A PRELIMINARY WORK TOWARDS THE DEVELOPMENT OF DIGITAL PROPERTY PRICE MODELING

Assoc. Prof. Dr. Tono Saksono[1]
Norshazwani Afiqah bt Rosmera

Abstract

This is a preliminary study towards the development of a spatial-based digital property price modeling (DPPM) that will be of primary importance as a genuine and comprehensive market guide for investors, industries, financers, brokers, insurance and regulator in real estate industry and market. The study commences with the development of a robust hybrid modeling of all independent social and economic variables to construct stochastic and functional model that determine a property price at specific geographical point. That is in essence a spatial-based econometric modeling. Data used to check the robustness of the model were collected by twenty six senior students in the Department of Real Estate Management, UTHM during students' mobilization program to Kuching, Sarawak that was partly financed by UTHM. Survey planning and affirmation were conducted using Google Earth during GIS class. Geographical coordinates were recorded by using students' smart mobile phones that are equipped with mobile GPS navigation system, whilst property characteristics were obtained from brief interview with the property owners. The preliminary results are quite encouraging. The average relative accuracy achieved is about 7% of the average property price within the predetermined window. The results encourage future work to incorporate the use market price from property transaction data available. Based on the constructed model, an estimated price of property at any point of interest can therefore be estimated using the developed DPPM. In addition, the magnitude of real estate industry market, forecast of financial and banking industry and labor work involved, and government's tax revenue prediction in this sector can be automatically assessed through the developed DPPM.

Keywords: spatial-based econometric modeling, digital property price modeling (DPPM).

[1] Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Johor. Email: tsaksono@uthm.edu.my

Introduction

Nature is created in analog form and therefore, is physically smooth and continuous. In the mean time, computer is only able to read data in discrete format. In order that a natural object is readable by computer, its physical form then has to be transformed into digital format, only then an automation process within a digital computer can be performed. This is what so called a digitization process.

Whilst *modeling* is a representation (normally mathematical) of a process, concept, physical, or operation of a system that is normally implemented in a computer program (slightly modified from definition by *dictionary.refernce.com*). In mapping science related disciplines, for example, engineers and planners are familiar with what termed by digital elevation model (DTM) which is basically 3-D discrete representation of a terrain surface (normally earth surface). DTM is therefore essentially a result of transformation process of earth surface, which is originally in analog format, into digital format in order to be readable by a digital computer. Upon which, all automation process involving the DTM data can be performed for various applications related with human economic activities.

Scientists have put a lot of thoughts and efforts to produce DTM as a discrete representation of the earth surface. Currently, DTM data can be generated automatically via different types of modern sensors deployed in airplanes, satellites or other flying platforms. Among other advanced technologies to generate DTM data are: autocorrelation techniques applied to overlapped remote sensing images, radar interferometry technology (e.g. Shuttle Radar Topographic Mission), and LIDAR (light detection and ranging) technology, and so forth. These technologies are capable in producing billions of accurate DTM data very fast. Again, the main objective of having DTM data is for automation process in the computer. By the availability of DTM data, the followings can be carried out automatically and fast:

1. Contour lines can be generated;
2. Earthwork volume can be calculated;
3. The most economical route of planned highway can be determined;
4. Transmitter location for cellular phone network providers can be planned and constructed.

Put in short, a lot of engineering tasks will be carried out a lot easier through automation process with the availability of DTM data.

The idea to develop *Digital Property Price Modeling* (DPPM) comes into picture as an attempt to automate all planning and economic processes of all tasks related to property price data, similar to what DTM has been functioning in engineering. DTM data is simply cloud discrete points of earth's topography consisting of 3-D spatial information (*easting*, *northing*, and *height above sea level*). Spatially, it takes no difference because the first-two information also uniquely identifies the location of certain property (*easting* and *northing*) on the surface of earth. What makes them different is in the third component of the spatial information. If the third component of the information of DTM is the height of the point above the mean sea level, for DPPM it is associated with the price of property at that particular coordinates of a point that is influenced by the complexity of social, economic, geographical, and psychological factors. Mathematically, subject to the availability of proper data, the construction of these two models should be the same. The accuracy is another issue because it is directly dependent on the resolution of the original data. A civil engineer who wants to design irrigation network, for example, he/she will not be able to obtain the required accuracy to ensure water flows to follow gravity force if he/she uses 1-km resolution DTM data generated from shuttle radar topographic mission, for example. Likewise, resolution of the generated DPPM will bound the accuracy of economic analysis of the data for various purposes.

The importance of DPPM

Housing authority, banks, insurance industry, investors, or even consumers of real estate industry at all levels actually need a robust guide for the market trend of the real estate price in specific area of interest. Among others are:

1. As a regulatory body, the government needs it as the basis before it produces regulations in governing the real estate development and its corresponding sectors. In addition to that, the availability of a robust guide (e.g. analytical model) will enable the government to forecast the projected tax collection in the related sector, and in return will better estimate the contribution of real estate sector to the overall national economic growth more accurately.
2. Banks, insurance and other financial industry sectors will be capable of accurately calculating the financial risks by the availability of a robust analytical model for real estate market trend in the area;
3. For investors, a robust analytical model will guarantee a better forecast of profitability of any investment in any real estate sector of interest.
4. Individual real estate consumer will be able to better estimate his/her long-term investment, better planning for their social life, investing in the most efficient way including the most efficient way of calculating the cost of money when he/she borrows it from the bank.

The problem of the matter is: a robust analytical model for an accurate estimate of mass property valuation is not available in real life, neither in Malaysia nor in any other country. Valuation and Property Services Department (Jabatan Penilaian dan Perkhidmatan Harta - JPHT) in Malaysia certainly records property transactions, provides information regarding the summary of state-based supply of residential units, and provides property stock report. However, an overall analytical model of property price is not available at all.

Development of robust model

The study commences with the development of a robust modeling of all independent social and economic variables that determine a property price at specific geographical point. When the study is very much associated with quantitative problems, the robust model would be defined as theoretical system with which, one will be able to describe physical phenomena or properties from a set of known events. This is called the *mathematical model* by many. The mathematical model does not have to be exhaustive, and can be improved to accommodate other social economic variables in the future. Whatever the level of complexity of the social economic variable involved, the mathematical model is basically composed of two main components: the *functional* and the *stochastic models* (Mikhail, 1976). The functional model describes the deterministic properties of the physical situation, whilst the stochastic model deals with the nondeterministic parts where variables involved are having probabilistic characteristics. In practice, the stochastic model generally embraces the observable quantities, and it represents the physical model.

Regardless of influencing social, economic or even political factors that might be bias, in general, property price transaction data has stochastic characteristics in nature. In order to construct a mathematical model for the development of DPPM, a complex mathematical model with a mixture of social, economics, and geographical data is required. An advanced numerical and statistical analysis will be required for the estimation technique of a complex mathematical model. This is going to be an improvement of the existing econometric modeling available in traditional econometrics textbooks.

A conventional multiple regression model generally used in property applications can be seen in (Ramanathan, 2002) e.g. see p. 144. However, the model given does not seem to be sufficient to represent up to date factors that influence property price nowadays. A slightly updated traditional non-spatial model is given here:

$$Price_a = \beta_1 + \beta_2 LandA + \beta_3 BuildA + \beta_4 Bedr + \beta_5 Bath + \beta_6 Garage + \beta_7 BuildAge + v_a \quad (1)$$

where:

$Price_a$	Price of the property
$\beta_1, \beta_2, .. \beta_7$	Deterministic parameters of the non-spatial model (to be determined)
LandA	Area of the land plot of the property
BuildA	Area of the building in the property
Bedr	Number of bedrooms in the house
Bath	Number of bathrooms in the house
Garage	Availability of car garage or carport
BuildAge	Age of the building
v_a	Error (i.e. residual error) in the non-spatial model.

The above conventional model generally applied in econometrics is not sufficient to model the geographical location of the property. Two houses that have exactly the same physical characteristics such as the same number of bedrooms, the same number of bathrooms, the same land plot area, and so forth might cost totally different if they are located in two different geographical areas. For instance, the first one is located just a few hundred meters from a business district whilst the second one is located two miles away. The prices of such houses will be totally different. Some researchers approach the location factor by introducing appropriate weighting system by taking into account the geographical effect of neighboring properties in the area (Christopher Bitter, 2007), (Gerkman, 2010), (Malpezzi, 2002), (Suriatini Ismail, 2008), (Dubin, 1998), and (Fingleton, 2007).

Spatial modeling in this study is approached differently from the abovementioned works. The concept is adopted similar to what has been applied in digital photogrammetry by introducing the so called *additional parameter* in order to enhance the above model (1). The *additional parameters* in photogrammetry are mathematically a *high degree polynomial* as given in (2). The benefit is twofold. Firstly, it has been proven to be excellent model in *bundle adjustment system* in photogrammetry (Gruen, 1980a, and Gruen, 1980-b). Secondly, it is also excellent model in *digital terrain modeling* (Saksono, 1984), and even to model very volatile stock market data (Saksono, 2011). It is strongly believed therefore that such a model will also be sufficiently good to cater the geographical parameters of the property data. The third degree polynomial model is given:

$$Price_b = p_1 + p_2 East + p_3 Nort + p_4 East.Nort + p_5 East^2 + p_6 Nort^2 + p_7 East^2 Nort + p_{\ldots} \quad (2)$$

$p_1, p_2 \ldots p_{10}$	Deterministic parameters of the spatial model (to be determined);
East	Easting or longitude of the location of the property;
Nort	Northing or latitude of the location of the property;
v_b	Error (i.e. residual error) in the spatial model.

Therefore, DPPM will be designated to merge the conceptual model of social economic factors (non spatial) and the geographical factors (spatial), although fundamentally, the mathematical models are originally separated.

Pilot project for checking the robustness of the model

Ideally, data used to check the wellness of theoretical models in (1) and (2) should come from real property transaction data. The data might come from recorded property transaction by JPPH (Jabatan Penialian dan Pehidmatan Harta) Malaysia. On the one hand, however, upon a brief inspection to its website, JPPH data might not be sufficiently detailed for checking the theoretical models. In addition to that, we were not prepared to cope with the red tape procedure as timeline was quite restricted. Upon discussions with the senior students in the GIS course for the first-half of 2011 class, we then decided to collect the data as a class assignment. The students planned and proposed a mobility program to Kuching, Sarawak through the Department of Real Estate Management, Faculty of Technology Management and Business, University Tun Hussein Onn Malaysia (UTHM). UTHM then approved to partly finance the trip with willingness for the students to bear the gap.

Field work was planned using Google Earth navigation engine. Taman BDC in downtown Kuching was selected as a target location of the field work, and the field work was carried out on 24-26 March 2011. Twenty six students were split into three groups, each group was responsible to collect at least thirty data required to be fed to the theoretical model in (2). Students were equipped with their own smart phone completed with GPS navigation to record coordinates (easting, northing) or (latitude, longitude) which were then transformed into map projection coordinates. Collected data was immediately recorded as red pins in the Google Earth image as shown in Figure 1. Characteristics of property surveyed in Window 2 are given in Table 1.

Figure 1: Layout of surveyed data for each window

(window1: yellow border; window2: pink border; window3: blue border; window4: green border; window5: purple border)

In the mean time, data pertaining stochastic variables required by theoretical model (1) was obtained by interviewing property owners. In spite of lacking professional experience, the fourth year students in real estate management have sufficient capabilities to extract information gathered from the property owners to estimate the current market price of the property surveyed.

Price (RM 000)	Area (square feet)	# of bedrooms	# of bathrooms	Garage	Building Age (y)
160	860	3	2	1	20
260	860	5	3	2	20
298	1098	3	3	2	22
300	1,240	4	3	2	22
400	880	3	3	2	22
300	775	3	3	2	22
260	775	3	3	2	22
300	775	3	2	2	22
300	775	3	3	2	22
300	775	3	2	2	22
300	775	3	2	2	22
450	1,395	4	3	2	22
350	775	3	2	2	22
350	775	3	2	2	22
400	904	3	2	2	22
280	1,098	3	3	2	22
400	775	3	2	2	22
500	1,628	3	2	2	22
300	775	3	3	2	22
450	1,243	3	2	2	22
280	775	3	2	2	22
293	775	3	2	2	22

Table 1: Characteristics of property for Window 2

(Garage = 2 is meant for permanent garage; Garage = 1 is meant for temporary carport)

Data processing

Equation (2) contains deterministic parameters p_i. They are essentially additional parameters to embrace spatial modeling to enhance the former model in (1). Combining (1) and (2) will produce a substantially enhanced model. The combined model is called the hybrid model for the development of digital property price modeling as the objective of this research, and is given in (3).

$$Price = \alpha + [\beta_{1}2\ LandA + \beta_{1}3\ BuildA + \beta_{1}4\ Bedr + \beta_{1}5\ Bath + \beta_{1}6\ Garage + \beta_{1}7\ BuildAge \quad (3)$$

Where:

Price	$Price_a + Price_b$
α	$p_1 + \beta_1$
v	$v_a + v_b$

Through graphical inspection of Figure 1, the collected data are then grouped into five different windows. In order to minimize the bias, the geographic location of the window is selected arbitrarily and independent from certain group of students who conducted the survey. A window might have combination of several groups so long as there is sufficient number of data required to solve for deterministic parameters in (3). As the data processing is carried out using least squares technique, the number of data should be sufficiently larger than the number of the deterministic parameters β_i and p_i. Respectively, the number of data in each window is shown in Table 2. Each window is processed independently.

If the data in each window is substituted into (3), for *n* number of data in each window, we will have a series of linear[2] equations:

$$
\begin{aligned}
Price_1 &= \alpha + \beta_2 LandA_1 + \beta_3 BuildA_1 + \beta_4 Bedr_1 + \beta_5 Bath_1 \ldots \\
Price_2 &= \alpha + \beta_2 LandA_2 + \beta_3 BuildA_2 + \beta_4 Bedr_2 + \beta_5 Bath_2 \ldots \\
Price_3 &= \alpha + \beta_2 LandA_3 + \beta_3 BuildA_3 + \beta_4 Bedr_3 + \beta_5 Bath_3 \ldots
\end{aligned} \qquad (4)
$$

. . . .

$$Price_n = \alpha + \beta_2 LandA_n + \beta_3 BuildA_n + \beta_4 Bedr_n + \beta_5 Bath_n + \beta_6 Garage_n + \beta_7 BuildAge_n + \ldots$$

Rearranging and collecting them in matrix form, we will have:

$$v_{(n,1)} = A_{(n,u)} x_{(u,1)} + l_{(n,1)} \qquad (5)$$

Where:

$v_{(n,1)}$	Residual error (v) vector with n rows and *1* column
$A_{(n,u)}$	Coefficient (design) matrix with n rows and u column
$x_{(u,1)}$	Vector of parameters to be determined (β and p) with u rows and 1 column
$l_{(n,1)}$	Observation vector with n rows and *1* column

Under Null Hypothesis (H_0), the distributions of l and v are defined as:

$l \sim N(Ax, \sigma_0^2 W^{-1})$ — l is normally distributed with expected value Ax and precision $\sigma_0^2 W^{-1}$; where σ_0^2 is the a-priori variance, and for W^{-1} is an inversion of weight matrix W of the observable values (for practical reason, W is an *identity matrix*);

[2] Although the above function is non-linear in nature, however in least squares principles, it is linear with respect to the deterministic parameters β and p.

$v \sim N(0, \sigma_0^2 Q_{vv})$ v is normally distributed with expected value *0* (zero) and precision $\sigma_0^2 Q_{vv}$; where Q_{vv} the variance-covariance matrix of *v*.

Least square principles require that the sum of the squared residuals v be constrained minimum. This leads to the construction of a normal equation for the solution of the estimated parameters $\hat{x}$ and the precision estimate of the estimation model with respect to the observed data $\hat{\sigma}_0^2$. Without proof, the minimum variance unbiased estimator for *x* and σ_0^2 are given in (6) through (8).

$$\hat{x} = (A^t W A)^{-1} A^t W l \quad (6)$$
$$v = (A(A^t W A)^{-1} A^t W - I) l \quad (7)$$
$$\hat{\sigma} = \left\{ \frac{1}{r} (A\hat{x} - l)^t W (A\hat{x} - l) \right\}^{1/2} \quad (8)$$

Where:
r = *(n-u)* or redundancy (degree of freedom)
v = residual error, that is the discrepancy of each observation from the theoretical model
$\hat{\sigma}$ = standard deviation of the theoretical model of each window.

Despite low computational burden of the pilot project, as part of an assignment for final-year student's project, the data processing was conducted in Matlab high level programming language environment. As such, the student is introduced to the arts of basic computing and statistics, although it is rather to premature to expect the student to do computer coding.

Results

Collected data for each window (one of the windows is shown in Table 1) was calculated independently using the theoretical model developed in (3). The estimated price was calculated using (6) and the standard deviation was calculated using (7). The results are tabulated in Table 2.

Window	Number of points	Average Price (RM 000)	Standard Deviation (RM 000)	Relative Accuracy (% wrt. average price)
Window 1	23	199.87	22.40	11.2
Window 2	22	328.68	26.69	8.1
Window 3	24	387.00	~ 0	~ 0
Window 4	33	379.00	10.28	2.7
Window 5	33	215.21	27.58	12.8
AVERAGE PRICE:		301,952		

Table 2: Results of the estimation model of all windows

Analyzing Table 2, Window 3 actually looks peculiar owing to the fact that the standard deviation is close to zero. This indicates that in that particular window, the theoretical model almost perfectly fit the physical model. This happens quite seldom in practice of course. However, so long as it does not occur quite frequently, then, nothing is suspicious.

If we take the average standard deviation from Table 2, we will come up with $\sigma_{average}$ of RM 17,390 which is actually bit too large. This simply means that the theoretical model developed in this study deviates from the physical model (i.e. the actual property price collected in the field). In other words, there is potential overpricing or under-pricing by about RM 17,000 in the valuation at Taman BDC. Partly, this is probably caused by the fact that fourth-year students are yet to be a professional valuer. It will take quite a number of years practicing as a valuer to become a professional valuer that is for sure. However, one should also bear in mind the fact that the average price of the property in Taman BDC Kuching is also quite high, that is RM 301,952. If we take the relative accuracy which is the standardized σ with respect to the average property price within the window, we will come up with a figure of 7%. For the average property price of RM 301,952 therefore, the estimation error is only RM 21,000. This is actually encouraging. Estimation error of RM 21,000 for RM 300,000 property is a quite moderate result especially as the physical model was generated by senior students. It is of firmed belief, that better data will improve the accuracy of the estimation model. Furthermore, better socio-economic model might have to be included (e.g. the luxury and psychological factors like *feng shui* in Chinese belief) in further study, in order to further enhance the theoretical model in (1).

Conclusion

The theoretical model of this research is a robust hybrid of the conventional econometric model with the spatial based model. It is shown that the additional spatial based model has substantially enhanced the conventional econometric model. With this better model, a more reliable DPPM data can be provided. The DPPM data will be of paramount importance not only for real estate industry and its related sectors, but it will be beneficial also for government, banks, insurance industry, investors, and even individual and traditional customer.

It is believed that neither Malaysia nor any other country has already developed digital property price model (DPPM) as an important element in its national development planning. Therefore, the study is really a breakthrough to some extents. Among other significance of the research outcomes to national development of any country are:

- Government will be able to calculate the potential assets of the property sectors more accurately. In return, the government will be able to estimate the potential earnings from the related tax and other levy;
- Government will be able to accurately reformulate policies regarding projected growth in the real estate industry and its related sectors, once an accurate digital property price model (DPPM) is accurately available;

- This research will provide underlying scientific groundwork if any government wants to extend the constructed model for the whole country.
- The research outcomes are expected to lessen dispute due to substantial difference in the valuation of property assets.

Acknowledgement

The authors thank Encik Muhammad Azwan and Encik Muhammad Hafidz, as initiators and group leaders, who had prepared all administrative work for the trip. They have led the survey quite successfully, and we have altered an originally planned touristic trip into a successful scientific activity. All GIS class students of 2011 enrollment who were involved in the data collection used for this preliminary study also have invaluable credits to the success of the field trip. They made the trip so cheerful and unforgettable. All Taman BDC residents in Kuching, Sarawak who had been very helpful in providing information required for the study are also most appreciated. Some of them were even so kind to provide the survey team some refreshments and drinks. Finally, the authors are very grateful for financial support of UTHM for the trip.

Bibliography

1. Christopher Bitter, e. a. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *J Geograph Syst* , 21.
2. Dubin, R. A. (1998). Spatial Autocorrelation: A Primer. *Journal of Housing Economics* , 24.
3. Fingleton, B. (2007). A multi-equation spatial econometric model, with application to EU manufacturing productivity growth. *J Geograph Syst* , 26.
4. Gerkman, L. (2010). Topics in Econometrics with Applications to House Prices. *Publication of the Hanken School of Economics* , 126.
5. Gruen, A. (1980a). Internal Reliability Models for Aerial Bundle Systems. *XIV Congress of International Society for Photogrammetry, Commission III* (p. 11). Hamburg: International Society for Photogrammetry.
6. Gruen, A. (1980b). Precision and Reliability Aspects in Close Range Phtogrammetry. *XIVth Congress for Photogrammetry, Commission V* (pp. 117-132). Hamburg: International Society for Photogrammetry.
7. Malpezzi, S. (2002, April 10). Hedonic Pricing Models: A Selective and Applied Review. (K. G. O'Sullivan, Ed.) *Housing Economics: Essays in Honor of Duncan Maclennan* , p. 45.
8. Mikhail, E. M. (1976). *Observations and Least Squares.* Washington D. C.: University Press of America.
9. Ramanathan, R. (2002). *Introductory Econometrics with Applications* (5 ed.). (T. Gay, Ed.) Mason, Ohio, USA: Michael P. Roche.

10. Saksono, T. (1984). *Combination of Single Photograph with Digital Eelevation Model for Planimetry dan Height Data Extraction.* Columbus: Ohio State University, M.Sc Thesis.
11. Saksono, T., e. a. (2011). Undisclosed Potential of Muslims' Contribution to Global Economy Development. *International Journal of Trade, Economics and Finance , 2* (6), 8.
12. Suriatini Ismail, e. a. (2008). Spatial Autocorrelation in Hedonic Model: Empirical Evidence from Malaysia. *International Real Estate Research Symposium (IRERS) 2008* (p. 15). Kuala Lumpur: International Real Estate Research Symposium (IRERS) 2008.