Property price separation between land and building components

by

Ünsal Özdilek

Paper submitted for publication in the *Journal of Real Estate Research*

**Contact address:**

Ünsal Özdilek, Business School, Department of Strategy, social and environmental responsibilities, University of Quebec, Montreal City, Canada, H3C 3P8
Phone: (519) 987-3000, extension: 4436
Fax : (514) 987-0422
Email : ozdilek.unsal@uqam.ca
Property price separation between land and building components

Abstract

Observed sales prices are direct references for the market value of properties, but they do not provide information about the separate values of land and building. There are different theories and methods, each one being limited in practice. This paper presents the troublesome issue of price separation and proposes a practical alternative, using detailed data from Montreal (Canada). The empirical results support the separability thesis in practice for the cases of residential properties.

Keywords: property price; land; building; improvements; value apportionment; separate evaluation
1. Introduction

Land valuation is quite easy when there are sufficient comparable vacant lands being sold on the market. This is hardly the situation nowadays in most built-up cities where more than ever before the land market is progressively disappearing, and land value is becoming an ever more elusive concept. What is negotiated and appears commonly in the deed of sales is a total price paid for the entire property. There is no direct indication of the separate components of prices for the land and the improvements.

Land value has traditionally been explained in terms of classical production rent, neoclassical utility, location rent, social dynamics and political decisions. Each of these concepts contributes of course to the explanations, but they are in general theoretical and restricted with various hypotheses. Even in the recent literature, based on the empirical modeling culture and data, the price apportionment between land and building components has always had a low priority. The analysis of the dialectic between the two components is missing; they are analysed in isolation or mixed. However, there are some confusing echoes from classical fields of revenue imputation and property tax impacts.

Arguments on the inseparability thesis have been presented by Ely (1922), Russell (1945), Ratcliff (1950), Fisher (1958) and Dorau and Hinman (1969), and are restated in some contemporary views (Mills, 1998; Robinson, 1999; Fischel, 2000; Kitchen, 2003; Hendricks, 2005). The consensus is that apportionment is not practicable (and is useless) because land and improvements are merged together like an “omelette” to form a new joint product. In contrast, several early authors proposed or defended separation (Smith, 1886; George, 1898; Marshall, 1922; Brueckner, 1986; Oates and Schwab, 1997). While explanations from both sides are essentially theoretical, some responses might come from the practice of appraisal using the three classical methods of price, cost and income. The similar theoretical debates can be confusing, but in practice what matters are total value estimations, notably for taxation purposes (Rice, 1997; Andelson, 2000).

The issue of apportionment thus remains troublesome. Despite the “silence” surrounding improvements and the general omission in practice, land value continues to exist and shape the dynamics of real estate markets. It even progressively catches up with the value of the improvements, sometimes going much further. While there is no empirical evidence of this, many practical situations require separate estimations, for example city and capital gains taxation, tax incidences (Gloudemans,
2001), challenges by contributors in courts (Keligan, 1994), applicability of the cost method for improvements only (Appraisal Institute, 2001), decisions about depreciation, amortization, renovations, and demolitions (Arnott and Petrova, 2002), management of insurances and mortgages (Guofang et al., 2003), to measure the speculative effects on land (agricultural) and its products (Stokes and Cox, 2014), sound land use policies and practices (Dancaescu, 2000) or addressing land value maps (Ohno, 1985).

The issue of apportionment is relevant everywhere and concerns all types of properties, especially housing in urbanized areas for which land market is scarce and classic methods are ineffective. This paper focuses on the issue of apportionment of price between the land and the improvements for residential properties in the context of mass valuations. It develops an original alternative of price separation using the framework of measure theory and a large database from the city of Montreal (Canada).

2. Price apportionment methods

There are three possible price apportionment theories supporting different techniques and methods in practice, listed in Table 1 and discussed in Özdilek (2012), in which we add the last alternative of geometric modeling elaborated in this paper. In the list, the approach that is closest to our proposed method is the OLS model which allows one to decompose the total price of properties between their multiple utility generating parameters (Lancaster, 1966; Rosen, 1974; Sirmans et al., 2005).

There have been a few attempts to estimate land value parameters, by considering data only on vacant land sales (Clapp, 1990; Isakson, 1997; Colwell, 1998). Shi (2014) more recently elaborated an approach to improve the “net rate” factor, classically used in the conceptual framework of separation by the extraction technique. Other studies used total prices, separated between land and improvements (Guerin, 2000; Gloudemans, 2002; and Plassmann and Tideman, 2003). These studies relax key contingencies of the separation issue in the hedonic framework by having recourse to both improved and unimproved land sales in the same model, meaning that separation is not possible otherwise. Özdilek (2011) presents a separation technique from a hedonic approach, providing conceptual as well as empirical details of price separation.

<Insert Table 1>
Classical methods in both FAT and RAT theories are approximate and do not provide reliable results. They are essentially subjective as they precede ad hoc almost by a rule of thumb and some simple assumptions (Oates and Schwab, 1997; Mills, 1998; Anas, 2003). The parametric approach, which essentially refers to the hedonic modeling tradition through multiple linear or nonlinear regression analysis (Breiman, 2001), proposes an interesting solution. However, it is limited to the independence hypothesis between the parameters and the accuracy of their isolated contributions. If there are errors of estimation per utility attribute of the properties, the sum of their individual errors might be significant in separate estimations. There is also the difficulty of separating the constant and the error terms of the models which increase, especially when the price determinant parameters are missing.

In the geometric approach presented in this paper, we assume that when lands are identical or very similar (given their proper multiple attributes), if there is a variation in total prices, it comes mainly from the building component. Inversely, the same hypothesis applies when the building components are identical. Based on this fundamental proposition, we maintain the simplicity and the objectivity of the sales comparison method in a straightforward manner, using a large set of observations. Our nonparametric approach achieves separation by starting from the attributes of the properties as given and then observing price differentials, whereas in the OLS technique, the process begins by breaking down the total prices between the parameters and then assembling their “estimated” utility contributions in the land and building values.

Both approaches are exposed, though in different ways, to the problem of interactions (for parameter variations in space, see Pavlov, 2000), because we cannot presume they will be entirely eliminated for the simple reason that lands and buildings go together. But, if we create similar conditions starting from the separate parameters and then use the price differentials, as in our approach, we can at least hope to keep any possible interactions constant. For example, if the lands are identical or very similar (same usage, sector, size, physical form, proximity to various location points such as a park or a school), then the interactions with the building components will be the same (though not eliminated). In those circumstances, the differences in price originate from the building. The same reasoning applies when the building parameters are constant (e.g., living area, number of stories, external recovery or types of floors) and the lands are different. Even the competition in the market encourages an optimal level of combinations; there are however some cases of over and under improvements, against the principle of the *Highest and Best Use* (HBU).
Another concern is that the sum of the values of the land and the building might be less than the total value of the property. From the modern view of finance, total value might include an Option value for unimproved lands (Quigg, 1993) as well as an option to redevelop improved lands (Williams, 1991; Clapp et al., 2012) or to renovate (Plaut and Plaut, 2010). This would be the case with another intricate agent known as Business value which originates from entrepreneurial efforts and managerial abilities (Martin and Nafe, 1996). There is a debate on that issue, but Miller et al. (1995) and Rabiansky (1996) consider them as being parts of the land and the building. In our approach, the values of these types of agents are held constant in time and space by creating similar conditions for the utility expressions of the improvements. Also, the redevelopment or the renovation options are explainable in the model by the inclusion of the various parameters of age, condition, sector, and location quality. The particular and inverse relation progressing in time between the land and building components deserves more attention in a context of uncertainty, using the option value framework. In fact, as land usually gains value over time while buildings always depreciate, it would be interesting to know when to consider redevelopment, for example by taking into account the imbalances in their ratios and the evolution of the construction costs.

On a technical note, there is an unexplained residual of value which is usually specified under the form of an error term in the equations of statistical approaches. Our geometrical approach, which is different in nature, does not use such a statistical measure, but it indirectly has the conditions of the error “sharing” following the parameters initially gathered in one or the other component. Using a high number of comparables with the appropriate parameters which should help to reduce the significance of that term (which also has treatments in the field of spatial autocorrelation – Anselin, 1995).

3. Main concepts of the model

The approach we have developed for separate price estimation uses the properties of measure theory in a geometric space representation (Halmos, 1974; Royden, 1988). Although the comparison between parametric OLS models and the nonparametric separation method proposed here may seem irrelevant because the underlying theory is different, we can argue that our model would perform well since the final price is estimated from observations that are very similar to the subject’s land and building components.
The main concepts and the mathematical expressions of that model are presented in the following sections in relation to total and separate price predictions. Some details of the equations and the practical algorithms are reported in the Appendix.

### 3.1 Prediction of the total price

We define a utility space $\mathbf{U}$ as a (locally Euclidean) subspace of the standard Euclidean space $\mathbb{R}^N$ for a finite or infinite $N > 0$. As a subspace, $\mathbf{U}$ has a number of observable coordinate functions $u_i$. We suppose that $i$ belongs to a finite set $\{1, \ldots, M\}$, and that $u_i$’s are global coordinates and express characteristics of properties (i.e., size, number of rooms, presence of a swimming pool, proximity to urban parks, etc.). The utility space $\mathbf{U}$ has to be appropriate with a sufficient number $M$ of observable functions $u_i$ allowing one to distinguish prices $P_i, P_j$ for different properties $i, j$.

Property price $P$ is a nonnegative function $P: \mathbf{U} \to \mathbb{R}^+$ and it is a **measure** on utility space $\mathbf{U}$, so that $(\mathbf{U}, P)$ is a measure of space in a proper sense. $P(A)$ expresses the total price of properties, identified with all points of the set $A \subset \mathbf{U}$.

The density (Radon-Nikodym derivative) $\frac{dP}{d\mu_L}$, where $d\mu_L := du_1 \wedge \cdots \wedge du_M$ is the Lebesgue measure on $\mathbf{U}$, exists and is the price $P(u)$ of property $u \in \mathbf{U}$. Measure $\mu_L$ is sigma-finite as the measure on a subspace of Euclidean space $\mathbb{R}^N$. $P << \mu_L$ is absolutely continuous since $\mu_L(A) = 0 \Rightarrow P(A) = 0$ and the derivative exists. The meaning of $P$ here is the point price $P(u), u \in \mathbf{U}$.

All events (observable cases) happen in the measure space $(\mathbf{U}, P)$, i.e., the class of observable properties $K = K_{\text{sold}} \cup K_{\text{unsold}}$ belong to the space $(\mathbf{U}, P)$ (where the set $K_{\text{sold}}$ denotes given statistical data). The set $K_{\text{sold}}$, similarly to the space $(\mathbf{U}, P)$, carries (at least) two different measures: $P_{\text{sold}}$, which assigns the sum of prices of all properties located in a subset of $K_{\text{sold}}$, and $\mu_{\text{count}}$, which counts points of a subset of $K_{\text{sold}}$. These measures are in a very similar relation to the measures $P$ and $\mu_L$ on $\mathbf{U}$.

The density (Radon-Nikodym derivative) $\frac{dP_{\text{sold}}}{d\mu_{\text{sold}}}$ exists and is the price $P(u)$ of property $u \in K_{\text{sold}}$. At this point we have two models: **continuous** $(\mathbf{U}, P, \mu_L)$ and **discrete** $(K_{\text{sold}}, P_{\text{sold}}, \mu_{\text{count}})$. Radon-Nikodym derivative $\frac{dP_{\text{sold}}}{d\mu_{\text{sold}}}$ of the discrete space $(K_{\text{sold}}, P_{\text{sold}}, \mu_{\text{count}})$ “converges” to Radon-Nikodym
derivative $\frac{dP}{d\mu_L}$ of the continuous space $(U, P, \mu_L)$ when cardinality (or the data size) of the set $K_{sold}$ is sufficiently high. To make this important observation clear the idea of convergence needs to be explained. It can be understood as just an enhancement of the continuous model with new points or in a proper way as a pointwise convergence in the space of Lebesgue measurable functions on $U$ when the statistical data are linearly interpolated on small pieces over all $U$. Let $\mathcal{F}$ be any filter of Lebesgue measurable sets on $U$ converging to a point $u \in U$. Then

$$P(u) = \lim_{\mathcal{F} \to u} \frac{P(O)}{\mu_L(O)} \approx \lim_{\mathcal{F}_{sold} \to u} \frac{P_{sold}(O_{sold})}{\mu_{count}(O_{sold})},$$

(1)

$O \in \mathcal{F}$, $\mathcal{F}_{sold} = \mathcal{F} \cap K_{sold}$, $O_{sold} = O \cap K_{sold} \in \mathcal{F}_{sold}$.

The expression in (1) shows that when the number of observed cases $O_{sold}$ (or the references in the discrete data) increases, this better reflects the behavior of the market. This idea of convergence in a continuous case supports our main proposition and method of separation with a discrete case. It holds a simple, but subtle point exemplified in Table 3. Using average point prices for the land and the building requires reliable “average total prices” from the filtered similar properties to the subject.

For example, when land characteristics are kept similar (or constant) in the data, the number of references should increase sufficiently to provide a reliable average total price. This total price approaches the real total price in the market by increasing the number of comparables, thereby justifying the use of the “average price of the buildings” (or that of the lands). Following that objective, we cannot be very rigid in creating high similarity which will end up with an insufficient number of comparables (and consequently move away from the real average prices). So, the precision depends mainly on the size of the data: the more it increases the more reliable are the estimates, as might be the case in the context of mass appraisals.

A practical realization of the algorithm for estimating the total price of a property is presented in the Appendix. This algorithm assigns to an unknown point the average value of the price of the known ones most similar to it. In this small neighbourhood the algorithm acts linearly (assigning the barycentre of the points around). So, the approximation is piecewise linear (the actual price function looks as if it is covered by small scales on the small pieces where the data $K_{sold}$ are missed). As the number of the data increases these scales become smaller and smaller, coinciding with pieces of tangent hyperplanes.
3.2 Prediction of separate prices

Let coordinates $u_i$ of the utility space $U$ be split in two groups: parameters related to the building (we will call them $x$’s) and parameters related to the land (respectively $y$’s) so that $u = (x, y) \in U$. When coordinates $u_i$ are global (which we assume here) it means that $U = X \times Y$. The price function $P(x, y)$ is called (globally) separable if:

$$P(x, y) = B(x) + L(y) \quad (2)$$

where $B(x)$, $L(y)$ are respectively prices of the building and the land. Note there is a local version of separability of (the price) function. It becomes important when $U$ cannot be represented globally as the product $U \neq X \times Y$. Since the price function $U \supset K_{\text{sold}}^p \to R^+$ is given on a discrete set, it is not obvious if $P$ is separable or not. The price function $P: U \to R^+$ is separable iff all the second derivatives (one with respect to a parameter of the building and the other with respect to a parameter of the land) of the linear distribution $P$ are equal to zero, i.e., $P_{x_iy_j} = 0$, where $x_i$, $y_j$ are parameters of the building and the land respectively. The proof is a direct generalization of the criterion “a smooth $P$ is separable iff $P_{x_iy_j} = 0$”. $P$ is separable iff

$$\int_U P(x, y) f_{x_iy_j} d\mu_L = 0, \quad (3)$$

for any smooth function $f(x, y)$ vanishing on the boundary of $U$ (the required integral is, by definition, the second derivative of $P$ determined on smooth sample functions). A practical criterion for separability of price function $P$ can be:

$$\int_U P(x, y) \frac{\partial^2}{\partial x_i \partial y_j} e^{-[k_1(x-x_0)^2+k_2(y-y_0)^2]} d\mu_L \approx 0 \quad (4)$$

where $k > 0$ is a sufficiently big number, and $x_0$, $y_0$ are arbitrary points of $U$. In the case of the discrete function $P: K_{\text{sold}} \to R^+$ is separable iff

$$\text{Average}(\sum_{(x,y) \in K_{\text{sold}}} P(x, y) \frac{\partial^2}{\partial x_i \partial y_j} e^{-[k_1(x-x_0)^2+k_2(y-y_0)^2]}) \approx 0. \quad (5)$$
Equations (2) to (6) provide a proof about the price separability with the interaction effects related to both components becoming almost nil in the empirical example under consideration. Note the average in the discrete function in (5) is taken in order to more easily decide if the required sum can be regarded as zero for one isolated property (indeed, compared to an average price of $145,881, we checked the data by discrete function and got small values between $30 and $40 for different exponential functions under the sum). If the size of the data becomes greater than that which was used here, these values will converge to zero.

According to (5), if \( P \) is separable then

\[
P = B \cdot \mu_{L_1} + L \cdot \mu_{L_2}
\]

\[
P(\mathcal{A}) = \int_{\mathcal{A}} B(x) d\mu_{L_2}
\]

\[
P(\mathcal{A}') = \int_{\mathcal{A}'} L(y) d\mu_{L_1}
\]

\( A \subset X, d\mu_{L_2} = dx_1 \wedge \ldots \wedge dx_n \) is an (infinitesimal) Lebesgue measure on the utility space \( X \) corresponding to the building and \( A' \subset Y, d\mu_{L_2} = dy_1 \wedge \ldots \wedge dy_m \) is an (infinitesimal) Lebesgue measure on the utility space \( Y \) corresponding to the land. The fractions of the following measures exist (and are unique) and present measures on the slices

\[
U_x := \{(x, y) \in U | x \text{ is fixed, } y \text{ varies}\} \quad \text{and} \quad U_y := \{(x, y) \in U | y \text{ is fixed, } x \text{ varies}\}
\]

\[
\begin{align*}
\frac{dP}{d\mu_{L_2}}(x) &= B(x) \cdot \mu_{L_1} + L \quad \text{(a measure on } U_x) \\
\frac{dP}{d\mu_{L_1}}(y) &= L(y) \cdot \mu_{L_1} + B \quad \text{(a measure on } U_y)
\end{align*}
\]

The first point is immediate from the formula:

\[
\int_{\mathcal{U}} P(x, y) dx_1 \wedge \ldots \wedge dx_n \wedge dy_1 \wedge \ldots \wedge dy_m = \int_{\mathcal{U}} (B(x) + L(y))dx_1 \wedge \ldots \wedge dx_n \wedge dy_1 \wedge \ldots \wedge dy_m.
\]

The result of the division of two differential forms is not unique, but by fixing the base differentials \( dy_1, \ldots, y_m \) (or \( dx_1, \ldots, dx_n \)) in the fraction it becomes unique by the rules of calculation in the exterior differential algebra \( \Lambda(x, y, dx, dy) \). Further, we suppose that Lebesgue measure \( \mu_L(U) \) is finite (e.g., when \( U \) is compact, which is a very practical assumption), and so are the Lebesgue measures \( \mu_{L_1}(Y) \) and \( \mu_{L_2}(X) \).
The separation formulas for given points \( x \in X \) and \( y \in Y \) are (the first one is obtained by integrating over \( Y \), and the second one over \( X \)):

\[
\begin{align*}
B(x) &= \frac{1}{\mu_L(Y)} \cdot \int_Y \frac{dP}{d\mu_B}(x) - \frac{1}{\mu_L(Y)} \cdot L(Y) \quad \text{(the building price)} \\
L(y) &= \frac{1}{\mu_L(B)} \cdot \int_X \frac{dP}{d\mu_L}(y) - \frac{1}{\mu_L(B)} \cdot B(X) \quad \text{(the land price)}
\end{align*}
\]  

(9)

For the discrete data, the separation formulas are:

\[
\begin{align*}
B(x) &\approx \frac{1}{|U_x \cap K_{sold}|} \cdot \sum_{u \in U_x \cap K_{sold}} P(u) - \bar{L} \quad \text{(the building price)} \\
L(y) &\approx \frac{1}{|U_y \cap K_{sold}|} \cdot \sum_{u \in U_y \cap K_{sold}} P(u) - \bar{B} \quad \text{(the land price)}
\end{align*}
\]  

(10)

where \( U_x, U_y \) are slices of comparable cases over points \( x \) and \( y \), \( |U_x \cap K_{sold}| \) denotes cardinality of the set \( U_x \cap K_{sold} \), and \( \bar{L}, \bar{B} \) are average prices provided from the market (and compared to that of the city appraisals). We can even start from arbitrary values and get real average prices for the land and the building by an improvement algorithm. Formulas (9) and (10) are the result of translation of the continuous model to the discrete one (we also take into account that if \( \mu_L = \mu_L \cdot \mu_B \) in the continuous case, then somewhat to the contrary \( \mu_{count, U} = \mu_{count, X} = \mu_{count, Y} \) in the discrete case). Note that even if we have a large amount of data we cannot practically expect that there will be enough properties in the slices \( U_x, U_y \) (or that they will not be empty at all). So, the slices should actually be presented by

\[
U_x := \{(\tilde{x}, y) \in U | \| \tilde{x} - x \| \leq \varepsilon, y \in Y\},
\]

(11)

where \( \varepsilon > 0 \) is a small number, and similarly for \( U_y \). In the filtering process of the similar comparables, by keeping constant the land or the building parameters, the algorithm brings values of \( \varepsilon \) within the same radius around zero. Instead of that critical value, which might retain different number of comparables, we can also require the same number of comparables. However, this request has the risk of enlarging \( \varepsilon \) slices of the similarities to a point where the results become less reliable.

The role of \( \bar{L} \) and \( \bar{B} \) as average prices needs more explanation. Let us consider here a fictional example of two very different sectors in Montreal, Westmount and Verdun. We will assume that the subject property for which separate estimations are needed is situated in Westmount and the question is
whether the averages should be taken from Westmount or the whole city (here as being formed by these two sectors in the following table).

When comparing various characteristics of the subject’s land in Westmount and those of all the lands in the same sector, by letting building characteristics free, four very similar cases are found in the data. Their average total price (e.g., $1.645 M) minus the average price of the buildings in the sector (e.g., $1 M) provides an estimate of $645 000. The estimation converges to the same value when the averages of the total prices and the buildings in the whole city are considered (8 cases in the example). The role of the averages here is just a fixed or “stable” point in the system. It allows separate value estimations, expressing a better reading of the particular relation between the two components with an increased number of comparable cases. Being too rigid with the averages in each sector (and even sub-sectors) risks there having an insufficient density of comparables around them, which is required in the method. The problem of interactions perhaps in favor of averages per sector also diminishes in the method by successively keeping their appropriate parameters constant. A mathematical exposition on that issue of a fixed point (and its improvement) is reported in the Appendix.

< Insert Table 2 >

In order not to repeat the prediction algorithm for separate values of the land and the building, we just indicate the differences here. Let us assume we want to predict the building price \( B(x) \), corresponding to the building utility parameters \( x \in X \). We proceed in exactly the same way in the prediction of the total price as in the algorithm in Appendix, with the only exception being for the metric in the utility space \( U \).

\[
\rho^2(x, y; \bar{x}, \bar{y}) = \sum_{i=1}^{n} k_i \cdot (x_i - \bar{x}_i)^2 + \sum_{j=1}^{m} \hat{k}_j \cdot (y_j - \bar{y}_j)^2
\]

(12)

Recall that coefficients \( k_i \) and \( \hat{k}_j \) are important for filtering the corresponding utility parameters (the bigger the coefficient the more important for this parameter). Here in the formula for \( \rho^2 \) all coefficients \( \hat{k}_j \) can be made equal to zero. It will give filtering neighbourhoods in the form of thick slices, converging to a practically small thick slice \( U_x \). As before, we take the average of the total prices of all properties occurring in \( U_x \) and subtract the average value of the land over all the data. This will
give the required predicted building price \( B(x) \). For the prediction of the price of the land \( L(y) \), \( y \in Y \) the algorithm is similar (for example, all coefficients \( k_i \) are made equal to zero).

For separate value estimations of each property, the filter uses the equation twice (12). In finding the most similar lands to the subject, the computation considers all the properties and the parameters of the land in the data, by letting go of the building attributes. The “geometrical similarity distance” building process each time compares all the properties in the whole data set and their characteristics to that of the subject. It runs the algorithm of computation 46 000 times, operating a formula of 21 parameters of separate estimations for only one subject property; this is repeated for additional cases (1025 subjects in our case).

In the similarity building process, we have the choice of whether or not to use the \( k \) coefficient. If a pure geometrical approach is required, the weights of \( k \) would be equal. In that intermediary process, the similarity distances become precise when the weights of the attributes are properly specified. Evaluations of the appraisers in the field are a possibility, but instead we used the relative importance of the parameter’s “\( t \) statistics” in an OLS model (last column of the Table 2). A non-linear optimization model, forcing the parameters between 0 and 1 is also another alternative. Such models do not guarantee that the results of the similarity building will always perform as most of the parameters precociously “evaporate” with nil weights. We compared the results anyway with equal \( k \) weights and observed that the filter is not very sensitive, usually tending to keep the same comparable properties.

The above deterministic procedure might be rendered stochastic by fitting a probability distribution function \( F \) to the distance random variable, choosing a significance error (a critical value for \( \varepsilon \)) and selecting all properties in the sample for which the Euclidian distance from the subject is smaller than the critical value. In this stochastic context, it is to be noted that the Euclidian distance is appropriate only if it is reasonable to assume that the attributes (as random variables) are independent; otherwise it would be worth considering, for example, the Mahalanobis distance, taking into account the correlations between attributes. It is represented by the formula:

\[
D_{ij}^2 = (x_i - x_j)' \Sigma^{-1}(x_i - x_j)
\]  

(13)
where \( x_i, x_j \) are the vectors of standardized parameters for properties \( i \) and \( j \) and \( \Sigma \) is the variance-covariance matrix of parameters for all properties. Vandell (1991) proposed an improved method, which has been revised with more insights about this interesting subject of comparable property selection (Gau et al., 1992, 1994; Lai et al., 2008).

4. A Practical example

Following the description of the model, we present in this section a practical example with 23 000 arm’s-length sales of single-family properties on the Island of Montreal during the year 2002. In this market, the total prices range between $41 800 and $610 000, with an average of $145 881. This data is considered for its convenience of being already operational, because many technical details of gathering, codification and computations have already been realized. More recent data would not change the practical issue of the model as the time effect is set as invariable.

As seen in Table 2, the full database contains detailed information about 21 attributes, distinguished between land and building components. There are three types and sources of information gathered on these properties: structural and financial data supplied by the city of Montreal, socio-economic data found from Statistics Canada and spatial data generated from the use of Geographical Information Systems (GISs). The application of GISs, based on a combination of various thematic maps (limits of boroughs, positions of important geographical objects like schools, rivers and commercial centers) made precise computations possible. For example, using X-Y coordinates, the position and proximity of the 23 000 comparable properties to the interesting points have been estimated.

Separate estimations using our approach are compared with those of the OLS model and the city of Montreal for the same 1 025 properties considered as a sample. We reconsidered the coefficients of the same OLS model, used in computing \( k \)-values. Based on the marginal contributions of the parameters, we were able to estimate separate values of the land and the building. In that model, the analysis of correlations and the test of the Variation Inflation Factor (VIF) did not suggest any significant interactions.
The second comparable base was the separate estimates produced by the City for the same properties in the sample. Professionals use essentially the cost and the sales method to derive these estimates. It is common for them to use a typical ratio in order to break down the observed total prices of properties between the two components. This ratio is based on the study of the historical and recently observed land prices, expressed per square foot for different areas throughout the city.

In the following practical example, we demonstrate how our method operates by considering only one case from the sample data. In the global data of 23,000 properties containing the 21 parameters of the reference properties, we used two distance metrics allowing us to “filter” by step the very similar lands and buildings in comparison to those of our subject (for which the observed price is $420,000).

< Insert Table 4 >

We first filtered the same lands (or very similar ones) according to their 10 defining parameters in the global data in comparison to those of the subject and identified 47 properties. For these properties on which only the building parameters change, we obtain an average total price of $186,484. Considering $95,000 as the average point price of buildings in Montreal during 2002 (based on the application of the cost method), the estimated separate price for the subject land is $91,484. Using the same global data and keeping the same radius of similarity in the filter, this time, we found 16 cases of comparable properties with identical buildings, but different lands. As the average price of lands in Montreal was $48,328 for the same year, the estimated value of the subject building was $332,414. This process of estimation in the first algorithm is repeated for 1,024 other cases in the sample. In the second algorithm of the fixed point improvement, we used the averages of separate estimations from the sample data. Table 4 presents the final results of the model, compared with those of the OLS and the city’s estimations. Considering the Root mean square error (RMSE) measure, our approach performs better than the OLS model and behaves very similarly to those of the city. Considering that the city integrates directly observed prices in the estimations and applies ratios uniformly for all the properties located in the same neighbourhoods, the separate estimates provided from our approach are very consistent.

< Insert Table 5 >
Even if the averages and the ratios of estimates from different models and sources are close for the land and the building, the individual estimates are significantly different. Figure 1 shows the results of our approach for all the estimates in the sample data. In the first step of the model, using averages from the market as fixed points, estimated prices are far from the observed prices. But, in the next step, using the improved averages from the results of the first algorithm, the estimations of the total prices are shifted and better fit observed prices. It is noticeable also in both steps that the errors in the estimates follow a small interval of errors, which slightly expand on the right side of the graphic (for expensive properties).

< Insert Figure 1 and Figure 2 >

In the case of the city’s estimates, we observe a similar pattern while distinguishing the distributions of the errors in a constant smaller interval everywhere. The OLS model performs less well for the same cases in the sample where the fluctuations of errors are mostly out of the tolerable range, for both total and separate estimations.

The results of the sample interpolated in the following map show interesting patterns. For example, as expected, the building values are higher around the well-known markets encircling Mount Royal, while also being high far away from downtown areas, especially in the west part of the city. Land values are visibly higher in the center of the city, decreasing progressively towards the peripheral regions. There are also higher land values in some sections along the border of the St. Lawrence River. The remarkable information from the results and the map is the negative correlation between the two components: in general, lands of higher value accommodate less expensive buildings, and vice versa.

< Insert Figure 3 >

These configurations are interesting for analyzing the specific dialectic between the land and the building, in space but also in time. Are higher quality buildings placed on expensive or less expensive lands? Is land value affecting the building or the inverse? What are the effects of different taxation systems on the total or separate values? How do the values of the two components evolve in space and time? When should buildings be demolished or improved? Accurate responses to such questions can be provided only with separate estimations in practice.
5. Conclusion

Separate estimations for the land and the building are difficult to achieve in practice using approximate methods. This situation raises more questions in the context of almost entirely built-up cities where land transactions are scarce compared with low density residential properties. Even when there are land transactions they are usually not reliable references for other usages with different characteristics.

In addition to practical difficulties, there is confusion from the theory and the academic fields where opinions diverge. For authors defending the inseparability thesis, land and building components are considered to be merged like an “omelette,” forming a new entity. For them, it is almost impossible and even useless in practice to get separate values. From the opposite side, authors defend theoretically the separability thesis and its practical necessity for different purposes without explaining how it can be done.

In this paper, we have presented the bases for the separability thesis and provided an empirical solution which can be used in practice, especially in the context of mass appraisals. To verify the applicability and accuracy of our approach, we used detailed data for the cases of residential houses, and compared them with the estimates of the city and the OLS model. From the strength of the results, it is encouraging to observe that our approach provides a reliable alternative. In our practical example, it is applied to the cases of single-family houses for which the issue is the most critical, but it can also be used for other types of properties if sufficient data exists.

Obviously, no modeling effort can pretend to trace a clear-cut line of separation between the land and the building values in practice. Debates on conceptual bases, interaction effects between the parameters, allocation of a part of value to other types of agents like the Option and Business values or the unexplainable part of the value in the model are some of the challenging aspects before the important issue of price separation.

Existing approaches have different solutions and all of them face, as we do, these same challenging questions. In comparison to the classical and OLS based methods, the use of the measure space theory as presented here, combined with the remarkable progress in the quality of data and the capacity of modern tools, increases expectations for accurate estimations of the land and the building values. It is clear that the method presented can also integrate some intermediary steps, for example in comparable
property selection and combining the advantages of the hedonic regression model in estimating the total or average values.

We approach the issue of separation in an objective way by proposing an original method that rests on the use of measure theory in a geometric space representation. Its conceptual bases, mathematical proofs and practical algorithms are presented by focusing on the price separation in the most direct way. It is hoped that it will bring light to that issue, by “relaxing” more the strong resistance of the inseparability thesis. The separate price estimation in practice will be of more concern in the future when lands become rare (and more expensive), in a perspective of increasing consciousness of social and environmental responsibilities.
Algorithm for estimating the total price:

- Step 1. Normalize the given data $K_{\text{sold}}$ in each variable $u_i$, $i = 1, ..., n$ (e.g., divide $u_i$ by the biggest occurring value of this coordinate). It will shrink or stretch each coordinate to the interval $[0,1]$ or $[-1,1]$ (if negative values are allowed) to make them “equal” in the sense of their importance.

- Step 2. Introduce an order of property parameters $u_1, ..., u_n$ with respect to their importance and assign to the parameters of higher importance the value $k_1 = 1$, to those of less importance a lower value, e.g. $k_2 = 0.9$, etc., up to the last parameters of importance with value, e.g. $k_m = 0.1$. All assigned parameters should be positive less than or equal to 1.

- Step 3. Make a weighted Euclidean metric on $U$, e.g., $\rho^2 = k_1 \cdot (u_1)^2 + k_2 \cdot (u_2)^2 + k_3 \cdot (u_3)^2 + \cdots + k_m \cdot (u_n)^2$. The coefficients will reflect the form of the elliptical neighbourhoods of a point $u$ in $U$ (radiuses to the direction of less important parameters will be big and to more important parameters they will be small). If $\rho$ converges to 0, then these neighbourhoods will filter properties from $K_{\text{sold}}$ to save the most similar to $u \in U$ with respect to the order of their importance.

- Step 4. For the point of interest, the subject property, $u \in U$ take its neighbourhood $O_r := \{w \in U \cap K_{\text{sold}} | \rho(u,w) \leq r\}$ (e.g., $r \in [0,1]$ and the first choice is $r = 1$), and calculate $P_{\text{sold}}(O_r)$ if $\mu_{\text{count}}(O_r) \neq 0$, i.e. if $O_r$ is nonempty.

- Step 5. If $O_r$ in the previous step is nonempty for that $r$, take a smaller $r$ and return to the previous step. If $O_r$ is empty in the previous step take a bigger $r$ and return to the previous step. In this way a “smallest” nonempty $O_r$ or $O_{\text{min}}$ can be found. Then take $P(u) = \frac{P_{\text{sold}}(O_{\text{min}})}{\mu_{\text{count}}(O_{\text{min}})}$ to find the required predicted value.

Fixed point improvement in separation:

Direct analysis of separation formulas gives rise to the fact that we have, rather, an iterated procedure

\[
\begin{align*}
\vec{B}' &= f(\vec{L}) \\
\vec{L}' &= g(\vec{B})
\end{align*}
\]

acting on the space $\{\vec{B}, \vec{L}\}$, where $\vec{B}, \vec{L}$ are casewise coordinates of buildings and lands.
over all given data. The prices being predicted (denote them \((\vec{B}_0, \vec{L}_0)\)) are a fixed point on this map. This is the unique fixed point, since the expressions for \(f\) and \(g\) are linear. The question is whether the point \((\vec{B}_0, \vec{L}_0)\) is stable (to make it possible to approximate it by the iterations starting from an arbitrary point). Regardless of the fact of linearity of \(f\) and \(g\), the huge size and uncertain structure of the corresponding matrix make it impossible to estimate its eigenvalues to solve the problem. However, we will generally see a negative answer.

Since the divergence of the “differential” \(\begin{pmatrix} f(\vec{L}) - \vec{B} \\ g(\vec{B}) - \vec{L} \end{pmatrix}\) is obviously negative (and consequently, Lebesgue measure on \({\vec{B}, \vec{L}}\) is decreasing under the map), the fixed point \((\vec{B}_0, \vec{L}_0)\) can only be hyperbolic with a nonempty stable manifold through it. This is all that we can say about \((\vec{B}_0, \vec{L}_0)\). The fortunate occasion, proven by a numerical experiment, is that initial points \((\vec{B}, \vec{L})\), such that \(\text{Average}(U_x \cap \tilde{K}_\text{sold}) = \vec{L}, \text{Average}(U_y \cap \tilde{K}_\text{tsold}) = \vec{B}\) (averages over all data), belong to the stable manifold. As seen in the practical example, even the second iteration with these initial data gives a significant improvement of the predictions. A practical implementation of the fixed point algorithm is presented as follows.

- Step 1. Take initial values \((\vec{B}, \vec{L})\), which are in the space of attraction of the fixed point \((\vec{B}_0, \vec{L}_0)\). As mentioned above, for the first iteration just interpret \(\vec{L}, \vec{B}\) in the separation formulas as total averages over all the data (or the city).
- Step 2. Substitute the resulting values \((\vec{B}', \vec{L}')\) := \((f(\vec{L}), g(\vec{B}))\) of the previous step to the same formula \((f(\vec{L}'), g(\vec{B}'))\). It gives the second approximation.
- Step 3. If necessary, repeat step 2.

Note that repeating the iterations will not increase the accuracy of the prediction up to 100%, because it is determined by the number of cases in the data. So, just two steps can be sufficient for a high accuracy prediction. The other reason that not too many iterations should be taken is the fact that we do not know for sure if the stable manifold \(M^-\) is the total space. If it is not, there is a nonempty unstable manifold \(M^+\). In this case the stable manifold plays the role of repeller and the unstable one that of attractor. So, if the initial point is not exactly on \(M^-\) (let it be an arbitrary small transversal shift from it) then after several steps of approaching the fixed point \((\vec{B}_0, \vec{L}_0)\) it will be unlimitedly going far away.
References
Anselin, L. Local indicators of spatial association – LISA. Geographical Analysis, 1995, 27, 93-115
Halmos, P. R. Measure Theory, Springer-Verlag, 1974.


Shi, S. The Improved Net Rate Analysis, *Journal of Real Estate Research* (Forthcoming paper).


Table 1 - Various price apportionment methods

<table>
<thead>
<tr>
<th>Apportionment theories</th>
<th>Techniques</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fractional Apportionment Theory (FAT)</td>
<td>1.1 Allocation</td>
<td>Cost, Price</td>
</tr>
<tr>
<td></td>
<td>1.2 Abstraction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.3 Extraction</td>
<td></td>
</tr>
<tr>
<td>2. Rent Apportionment Theory (RAT)</td>
<td>2.1 Subdivision</td>
<td>Income</td>
</tr>
<tr>
<td></td>
<td>2.2 Land residual</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.3 Ground rent</td>
<td></td>
</tr>
<tr>
<td>3. Price Apportionment Theory (PAT)</td>
<td>3.1 Direct comparison</td>
<td>Price</td>
</tr>
<tr>
<td></td>
<td>3.2 OLS modeling</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>3.3 Geometric modeling</strong></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Example of average values

<table>
<thead>
<tr>
<th>Sales of comparable properties</th>
<th>Westmount</th>
<th>Verdun</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sale 1</td>
<td>1,980,000</td>
<td>990,000</td>
<td>1,485,000</td>
</tr>
<tr>
<td>Sale 2</td>
<td>1,400,000</td>
<td>700,000</td>
<td>1,050,000</td>
</tr>
<tr>
<td>Sale 3</td>
<td>1,600,000</td>
<td>800,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>Sale 4</td>
<td>1,600,000</td>
<td>800,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>Averages</td>
<td><strong>1,645,000</strong></td>
<td><strong>822,500</strong></td>
<td><strong>1,233,750</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average values</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Land</td>
<td>300,000</td>
<td>30,000</td>
<td>165,000</td>
</tr>
<tr>
<td>Building</td>
<td>1,000,000</td>
<td>177,500</td>
<td>588,750</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Separate land value estimation</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average price of properties when lands are the same*</td>
<td>1,645,000</td>
<td></td>
<td><strong>1,233,750</strong></td>
</tr>
<tr>
<td>Average price of buildings</td>
<td>1,000,000</td>
<td></td>
<td>588,750</td>
</tr>
<tr>
<td>Predicted price of the subject land</td>
<td>645,000</td>
<td></td>
<td>645,000</td>
</tr>
</tbody>
</table>

* 4 References in each sector (8 references in the whole city)
Table 3. Definition, codification and statistics of variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition and codification</th>
<th>Average</th>
<th>Min.</th>
<th>Max.</th>
<th>k-values from OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Selling price of properties in 2002 (in can. $)</td>
<td>145,881</td>
<td>41,800</td>
<td>610,000</td>
<td>-</td>
</tr>
<tr>
<td><strong>Building attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Livingarea</td>
<td>Living area in sq. feet</td>
<td>1,543</td>
<td>502</td>
<td>4,904</td>
<td>28.2%</td>
</tr>
<tr>
<td>2. Basemntsize</td>
<td>Finished basement in sq. feet</td>
<td>371</td>
<td>0</td>
<td>2,413</td>
<td>6.6%</td>
</tr>
<tr>
<td>3. Detached</td>
<td>Type of property is detached = 1; 0 otherwise</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4.4%</td>
</tr>
<tr>
<td>4. Age</td>
<td>Age of properties in years</td>
<td>31</td>
<td>0</td>
<td>132</td>
<td>4.4%</td>
</tr>
<tr>
<td>5. Garajsize</td>
<td>Size of the garage in sq. feet</td>
<td>191</td>
<td>0</td>
<td>1,350</td>
<td>4.4%</td>
</tr>
<tr>
<td>6. Fireplace</td>
<td>Number of fireplaces</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4.1%</td>
</tr>
<tr>
<td>7. Bungalow</td>
<td>Model of property is detached = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.4%</td>
</tr>
<tr>
<td>8. Swimpolsiz</td>
<td>Size of the swim pool in sq. feet</td>
<td>14</td>
<td>0</td>
<td>880</td>
<td>2.7%</td>
</tr>
<tr>
<td>9. Bathrom</td>
<td>Number of bathrooms</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2.4%</td>
</tr>
<tr>
<td>10. Brick</td>
<td>Exterior recovery of the property is brick =1; 0 otherwise</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.9%</td>
</tr>
<tr>
<td>11. Electric</td>
<td>Principal heating system is electric =1; 0 otherwise</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.7%</td>
</tr>
<tr>
<td><strong>Land attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Distance</td>
<td>Distance of the subject to the references (x-y coordinates)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.6%</td>
</tr>
<tr>
<td>2. Landarea</td>
<td>Land area in sq. feet</td>
<td>5,493</td>
<td>714</td>
<td>30,339</td>
<td>6.3%</td>
</tr>
<tr>
<td>3. River100</td>
<td>Less than 100 m. from the river = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>4. Commerces</td>
<td>Dist. In km to the shopping center in m.</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2.2%</td>
</tr>
<tr>
<td>5. Railway60</td>
<td>Less than 60m. from the railway = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.7%</td>
</tr>
<tr>
<td>6. Highway60</td>
<td>Less than 60m. from the highway = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.5%</td>
</tr>
<tr>
<td>7. Subway</td>
<td>Dist. In km to the subway station in m.</td>
<td>9</td>
<td>0</td>
<td>24</td>
<td>1.2%</td>
</tr>
<tr>
<td>8. Indst100</td>
<td>Less than 100m. from heavy industrial zone = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.7%</td>
</tr>
<tr>
<td>9. Parc100</td>
<td>Less than 100m. from park = 1; 0 otherwise</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.7%</td>
</tr>
<tr>
<td>10. School</td>
<td>Dist. in km to an educational institution</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Table 4. An example of separate estimations

<table>
<thead>
<tr>
<th>Subject property</th>
<th>Estimations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average (total) price of filtered properties when lands are the same</td>
<td>186,484</td>
<td>47</td>
</tr>
<tr>
<td>Average price of buildings in Montreal</td>
<td>95,000</td>
<td></td>
</tr>
<tr>
<td>Predicted price of the subject land</td>
<td>91,484</td>
<td></td>
</tr>
<tr>
<td>Average (total) price of filtered properties when buildings are the same</td>
<td>380,742</td>
<td>16</td>
</tr>
<tr>
<td>Average price of lands in Montreal</td>
<td>48,328</td>
<td></td>
</tr>
<tr>
<td>Predicted price of the subject building</td>
<td>332,414</td>
<td></td>
</tr>
<tr>
<td>Total price prediction</td>
<td>423,898</td>
<td></td>
</tr>
<tr>
<td>Observed price</td>
<td>420,000</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>3,898</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Estimates

<table>
<thead>
<tr>
<th>Measure space model</th>
<th>Average price of lands</th>
<th>Average price of buildings</th>
<th>Ratio building over land</th>
<th>Average total prices</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial averages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48,328</td>
<td>95,000</td>
<td>1.966</td>
<td>143,328</td>
<td>41,569</td>
<td></td>
</tr>
<tr>
<td><strong>Measure space - improved averages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51,585</td>
<td>102,035</td>
<td>1.978</td>
<td>153,620</td>
<td>18,676</td>
<td></td>
</tr>
<tr>
<td><strong>Hedonic - OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50,396</td>
<td>108,278</td>
<td>2.149</td>
<td>158,674</td>
<td>30,723</td>
<td></td>
</tr>
<tr>
<td><strong>City evaluations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50,982</td>
<td>97,891</td>
<td>1.920</td>
<td>148,873</td>
<td>17,636</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Results of estimations from the measure space model
Figure 2. Results of estimations from OLS and City
Figure 3. Separate building and land values