MAPPING OF REAL ESTATE PRICES USING SOFTWARE TOOLS

Eduard Hromada[1]*

[1] Czech Technical University in Prague, Faculty of Civil Engineering, Thákurova 7, 166 29 Prague 6, Czech Republic

Abstract

The paper informs about an innovative software tool called EVAL that can be used for mapping and analysing of real estate prices. The presented software tool systematically assembles, analyses and assesses property price quotations in advertisements. For each half year, the software assembles over 650,000 price quotations concerning sale or rental of apartments, houses and allotments. All offers are continuously stored in a database and are thoroughly analysed for credibility. Thus every price quotation is assessed in the light of the objectivity and accuracy of presented information, and compared with older quotations. There is a range of approximately 250 error checks to verify each quotation and eliminate purposely misrepresented information, or manipulation malpractices. In case of the slightest discrepancy (for example designed manipulation of information by real estate agent, purposely incomplete information on the presented property, duplicity, etc.) quotation is discarded from the database. EVAL proposes an innovative way of property market evaluation. It is useful for determining the market value of apartments, family houses and allotments. It may also be used for the determination of the usual rate of apartment or house rent and for the determination of a property market value at an older date of evaluation.

Key words

Evaluation of properties; market value; real estate; software tools

*Corresponding author: Tel.: +420-224-353-720, Fax: +420-224-355-439
E-mail address: eduard.hromada@fsv.cvut.cz

1   INTRODUCTION

Due to current economic recession, real estate has become an attractive option of investment especially for natural persons. Future development of economic situation is highly uncertain, and real estate is a way of investment which does not decline in value rapidly. However, latest recession development shows that real estate prices cannot keep rising all the time. Moreover, risks are connected to real estate management, and they need to be constantly monitored. The potential buyer, therefore, has to evaluate real estate property critically, and see to its pertinence in long-term prospect, in other words decide whether real estate is fit to become an investment.

Real estate's importance in various commercial activities keeps rising, it is used as a pledge of security, means of securing investment plans, safeguard of value, etc. Due to these, and other, factors there have been numerous articles concerning real estate market analysis in both mass media and scholarly publications. Unfortunately, not all presented information is objective and unbiased. Many cases have shown "independent specialists" stating information with no verifiable background. We have witnessed manipulation of information by lobbies (such as banks offering mortgage, real estate companies, building companies, developers, etc.). We have also witnessed self-censorship in several magazines and newspapers. Independent authors, professionals in the given field, do not publish negative information about real estate market development. Such information, though true, is not communicated because income of the mentioned media is formed largely by advertising space for lobbying groups. The author of this paper offers objective and unbiased evaluation of price development in real estate market because he is a member of an independent institution - Czech Technical University in Prague. Author brings forward information based on extensive research.

2   LITERATURE REVIEW

Many studies analyse real estate prices, especially at the national level. The study [1] analyzes the long-run integration relationship between equity and real estate prices in 30 developed and emerging economies divided into four subpanels according to income levels and financial market structure. The paper [2] describes method of creation of real property database to determine capitalization rate of real estate. The article [3] discusses approaches and options for identification of disequilibrium asset prices movements. The paper [4] presents a method of applying the ensembles of genetic fuzzy systems to build reliable predictive models from a data stream of real estate transactions. The paper [5] presents a model that includes a database storing hedonic characteristics and coefficients affecting the real estate price level, and uses information from recently sold projects. The results show that fuzzy neural network prediction model has strong function approximation ability and is suitable for real estate price prediction.

The issue of real estate database creation is also solved by the team of researchers from Universidad de los Andes, Colombia [6]. The research team created unique database of residential real estate prices in Caracas containing 17,526 transactions. The statistical results of the database are used for testing the changes in housing prices in the case of occurrence of risks such as land invasions and expropriations. The research team examines the microeconomic determinants of residential real estate prices (the number of parking spaces, the age of the property, the incidence of crime, the average income in the neighbourhood, etc.) that significantly affect the prices.

The paper [7] describes the data mining methods of real sales data from the assessment office in a large US city. The results are used for predicting the value of residential real estate based

on past comparable sales transactions. The paper introduces an approach for improving predictions using an adaptive, neuro-fuzzy inference model.

The paper [8] presents the utilization of the database that contains 21,902 units of flats in real estate market of Santiago de Chile that is used for introducing improvements in terms of summer thermal conditions according to the specific requirements of the apartment typologies. Based on the database results, it was found that the best performance in terms of summer comfort can be obtained from the combination of diverse parameters that would be significant in respect to the reduction of overheating, such as solar protection and night ventilation.

The paper [9] describes the results of case study that uses representative database of sales in urban real estate for a medium size city in the South of Spain. There are used Artificial Neural Networks that provide better dwelling prices estimation, avoidance of bias at different market segments, direct use of categorical data and full use of the information available.

The paper [10] presents prototype system for real estate information collection and visualization. There are used some techniques like data indexing, data compression, distributed file system and highly efficient structured query language mechanism. There are used visualization techniques such as Treemap, StreamGraph and Line Chart to display statistical data of real estate market.

The paper [11] focuses on presentation of the real estate appraisal system based on GIS and BP neural network. The system includes appraisal model, trade case, GIS database and query analysis module. The system allows to construct the hedonic price estimate model. The system can improve the efficiency of the real estate after-sale service and improve the service level.

At this moment, there are only two scholarly institutions in the Czech Republic, except for the author's work, which focus on real estate problematic in long-term and systematic manner. These are Institute for Regional Information, Ltd. and the department at Czech Technical University in Prague led by doc. Ing. Václav Dolanský, CSc. Both places collect advertisements of real estate property for sale and perform comparative analysis of real estate price development. However, they collect advertisements manually in both workplaces, which only allows for a limited scope of data retrieval. Statistical processing of such data volume is only sufficient for basic and aggregated analyses, and also causes distortion of results that can be obtained. Both departments research only a handful of specific areas, a very limited scope of real estate categories, and factors that determine real estate's market prices.

Czech Statistics Authority is also an institution which gathers limited data in the field of real estate. This Authority annually issues a statistical analysis of common real estate prices. Common price is established by authorized experts at the moment of sale, and needs to be assessed for the purposes of statement of taxable income. These analyses are sourced from the statements of taxable income handed in by real property owners. Such data, however, only offers basic aggregated information on price development. Statements of taxable income are often formed by disfigured information about real estate value. Taxpayers frequently coerce authorized experts towards low numbers, because it means lower income tax to be paid. However, there has been a substantial change in these politics since 1st January 2014. Taxpayer may now choose a different form of establishing tax base – they may still ask an authorized expert to state common price, but they may state informative value based on market prices of comparable real estate in given place and time. Such numbers are ever less realistic, and information submitted by Czech Statistical Authority loses its value and credibility entirely [12].

3   EVAL SOFTWARE CONCEPTION

EVAL is a software application which systematically gathers, analyses, and evaluates real estate price offers advertised via real estate servers. At this point, the scope of the software application covers and records the majority of real estate advertisements on the Internet in the Czech Republic. Information has been automatically gathered from servers each month since 2007. In the first half of 2014, software database gathered more than 650,000 new entries with price offers for purchase or rental of flats, houses, commercial objects, and allotments. The software tool enables detailed analysis of market price development in monthly periods, in all municipal areas in the Czech Republic, and in all real estate categories as sorted by servers which offer real property for sale or rent.

It should be noted that there is a difference between "list prices" and the "final prices" paid. According to the author's own research, the difference is between 10-15 percent. Some offers on real estate websites are only wishes of property owners. Though not precise in every detail, average values of real estate market show current trends of property prices quite accurately.

EVAL software is being further developed and adjusted by its author. Currently, a module has been added to gather information from land registry. The tool is built using *Borland Delphi for Microsoft Windows* programming. For optimal functioning it requires stable and fast internet connection, Microsoft Windows XP operation system, or a higher version, large storage space on hard disc (numbers of several TB of data), and an efficient processor providing for processing of large data files.

Software is structured as several partial modules which co-operate towards a common end:

A. Module for gathering and download of links on the Internet

This module systematically searches real estate servers and webpages of real estate companies, and logs current links to real estate purchase or rental into a database. For full functioning of the module, it was necessary to discover the methods of storing data on each server where EVAL gathers information. This module of the EVAL gathers over 110,000 individual entries each month – full texts of individual advertisements.

Data acquisition is performed in following real estate categories:

- Flats for sale.
- Flats for rent.
- Allotments for sale.
- Houses for sale.
- Commercial objects for sale.
- Commercial objects for rent.

This module requires author's adjustments and interventions on a non-scheduled basis because the providers of real estate servers change the structure of data storage from time to time. Should such risk arise, it is necessary to adjust EVAL's source code as soon as possible. Otherwise data from that given server may not be downloaded in that particular month. Author uses automated control mechanisms to manage the risks of such situation; after each monthly download, the volume and structure of data is compared with previous month's log. If there is an impermissible divergence (lacking data continuity) or if no data is downloaded, the server is analysed closely, adjustments are made, and data is downloaded correctly.

As already said, data download is repeated monthly. In each month, the volume of gathered data is circa 25 GB. EVAL database is enriched by these new entries each month. That is the

reason why EVAL has extensive requirements concerning data storage and Internet connection. However, these requirements are easily met due to massive spread of memory media and Internet speed development over the past few years. Despite the size of database acquired since 2007, the author still keeps full textual versions and visual attachments of all noted advertisements. It is therefore possible to trace data, and subject them to further analyses.

The author presently focuses on broadening the scope of data acquisition – data made public by land registry should also be retrieved. But download from land registry is complicated because of the high level of data protection in systems of Czech Land Planning and Land Registry Office. There are obvious differences between land registry and real estate advertisements – data stored in real estate companies and their servers are essentially commercial advertising. There are no limits or restrictions, data need to be seen, accessed, and assessed easily. Such is the main strength of public servers, and at the same time their major weakness. With land registry data, situation is altogether different. At this moment, EVAL only contains complete land registry information on Prague as a region, and a part of Central Bohemia. To obtain this amount of data, 2 months of non-stop computer time of the application were necessary. Although acquired results are very interesting, their publication is problematic. Current Land Registry Law (§ 53) establishes high sensitivity of the data. Data in land registry contain detailed information about ownership, legal titles connected to real property, debtor distrait procedures, information on chattel mortgage, all real property transfers, etc. Entries also contain full names and addresses of all real estate owners. The potential for abuse of data is very high, and so Land Registry Law heavily restricts publication.

B. Data export module

This module is the most complex part of the software tool. Its task is to automatically export specific data from the downloaded advertisements. Then the quantified data can be further used for statistic evaluation of real estate market development. Data structure is defined for each real estate category obtained from advertisement texts. As an example, this paper states the structure crafted for category "Flats for purchase":

- Region.
- Municipality.
- Street.
- Postcode.
- Real estate price.
- Price per 1 m^2 of floor space.
- Advertisement number.
- Date of last modification.
- Date of advertisement entry.
- Space arrangement.
- Floors.
- Number of object's underground floors.
- Number of object's above-ground floors.
- Ownership.
- Floor space.
- Loggia space.
- Balcony space.
- Terrace space.
- Cellar space.
- Type of building.
- Condition of building.
- Path and file number of file with advertisement text.

There are similar structures for other categories of real estate, they follow from common features in each category. For advertisements that do not state certain parameters of given real estate (e.g. terrace space, cellar space), database states a blank space for that parameter. However, it is in the interest of the advertising party to state as much information as possible.

The output of *Data export module* are database files which contain structured data about advertised real estate.

Data export module also requires the author to monitor any potential changes in data structure and formatting across individual real estate servers (e.g. names of smaller municipal units in the vicinity of Prague have recently been replaced by more generalized locations "Prague-east" or "Prague-west"). In such case, when data structure changes, EVAL's source code needs to be readjusted accordingly. However, these changes do not need to be reflected immediately, as data in these cases are already stored in an offline database in storage media.

C. Data filter module

Database files created by *Data export module* are further analysed and verified. Every newly acquired price offer is evaluated from the viewpoint of objectivity and truthfulness of presented information, it is compared with older entries, repeatedly advertised real estate, checked for completeness of information, etc. There are roughly 250 potential mistakes to evaluate in each advertisement. If crude discrepancies or potential manipulation of information are discovered (e.g. advertising large numbers of similar offers of fictitious real estate for higher prices than usual in given street or location which may influence potential client's decision making process), such entry is discarded from further evaluation.

For each location in the Czech Republic, and each real estate category, a certain confidence interval of prices is suggested. Real estate transactions are usually made within this confidence interval. If advertised price is not within the confidence interval, the advertisement is discarded. The confidence interval was determined on the basis of long-term monitoring of property prices. Acceptable deviation from the average real estate prices can be tolerated. Price data within this acceptable range are downloaded by EVAL, its safety mechanisms may not interfere with its function. Rather rule out untrustworthy advertisements so that they cannot be used in statistical analyses (such as those where the advertiser falsely states the price of real estate as 1 CZK in order to place their real estate at the top of real estate servers' offers, another example is the advertisement of the flat with floor area of 1 m^2 only). This system also treats advertisements with false statements of built-up area/floor area/number of floors, which is unrealistic in given category.

At this stage, a difficult, yet vital task of the application is to identify duplicity of advertisements. That is often a problem even for a human researcher, and it is even more problematic for a computer code. EVAL evaluates the extent of match in key parameters of advertised real estate (one of the many intercepting parameters may be for example telephone number or e-mail of the advertising party). Duplicate advertisements have to be removed from the database so that they do not distort statistic data. If there happen to be several duplicate advertisements for one real estate, only the one with the lowest offer price is stored in the database. If the demanding party (buyer, tenant) finds duplicate advertisements, they will strictly stick to lowest offer price, not an average value.

D. Data evaluation module

This part of the software tool creates statistics schemes to describe the development of real estate market according to user's demands. It examines development on a timeline, and scrutinizes the relations among the many monitored variables of real estate. These schemes may be used for the needs of real estate development analysis, evaluation of regional disparities concerning financial accessibility of housing, and creation of prognoses concerning market development. Following table 1 shows examples of statistics schemes which can be created by EVAL.

Tab. 1: Examples of EVAL-generated statistics schemes.

Real estate category	Scheme example
Flats for sale	Temporal development of market prices in relation to ownership (personalty, co-operative ownership)
	Temporal development of market prices in relation to apartment category (garconniere, 1 bedroom to 5 bedrooms, kitchen or kitchenette)
	Temporal development of market prices in relation to material characteristics (brick, concrete, wood)
	Temporal development of market prices in relation to location (region, municipality, possibly even neighborhood, or street if data are sufficient)
	Temporal development of market prices in relation to the size of municipality
	Relation between flooring, market price for 1 m^2 or rental
	Relation between market prices of flats and houses
	Temporal development of the number of real estate on offer (advertising party) in relation to flat category (number of bedrooms, kitchen or kitchenette) and location
Allotments for sale	Temporal development of market prices in relation to allotment category (building ground, commercial, garden, agricultural, wood, etc.)
	Temporal development of market prices in relation to location (region, municipality, possibly even neighborhood if data are sufficient)
	Temporal development of market prices in relation to size of municipality
	Relation between allotment size and market price for 1 m^2
	Relation between market price and distance of allotment from road/ mass transportation
	Temporal development of the number of real estate on offer (advertising party)

Many other schemes like these may be created according to the needs or demands of researchers, and professionals.

4   CONCLUSION

The output of EVAL software enables professionals and researches in the field of real estate to gain insight on actual development of market prices of real estate in the Czech Republic. This output may be used as grounds for appropriate investments or housing decisions for both natural persons and companies.

We have witnessed a steady long-term decrease of real estate market prices since the second quarter of 2008. This negative trend does not seem to be changing. Although media strive to present positive information, there is no sign that prices should start rising. Only a handful of real estate categories in very distinct locations display stagnation of prices or ever so slight positive development.

It is not easy to answer the question how much market prices have actually dropped since the peak in the second quarter of 2008. Author of this paper assumes that prices of the second quarter of 2008 were not objective market prices (there is a discrepancy in perception/definition of market price and objective price/value of real estate). Year 2008 could probably be identified as an artificially created price bubble of real estate market aided by macroeconomic, financial, marketing and medial tools. Such development may have been caused also by availability of mortgage for a large portion of the population.

REFERENCES

[1] Čeh Časni, A., Vizek, M. (2014). Interactions between real estate and equity markets: An investigation of linkages in developed and emerging countries. *Finance a Uver - Czech Journal of Economics and Finance*. Prague. **64**(2) , pp. 100-119. ISSN: 0015-1920.

[2] Ardielli, J.; Janasova, E. (2012). Creation of real property database for determination of capitalization rate of real estate. *12ᵗʰ International Multidisciplinary Scientific GeoConference and EXPO, SGEM 2012,* Varna; Bulgaria. **4**, pp. 877-882.

[3] Komárek, L.; Kubicová, I. Methods of identification asset price bubbles in the Czech economy. *Politicka Ekonomie*. Prague. **59**(2), pp. 164-183. ISSN: 0032-3233.

[4] Trawiński, B.; Smętek, M.; Lasota, T.; Trawiński, G. Evaluation of fuzzy system ensemble approach to predict from a data stream. *6ᵗʰ Asian Conference on Intelligent Information and Database Systems, ACIIDS 2014,* Bangkok; Thailand. **8398**(2), pp. 137-146. ISSN: 1611-3349.

[5] Liu, J. G.; Zhang, X. L; Wu, W. P. Application of fuzzy neural network for real estate prediction. *3ʳᵈ International Symposium on Neural Networks*, ISNN 2006 - Advances in Neural Networks. Chengdu. China. **3973**, pp. 1187-1191. ISSN: 0302-9743.

[6] Contreras, V.; Garay, U.; Santos, M.A.; Betancourt, C. Expropriation risk and housing prices: Evidence from an emerging market. *Journal of Business Research*. **67**(5), May 2014, pp. 935-942. ISSN: 0148-2963.

[7] Guan, J.; Shi, D.; Zurada, J. M.; Levitan, A. S. (2014). Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration. *Journal of Organizational Computing and Electronic Commerce*. **24**(1), pp. 94-112. ISSN: 1091-9392.

[8] Encinas, F.; De Herde, A. (2013). Sensitivity analysis in building performance simulation for summer comfort assessment of apartments from the real estate market. *Journal Energy and Buildings*. **65**, pp. 55-65. ISSN: 0378-7788.

[9] Núñez Tabales, J.; Rey Carmona, F.; Caridad Y Ocerin, J. M. (2013). Implicit prices in urban real estate valuation. *Journal Revista de la Construccion*. **12**(2), pp. 116-126. ISSN: 0717-7925.

[10] Yu, D.; Yu, M.; Ye, L.; Liang, R. (2012). *Method of real estate information services based on Hadoop. Journal of Huazhong University of Science and Technology (Natural Science Edition).* **40**(1), pp. 66-69. ISSN: 1671-4512.

[11] Zhang, S.; Yu, Y.; Sun, J. (2011). Design and development of real estate after-sale service system based on networking. *Journal of Shenyang Jianzhu University (Natural Science).* **27**(2), pp. 403-408. ISSN: 1671-2021.

[12] Schneiderova, R. (2007). The Building's Value Assessment using the Utility and the LCC. *Central Europe towards Sustainable Building 07 Prague*. Prague: CTU, Faculty of Civil Engineering, **1**, pp. 126-131.