

The overall quality of models was evaluated using a combination of widely used metrics. Specifically it is combination of *correlation coefficient* (R), *mean square error* (MSE), together with number of epochs necessary to train the network for a given precision. This metric comes out of metrics generally used in regression analysis and forecasting (correlation coefficient). It also includes mean square error, which is a widely used method of evaluating neural network performance. These criteria are combined because in case of neural networks the correlation coefficient might not have enough explanatory power. All of these partial measurements are implemented directly in the core of MATLAB environment (2011a, The MathWorks, Natick, Massachusetts, U.S.A.), which was used for the experiments.

Correlation coefficient (R) was computed using the equation (1), which is based on standard normalized correlation coefficient implemented in MATLAB.

$$R^2 = 1 - (SS_{res} / SS_{tot}) \times ((n-1) / (n-d-1)) \quad (1)$$

SS_{res} is the sum of squares of residuums of a given regression function. SS_{tot} is sum of square errors of mean value of dependent variable (total sum of squares), n is number of observations in the dataset (i.e. data set size) and d is polygon degree. Correlation coefficient R describes the dependency strengthens between input data (values computed by a given neural network model) and target data (sample values from the dataset). When dealing with forecasting it describes prediction precision. If $R = 1$, then the dependency between input and target is strong. The objective of experiments is to find a model whose value of correlation coefficient is approaching 1. If $R = 0$ then there is no dependency between the data and the model is unusable. The proposed definition of normalized correlation coefficient overcomes differences between various used models and is thus more objective when comparing them.

$$NMSE = \frac{1}{n\sigma^2} \sum_{i=1}^n (x_i - d_i)^2 \quad (2)$$

Computation of normalized MSE was performed using the equation (2), where σ^2 is scatter of target values d_i and n is the size of data set.

The third measured value is number of epochs necessary for learning for an accuracy of 0.01 gradient value. This value was determined experimentally and proved to be optimum value for ensuring good approximation of the data set while still maintaining good forecasting capabilities for four following steps (quarters). For experiments with RBF the actual prediction error was also tracked, since in the implementation in MATLAB does not allow tracking of actual values of *R* and *MSE* when training the network. Also it is important to note that all predictions made are point forecasts.

RESULTS

Table 1 shows results obtained by various topologies of MLP neural network. All tested networks were trained using Levenberg-Marquardt algorithm and using the metrics described above. Each value in the topology column represents the number of neurons in *input*, *hidden* and *output* network layer, where (2) denotes over-fitted topology.

Since the dataset is not particularly large, networks with smaller topologies perform slightly better. When more than ten neurons were used in the hidden layer, the effect of over-fitting was becoming apparent. That is, when in training mode, the overall error of the network decreases; but in testing mode, the predicted values have higher deviations. Also it is interesting to note that both values of the comparison metric report different best results.

Overall it can be seen that there is no single best topology for prediction of a given time-series. Therefore the result confirms previous experience – that MLP network is a will not find a single solution for the input data, which produces oscillations and instability. Given the chosen metric, we can

see the last 10 neuron topology as a best result. The 10-neuron topologies also have much higher number of training epochs which implies that the approximation elasticity of the network is also higher.

Results obtained using RBF neural networks only confirmed previous experiments, that the time-series is not long enough for proper training. Therefore the results of RBF networks were much worse than those of MLP, the only positive being quicker training.

Last part of the experiment was using non-linear autoregressive networks (NAR). NAR networks are also implemented in MATLAB environment (using *narnet* function). Topology of a NAR network is defined using function *preparets*, this function prepares input and target data for training and testing (simulation) of the network. This function normalizes both input and target data with normal distribution and constant level of noise, so that they can be used in either training or simulation phase. Also this function ensures correct switching between input and target data as many times as necessary for correct fill-up of the *input layer* and *delay states layer*. We have again used the Levenberg-Marquardt learning algorithm in its' standard implementation in MATLAB function *trainlm*.

The results obtained for NAR models show that correlation coefficient itself is incapable of evaluating neural networks. In all tests performed on NAR networks, the values of correlation coefficient were almost identical.

In this section we have presented most important results of our experiments with chosen neural network types. The selection of tested neural networks was driven by experience of the authors and by the research conducted earlier and briefly described in the beginning of this paper. The

I: Results obtained for MLP networks

Topology	Learning stage			Testing stage	
	MSE	R	Epoch	MSE	R
1-3-1 (1)	1,37E-04	9,93E-01	14	7,25E-05	1,00E+00
1-3-1 (2)	1,09E-04	9,97E-01	7	3,09E-04	9,62E-01
1-5-1	1,77E-04	9,09E-02	7	1,86E-04	9,96E-01
1-7-1	8,36E-05	9,96E-01	12	2,01E-04	9,74E-01
1-10-1 (1)	1,87E-04	9,91E-03	32	3,31E-04	1,00E+00
1-10-1 (2)	1,02E-04	9,95E-01	28	3,90E-06	9,99E-01

II: Results obtained of NAR model

Topology	Learning stage			Testing stage	
	MSE	R	Epoch	MSE	R
1-3-1	1,44E-04	9,89E-01	9	1,94E-04	1,00E+00
1-5-1 (1)	1,15E-04	9,92E-01	10	1,93E-04	1,00E+00
1-5-1 (2)	1,12E-05	9,99E-01	18	1,36E-04	1,00E+00
1-7-1	7,41E-05	9,95E-01	10	1,71E-04	1,00E+00
1-10-1	8,02E-05	9,89E-01	8	5,34E-05	1,00E+00

metric used for evaluation the networks is based on combination of existing methods from the field of neural networks (MSE) and economic application (R). Our tests have shown that it is not appropriate to rely on a single method. It is actually the combination of different criteria (while taking

into account another ones – such as the number of epochs), which enables us to more precisely evaluate the results. Also the experiments with NAR networks have shown that the conclusiveness of correlation coefficient (R) is lower.

SUMMARY

Overall the best prediction results were obtained using the NAR neural networks. This is not a surprising conclusion, since the complexity of the NAR method introduces a tremendous potential in improving prediction precision. Another important property of NAR is the capability to work with larger number of highly correlated variables. Quality of universal method was also proven by MLP neural network, which in some cases outperformed the NAR model. On the other, higher sensitivity to input data observed in MLP models was also confirmed.

Construction of a simple prognostic model is a very complex task. Among the most used method are statistical methods based on moving averages or methods based on Box-Jenkins methodology. The resulting models can be either widely known linear statistical models (AR, MA, ARIMA), or less known non-linear models. As shown at the beginning of this paper, there are also published results where neural networks are used for this kind of tasks. The objective of this paper was to compare different types of artificial neural network models, with the verification of the results performed on a real development of a given index of household consumption.

Acknowledgement

This work was supported by research design No. MSM 6215648904/03/02.

REFERENCES

- ANDERSON, J. M., 2000: *Computer Models for Retirement Policy*. Paper presented at the Spring Retirement Meeting, Las Vegas.
- ARTL, J., ARTLOVÁ, M., 2007: *Ekonomické časové řady*. Grada Publishing.
- CZECH STATISTICAL OFFICE, 2009: *Statistika rodinných účtů (vydání a spotřeba domácností)*. [cit. 2009-06-24]. Cited from <http://www.czso.cz/csu/redakce.nsf/i/rodinne_ucty>.
- DE GOOIJER, J. G., KUMAR, K., 1992: *Some recent developments in non-linear time series modelling, testing, and forecasting*. International Journal of Forecasting 8, pp. 135–156.
- DREYFUS, G., MARTINEZ, J. M., SAMUELIDES, M., GORDON, M., BADRAN, F., THIRIA, S., HÉRAULT, L., 2005: *Neural Networks – Methodology and Applications*. Springer, Berlin.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., 2009: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edition. Springer Series in Statistics. Springer.
- HORNIK, K., 1993: *Some new results on neural network approximation*. Neural Networks 6, pp. 1069–1072.
- KHASHEI, M., BIJARI, M., 2010: *An artificial neural network (p, d, q) model for timeseries forecasting*. Expert Systems with Applications, 37 (1).
- LAWRENCE, M., GOODWIN, P., O'CONNOR, M., ONKAL, D., 2006: *Judgemental Forecasting: A Review of Progress Over the Last 25 Years*. International Journal of Forecasting, 22, pp. 493–518.
- MAKRIDAKIS, S. G., WHEELWRIGHT, S. C., HYNDMAN, R. J., 1998: *Forecasting: Methods and Applications*. New York: John Wiley & Sons.
- MILLS, A., 2009: *Introduction to Forecasting Methods for Actuaries*. Forecasting & Futurism.
- MCNELIS, P. D., 2005: *Neural networks in finance – gaining predictive edge in the market*. Elsevier Academic Press.
- POPELKA, O., ŠTASTNÝ, J., 2009: *Uplatnění metod umělé inteligence v zemědělsko-ekonomických predikčních úlohách*. Brno, Mendelova zemědělská a lesnická univerzita v Brně.
- SENAJ, M., 2007: *Odhad spotřebnej funkcie pre Slovensko a prognóza spotreby*. Výskumná štúdia, 1/2007. Národná banka Slovenska.
- SENENSKY, B., 2008: *Predictive Modeling*. CompAct (SOA Technology Section newsletter).
- SINGER, M., 2007: *Zadluženost domácností v ČR podle poznatků ČNB*. Euro setkání „Život na dluh“. Česká národní banka.
- ŠÍMA, J., NERUDA, R., 1996: *Theoretical Issues of Neural Networks*. Prague: MATFYZPRESS, MFF UK. [cit. 2011-09-24]. Cited from <<http://www2.cs.cas.cz/~sima/kniha.pdf>>.
- ŠTENCL, M., ŠTASTNÝ, J., 2010: *Neural network learning algorithms comparison on numerical prediction of real data*. In: MENDEL 2010, 16th International Conference on Soft Computing. Brno University of Technology.
- WANG, Y.-H., 2009: *Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach*. Expert Systems with Applications, 36 (1).

- YU, L., WANG, S., KEUNG LAI, K., 2009: *A neural-network-based nonlinear metamodeling approach to financial time series forecasting*. Applied Soft Computing, 9 (2).
- ZHANG, G. B., PATUWO, E., Hu, M. Y., 1998: *Forecasting with artificial neural networks: The state of the art*. International Journal of Forecasting, 14 (1), pp. 35–62.

Address

Ing. Michael Štencl, Ph.D., Ing. Ondřej Popelka, Ph.D., doc. RNDr. Ing. Jiří Štastný, CSc., Ústav informatiky, Mendelova univerzita v Brně, 613 00 Brno, Česká republika, e-mail: michael.stencl@mendelu.cz, jiri.stastny@mendelu.cz, ondrej.popelka@mendelu.cz