INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 2(52) • 2019
ISSN 1507-3858
e-ISSN 2450-0003

Tomasz Jasiński
Lodz University of Technology, Faculty of Management and Production Engineering
e-mail: tomasz.jasinski@p.lodz.pl
ORCID: 0000-0001-9367-0086

Anna Bochenek
Lodz University of Technology, Faculty of Civil Engineering,
Architecture and Environmental Engineering
e-mail: anna.bochenek@p.lodz.pl
ORCID: 0000-0003-3396-2006

MODELLING OF ARREARS IN PAYMENTS FOR DWELLING USING ARTIFICIAL NEURAL NETWORKS

MODELOWANIE ZALEGŁOŚCI W OPŁATACH ZA MIESZKANIA PRZY UŻYCIU SZTUCZNYCH SIECI NEURONOWYCH

DOI: 10.15611/ie.2019.2.04
JEL Classification: C4, D1, E17

Summary: The study presents the construction process of a model that forecasts arrears in dwelling payments in individual municipalities in Poland depending on the values of economic data from previous years. This enables to estimate arrears that will arise in the investigated municipalities in the year of analysis. The model constructed on the basis of artificial neural networks, which is a tool from the area of artificial intelligence, was used to carry out forecasts. More than one hundred thousand networks with multilayer perceptron (MLP) and radial basis function (RBF) architectures were tested. The MAPE for prediction of the number of indebted dwellings in municipalities with at least 50 indebted premises was 6.08%. The correctness of forecasts in the area of the direction of changes of household debt in municipalities between 2014 and 2015 was 76.84%.

Keywords: payment arrears, household debt, forecasts, municipalities, artificial neural networks.

Streszczenie: Opracowanie przedstawia proces budowy modelu prognozującego zaległości w opłatach mieszkaniowych w poszczególnych gminach w Polsce w zależności od wartości danych ekonomicznych z lat poprzednich. Umożliwia to oszacowanie zaległości, jakie będą kształtować się w badanej gminie w roku dokonywania analizy. Do przeprowadzenia prognoz został użyty model zbudowany na bazie sztucznych sieci neuronowych będących narzędziem z obszaru sztucznej inteligencji. Testom zostało poddanych ponad sto tysięcy sieci o architekturze perceptronu wielowarstwowego (MLP) oraz o radialnych funkcjach bazowych (RBF). MAPE dla prognozy liczby zadłużonych nieruchomości (z zaległościami w opłatach mieszkaniowych) w poszczególnych gminach wyniósł 6,08% (badaniu poddano gminy, w których zadłużenie występowało w minimum 50 nieruchomościach). Poprawność prognoz kierunku zmian zadłużenia gospodarstw domowych w gminach w latach 2014-2015 wyniosła 76,84%.

Słowa kluczowe: zaległości płatnicze, zadłużenie gospodarstw domowych, prognozy, gminy, sztuczne sieci neuronowe.

1. Introduction

The possibility of forecasting households (HHs) debt is an important element of economic policy, not only at central level, but also at the level of municipal activities. It should be noted that the knowledge of future arrears in housing payments, which de facto also represent a loss of revenue for municipal budgets, allows both for financial decisions to be taken in advance in order to balance the budgets of individual municipalities and for measures to be implemented that shape the negative trends in the arising HHs debt.

HHs debt is undoubtedly one of the most important factors affecting the quality of life of people. HHs liabilities to a large extent are financed from credits, loans and other liabilities (over 40% of liabilities were financed from external sources – data for 2016). Thus, HHs are charged with additional fees (commissions and bank charges). Every year, the level of financial liabilities of the owners of HHs increases. In 2014, HHs' average debt amounted to PLN 57,700, and two years later it increased to PLN 64,400 [National Bank of Poland 2017]. Already in the first quarter of 2017, the ratio of HHs' financial liabilities to their annual disposable income reached 62.2%, and the rate of voluntary savings decreased to –0.5% [Kolasa 2017]. According to data from the Central Statistical Office [2018], expenditure on the use of flat or house, water supply, electricity, gas and other fuels in 2008-2016 ranged from 21.2% up to 22.8% of total consumption in the HH sector. Obviously, the financial problems of the HHs are reflected in the arrears in dwelling payments.

The aim of this study was to develop a model to forecast the number of indebted premises and the direction of annual change in the value of debt in municipalities. For this purpose, information from the digital databases of the Central Statistical Office was used. The information was obtained for the basic territorial units – urban municipalities, urban-rural municipalities and rural communes. The authors used artificial neural networks (ANNs), which are one of tools in the field of artificial intelligence, to construct forecasting models.

2. Bases of artificial neural networks

The beginning of ANN dates back to 1943, when McCulloch and Pitts [1943] created an artificial neuron model. It constituted a structure with a high degree of analogy to its biological equivalent. A schematic comparison of the artificial and biological structure of a nerve cell is presented in Figure 1, as x_1 to x_n were presented in the so-called inputs equivalent to structures called dendrites, of which the main purpose is to introduce signals into the cell. A feature of an artificial neuron is that it is linked to each of inputs of a single real number, called a weight. For each of the inputs, the value of the input signal is multiplied by the weight associated with it. The resulting products are added together and constitute the so-called total stimulation of neuron(s), which is then is transformed by the use of the so-called activation function. To some extent, this is equivalent to the nucleus of a biological nerve cell. Its function is one of numerous transformations, among which the following functions should be mentioned: logistic, hyperbolic tangent, identity (relatively rarely used in economic analyses), modified sine, and Gaussian. The neuron output determined y is the equivalent of a biological axon and is responsible for sending the signal from the nerve cell to next neurons, or it is an output of the entire neural network.

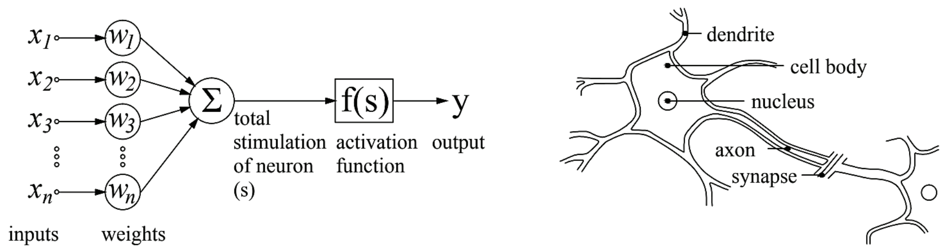

Fig. 1. Comparison of the artificial (left) and biological structure (right) of a nerve cell

Source: own elaboration.

2.1. Types of artificial neural networks

The connected neurons form the ANN. Depending on technical aspects such as the way they are connected, the direction of signal flow between them and the used training methods, many types of ANNs can be distinguished. The basic division comprises two categories of networks: (i) trained in supervised mode (with a so-called teacher) and (ii) operating in unsupervised mode. In the second case, it is not necessary to know the output patterns. The following part of the description of the construction and operation of ANN refers only to the network trained with the teacher, with full knowledge of actual values of the output patterns for all data used in the ANN training process.

Another division of the network is carried out based on the direction of signal flow between neurons. A distinction is made between (i) feedforwarded and (ii) recurrent ANNs. In case of the former, the signals are transmitted only in one direction, i.e. from the input to output of the network. The cells are grouped into so-called layers and the signals move from the nerve cells of a given layer to the neurons grouped in the next layer (counting from the entry). Recurrent networks contain feedback, i.e. connections that transmit signals to neurons grouped in earlier layers (closer to the network input) or to nerve cells of the same layer. This approach results in some time lag (in case of analyzing data in time series). Forecasts are then carried out using pre-processed data from earlier periods. As the data used in empirical studies described in this paper were not time series, it would be unreasonable to test recurrent ANNs.

Two types of feedforwarded neural networks were used in the study: multilayer perceptron (MLP) and radial basis function (RBF). The first one may occur in many variants differing not only in the number of used nerve cells, but also above all in their different grouping into aforementioned layers. MLP has only three types of layers. A single input layer (which neurons do not gain knowledge, but perform only technical functions in a form of introducing external signals to the model), optional hidden layers (their number in theory is unlimited, but in real applications it is usually limited to two, sometimes three) and one output layer. The diagram of MLP is presented in Figure 2 (left).

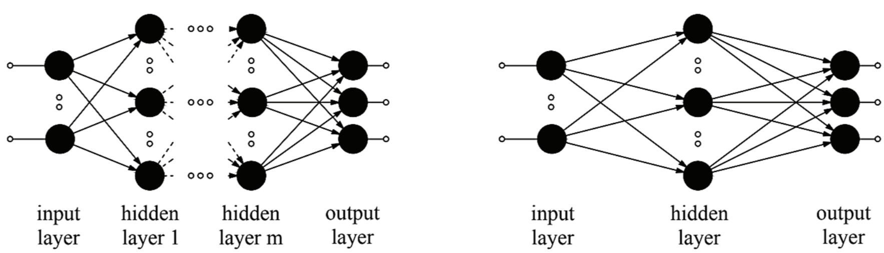

Fig. 2. The diagram of MLP (left) and RBF (right) network

Source: own elaboration.

The RBF network concept is derived from the MLP. It always has three layers. The nerve cells of a hidden layer are built on the basis of radial activation functions (hence the name of network). An example is the Gaussian function, whose values change symmetrically around the center point. Neurons of the output layer, by calculating the weighted sum of output signals of neurons of the hidden layer, aggregate their values. This implies the use of linear activation functions. The diagram of construction of RBF-type networks is presented in Figure 2 (right).

2.2. Supervised training artificial neural networks

In the ANN training using a program called teacher, which is used during the study, it is necessary to have output patterns. On their basis, the error generated by individual nerve cells is calculated, and then changes are introduced, resulting in an increase in the quality of the model's functioning. These modifications concern the weights assigned to inputs of the learning neuron layers, i.e. hidden and output layers. There are many ways to train ANN. The most popular includes those created in 1974, i.e. the method of error back propagation and its modified version with momentum. Amongst other known training algorithms one can include the conjugate gradient method and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. Their common feature is the necessity to separate from the set of all patterns (input and output) the training subset. Its multiple presentation of ANN along with the activation of training mechanisms (weight modification) results in the network acquiring knowledge. The model error decreases asymptotically to zero with the training time.

Since too long training could lead to so-called overlearning ANNs, it is necessary to determine the moment at which the training process should be stopped. For this purpose, another subset (less numerous) called the validation set is distinguished (in the subject literature, this term is used interchangeably with the test word). During the ANN presentation, this set of data does not run algorithms that modify the weights, which means that the network cannot learn on the basis of these data. They are only used to determine the direction of the model error change. When it reaches its minimum value and starts to grow, it means that the training process has to stop, as its continuation would consequently lead to ANN learning 'by heart' the answers from the training set, without being able to generalize them to other cases. The model created in this way should be subjected to further tests with the use of the test set (as mentioned earlier in the subject literature, the set nomenclature is not unambiguous). Maintaining this procedure significantly increases the reliability of the results generated by the ANN.

3. Determinants of households debt

The literature usually considers two groups of factors that affect the level of HHs' debt of credit and loans. These are (i) demand factors and (ii) associated with the force of supply [Coletta et al. 2014].

Among the demand factors, Friedman's theory of permanent income deserves detailed analysis. It assumes that the level of consumption depends on the average income that HHs expects to obtain in the long run. If demand cannot be met due to insufficient temporary income, savings are used as well as external sources of financing, mainly in the form of credit products. In this case, smoothing consumption takes place [Coletta et al. 2014]. It should be noted that the generation of arrears in dwelling payments is de facto a form of raising funds for the part of consumption outside the HHs.

A similar approach was followed in the Life-Cycle Hypothesis. It assumes that the HHs draw up consumption plans taking into account projected lifetime income adjusted (increased) by the initial wealth or inheritances [Begg et al. 2002]. This theory is confirmed by empirical studies. Alfaro and Gallardo [2012] observed a peak level of debt (and thus indirectly the risk of its non-repayment) for people of middle age (42 years). When this age is exceeded, the debt decreases.

The age of consumers is one of the factors influencing their willingness to borrow. Analyses of data from the household budget survey showed that those HHs in which the head of household reached the age of 45 years are less likely to have debt under a loan or credit. For HHs where the head of household has reached or exceeded the age of 55 years, the decline is even more visible [Wałęga 2010].

HHs debt also depends on their rate at which they discount the future. On this basis, two types of HHs can be distinguished: patient and impatient. The problem of maximizing their lifetime utility is similar in both groups. Impatient households are subject to an additional borrowing constraint. Depending on monetary policy and legal regulations, this translates into an increase in HHs debt [Zhang 2019; Turdaliev, Zhang 2019; Alpanda, Zubairy 2017].

Meng et al. [2011] in the model of HHs debt as variables used: HHs debt – seasonally adjusted quarterly data, measured as at the end of quarter (ii) gross domestic product (GDP) – seasonally adjusted quarterly data, (iii) consumer price index (CPI), (iv) house price index (HPI), (v) interest rate – quarterly average monthly data, (vi) unemployment rate – quarterly average monthly data, (vii) population size (quarterly estimated), (viii) number of new dwellings (houses and flats). Nomatye and Phiri [2018] also included macroeconomic data in their research such as: (i) the ratio of gross fixed capital formation (GFCF) to GDP, (ii) repo rate and, (iii) increase in the HPI for medium sized houses.

Another group of factors shaping the debt amount are the institutional and legal regulations conditioning the demand and supply on the market of consumer loans and credits. These are in particular the following: (i) level of creditor's and debtor's protection (among others, in the context of consumer bankruptcy, effectiveness of recovery), (ii) availability of information on the reliability of borrowers, (iii) tax regulations (e.g. possibility to deduct part of the mortgage tax rate in the Netherlands) [Coletta et al. 2014], (iv) regulatory loan-to-value (LTV) [Alpanda, Zubairy 2017].

4. Empirical research method

The number of indebted premises and the direction of annual change in the value of debt were subjected to modelling with the use of ANN. The data presented in Table 1 were used as independent variables.

Table 1. The set of independent variables

Described area	Independent variable
population	total population and in a division into pre-production, production and post-production age (data for the previous year)
	change in the total population and in a division into pre-production, production and post-production age (between the previous year and two years ago)
	feminization factor (data for the previous year)
labour market	number of employed per 1000 inhabitants (data for the previous year)
	number of employed per 1000 inhabitants (2 years earlier)
	change in the number of employed per 1000 inhabitants (between the previous year and two years ago)
	total number of unemployed (data for the previous year)
	number of long-term unemployed (data for the previous year)
	percentage change of total unemployed and long-term unemployed (between the previous year and two years ago)
	number of deregistered business entities (data for the previous year)
housing	total number of premises with arrears (2 years earlier)
	number of premises with arrears of more than 3 months (2 years earlier)
	difference between the total number of arrears and more than 3 months (2 years earlier)
	number of housing allowances paid to users of communal premises and in total (data for the previous year)
	change in the number of housing allowances paid to users of communal dwellings and in total (between the previous year and two years ago)
	number of social housing premises (data for the previous year)
incomes of municipal budgets	value of income from personal income tax (data for the previous year)
	value of income from corporate income tax (data for the previous year)
	percentage change of income from personal income tax, corporate income tax, agricultural tax, real property tax, total taxes (between the previous year and two years ago)

Source: own elaboration.

The construction and operation of ANN implies the need for a far-reaching reduction in the number of independent variables. In the research, the authors empirically – comparing the quality of models – selected such subset of input variables that resulted in the creation of the best possible network. Due to the assumed one-year forecast horizon, it was necessary to build a set of independent variables on the basis of data from 2011, 2012 and 2013, 2014 (by one and two years earlier than the dependent variable, i.e. the variable for 2013 and 2015, respectively).

From this set of data, municipalities with less than ten thousand inhabitants were removed. In their case, it was considered that the forecasted debt would be too dependent on the behavior of individual HHs and clearly could not be effectively predicted. The developed results could be random and could distort the actual quality of model functioning.

Next, three subsets were created. Data from 2011-2013 were used to build the training (600 elements) and validation (125 elements) set. Data for 2013-2015 (729 elements) were used in their entirety to check the quality of network operation (they were a test set). The applied approach potentially enables the use of models in economic practice (with a one-year prediction horizon).

5. Results

More than one hundred thousand ANNs with MLP and RBF architecture, differing in the number of neurons, used activation functions, training method and a set of independent variables, were tested. The highest level of forecasts' accuracy was developed by MLP networks.

Figure 3 presents a construction diagram of the best network forecasting the number of indebted dwellings in municipalities, along with the used independent variables. This was ANN of MLP type with a structure of 11-18-1 (number of neurons in subsequent layers, i.e. input – hidden – output). The neurons of learning layers, i.e. hidden and output, were using the logistic activation function. The training was carried out with the BFGS algorithm. MAPE was used to determine the quality of the model's operation. As for 264 municipalities the number of indebted premises was relatively small (less than 50), even the slightest forecast error would have resulted in a high MAPE value. For this reason, in order to maintain the objectivity of research, this was determined for 465 municipalities with at least 50 indebted premises and amounted to 6.08%. Figure 4 shows the scatter diagram between the actual values of dependent variable and its forecasts.

Figure 5 shows construction scheme of the best network forecasting the direction of changes of household's debt in municipalities between 2015 and 2014. Eight independent variables were used. MLP had 8-7-1 architecture. Neurons of hidden layer were built using an exponential activation function. The logistic function was applied in output layer cells. Training was conducted using the BFGS algorithm. The correctness of forecasts for the entire test set was 76.84%. It should be noted that it

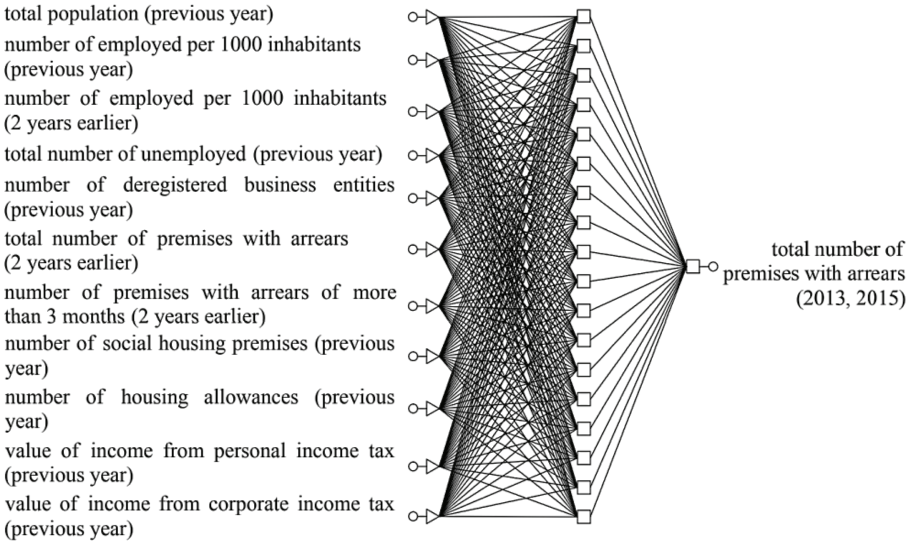

Fig. 3. The best network forecasting the number of indebted dwellings in municipalities

Source: own elaboration.

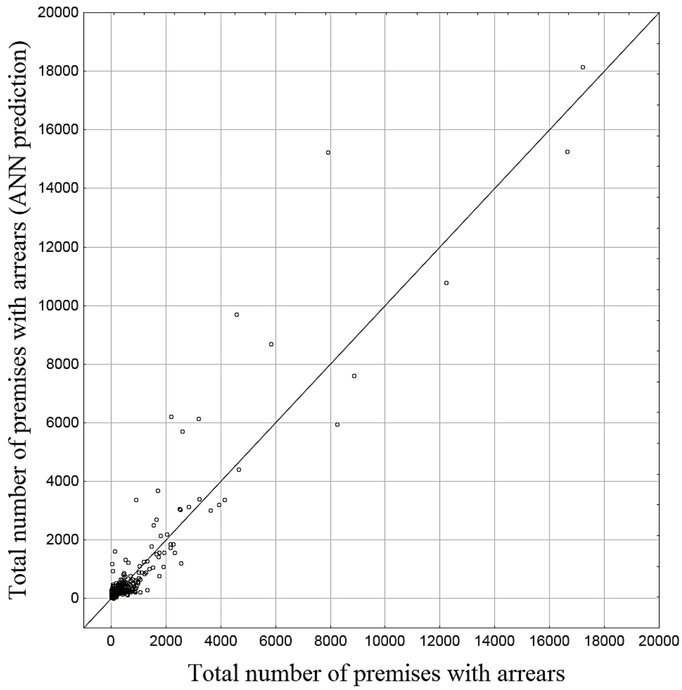

Fig. 4. The number of premises with arrears in municipalities and their forecasts

Source: own elaboration.

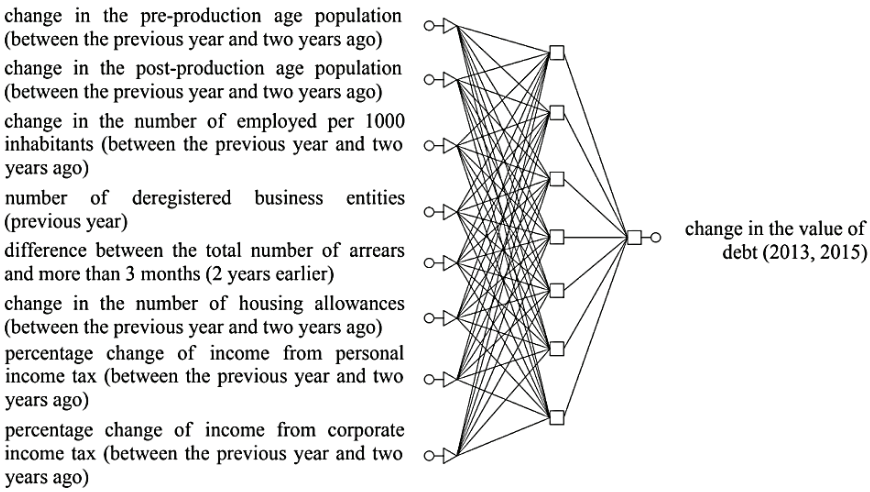

Fig. 5. The best network forecasting the direction of changes of household's debt in municipalities

Source: own elaboration.

was slightly higher (76.98%) for municipalities with at least 1500 indebted dwellings. In the case of municipalities with a lower number of indebted premises, it was (76.59%).

6. Conclusions

The studies confirmed both the possibility and the adequacy of using ANNs to forecast both the direction of HHs' debt changes in residential property and the future number of dwellings in debt. Forecasts for municipalities with low indebted premises have the potential to result in unacceptable network failure. This is due to the fact that it is not possible to model the behavior of individual HHs using the available independent variable. The lack of many desirable input variables (e.g. income per capita on the level of municipalities) can to some extent be compensated by other, derived data which are available at the required level (e.g. amounts of tax). The problem is not only the availability of a particular type of data, but also the frequency of data provided by the Central Statistical Office. The method used requires access to data preferably from three consecutive years (to learn the model) and the last two years to make predictions in real conditions. Forecasts based on MLPs turned out to be definitely more precise than using RBF networks. The BFGS algorithm proved to be better in the analyzed problem than the method of error back propagation (with the momentum factor).The presented models are predisposed to forecasts in municipalities where the debt of HHs is a noticeable problem and is of a mass nature.

The proposed solutions can be an important element of a modern system of management and financial planning in municipalities.

Future research to improve the accuracy of the model should be carried out in two areas: (i) extending the set of input variables with further data (which requires their availability in the Central Statistical Office), (ii) using deep learning. Deep neural networks (DNNs) are particularly predisposed in analyzing large data sets and enable the proper functioning of ANNs with higher complexity. The extension of input data may mean the need to use DNNs, whose algorithms work more efficiently enabling the use of the potential inherent in networks with many layers and a greater number of neurons than MLP.

References

Alfaro R., Gallardo N., 2012, *The determinants of household debt default*, Revista de Análisis Económico, vol. 27, no. 1, pp. 55-70.

Alpanda S., Zubairy S., 2017, *Addressing household indebtedness: Monetary, fiscal or macroprudential policy?*, European Economic Review, vol. 92, pp. 47-73.

Begg K.H., Fischer S., Dornbusch R., 2002, *Economics*, MGH Pub. Co., Berkshire.

Central Statistical Office, Macroeconomic Data Bank, 2018, *Living conditions of population*, https://bdm.stat.gov.pl (10.12.2018).

Coletta M., De Bonis R., Piermattei S., 2014, *Household debt: A cross-country analysis*, Bank of Italy Temi di Discussione (Working Paper) 989, https://ssrn.com/abstract=2571267 (12.12.2018).

Kolasa A., 2017, *Sytuacja finansowa sektora gospodarstw domowych w I kw. 2017 r*, National Bank of Poland, https://www.nbp.pl (10.12.2018).

McCulloch S., Pitts W.S., 1943, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, vol. 5, pp. 115-133.

Meng S., Hoang N.T., Siriwardana M., 2011, *The Determinants of Australian Household Debt: A Macro-level Study*, Business, Economics and Public Policy Working Papers, School of Business, Economics and Public Policy Faculty of the Professions, University of New England, vol. 4, pp. 1-24.

National Bank of Poland, 2017, *Zasobność gospodarstw domowych w Polsce. Raport z badania 2016r*, https://www.nbp.pl (14.12.2018).

Nomatye A., Phiri A., 2018, *Investigating the macroeconomic determinants of household debt in South Africa*, International Journal of Economics and Financial Issues, vol. 8, no. 2, pp. 62-69.

Turdaliev N., Zhang Y., 2019, *Household debt, macroprudential rules, and monetary policy*, Economic Modelling, vol. 77, pp. 234-252.

Wałęga G., 2010, *Determinanty zadłużenia gospodarstw domowych w Polsce w świetle wybranych teorii konsumpcji*, [in:] Z. Dach (ed.), *Otoczenie ekonomiczne a zachowania podmiotów rynkowych*, PTE, Kraków, pp. 114-133.

Zhang Y., 2019, *Household debt, financial intermediation, and monetary policy*, Journal of Macroeconomics, vol. 59, pp. 230-257.