FUZZY MODEL DEDICATED TO THE POLISH INTERNET MORTGAGE MARKET

ALEKSANDER ORŁOWSKI¹, EDWARD SZCZERBICKI¹

¹ Gdansk University of Technology, Faculty of Management and Economics, Gdansk, Poland

This paper presents the fast-growing mortgage market in Poland. It sums up research done over the last 4 years, also showing the complete process of building and verifying a model of the market consisting of three sub-models. The sub-models are described for three market situations, i.e. stable market, crisis, and boom. Afterwards, the variables within the models are analyzed statistically (e.g. correlation level) and finally, the sub-models are verified. The paper also presents the process of the replicative and prognostic verification of the model. It concludes with suggestions for further research in this area, leading to the ultimate application of the proposed model in a real life prognosis process.

KEYWORDS fuzzy modeling, Internet mortgage market, fuzzy logic
INTRODUCTION \& MARKET DESCRIPTION

The Polish Internet mortgage market is an example of a fast growing and even faster changing market. The authors have been observing the market and have actively participated in it (being one of the leading sellers) for more than 7 years. The changeability of the market and lack of tools supporting the management processes on it (especially those for prognosis) pushed the authors towards the idea of creating a model dedicated to supporting decision-making on the market.

Later on, a description of the market will be presented together with the definition of its main problems.

The Polish Internet mortgage market consists of 4 main sectors: banks, brokers, partner web pages and customers (Orłowski 2008, 12). Banks are institutions that sell mortgages and offer the option to apply for one on their own web pages (Czekaj 2008, 50). Because the market is substantial in size and is characterised by fierce competition, banks allow their partners to sell their financial products on the partners\' web pages to generate bigger sales. These companies are typical brokers that receive commission for every product that they sell. Moreover, they promote and sell bank products on their pages but also create a network called a \"partner system\" which allows the owners of small web pages to sell bank products on their own pages. The owners of private web pages do not sell enough products to cooperate directly with banks, which is the reason for their association with brokers. For each product sold on a partner web page, its owner receives commission from the broker, who in turn gets his from the bank.

The authors decided to describe the market from the customers\' point of view and use it for working out a model which presumes numerous customers looking for mortgages using web
searchers to find the best offer. In doing so, they reach for partner web pages offering different mortgages. It needs to be pointed out that top search results are bought by the owners of private web pages. A customer chooses one of the web pages (usually one of the top ones in the web search) and moves to it. This web page provides information about the loan and often adds calculators that enable the customer to find out his/her credit rating. There is also a special link with an application template for a loan in the chosen bank. When visiting the web page, a customer fills out the application for the mortgage, which is subsequently sent to the broker with whom the partner has signed a contract. Then the broker sends this application to the bank with which the broker signed the contract in the first place. The bank begins the credit procedure, contacts the client, checks the credit rating, and makes a decision.

The main challenge of a strategic nature for this briefly described market is making accurate predictions about the number of mortgages sold, especially concerning mortgages sold by web partners. Due to the fact that there are several variables of a different nature (quantitative and qualitative) influencing the market, traditional statistical methods cannot be used here because they do not work properly. This is an important modeling aspect in the case where a number of variables, such as "current market feelings", cannot be expressed numerically. As stated in (Orłowski, Szczerbicki 2010, 520) the complete description of a real-life system often requires far more detailed data than a human being could ever simultaneously recognize, process, and understand (Zimmermann 2001, 3).

To address the challenges presented above, a dedicated rule-based model for predicting the number of mortgages sold using web partners has been proposed and developed. The proposed model is described in the following section.
THE PROCESS OF MODEL DEVELOPMENT

To deal with the problem existing on the market, a dedicated model was to be created. Its creation took several years, over which several steps were made, the most important of which will be presented in the subsequent parts of the current paper.

In 2008 the first version of a rule-based model was created, based on the experience of one of the authors of the Internet mortgage market encompassing the period 2003-2008. As such, it represented market conditions from that time period (a fast-growing market with highly positive prospects for the future). The model consisted of 240 rules divided into two scenarios: a positive and a negative one. The general production rule in the model was as follows:

Production Rule:

IF variable_1 is value_1 AND variable_2 is value_2 AND variable_3 is value_3 AND variable_4 is value_4 AND variable_5 is value_5 AND variable_6 is value_6 THEN the result will increase value_7

The rule consists of variables and their values presented below:

- Variable_1 = [Commission]
- Variable_2 = [Interest rates]
- Variable_3 = [Advertising]
- Variable_4 = [WIG]
- Variable_5 = [IbnGR]
- Variable_6 = [WNE]
- Result = [Selling mortgages in the Internet]
- Value_1 = [Small, medium, high]
• Value_2 = [Very small, small, medium, high]
• Value_3 = [Very small, small, medium, high, very high]
• Value_4 = [bad, average, good, very good]
• Value_5 = [bad, average, good, very good]
• Value_6 = [very bad, bad, average, good, very good]
• Value_7 = [Very small, small, medium, high, very high]

After analyzing, it turned out that the correlation of variables in the model might be problematic. Generally, the pre-defined initial group of 6 variables had an unnecessarily strong representation of very general economic indicators which highly correlated with each other. The correlation was checked and the variables in the model were optimized. Having dealt with the correlation issue, the market situation of the 2008 financial crisis exposed the authors to another problem. The worldwide economic situation changed rapidly and the deep financial disturbance reached the mortgage market. The model, which was developed and embedded in times of a fast-growing market and economy, did not work properly in the times of financial crisis that came later, so it was necessary to introduce changes to the developed model.

As a result of the introduced changes, the number of general economic variables was reduced to include 4 variables.

**FUZZY MODEL**

Unfortunately, the redefined rule-based model did not work as expected mostly due to the specific character of the mortgage market, with several variables influencing the market, which, however, cannot be described in terms of numbers using crisp values needed for hard mathematical modeling. Thus, it turned out necessary to try out a different approach, which is also the reason why fuzzy modeling was suggested. The process of building a fuzzy model was

In the process of fuzzy model development, a rule-based model consisting of 81 production rules was used (Orłowski, Szczepanicki 2010, 56). The number of rules came from the number of variables (four) in the model and the values of these variables (three linguistic values for each variable): \(3^4 = 81\) (2)

Each rule in the rule base is developed using the IF... THEN logical construct consisting of four variables, as in the following example (Ruan, Kerre 2000, 10):

*Production Rule:*

IF Commission is *small* AND Interest rates are *small* AND Advertising is *small* AND WIG is *small* THEN Selling mortgages in the Internet is *small*  (3)

As it was not possible to automatically generate the output values, expert market knowledge was employed. For defuzzification, the Height Method (also known as the Max-membership principle) was used (Szczepanicki, Lisboa and Kacprzyk 2000, 69). Finally, a fully developed fuzzy model was created and basic tests were made. After analyzing the results, the idea of proposing sub-models that could account for different market situations was stated.

**SUB-MODELS**

First, as presented before, a general fuzzy model was created; a model which was based on the original idea of developing a single model to represent the market reality (presented in the first section). At first, problems appeared during the process of selecting the variables for the model; it seemed impossible to find variables representing the whole market described by data from different time periods. In each period, different variables represented the market conditions. Based on past results (data from the years 2003-2010), one can conclude that the creation of one
comprehensive model does not give the proper results, especially when the market changes. Due to this fact, it was decided to divide the model into three single sub-models suitable for different market conditions.

Referring to the theory of economic cycles, it was decided to compose a general model of three sub-models representing three main market conditions: a fast-growing market (boom), a recession, and moderate/stable growth, which are the most distinct stages of the economic cycle. As it was important to keep a proper balance between the number of sub-models proposed and the level of differences influencing the market, the three sub-models seemed a proper solution.

In the next steps, each sub-model was fully described and work on implementation and verification started. It took five months to prepare the final version of the sub-models.

**REPLICATIVE VERIFICATION**

Having worked out the model, the time for its verification came. The authors decided to divide the process into three parts as there were 3 sub-models created for three market situations. Each verification was based on data recorded in the market situation for which the sub-model was designed. Later, proper weighting of variables, and granulation levels, were added.

The model was verified using a dedicated software interface developed in MS Excel. The interface allows for easy changes of the number of variables, weights, and granulation levels.

**SUB-MODEL 1: STABLE MARKET**

Sub-model nr 1 is dedicated to the conditions of a stable market - the market situation observed on the Polish mortgage market in the years 2005-2006. Based on the previously-selected variables for the sub-model, it was necessary to verify the model against the real, historical data. This paper will later present two cases for that sub-model, the first case aims both
to check the influence of each input variable on the output variable and to verify the way the sub-model works on real data.

**Case 1**

Case 1 is built based on real data observed on the Polish Internet mortgage market in February 2005 (Table 1). In the first attempt, the granulation level in the presented sub-model was set at level 3, which produced 81 rules generated by the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0.55%</td>
</tr>
<tr>
<td>Interest rate</td>
<td>6.50%</td>
</tr>
<tr>
<td>Advertising</td>
<td>181.96 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>26864</td>
</tr>
</tbody>
</table>

As stated above, the first priority was to check the reactions and tendencies of the sub-model based on the following input-output variable changes:

**Commission**

First, the influence of Commission on the output variable was checked. The minimum and maximum possible commission levels were used in the model, while all the other variables remained constant. It was observed that when the commission changed from 0.55 up to 0.8 there were no changes in the output variable because the maximum level of commission recorded in all the data was never higher than 0.6, so higher values never in fact existed.

Then, the level of commission was lowered from 0.5, 0.4 to 0.3. In each case, the output variable generated by the model was also lower - so the reaction of the model was correct.

**Interest rate**
Next, the influence of the interest rate on the output variable was checked. In theory, the higher the interest rate, the lower the amount should be of mortgages sold (output variable).

The first test was prepared for the interest rate 6.5% with other input variables constant - the model generated the output variable (value of mortgages sold) of 107.55 PLN. In the second test, the interest rate was changed to 4.5% - the model generated the output variable of 660.13 PLN, as theoretically assumed. In the next test, the interest rate was 5%, for which the model generated the output variable of 235.76 PLN.

The above results proved that the model reactions to the changes of the input variable “interest rate” were compliant to theoretical assumptions.

**Advertisement**

The input variable “advertisement” presents the amount of money spent by the owners of private web pages on advertising mortgages. According to the theory - the more money is spent on advertising, the higher the value is of mortgages sold.

For the first test, the input variable “advertisement” was set to 152.21 PLN (the lowest amount recorded within the time of observation). For the presented input variable, the model generated the output variable of 235.76 PLN. Next, the input variable was changed to 285.43 PLN and the model generated the output value of 347.97 PLN, which is congruous with theoretical assumptions. In the third step, the tested input variable was increased to 982.41 PLN. The model generated the output variable of 1292.084 PLN, which is also congruent with theoretical assumptions.

**WIG**
WIG is Warsaw Stock Exchange’s main indicator. It was assumed that the higher its value is and the more its tendency to grow is, the higher is the expected value of mortgages sold (output value).

The first test was prepared for the input value (WIG) of 25617 points, which was the lowest result recorded in the whole observation period. The model generated the output value (value of sold mortgages) of 1292,084 PLN. In subsequent tests, the input value was increased, while the other input values were stable (ceteris paribus).

In the second observation, the input value (WIG) was increased to 35000 points which was the average level of all observed records (25617–51961). With the new input value and ceteris paribus, the output value generated by the model was 1484,162 PLN, which is congruous with theoretical assumptions. It was then decided to prepare a third test, with the input value (WIG) set at 48000 points, in which case the model generated the output value of 1650,335 PLN, which agreed with the accepted assumptions.

Having tested the influence of input variables, it seemed necessary to test sub-model 1 itself.

Case 2

In the second case, new input data was introduced and it was decided to use real data from August 2005. For the given input data, the granulation level was set to 3, due to which 81 rules were generated by the model. The model generated the output variable of 675,44 PLN, which was 22,6% lower than the real value recorded in the presented time.

When the granulation level was changed to 4, 256 rules were generated and the model generated the output variable of 1072,13 PLN, which was 22,8% higher than the real value from August 2005.
A third test was carried out with the same input variables but the granulation level changed to 5; 625 rules were generated and the output variable of 825.17 PLN was only 5.5% lower than the real value.

**SUB-MODEL 2: RECESSION**

Sub-model nr 2 is dedicated to the conditions of the market in recession, observed on the Polish mortgage market in the period between the beginning of 2008 till the fourth quarter of 2009 and it has been verified based on the data recorded at that time.

**Case 1**

In the first case, data from March 2008 was selected (Table 2) to verify the sub-model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0.55%</td>
</tr>
<tr>
<td>WIBOR6M</td>
<td>6.3</td>
</tr>
<tr>
<td>Advertising</td>
<td>830 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>44900</td>
</tr>
</tbody>
</table>

For the presented input values, three evaluations were made. For the first one, the granulation level was set at 3 and 81 rules were generated. The model produced the output variable (Selling Mortgages in the Internet) of 1040.35 PLN, which was 38% lower than the real value recorded at that time (1704 PLN).

A second test was carried out on the same input variables but the granulation level was set at 4 and 256 rules were generated. The model generated the output variable of 1306.29 PLN, which was 23% lower than the real value recorded in that time.
A third test was based on the same input variables but the granulation level was set at 5 and 625 rules were generated. The model gave the output variable of 1154 PLN, which was 32% lower than the real value.

It can be concluded that the difference of about 30%, as presented above, is significant, so for complete verification some other input data are needed (case 2).

Case 2

The data from December 2008 (presented in Table 3) was used in the second case. The methodology of verification was the same as that presented in case 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0%</td>
</tr>
<tr>
<td>WIBOR6M</td>
<td>6.42</td>
</tr>
<tr>
<td>Advertising</td>
<td>0 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>27725 points</td>
</tr>
</tbody>
</table>

Three evaluations were made; for the first one, the granulation level was set at 3 and 81 rules were generated. The model gave the output variable (Selling Mortgages in the Internet) of 0 PLN, which was the same as the real value recorded in that time.

The second test used the same input variables but the granulation level was set at 4 and 256 rules were generated. The model generated the output of 0 PLN, which was the same as the real value recorded in that time.

The third test used the same input variables but the granulation level was set at 5 and 625 rules were generated. The model gave the output variable of 0 LN, which was the same as the real value recorded in that time.
Case 2 was chosen for specific data - the real recorded value of 0 PLN, typical of a recession. Sub-model 2 produced the proper results (the output value produced by the model was the same as the real value recorded in that period) for all granulation levels.

The results received in case 1 tend to be rather imprecise, especially when the results are compared with those received in case 2, which perfectly match the real values. It seems to be necessary to do detailed research (including changes in the weights of variables) and add other cases with new data to verify the sub-model.

**SUB-MODEL 3: FAST-GROWING MARKET**

Sub-model nr 3 is dedicated to the conditions of the fast-growing market. The market situation called "fast-growing market" was observed on the Polish mortgage market between the third quarter of 2006 till the end of 2007. Below, it will be verified based on data recorded at that time.

**Case 1**

In the first case, data from October 2006 was selected (Table 4) to verify the sub-model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0.58%</td>
</tr>
<tr>
<td>Interest rate</td>
<td>4%</td>
</tr>
<tr>
<td>Advertising</td>
<td>600 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>47393</td>
</tr>
<tr>
<td>Average mortgage value</td>
<td>155524 PLN</td>
</tr>
</tbody>
</table>

For the presented input values, three evaluations were made. In the first one, the granulation level was set at 3 and 243 rules were generated. The model produced the output variable
(Internet Mortgage Sales) of 1443.65 PLN, which was 15.7% lower than the real value recorded at that time.

The second test was based on the same input variables but the granulation level was set at 4 and 1024 rules were generated. The model generated the output variable of 1560.87 PLN, which was 8.8% lower than the real value recorded in that time.

The third test was based on the same input variables but the granulation level was set at 5 and 3125 rules were generated. The model gave the output variable of 1656.85 PLN, which was 3.3% lower than the real value.

It can be concluded that a difference of 10% and lower is quite acceptable, which means that the granulation levels of 4 and 5 seem to be proper for future research. For completeness, the verification should also be performed based on some other, different input data (case 2).

**Case 2**

Data from October 2007 (presented in Table 5) was used in the second case. The methodology of verification was the same as presented in case 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0.55%</td>
</tr>
<tr>
<td>Interest rate</td>
<td>4.75%</td>
</tr>
<tr>
<td>Advertising</td>
<td>598.3 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>64292 points</td>
</tr>
<tr>
<td>Average mortgage value</td>
<td>196948 PLN</td>
</tr>
</tbody>
</table>

Three evaluations were made. For the first one, the granulation level was set at 3 and 243 rules were generated. The model gave the output variable (Internet Mortgage Sales) of 1607.68 PLN, which was 5.5% lower than the real value (1701.26 PLN) recorded in that time.
The second test used the same input variables but the granulation level was set at 4 and 1024 rules were generated. The model generated the output variable of 1925.39 PLN, which was 13.1% higher than the real value recorded in that time.

The third test used the same input variables but the granulation level was set at 5 and 3125 rules were generated. The model gave the output variable of 2062.92 PLN, which was 21.2% higher than the real value.

Case 2 also produced acceptable results for the difference between the verification and real result of 5.5% but only when the granulation was set at 3, whereas this granulation level was the worst one in the previous case. Due to this fact, a third verification run had to be made (case 3).

**Case 3**

In case 3, data from February 2007 was selected (Table 6) to verify the sub-model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commission</td>
<td>0.43%</td>
</tr>
<tr>
<td>Interest rate</td>
<td>4%</td>
</tr>
<tr>
<td>Advertising</td>
<td>1054.25 PLN</td>
</tr>
<tr>
<td>WIG</td>
<td>53896 points</td>
</tr>
<tr>
<td>Average mortgage value</td>
<td>184000 PLN</td>
</tr>
</tbody>
</table>

Three evaluations were made in case 3. For the first one, the granulation level was set at 3 and 243 rules were generated. The model gave the output variable (Internet Mortgage Sales) of 2531.29 PLN, which was 72% worse than the real value (1473.32 PLN) recorded in that time.

The second test was made using the same input variables but the granulation level was set at 4 and 1024 rules were generated. The model generated the output variable of 2413.15 PLN, which was 64% worse than the real value recorded in that time.
The third test was made using the same input variables but the granulation level was set at 5 and 3125 rules were generated. The model gave the output variable of 2475.5 PLN, which was 68% worse than the real value.

All the results obtained in case 3 tend to be rather poor, especially when compared with those obtained in cases 1 and 2.

Consequently, the detailed research was done (including changes in the weights of variables which did not change the results) and it turned out that the data used in case 3 was very rare and not typical for the fast-growing market (very small total value of mortgages sold). It seems statistically possible that this kind of situation may happen but generally, the situation on the market differs from the one presented in case 3.

CONCLUSIONS

Based on the real input data, several verification processes were made for each sub-model (including the examples presented above). For the sub-models of stable market and boom the results might be called well-rounded with just a few worse results but generally, an average mistake amounted to around 10-12%. The sub-model for a recession still needs work because the average difference between the real result and the model prognosis was much higher.

In the future, "special cases", like case 3 in boom, will need to be considered, which means cases with input data which is rare for that kind of a defined market, but which happened in real market situations. That is also a point of future works.

PROGNOSTIC VERIFICATION

Besides verification based on historical data, the complex creation of the model required the usage of prognostic verification.
The process of prognostic verification was divided into two stages. First, the main manager responsible for managing the Partner System (85% of partner web pages existing on the market), being the best market expert in the country, was asked to chose web pages representative of the whole market. The owners of selected web pages took part in the first part of the verification process by answering questions about input variables and specially created business scenarios. In the second part, the model was assessed by the main manager of the Partner System.

At this point, the question arises why the selected web page owners did not receive the ready-made model for testing. The main reason was the fact that the authors had no tools to check if the selected page owners acted entirely in the way suggested by the sub-model. For example, if the model suggests optimal advertising expenses, one will need to be certain that the user has acted in the way suggested by the model. In fact, users tend to mistrust a model which is in the testing phase and will be reluctant to invest their money for fear of the model being inaccurate. To overcome this difficulty, special real input data business scenarios were created, to generate prognoses by the model. The owners of web pages, being market experts, evaluated the scenarios and the input variables used in the model. To reduce the risk of untypical expert experiences, several experts took part in this process.

**THE RESULTS**

Two out of three presented business scenarios were assessed as very reliable by most of the owners of the web pages. The third business scenario was considered reliable by 1/3 of verified experts, which means that this sub-model needs future development.

The general input variables used in the model were mostly assessed: useful by the experts, with a few exceptions like WIBOR6M. It must be added that the web page owners were ranking variables from the point of view of their own pages, not referring to the market as a whole.
whereas variables like WIBOR6M represent the whole market and as such this is included in the model.

CONCLUSIONS

The new, fast-growing Internet mortgage market seemed to need a dedicated model to support the owners of partner web pages existing on it. When a rule-based model was created, changing market conditions and the model’s improper reactions resulted in several changes that had to be made. Eventually, a fuzzy model was created, which subsequently was divided into three sub-models.

The ready model had to be verified, which was done in two steps. Firstly, the model was verified based on real historical data. For each sub-model, several cases were made in which model prognoses were compared with real results. Then a prognosis verification was carried out in which dedicated experts, both representatives of partner web pages and brokers, verified the input values used in the model and business scenarios created for this process.

Generally, the results were satisfactory but one of the sub-models (recession) has to be worked on, and special cases that rarely appear in the market must be taken into consideration. It is also necessary to work on a dedicated tool that will automatically choose a proper sub-model according to the current market conditions. Finally, the model will have to be tested against more input data.

REFERENCES


