Electronic Journal of Polish Agricultural Universities is the very first Polish scientific journal published exclusively on the Internet, founded on January 1, 1998 by the following agricultural universities and higher schools of agriculture: University of Technology and Agriculture of Bydgoszcz, Agricultural University of Cracow, Agricultural University of Lublin, Agricultural University of Poznan, Higher School of Agriculture and Teacher Training Siedlce, Agricultural University of Szczecin, and Agricultural University of Wroclaw.

ELECTRONIC JOURNAL OF POLISH AGRICULTURAL UNIVERSITIES

2006
Volume 9
Issue 2
Topic
GEODESY AND CARTOGRAPHY

WIŚNIEWSKI R.. 2006. EVALUATION OF QUALITY AND GENERALIZATION PROCESSES IN NEURAL NETWORK SETS USED TO FORECAST THE VALUE OF SPACE **Electronic Journal of Polish Agricultural Universities**, Geodesy and Cartography, Volume 9, Issue 2.
Available Online http://www.ejpau.media.pl

EVALUATION OF QUALITY AND GENERALIZATION PROCESSES IN NEURAL NETWORK SETS USED TO FORECAST THE VALUE OF SPACE

Radosław Wiśniewski
Department of Land Management and Regional Development,
University of Warmia and Mazury in Olsztyn, Poland

ABSTRACT

Forecasting the value of real estate is an essential element that should be taken into account by the investor in the process of financing an investment. A similar situation can be observed in the process of land management. In such cases, the reliability of the model used for real estate value prediction becomes a key issue.

The geostatic model is designed to be used for diagnosing the land market system in the past and in the present (at the moment the forecast is generated). It then becomes a prognostic geostatic model used for forecasting. Geostatic models can be developed based on a set of artificial neural networks. A set of neural networks is a set of many trained monolithic neural networks, which are combined into one set to eliminate faults assigned to single network models, as well as to improve generalization capability and resistance.

The aim of the present study was to develop and test in practice a set of measures enabling to evaluate the quality of a forecasting model as well as its generalization capability.

Key words: neural networks, sets of neural networks, evaluation of quality, generalization error, forecast the value of space

1. QUALITY OF NETWORKS AND SETS

In order to assess the quality of a network as well as a set of networks, various measures may be used that are based on the error function value, generated by a network or a set of networks at output.

1.1. Quality evaluation of a single network

The quality evaluation of a single network may be conducted based on the following measures [2,4,5,6,7,10]:

a) Standard deviation ratio - IO

$$IO = \frac{\overline{s}_B}{s_C} \tag{1.1}$$

where: $\overline{s}_B$ mean standard deviation of network errors, s_C - standard deviation calculated for an output variable (e.g. real estate prices).

b) Sum of squares error – SSE

$$SSE = \sum_{i=1}^{N} (c_i - y_i)^2 \tag{1.2}$$

where: c_i - real estate price, y_i value calculated by the network.

c) Mean square error – MSE

$$MSE = \frac{SSE}{N} \tag{1.3}$$

where: N – number of elements in a given set; the remaining symbols as above.

d) Root mean square error – RMSE

$$RMSE = \sqrt{MSE} \tag{1.4}$$

e) Normalized RMSE - NRMS

$$NRMS = \frac{RMSE}{s_c} \tag{1.5}$$

where: s_c - standard deviation calculated for a set of real estate prices.

f) Average relative variance error - ARV

$$ARV = \frac{SSE}{\sum_{i=1}^{N} (c_i - \overline{c})^2} \tag{1.6}$$

where: $\overline{c}$ - mean real estate price in a given set.

g) Inequality coefficient squared – IC

$$IC = \frac{MSE}{\frac{1}{n}\sum_{i=1}^{N} c_i^2} \tag{1.7}$$

1.2. Quality evaluation of a network set

The quality evaluation of a network set may be conducted using the same measures, assuming that value d_i replaces y_i in formula (1.2). d_i is the vector value at output of the network set. The measures from (1.1) to (1.7), used to assess the quality of network sets, will be marked (for the purpose of distinguishing) with the letter “z”, e.g. sum of squares error for network sets will be identified as *SSEz.*

2. EVALUATION OF THE GENERALIZATION ERROR OF NETWORKS AND SETS OF NETWORKS

Various methods may be applied to evaluate the generalization error, which usually use the information in model residuals. They are primarily based on the sum of squares error, i.e. on the value, which expresses the total squared deviations of the estimated values (y_i - real estate value) from the input values (c_i - real estate price). Additionally, these measures take into account the information regarding the size of the input set, as well as the number of weights in the network or network set structure.

2.1. Evaluation of network generalization error

The network generalization error evaluation measures based on the log likelihood [1,3] are used in the paper:

a) Akaike information criterion – AIC

$$AIC = -2\frac{L}{N} + 2\frac{W}{N} \tag{1.8}$$

where: L – logarithm value of log-likelihood function – $L = -\frac{N}{2}\left(1 + \ln(2\pi) + \ln\left(\frac{SSE}{N}\right)\right)$, W – number of weights in the network.

b) Schwarz criterion – SC

$$SC = -2\frac{L}{N} + W\frac{\ln(N)}{N}. \tag{1.9}$$

c) Hannan-Quinn criterion – HQ

$$HQ = -2\frac{L}{N} + 2W\frac{\ln(\ln(N))}{N}. \tag{1.10}$$

d) Normalized Bayesian information criterion – NBIC (Lula 1999)

$$NBIC = \ln\left(\frac{SSE}{N}\right) + \frac{2W}{N}\ln(N). \tag{1.11}$$

The Pearson's correlation coefficient (R), determined between a series of real estate prices and a series of network output values [4], may be additionally used. The square of the Pearson's correlation coefficient (R^2) may be also applied.

2.2. Evaluation of network set generalization error

The evaluation of the network set generalization error is not as simple as the evaluation of the network generalization error, since the number of weights in the entire network set is a derivative of the number of weights in particular networks combined into a set. In addition, each network may have a different structure and a different number of weights. In such cases, the measures provided in formulas (1.8) – (1.11) should be modified taking into account the set structure.

The modified (1.8) – (1.12) measures may be presented as follows:

a) Modified Akaike information criterion – ZAIC

$$ZAIC = -2\frac{L_z}{N} + 2\frac{M}{N}, \tag{1.12}$$

while

$$L_z = -\frac{NM}{2}\left(1 + \ln(2\pi) + \ln\left(\frac{SSK_z}{N}\right)\right), \tag{1.13}$$

where: L_z – logarithm value of the modified log likelihood function, M – number of networks in a set, SSK_z - mean value of sum of squares error for the *k-th* network in the set:

$$SSK_z = \sum_{i=1}^{N}\frac{\sum_{k=1}^{M}(c_i - y_{ik})^2}{M}, \tag{1.14}$$

where: i – subsequent observation in the data set.

b) Modified Schwarz criterion – ZSC

$$ZSC = -2\frac{L_z}{N} + M\frac{\ln(N)}{N}. \quad (1.15)$$

c) Modified Hannan-Quinn criterion – ZHQ

$$ZHQ = -2\frac{L_z}{N} + 2M\frac{\ln(\ln(N))}{N}. \quad (1.16)$$

d) Modified normalized Bayesian information criterion – ZNBIC

$$ZNBIC = \left[\ln\left(\frac{SSK_z}{N}\right) + \frac{2M}{N}\ln(N)\right]M. \quad (1.17)$$

e) Set generalization index – WG1

$$WG1 = \frac{\sum_{i=1}^{N}(c_i - d_i)^2}{SSK_z}. \quad (1.18)$$

f) Set generalization index – WG2

$$WG2 = \frac{SSK_{zs}}{SSK_z} \quad (1.19)$$

where: SSK_{zs} - mean value of sum of squares error for the *k-th* network in the set, calculated based on the output vector of the network set (d_i):

$$SSK_{zs} = \sum_{i=1}^{N}\frac{\sum_{k=1}^{M}(d_i - y_{ik})^2}{M}, \quad (1.20)$$

where*:* *i* – subsequent observation in the data set, d_i – values of the output vector of the network set.

3. TESTING THE MEASURES USED TO EVALUATE QUALITY AND GENERALIZATION ERRORS

Much information on the subject of transaction is utilized in the process of building a model that provides the basis for space value prediction. Taking into account the difficulty in the physical acquisition of data (field inspection), the object of research shall be the undeveloped land plots available on the Olsztyn real estate market, designated for single family housing with accompanying facilities (ORNGN). This decision was dictated not only by the access to relevant information, but primarily by broad theoretical knowledge about the relations on the land market system in Olsztyn. The descriptive statistics of the variables for ORNGN are presented in Annex 1.

Such a research object – ORNGN – allows to adopt the predicted value of space as a variable, simultaneously assuming that in the process of listing the output data for forecasting, this role would be fulfilled by the observed unit transaction price denoted *C_gr_ol*.

The forecasting variables follow Wiśniewski [9]. The base variable set (BZZ) was established on this basis and then used to develop the forecasting model. BZZ encompasses the variables listed in Annex 1 to Wiśniewski [9].

3.1. Base reference network set

The evaluation of quality and generalization errors was performed based on the so-called base reference network set (BZSO). BZSO will be applied to evaluate the measures proposed in item 2 of this paper, which assesses the quality and generalization errors of networks and network sets. BZSO may be also used as a reference model in the following processes: selecting variables, defining the network structure, defining the structure and size of a network set.
BZSO was developed based on the following assumptions:

a) Forecast variable: *C_gr_ol* one observation in advance (one period).
b) Maximum delay for all of variables: 10 observations (around 2 weeks).
c) Independent variables: BZZ.
d) Number of tested networks: 1000.
e) Number of selected networks: 100.
f) Selection criteria for networks: balance between the error and network diversity.
g) Network types/[training method]: RBF/[KM, KN, PI], GRNN/[SS], MLP3/[BP100,CG50], MLP4/[BP100, CG50] (symbols as in Annex. 2a and 2b).
h) Activation function: linear, logistic.
i) Number of hidden neurons: minimum -1, maximum in the first hidden layer - 25, in the second hidden layer - 13, for the RBF network - 397.
j) Number of cases: 1104, applied in network processes: 1094.
k) Network verification technique: UWT 794/200/100 (symbols as in Annex 2c).
l) Operationalization of continuous variables presented on an interval and ratio scale - minimax (0,1), of discrete variables presented on an ordinal scale - minimax (0,1) and of nominal discrete variables – binary.
m) The quality of network training, validating and testing: standard deviation ratio – IO (formula 1.1).
n) Network training, validating and testing error: *RMSE*.
o) Establishing a set: arithmetic mean of the output of particular networks.
p) Number of replications: 3.

Table 1 presents the measures of quality and the generalization errors of base reference network sets (BZSO) for the entire observation set as well as for the following subsets: training, validating and testing.

Table 1. Measures of quality and generalization errors of base reference network sets (BZSO)

Set	Measures	BZSO_1	BZSO_2	BZSO_3
1	2	3	4	5
Training	IO	0.7412	0.7591	0.7713
	SSEz	515306.9590	526226.4894	511401.3324
	MSEz	649.0012	662.7538	644.0823
	RMSEz	25.4755	25.7440	25.3788
	NRMSz	0.7138	0.7291	0.7267
	ARVz	0.4941	0.5302	0.5456
	ICz	0.1392	0.1500	0.1514
	ZAIC	947.4950	947.0125	942.4648
	ZSC	948.0840	947.6016	943.0539
	ZHQ	947.7214	947.2389	942.6912
	ZNBIC	831.6439	831.1614	826.6137
	WG1	1.1725	1.1426	1.1235
	WG2	0.1471	0.1248	0.1099
	R	0.7365	0.7104	0.6859
	R^2	0.5425	0.5047	0.4705
Validating	IO	0.8358	0.8366	0.8250
	SSEz	152195.4162	171475.0146	178910.2799
	MSEz	760.9771	857.3751	894.5514
	RMSEz	27.5858	29.2810	29.9091
	NRMSz	0.7730	0.8293	0.8564
	ARVz	0.6411	0.6666	0.6332
	ICz	0.1773	0.1703	0.1760
	ZAIC	957.2367	967.3959	971.8446
	ZSC	958.8858	969.0451	973.4938
	ZHQ	957.9040	968.0633	972.5120
	ZNBIC	1202.2807	1212.4399	1216.8886
	WG1	1.0940	1.0749	1.0771
	WG2	0.0860	0.0697	0.0716
	R	0.6034	0.5984	0.6221
	R^2	0.3641	0.3581	0.3870

Table 1 cont.

1	2	3	4	5
Testing	IO	0.8818	0.8010	0.9002
	SSEz	63758.5273	71417.4615	114389.9874
	MSEz	637.5853	714.1746	1143.8999
	RMSEz	25.2505	26.7240	33.8216
	NRMSz	0.7075	0.7569	0.9684
	ARVz	0.7037	0.5970	0.7663
	ICz	0.1799	0.1789	0.2203
	ZAIC	943.2392	953.3234	995.8456
	ZSC	945.8444	955.9285	998.4508
	ZHQ	944.2935	954.3777	996.9000
	ZNBIC	1578.4855	1588.5697	1631.0919
	WG1	1.1239	1.1099	1.0601
	WG2	0.1103	0.0990	0.0567
	R	0.5578	0.6588	0.4900
	R^2	0.3112	0.4341	0.2401
Entire set	SSEz	731260.9025	769118.9654	804701.5997
	MSEz	668.4286	703.0338	735.5590
	RMSEz	25.8540	26.5148	27.1212
	NRMSz	0.7292	0.7479	0.7650
	ARVz	0.5318	0.5593	0.5852
	ICz	0.1488	0.1565	0.1638
	ZAIC	948.6066	951.2427	953.9413
	ZSC	949.0634	951.6995	954.3981
	ZHQ	948.7795	951.4156	954.1142
	ZNBIC	792.5629	795.1990	797.8976
	WG1	1.1519	1.1245	1.1042
	WG2	0.1319	0.1107	0.0943
	R	0.7005	0.6837	0.6560
	R^2	0.4907	0.4674	0.4303
	Structure*	17/11/46/26	17/12/37/34	15/51/2/32

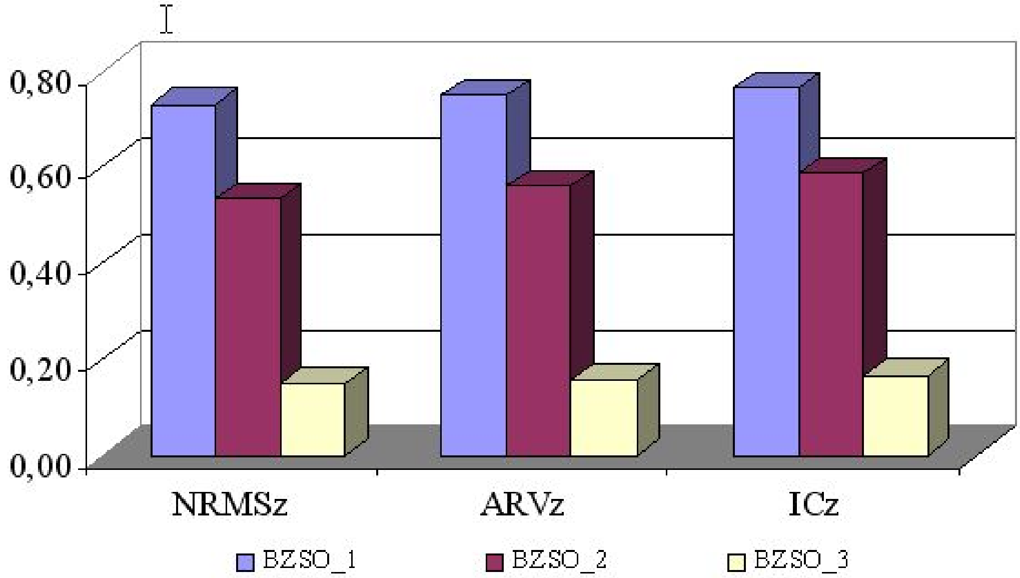
* - *subsequent network number GRNN/MLP3/MLP4/RBF in the set.*

Source: Own study

The base reference network set that may be adopted, based on the applied measures, is the set marked in Table 1 as BZSO_1 (column 3), (Fig. 1 and 2) . This set is characterized by the lowest indices among those that should be minimized (SSEz, MSEz, RMSEz, NRMSz, ARVz, ICz, ZAIC; ZSC, ZHQ, ZNBIC) as well as the highest among those that should be maximized (WG1 and WG2, R and R^2). Upon analyzing Figures 3 and 4, the following measures should be recognized as best: NRMSz and ZNBIC. These measures best reflect the quality and generalization capabilities of neural network sets.

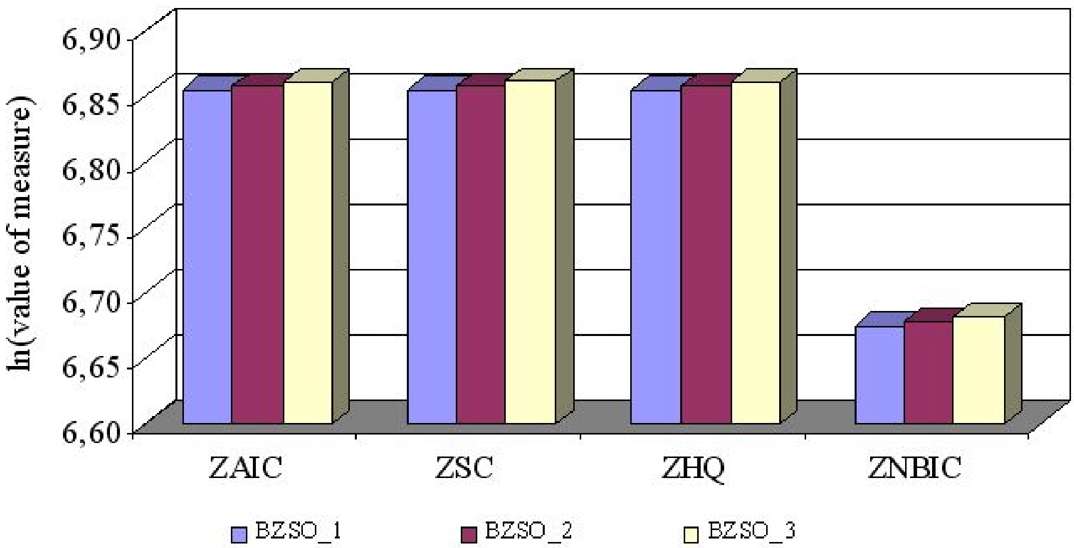
3.2. Selection of a verification technique

Fig. 1. Selected measures evaluating the quality of network sets for the entire observation set – NRMSz (1.5); ARVz (1.6), ICz (1.7)



Source: Own study.

Fig. 2. Selected measures (ZAIC (1.12); ZSC (1.15), ZHQ (1.16), ZNBIC (1.17)) evaluating the generalization of network sets for the entire observation set

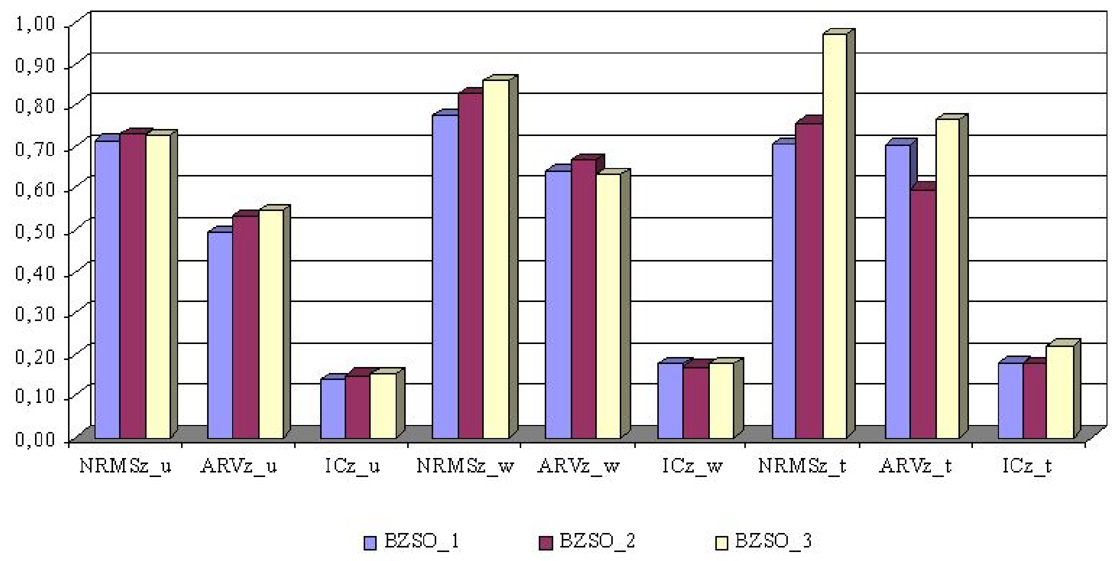


Source: Own study.

The purpose of neural model (networks in a set) verification processes is to define the estimation and generalization capabilities of training sets. They are additionally used as tools to verify the capacity to generalize knowledge acquired through network processes [4]. Verification techniques should take into account the size of gathered data sets. They should also be adequate in processes of forecasting real estate price series. The capability of a network to generalize knowledge as well as its overall generalization capability, is confirmed by its ability to generate proper results upon entering data that did not participate in the network training process, as well as to generate "correct forecasts". The basic network verification techniques include: **sample division methods,** [7,8,10], enabling to establish separate training, validating and testing sets, **cross-validation procedures**, **bootstrap** methods [4], **mixed methods** [7].

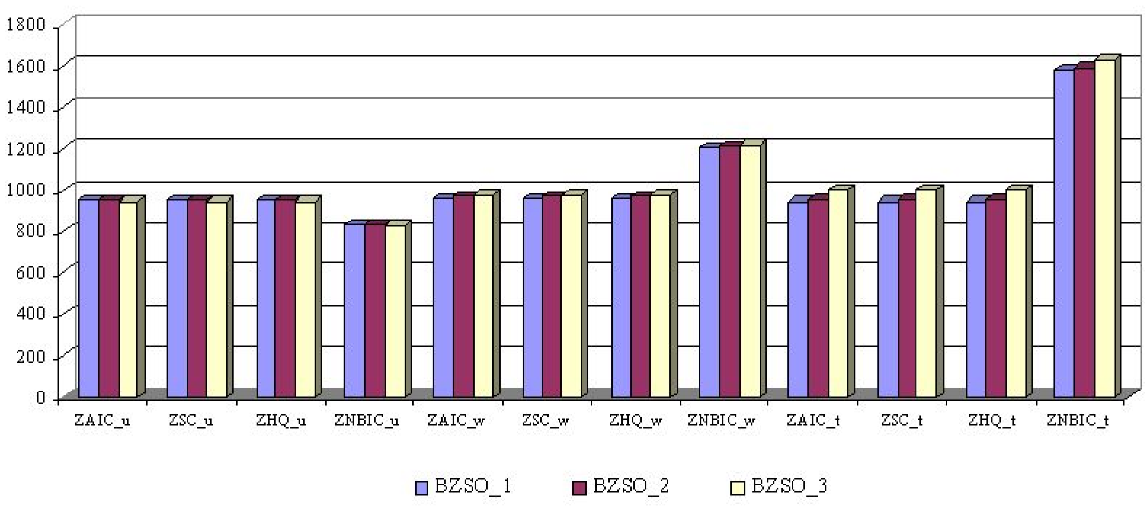
A detailed analysis revealed that mixed network verification methods provide the best results of forecasting real estate value within the framework of neural network processes. At the first stage, mixed methods apply the sample division method in order to indicate the number of cases for the training, validating and testing subsets. The second stage includes the use of a random selection technique (most often the Monte Carlo method), which performs the selection within the indicated number of cases for particular subsets.

Fig. 3. Selected measures evaluating the quality of network sets for the following subsets: training (u), testing (t) and validating (w)



Source: Own study.

Fig. 4. Selected measures evaluating the generalization of network sets for the following subsets: training (u), testing (t) and validating (w)



Source: Own study.

In accordance with Annex 2c, various verification techniques may be applied in network processes: *UWT*, *UWTL*, *UWTLL*. *UWT* is a typical sample division method, in which the establishment of the number of cases in particular subsets takes place only once, prior to the commencement of network processes. *UWTL* and *UWTLL* are examples of mixed method verification techniques. At the first stage, the algorithm applied is the same as in the *UWT* method. The second stage includes a random selection of cases (within the defined number) to particular subsets for each trained network. In the *UWTL* technique the testing set remains fixed, whereas in the *UWTLL* technique all sets are variable.

This section of the paper comprised research pertaining to the selection of a network verification technique. The listed and developed measures that enable to evaluate quality and generalization capabilities were applied for this purpose. In the process of searching for the best verification method, a study was conducted using the network set establishment method proposed in item 3.1. The verification technique was changed each time and the experiment was repeated once only. Table 2 presents the measures of quality and generalization errors of network sets (ZS) for the entire observation set, acquired depending on the adopted network verification technique. It was assumed that the testing set will include 100 cases, the validating set - 200, while the training set - 794.

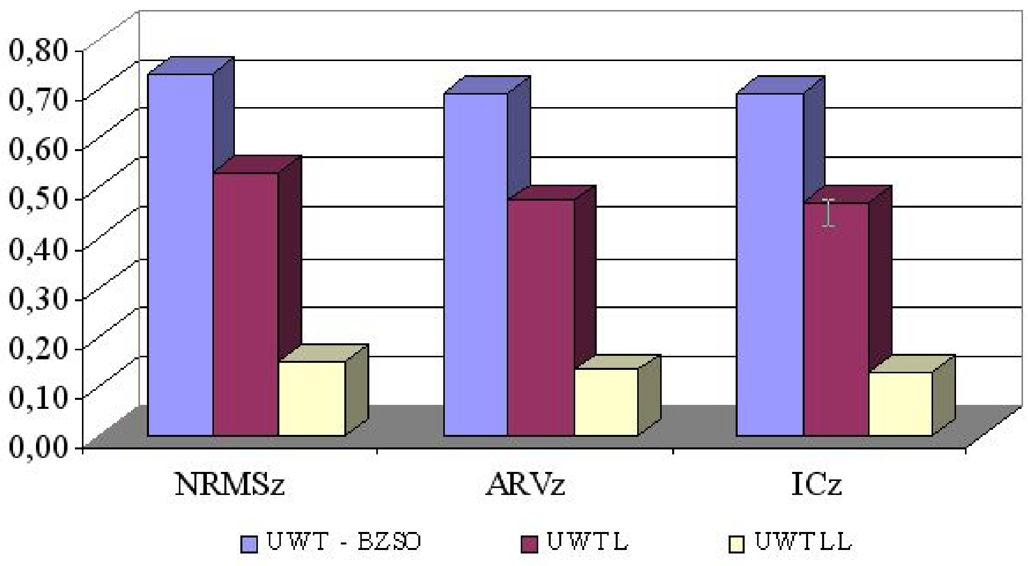
Table 2. Measures of quality as well as evaluation of the generalization errors of network sets (ZS) depending on the network verification technique

Set	Measures	*UWT - BZSO*	*UWTL*	*UWTLL*
1	2	3	4	5
Entire set	SSEz	731260.9025	652347.3823	650909.4738
	MSEz	668.4286	596.2956	594.9812
	RMSEz	25.8540	24.4192	24.3922
	NRMSz	0.7292	0.6888	0.6880
	ARVz	0.5318	0.4744	0.4734
	ICz	0.1488	0.1328	0.1325
	ZAIC	948.6066	947.5616	945.4903
	ZSC	949.0634	948.0184	945.9471
	ZHQ	948.7795	947.7344	945.6632
	ZNBIC	792.5629	791.5178	789.4466
	WG1	1.1519	1.2778	1.2544
	WG2	0.1319	0.2179	0.2339
	R	0.7005	0.7530	0.7519
	R^2	0.4907	0.5670	0.5654
	Structure*	17/11/46/26	20/29/21/30	20/31/17/32

** - subsequent network number GRNN/MLP3/MLP4/RBF in the set.*

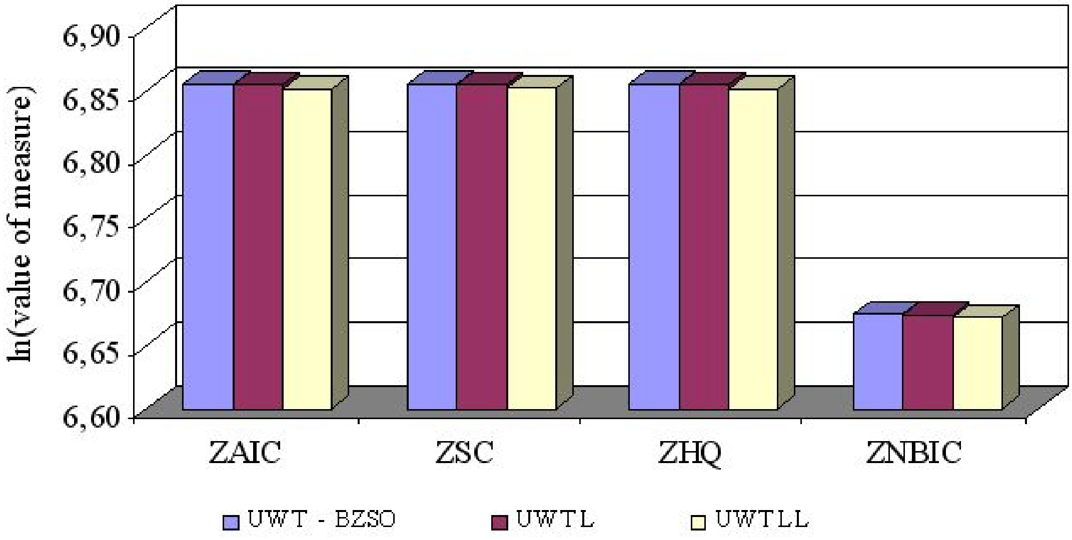
Source: Own study.

Fig. 5. Selected measures evaluating the quality of network sets for the entire observation set – verification technique selection



Source: Own study.

Fig. 6. Selected measures evaluating the generalization of network sets for the entire observation set – verification technique selection



Source: Own study.

An analysis of the data in Table 2 and in Fig. 5 and 6 shows that the best results, based on the applied measures, were generated with the use of the *UWTLL* verification technique. As a standard, the use of *UWTLL* as a verification technique is recommended based on the analysis conducted during further research.

4. CONCLUSIONS

1. Various measures may be used to evaluate the quality presented by neural network sets with respect to reproducing the model represented by data. The best results were achieved by applying the following measures: normalized RMSE – NRMSz as well as the mean error of the average relative variance – ARVz.
2. The best measure to apply in order to identify the generalization capabilities of neural network sets is the modified normalized Bayesian information criterion – ZNBIC.
3. It should be stressed that applying many measures should lead to constructive conclusions, since each of the measures is partially based on a different model used to verify quality with respect to reproducing the standard, or the generalization capability of neural network sets.

5. REFERENCES

1. Brooks Ch. 2002. Introductory Econometrics for Finance. Cambridge University Press. [in English]
2. Gately E. 1999. Sieci neuronowe. Prognozowanie finansowe i projektowanie systemów transakcyjnych. [Neural Networks for Financial Forecasting]. WIG PRESS, Warszawa. [in Polish]
3. Grasa A. A. 1989. Econometric Model Selection: A New Approach. Dordrecht: Kluwer Academic Publishers. [in English]
4. Lula P. 1999. Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych [Feed-forward Neural Networks in the Modeling of Economic Phenomena]. Wydawnictwo Akademii Ekonomicznej w Krakowie. [in Polish]
5. Masters T. 1995. Neural, Novel & Hybrid Algorithms for Time Series Prediction. Wiley, John & Sons. [in English]
6. Masters T. 1996. Sieci neuronowe w praktyce. Programowanie w języku C++ [Practical Neural Network Recipes in C++]. Wydawnictwo Naukowo-Techniczne, Warszawa. [in Polish]
7. StatSoft, Inc. 2004. STATISTICA (data analysis software system), version 6. www.statsoft.com. [in Polish]
8. Wiśniewski R. 1999. Metodyczne i praktyczne aspekty zastosowania sztucznej inteligencji w taksacji powszechnej [Methodic and practice aspects of implementation artificial intelligence in mass appraisal]. PhP thesis. [in Polish]
9. Wiśniewski R. 2005. Wpływ redukcji wymiaru przestrzeni danych wejściowych na jakość oraz błędy generalizacji zespołów sieci neuronowych [Influence of reduction input data dimension on the quality and generalization processes in neural network sets]. Materiały V Konferencji CMS'05 Computer Methods and Systems. AGH University of Science and Technology, Jagiellonian University and Cracow University of Technology, str. 319-324. Kraków. [in Polish]
10. Wykorzystanie sztucznych sieci neuronowych w modelowaniu ekonomicznym [Use artificial neural networks in economic modelling], red. J. Siedleckiego. 2001. Wydawnictwo Akademii Ekonomicznej im. O. Langego we Wrocławiu, Wrocław. [in Polish]

6. ANNEXES

Annex 1. Descriptive statistics of variables for ORNGN

Variable*	N	Mean	Median	Min.	Max.	Range	Standard Deviation	Skewness	Kurtosis
C_gr_ol	1104	56.49	50	2.64	208.38	205.74	35.54	0.89	0.99
data_d	1104	1751	2055	6	3180	3174	980.27	-0.29	-1.34
data_t	1104	251	295	2	456	454	140.25	-0.29	-1.34
data_m	1104	58	68	1	105	104	32.17	-0.29	-1.34
pow	1104	912	720	84	9712	9628	831.73	4.89	36.70
odl_c	1104	4.327	4.177	0.277	7.133	6.856	1.27	-0.09	-0.27
pzp	1104	2	2	1	3	2	0.50	-0.26	0.39
e	1104	2	2	1	3	2	0.51	-0.34	-0.41
w	1104	2	2	1	3	2	0.47	-0.54	0.32
g	1104	2	1	1	3	2	0.54	0.36	-1.03
t	1104	1	1	1	3	2	0.52	0.22	-1.45
k	1104	2	2	1	3	2	0.53	-0.15	-0.75
f_w	1104	0.9	1.0	0.6	1.0	0.4	0.19	-0.82	-1.32
d_d	1104	2	2	1	5	4	0.80	1.07	2.12
fr	1104	23	21	2	96	93	12.16	1.71	5.62
f_t	1104	2	2	1	2	1	0.44	-1.13	-0.73
gl	1104	35	32	3	162	160	16.19	2.87	14.44
ksz	1104	2	2	1	2	1	0.41	-1.42	0.01
l_dz	1104	1	1	1	12	11	1.15	4.35	21.30
usyt	1104	2	2	1	2	1	0.43	-1.14	-0.70
top	1104	4	4	1	5	4	0.97	-0.94	0.74
u_d	1104	2	1	1	5	4	0.67	1.22	2.09
u_k	1104	1	1	1	5	4	0.53	4.93	27.54
u_i	1104	1	1	1	5	4	0.64	2.95	8.75
atr_w	1104	1.136	0.985	0.037	3.513	3.476	0.80	0.46	-0.92
atr_l	1104	0.664	0.423	0.004	2.977	2.973	0.66	1.85	2.89
odl_h	1104	4.497	4.822	0.284	8.118	7.834	1.27	-0.29	0.17
odl_k	1104	4.922	4.842	0.554	8.211	7.657	1.38	-0.42	0.43

* - symbols of variables in accordance with Annex 3 in Wiśniewski [8].

Source: Own study.

Annex 2a. Network architecture – code definitions

Code	Network Architecture
GRNN	Generalized Regression Neural Network
MLP3	Multi-Layer Perceptron – three layers
MLP4	Multi-Layer Perceptron – four layers
RBF	Radial Basis Function network

Source: Own study.

Annex 2b. Training algorithms – code definitions

Code	Training Algorithm
BP	Back Propagation
CG	Conjugate Gradient
QN	Quasi-Newton
LM	Levenberg-Marquardt
QP	Quick Propagation
DD	Delta-bar-Delta
SS	SubSample
KM	K- Means – defining radial neuron weights
IS	Isotropic – defining radius (deviation)
KN	K- Nearest Neighbor - defining radius (deviation)
PI	Pseudo-Inverse (least squares linear optimization)
KO	Kohonen - defining radial neuron weights
GR	General Regression
PC	Principal Components

Source: Own study.

Annex 2c. Network verification techniques – code definitions

Code	Network Architectures
UWT N/N/N	Random division of the data set into three subsets: *U* – training, *W* – validating and *T* – testing.
UWTL N/N/N	Random division of the data set into three subsets: *U* – training, *W* – validating and *T* – testing. *U* and *W* subsets are variable during training.
UWTLL N/N/N	Random division of the data set into three subsets: *U* – training, *W* – validating and *T* – testing. *U*, *W* and *L* subsets are variable during training.
UWTB N/N	Bootstrap sampling. First stage – random division of the data set into two subsets: *U* – training, *W* – validating (fixed). Second stage – bootstrap sampling of *U* set, the remaining cases form set *T*.

N – respective number of cases in subsets *U*, *W* and *T*.
Source: Own study.

Radosław Wiśniewski
Department of Land Management and Regional Development
University of Warmia and Mazury in Olsztyn, Poland
danrad@uwm.edu.pl