
WORKING PAPERS

A residential land price index for Luxembourg: Dealing with the spatial dimension

Brano **GLUMAC**¹
Marcos **HERRERA-GOMEZ**²
Julien **LICHERON**¹

¹ LISER, Luxembourg

² National University of Salta, Argentina

LISER Working Papers are intended to make research findings available and stimulate comments and discussion. They have been approved for circulation but are to be considered preliminary. They have not been edited and have not been subject to any peer review.

The views expressed in this paper are those of the author(s) and do not necessarily reflect views of LISER. Errors and omissions are the sole responsibility of the author(s).

A residential land price index for Luxembourg: Dealing with the spatial dimension

Brano Glumac

Urban Development and Mobility, Luxembourg Institute for Socio-Economic
Research Institution, Luxembourg

Marcos Herrera-Gomez

CONICET – IELDE, National University of Salta, Argentina

Julien Licheron

Urban Development and Mobility, Luxembourg Institute for Socio-Economic
Research Institution, Luxembourg

April 2018*

Abstract

Urban development projects have many effects on society, such as improving mobility, health, education, and sustainability. For policy-relevant measures, it is important that policymakers are able to foresee how quality improvements influence the price of land. Therefore, our objective is to collect a set of variables able to account for the effects of a multitude of land quality improvements. In addition, surrounding plots and the natural and built environment might also influence urban land prices. However, most house price and land price indices do not control for any potentially related spatial effects. The urban land price index detailed here is based on land transaction prices for Luxembourg between 2010 and 2014 recorded in notarial deeds and cadastral data, together with geo-spatial characteristics. The proposed index includes many aspects in an initial hedonic model specification, the index also operates on a spatial model.

Keywords: land value; hedonic regression; spatial Durbin error model; Luxembourg

JEL classification codes: C23 ; E31 ; R31 ; R32

* This paper has been accepted on 21st of July 2017

1. Introduction

The housing market represents the largest proportion of household expenditure in developed economies, and changes to house prices can have important macroeconomic effects. The importance of this market sector was clearly exposed by the collapse of the U.S. housing market and the subsequent global financial crisis that began in 2007. As a consequence, nowadays there is a growing demand from government policymakers and independent sectors for reliable house price indices as economic indicators.

The general purpose of a house price index is to monitor economic activity. There is a strong correlation between house prices and economic activity, and therefore the house price index can be regarded as a barometer for the national economy (Hill 2013). In this article, we argue that an urban land price index can be used in the same way, because there is a high correlation between land prices and house prices and their causal correlation has been well explained (E. C. Hui, Leung, and Yu 2014). Similarly to a house price index, a land price index indicates changes in prices over time of land of comparable quality. Therefore, to design adequate housing policies and to understand the dynamics of the housing market, a precise estimation of an urban land price index across different regions or spatial units would be a very useful piece of information.

This article presents the first hedonic urban land price index with a national coverage. In addition, it incorporates a spatial dimension and extensive controls to capture price dynamics. It is apparent that surrounding land and the natural and built environment might influence the price of a plot of land over time (Garza and Lizieri 2016). However, most existing house price indices and land price indices do not control for any of the potentially related spatial effects. After performing a spatial dependence test, we constructed an index that is based on a hedonic model in order to cope with several identified spatial effects. The assumed key advantage is that such an index will better account for issues related to spatial and temporal dependence. In addition to a sufficient number of land transactions or basic data, it is necessary to collect a set of variables that accounts for the multitude of land quality improvements related to any urban development. On the assumption that an urban development project will have a multitude of impacts on society, such as improving mobility, health, education, and

sustainability, it is essential for policymakers to be able to monitor how these improvements influence the price of the land itself. Therefore, the reported index includes many aspects in the initial hedonic model specification.

One of the major difficulties in developing a land price index is the challenge of collecting basic data, which results in there being very few reported land hedonic indices (Nichols, Oliner, and Mulhall 2013). This paper presents an urban land price index for Luxembourg based on the land transaction prices reported by notarial deeds (Administration de l'Enregistrement et des Domaines, AED) and cadastral data (Plan Cadastral Numérisé, PCN), together with geo-spatial characteristics (GIS).

The rest of this article comprises five sections. A background section reviews existing understandings of land price valuations and indices. The methodology section explains in detail the selected functional form, initial model specification, and the calculation of the index. We develop the analysis to a fully parametric model in order to keep the results interpretable for practical purposes. Special attention is paid to describing data sources and the selection of treatments taken into account for selecting the most appropriate outliers. The results section reports on the hedonic model and double imputed hedonic index, and we complete the article with a general conclusion.

2. Background

Land is a heterogeneous good, because of its different characteristics and different uses, triggered by different dynamic processes. Furthermore, it can be stated that the appreciation and depreciation of land prices within a given region are a function of dynamic social, economic, legal, and environmental processes continuously interacting over time and across the space (Grigsby 1986). Similar to indexing any heterogeneous good, a land index illustrates changes in the prices of land of comparable quality over time.

From a historical perspective, real estate and land valuation or appraisal were introduced into the scientific community in the second half of the nineteenth century (Moore 2009). The first land price study based on a hedonic approach (Eurostat 2013) was a master's thesis on agricultural land values in 1922 at the University of Minne-

sota, as suggested by Colwell and Dilmore (1999). In 1970, the first automated valuation – a mathematically based software estimating a market value (IAAO 2003), also emerged from land valuations (Gwartney 1970). These facts are not a coincidence, and perhaps the first valuation developments occurred based on land observations because land features tended to be much more uniform than any real estate property. However, land valuation provides a set of unique problems, such as the highly speculative nature of transactions and few recorded transactions for analysis and modeling (IAAO 2003).

With regard to the prospects for urban land price indices, the applications are as numerous as for house price indices, although they do not provide exactly the same insights. It is important to highlight first the possibility to monitor national economic activity, similarly to the house price index. A strong correlation exists between house prices and economic activity, and therefore the house price index can be regarded as a barometer for the national economy (Hill 2013). For example, in many countries in the period leading up to the financial crisis, house prices rose faster than rents (Hill 2013). The urban land index can be used in a similar way, because there is also a high correlation between land prices and house prices. In addition, an accurate measurement of land price changes (across time and space) is important for local policy development such as institutionalizing value capturing (E. C. M. Hui, Ho, and Ho 2004). Further, for policy-relevant measures, it is important that policymakers are able to foresee how quality improvements influence the price of land as urban development projects have many effects on society, such as improving mobility, health, education, and sustainability. A land price index can also help us better understand real estate development processes, because the land development process can be regarded as an outcome of interrelated actors' decisions, in which the land price stands for a considerable part of a payoff function (Samsura, van der Krabben, and van Deemen 2010; Glumac et al. 2015). Furthermore, land supply and demand ratio differs for agricultural and urban uses, which implies different price tendencies over the years. Because the total land supply is shared and limited, all land prices correlate and closely lead to a zero-sum game outcome. Therefore, comparing the urban and the agricultural land price index might also reveal the trend of the conversion rate from agricultural to urban land. Lastly, the prospect of having a national urban land price index can be seen in the methodological advancements in the decomposition of land and structure values. This urge is evident from today's trend to derive house indices from decomposing a real

estate property value into structure and land (Francke and van de Minne 2016; Davis and Heathcote 2007; Bostic, Longhofer, and Redfearn 2007; Eurostat 2013; Diewert, Haan, and Hendriks 2015), where the structure part should capture technological advancement (i.e., smart home systems) and housing preferences (i.e., house style), and the land part should capture more macro-economic trends and events.

The most general differentiation of vacant land is between urban and non-urban plots (Northam 1971). They vary in many aspects, and thus different valuation approaches should be used. For example, for land that is remote from urban areas it would be more suitable to rely on the income potential (IAAO 2003) rather than using sales data to estimate the value. This article only details urban land plots, which can be also further divided into residential and mixed-use land. These plots are located in areas that can either be used for the construction of residential buildings or offices or commercial buildings. Although commercial urban land property is acquired to build income-producing properties, we can say the same for residential urban land, because in most cases it is bought by real estate developers. This implies estimating value via an income approach and methods such as discounted cash flow (French 2013). However, having adequate commercial data is rarely the case and estimating value by comparison or a hedonic approach is an alternative.

Since Gwartney (1970), many authors have reported on automated valuation models for land property in past decades: in seventies (Holland 1970), in eighties (Burrough 1989), in nineties (Clapp 1990; Des Rosiers, Thériault, and Recherche 1999) in two-thousands (Yamazaki 2001; Clapp 2003; Davis and Heathcote 2007; Bostic, Longhofer, and Redfearn 2007; Des Rosiers and Thériault 2008; Haughwout, Orr, and Bedoll 2008) and recently (Oliner, Nichols, and Mulhall 2010; Spinney, Kanaroglou, and Scott 2011; Nichols, Oliner, and Mulhall 2013; Liu, Wang, and Zha 2013; Costello 2014; Wu et al. 2014; Wade Brorsen, Doye, and Neal 2015; Glaesener and Caruso 2015; Demetriou 2016; (Demetriou 2018). Some patents have also been filed (Matsuo et. al 1999). However, very few of them are land price indices (Yamazaki 2001; Clapp 1990; Spinney, Kanaroglou, and Scott 2011).

The underlying differences between all the price indices are linked to the different objectives of constructing them. It is not strange then, that “there remains a distinct lack of consensus on the most appropriate index methods” (Goh, Costello, and

Schwann 2012). Therefore, different approaches and related methods for compiling constant quality indices have been developed over time (e.g., Eurostat 2013): appraisal based methods, repeated sales methods, median indices, and hedonic methods. Bearing in mind that land is not frequently traded (compared with real estate property) and that land appraisal reports are not necessarily centralized, it is not always possible to collect a sufficient number of land appraisal reports. In addition, the appraisal-based approach has also been greatly criticized for different types of biases (Shimizu and Nishimura 2006). Similarly, repeated sales methods are not considered, because the land is not frequently traded and it is even harder to collect repeated sales observations over time. Lastly, median indices are not able to capture the impact of quality changes. Briefly, the shortcomings of these methods point toward hedonic methods as the most appropriate for land indices if there is sufficient data available.

Important countrywide overviews of house price indices have been introduced by Hill (2013) and Eurostat (2013). However, land price indices have been used far less frequently internationally. Usually they are not a part of standard statistics reports, mostly because of unavailable data regarding either the dataset extent or quality. There is an index for apartment prices in Luxembourg (Lamboray 2010), and the results are presented on a regular basis. More recently a hedonic index was developed by Preclin (2015) to follow the prices of “old” houses in Luxembourg, based on the data available from AED and a survey of the dwellings’ characteristics. For the purpose of comparing price evolutions and further understanding the real estate market, the construction of a hedonic price index for urban land prices is complementary to the work undertaken so far in Luxembourg, and as mentioned, it is the first nationwide hedonic land price index.

3. Methodology

The hedonic regression model is accredited to Rosen (1974), and based on Lancaster’s (1966a) approach in consumer theory, although the development of hedonic models can be traced back to as early as 1922, as mentioned. Rosen (1974) also considered land a composite good, composed of a bundle of attributes from which consumers obtain utility. The price paid for this good can be decomposed by regression analysis, for example, to the implicit prices of the different attributes, hence allowing insights into

consumers' preferences and their willingness to pay for the heterogeneous components of a good. Several advantages of hedonic methods influence its choice for a land index. First, in principle the hedonic approach allows us to adjust for changes to both the sample mix and quality when the land characteristics are sufficiently covered by data (Eurostat 2013). Second, indices based on the hedonic method are relatively easy to stratify by different locations and products. Third, hedonic based indices are also best suited to test spatial price effects. A very detailed view on the strengths and weaknesses of different hedonic house price indices (time dummy, characteristics, imputed method, and hybrid) has also been produced (Hill 2013). Among these, the features of a double imputed index best fit the Luxembourg case. An important feature in the choice of index is that the reported index values for previous periods do not change when adding data from a new period. By contrast, estimating a model on a data sample that covers all periods (e.g., time dummy) will consequently change coefficients every time a new sample is added to the dataset. The imputed price method has been shown as a convenient method to update hedonic indices at different time intervals. This feature would provide policymakers with consistency in reporting from year to year.

3.1 Hedonic functional form and the spatial Durbin error model

In order to estimate the contributions of land characteristics using standard regression techniques, the hedonic model has to be specified as a parametric model. A Box-Cox test indicates the log-linear form as appropriate:

$$\ln y_n^t = \alpha_0 + \sum_{k=1}^K \beta_k^t x_{nk}^t + \varepsilon_n^t, \quad [1]$$

where $\ln y_n^t$ is the logarithm of the price of land plot n ($1, \dots, N$) for the period t ($1, \dots, T$), α_0 is the coefficient estimate of constant price quality for all periods t , β_k^t is the coefficient estimate of a characteristic k ($1, \dots, K$) for the period t ($1, \dots, T$), x_{nk}^t is a characteristic k ($1, \dots, K$) of land plot n ($1, \dots, N$) for the period t ($1, \dots, T$), and ε_n^t is the error term component of land plot n ($1, \dots, N$) for the period t ($1, \dots, T$) assumed as independent and identically distributed.

As has been noted, one of the most important determinants of land prices is location. This characteristic means that the hypothesis of uncorrelated error terms in equation [1] is probably violated by the presence of geographical spillovers. This omitted effect could generate biased estimates from the parameter model. Literature usually takes the simplest way of accounting for this omission through the inclusion of neighborhood dummy variables. However, when the exact geographical positions of land plots are available—and it is relatively straightforward to use a geographic information system (GIS) to calculate each land’s exact longitude and latitude—advanced approaches can be applied.

This increasing availability of spatial information at the land level allows the impact of location to be modeled in very different ways using spatial effects. There are three possible spatial effects: an endogenously spatial autoregressive term that reflects the importance of prices of neighboring land, an exogenous effect that reflects the impact for each unit of the change in exogenous variables, or an exogenous effect that might affect the unexplained part of the model through a spatial autocorrelation structure due to unobserved environmental effects common to neighboring land (Elhorst 2010).

A spatial exogenous observed effect implies that there is a significant influence from a known variable in the neighborhood of land plot n ; it is formally considered as a spatial lag term in equation [2]:

$$\ln y_n^t = \alpha_0 + \sum_{k=1}^K x_{nk}^t \beta_k^t + \sum_{k=1}^K \left(\sum_{j \neq n}^N w_j^t x_{jk}^t \right) \gamma_k^t + \varepsilon_n^t, \quad [2]$$

where w_{jn}^t is a weight different to zero if the land plot j ($j \neq n$), from 1 to N , is a neighbor of the land plot n for every period t . In our case, the neighborhood is defined using the inverse distance criteria with a threshold of 6km. This threshold guarantees at least one neighbor for every observed land plot n in all years, the coefficient γ_k^t captures the impact of average neighborhood characteristic k on the logarithm of land price n for the period t ($1, \dots, T$).

An unobserved environmental effect implies that there is a significant influence of an unknown variable in the neighborhood. It is formally considered as a spatial error term in equation [3] and is tested with a robust Lagrange multiplier:

$$\ln y_n^t = \alpha_0 + \sum_{k=1}^K x_{nk}^t \beta_k^t + \sum_{j \neq n}^N w_j^t \varepsilon_j^t \rho^t + v_n^t, \quad [3]$$

where ρ^t is a spatial coefficient that captures the unobserved average neighborhood influence that impacts on land plot n ($1, \dots, N$) for the period t ($1, \dots, T$), and v_n^t is an error term component of land plot n ($1, \dots, N$) for the period t ($1, \dots, T$).

A spatial endogenous effect implies that there is a significant influence of the average of the logarithm of prices in the neighborhood of land plot n ; it is formally considered as a spatial lag term. However, this effect was not tested because a spatial endogenous effect can be compensated for by a temporal lag in prices, which is the final purpose of a housing index (LeSage and Pace 2009, p. 26).

Spatial dependency tests indicate the existence of observed and unobserved exogenous factors. The discovered spatial lag in independent variables and spatial error can be captured by a spatial Durbin error model (SDEM), which is used as our final model [4]. To support this selection, we conducted Moran's I and robust Lagrange Multiplier tests (Anselin et al. 1996) on the linear model equation [1], and the test results (Table A2) provide strong evidence for the necessity to introduce a spatial Durbin error model.

$$\ln y_n^t = \alpha_0 + \sum_{k=1}^K x_{nk}^t \beta_k^t + \sum_{k=1}^K \left(\sum_{j \neq n}^N w_j^t x_{jk}^t \right) \gamma_k^t + \sum_{j \neq n}^N w_j^t \varepsilon_j^t \rho^t + v_n^t. \quad [4]$$

3.2 Initial Model Specification

In addition to the choice of a functional form, selecting the variables plays a major role in the quality of estimates. As mentioned, the appreciation and depreciation of land prices within a given region is a function of dynamic social, economic, legal, and environmental processes continuously interacting over time and across the space (Grigsby 1986). Therefore, the selected variables are "classed" within these four essential elements. In addition, two large classes that dominate any real estate property value estimation (French 2001) are added: location (accessibility and proximity) and the physical characteristics of the land and its surroundings. The initial list of variables

(Table 1) is additionally tested (Andersson 2000). In order to provide a consistent overview of the variables included in the hedonic price model for Luxembourg, three explanatory columns are shown: 1) “Class” is a group of variables derived from the mentioned definition; 2) “Variable” is a constructed measurement or proxy of a certain attribute that is characteristic of certain products such as land (Lancaster 1966a, 1966b); and 3) “Unit” is the measurement unit of the variable.

TABLE 1 Overview of classes, variables, and units

Class	Variable	Unit
Accessibility	Distance to road	m
Accessibility	Distance to motorway	m
Accessibility	Distance to bus station	m
Accessibility	Distance to train station	m
Accessibility	Distance to airport	m
Proximity	Car time to CBD*	min
Proximity	Public transport time to CBD*	min
Proximity	Medical amenities	0-1
Proximity	Commercial amenities	0-1
Proximity	Cultural amenities	0-1
Proximity	Proximity to border	dummy
Physical	Land plot size	are
Physical	Shape of land plot	1-7
Physical	Hillshade	0-255
Environmental	Nearby industry	are
Environmental	Nearby public green space	are
Environmental	Nearby agriculture	are
Environmental	Nearby forest	are
Environmental	Nearby water	Are
Legal/Physical	Highly buildable zone	dummy
Legal	Municipality territory	dummy
Legal/Economic	Baulücke**	dummy
Social	Population density	pop/km ²

Economic	Year	Y
Economic	Price per are	€/are

*Central Business District; ** Small neglected urban land plot

Location attributes are usually divided into accessibility and proximity components. With regard to accessibility, each transport mode has been captured by including distances to a road, motorway, bus station (Mulley 2013), train station, and airport. Each distance variable uses meters as a measurement unit. Proximity accounts in the first place for the distance to Luxembourg City or the central business district (CBD). In theory and in empirical studies this variable has been shown to be an important determinant of land prices, and it is also considered in existing price indices in Luxembourg, thus making it comparable with existing housing statistics for the country, such as apartments (Lamboray 2010) and old houses (Preclin 2015). The proximity to the CBD was measured by the time needed to reach it either by car (Osland 2010) or public transport. Within this class, proximity to health, commerce, and leisure services are also observed. Each of these variables was generated in a similar fashion. For example, to compose proximity to health, the total number of hospitals and pharmacies in one district was summed and normalized by district, with the highest total number equal to 1, and with 0 representing the district with the smallest number of hospitals and pharmacies. The last within this class is border proximity, measured as dummy variable equaling 1 for land plots within a distance of 100m of the border. Since Luxembourg is a small country in area, it is assumed that land plots closer to the border will have an impact on the country scale because of their higher percentage of the total area. From the physical characteristics, only two variables are included in the initial model specification. Size of the land plot measured in are, and hillshade. The latter indicates how bright or dark a plot is. Given the land slope and orientation toward the sun it is possible to calculate hillshade (ESRI ArcGIS 2011). Hillshade (Soguel, Tangerini, and Pictet 2007) is described by the integer variable with the value 0 representing the darkest and 255 the brightest land plots in the Luxembourg in the evening (7pm – 8.30pm) on 20 June. Last is the measurement of the land’s shape, captured by a Gravelius index. In this index (Gravelius 1914), the value 1 represents the most ideal shaped plot (a circle) with higher numbers standing for more extended plots. Five environmental variables are included in the initial specification, nearby industry (Anselin

and Lozano-Gracia 2008; Din, Hoesli, and Bender 2001) within 3km, and nearby public green spaces (Jones and Reed 2018) within 50m, as well as agriculture, water, and forest. These are computed as the total sum of land plots within the mentioned range and measured in total surface area. With regard to the legal variables, three are important for the land model specification. “Highly buildable zone” is a dummy variable with the value 1 indicating that at least 100m² of the land plot can be used for either residential or mixed-use purposes. This variable can be also considered as a physical class. The second variable indicates the municipality a land plot is in. Lastly, Baulücke was introduced as a Luxembourg policy tool to increase underused city space. A Baulücke plot is a small and neglected piece of land, expressed as dummy variable. With regard to the “social” variable class, there is empirical evidence of the importance of the influence of population density (p/km²) on land prices. The density range was set empirically at 3km. Lastly, the economic characteristics of a land plot are taken as the price per are (€/are) and the time of sale (year) covering the period from 2010 to 2014.

This initial specification (Table 1) was further reduced with a correlation test and a final set of variables is provided (Table 3). The effects caused by the different relationships that variables might have with the dependent variable are captured by the functional form of the hedonic model. For non-linear effects, we compared the significance level of a variable with quadratic and non-quadratic expressions in the full model specification. This procedure resulted in a squared expression for some of the variables (Table 4).

3.3 Hedonic Double Imputed Fischer Index

A very common methodological problem in constructing price indices is the impossibility of observing the same good, and its characteristics, in different periods. For example, the characteristics of a house sold in period 0 cannot be observed in period t because once sold, the house (with a set of unique characteristics) will not usually be resold in the same period t. Therefore, we chose to use an imputed index that imputes the missing prices predicted by a hedonic model (Eurostat 2013; Hill 2013). Specifically, we apply a geometric double imputation index using the Fisher formula (Eurostat 2013):

$$P_{HDI}^{0t} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0) \exp \left[\sum_{k=1}^K (\hat{\beta}_k^t - \hat{\beta}_k^0) \bar{x}_k^{0t} \right], \quad [5]$$

where $\bar{x}_k^{0t} = (\bar{x}_k^0 + \bar{x}_k^t)/2$ denotes the mean average characteristics in periods 0 and t, $\hat{\beta}_k^t$'s ($t = 0,1$ and $k = 0, \dots, K$) are the estimated coefficients using the equations from the previous section. For more details of this index, see Eurostat (2013, pp. 54–55). We use the equation [5] to calculate both a Fisher Index, based on the estimation of a non-spatial model [1], and a Spatial Fisher Index that uses the prediction of the SDEM model [4].

4. Data and outliers

4.1 Detection of outliers

The dataset was generated by merging data from notarial deeds (from AED), cadastral data (from PCN), and GIS. It includes the purchases of land plots between 2010 and 2014, comprising 9910 individual transactions. It is a relatively rich dataset with variables representing accessibility, proximity, and the physical, legal, and economic characteristics of land plots (Table 1). After removing mostly agricultural land plots, the dataset includes 4297 individual transactions of residential and mixed-use land, located in urban areas and suitable for the construction of residential dwellings or for the construction of offices or commercial buildings.

Special attention was paid to outlier selection. Characteristically for land price datasets, there are fewer observed transactions and a greater spread in dependent variable values compared with house price datasets. If not selected correctly, outliers can seriously misrepresent the land index. Three criteria were used to obtain a clean dataset. First, it is required for there to be at least two observations in one year in one municipality to have a credible number of observations and to test for spatial dependencies. We took a conservative position to avoid potential bias resulting from only one observation in one municipality. Municipalities having only one observation per year do not fulfill this criterion, thus 47 observations were removed. Second, using the distribution of the logarithm of price for each municipality, we drop values higher than three times the interquartile range, resulting in 44 fewer observations. Third, a specific criterion to

account for irregular spatial distribution of observations was introduced. It is a combination of an outlier map and a box-plot technique, using the software GeoDa. This software allows an analysis to identify land with extreme values, taking into account the geographical neighborhood (Anselin, Syabri, and Kho 2006). We accordingly removed observations for geographically close plots with extreme values, because although registered as separate transactions they were most probably traded as unique transactions. Including these would misrepresent betas for different variables (i.e., plot size), therefore 19 additional observations were removed. The final clean dataset consists of 4187 observations. Figure 1 shows the difference between the mean indexed values of the non-cleaned dataset, or all data index, and the clean dataset called the raw index. To avoid doubts of misrepresenting price trend, we provide detail insight for the outlier in the last reported year. It should be noted that the total difference of indexed values in 2014 includes the removal of 12 transactions falling under the first outlier criterion, 17 transactions falling under the second outlier criterion (values higher than three times the interquartile range for 2014 on municipal average), and four transactions under the third criterion.

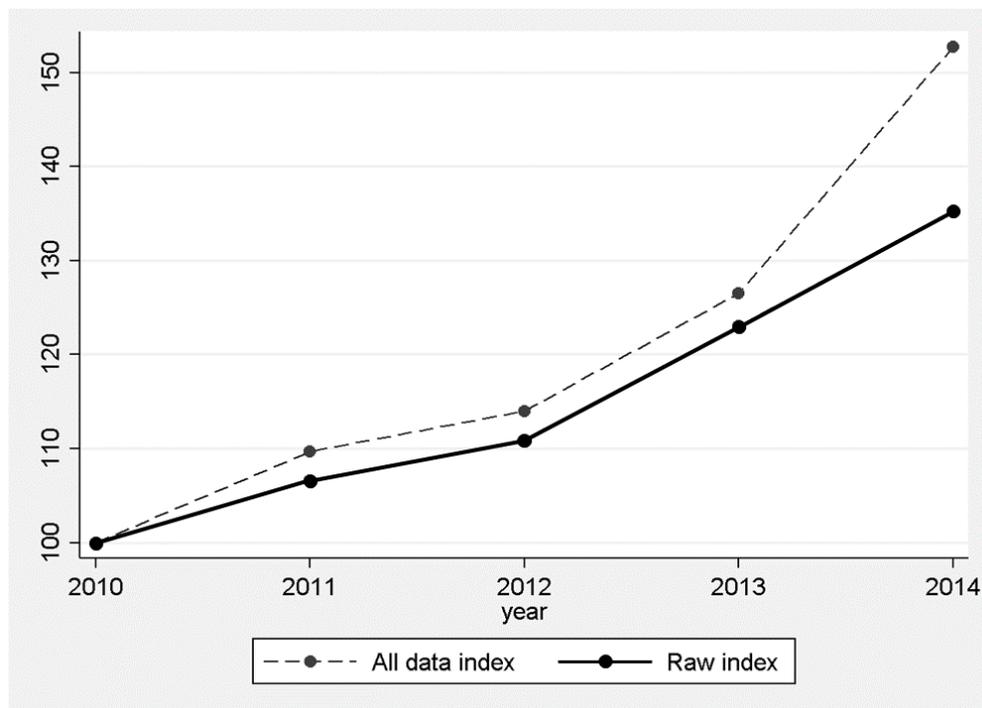


FIGURE 1 Difference between all data and cleaned data (raw) index

In addition, removed observations have to be reasonably well spread over the sampled years, otherwise the reduction of observed data in one specific year might substantially change the index values. The cleaned dataset in our result is valid because the outliers are well distributed (Table 2).

TABLE 2 Data removal over the years

Year	Initial	Cleaned	Outliers
2010	949	914	35
2011	870	860	10
2012	859	842	17
2013	748	733	15
2014	871	838	33

4.2 Descriptive statistics after outlier removal

Table 3 introduces the selected variables that are used in the model specification. It provides a basic insight into land prices and price trends. In addition, Table A1 also shows the descriptive statistics after outlier removal, but broken down into yearly statistics.

TABLE 3 Descriptive statistics after outlier removal (N=4187)

Variable	mean	SD	p25*	min	max
Price	388682.79	1.84e+06	125932.59	0.43	6.19e+07
Price per are	64705.68	107366.45	25000.00	139.53	2.83e+06
Price per are (ln**)	10.45	1.41	10.13	4.94	14.85
Distance to road	15.42	24.26	0.56	0.00	224.08
Distance to bus station	199.97	144.97	100.42	0.81	1464.31

Distance to train station	4214.61	3880.62	907.51	25.49	17551.00
Car time to CBD	25.37	14.67	15.34	0.00	73.10
Land plot size	6.21	11.44	2.89	0.00	385.27
Nearby industry	2.19	3.12	0.22	0.00	18.46
Shape of land plot	1.40	0.60	1.17	1.08	10.37
Population density	307.21	400.84	73.81	9.54	2411.41
Hillshade	177.48	15.65	170.30	86.81	242.70
Highly buildable zone	0.89	0.31	1.00	0.00	1.00
Baulücke	0.21	0.40	0.00	0.00	1.00

* The lowest 25% values; ** Normal logarithm values

5. Results

5.1 Land price spatial Durbin error model estimates

The reported estimates (Table 4) of the land price spatial Durbin error model (SDEM) show the regression for all observations (n=4187) or transactions made on residential and mixed-use land in the period between 2010 and 2014.

As a general remark concerning the selected specification, we initially tested for multicollinearity and non-linear effects for all variables in Table 1 under a simple model (equation 1) and only significant variables from the initial specification were included in the final model. However, under the spatial model, variables that are only significant for some of the year samples are also included in the hedonic regression model and in the price index. The reported model is stable over the years and the coefficient signs can be interpreted. The variable names, coefficients, and significance levels are shown in Table 4 for each year separately. The selected models are well fitted to the data, with reported pseudo-R² rising from 0.557 to 0.646. Additionally, Table A3 shows that the selected SDEM performs better than a standard ordinary least squares (OLS) model when comparing the calculated yearly pseudo-R².

The chosen dependent variable is the normal logarithmic value of a land plot price per are (Price per are (ln)). Commonly, researchers report only the influence of the plot size on the total transaction price, where the plot size is usually the most important determinant and thus disputably increases the model fit parameters. Including the price

per are as the dependent variable allows us to test the influence of plot size. With regard to the accessibility variables, distance to airport and distance to motorway were removed due to a high correlation with other variables. A minus coefficient sign (Distance to road; Distance to train station) indicates that a shorter distance has a positive impact on price increase. By contrast, distance to bus stops (Distance to bus station) has a concave quadratic function indicating that a close distance has a negative impact on price, as does a long distance. Close proximity to CBD, measured in minutes travel by car (Car time to CBD), also has positive impact on the price. Other proximity variables were omitted due to not being significant or having a high correlation with other variables. The first negative coefficient of plot size (Land plot size), as expected, indicates that the price per are will concavely drop with the increased size of land plots and stabilize afterwards. The brightest land plots (Hillshade) will have higher price values, however, the function is concave and represents the fact that the measurement of brightness is taken at one time of day and on one date, and therefore cannot be linear. A land plot within 3km of a higher percentage of industrial sites (Nearby industry) has a varying impact from year to year. A negative coefficient indicates the potential exposure to pollution and a positive impact might reflect value relating to several brown-field redevelopment projects. It is also important to note the impact of the spatial neighborhood (wx Nearby industry) that works as a counterbalance. Land plots with an extended shape (Shape of land plot) will have a lower price on the market. By contrast, land plots in areas only dedicated to residential purposes will have a price premium. In addition, a higher price is expected if a surrounding land plot also has at least 100m² of residential areas as indicated by the estimated coefficient of spatial variable (wx Highly buildable zone). The municipal dummy variables (Municipality dummies) are used to estimate the fixed effect on the price of a land plot depending on which of the 106 municipalities it is located in. These variables are not reported in Table 4, although numerous municipalities had a significant impact on a land price, usually positive. Higher population densities (Population density; wx Surrounding density) also suggest a higher price per are, although this variable is not that stable over the models in different years. Lastly, Baulücke does not show any significant influence on the price of the land plots, but it is important to report it, because it was a national policy to promote neglected and small land plots that could be potentially used for building.

TABLE 4 Yearly “semi-log” SDEM for residential and mixed-use land in Lux.

Variable	2010	2011	2012	2013	2014
Price per are (ln)					
Distance to road	-0.006***	-0.002	-0.002	-0.006***	-0.003**
Distance to bus station	-0.000	0.002***	0.001	0.001*	0.000
(-II-) squared	0.000	-0.000***	-0.000	-0.000*	-0.000
Distance to train station	-0.000**	-0.000**	0.000	-0.000	0.000
Car time to CBD	-0.016*	-0.021**	-0.015***	-0.019**	-0.022**
Land plot size	-0.039***	-0.012***	-0.026***	-0.022***	-0.028***
(-II-) squared	0.000**	0.000	0.000	0.000***	0.000*
Hillshade	0.071**	0.037	0.011	0.078**	0.036
(-II-) squared	-0.000**	-0.000	-0.000	-0.000**	-0.000
Nearby industry	0.109*	-0.131*	-0.022	0.162**	-0.119*
Shape of land plot	-1.064***	-1.723***	-1.170***	-1.671***	-1.443***
(-II-) squared	0.096	0.149***	0.101***	0.183***	0.179***
Highly buildable zone	1.596***	0.959***	1.285***	1.214***	1.171***
Population density	-0.001**	0.001*	-0.000	0.001	0.001**
Baulücke	0.007	-0.011	0.058	-0.013	-0.025
wx ^a Nearby.industry	-0.225**	0.185*	0.061	0.002	0.202**
wx ^a Surrounding.density	0.001**	-0.002**	0.000	-0.000	-0.001
wx ^a High.buildable.zone	2.001***	0.631**	1.589***	1.677***	1.673***
Municipality dummies	inc.	inc.	inc.	inc.	inc.
cons	4.174	8.543***	8.103***	3.536	6.652**
rho_cons ^b	0.297***	0.135	0.133	0.259***	0.401***
Statistics					
N	914	860	842	733	838
Pseudo-Rsq	0.646	0.633	0.557	0.624	0.571

*p<.1; ** p<.05; *** p<.01;

^a weighted average influence of neighboring land plots of different characteristics;

^b spatial coefficient that captures the unobserved average neighborhood influence

Figure 2 shows an overview of land index prices in Luxembourg between 2010 and 2014. The three lines represent the following: the raw index (Figure 1), the non-spatial double imputed Fisher index (OLS model with the same specification), and the spatial double imputed Fisher index (Section 3).

When observing the selected indices, there is an evident overall increase in the prices of residential and mixed-use land in Luxembourg from 2010 to 2014. The peak was reached in 2013, with 28 percent higher prices than in 2010, while in 2014 a stable trend appeared. Moderately steep growth (a 7 percent yearly average), went together with demographic and GDP growth. Furthermore, it is important to note the difference between the spatial and the non-spatial index. Two index lines show the pattern having a tapering form with lines converging to a common point; the starting (normalized) index year. As with any index count for price changes that are comparable in quality over time, we can say that the reported spatial index captures additional quality from observed and unobserved “land” characteristics in neighboring areas and thus reports lower price changes. Lastly, when comparing the selected index with the raw index, the following can be concluded: In December 2011 and 2014, an increase in price was higher than an increase in quality. By contrast, in December 2012 and 2013, the price increase was lower than the anticipated price for quality increase. Such differences are expected, especially in relatively limited land markets, because large-scale urban development projects have a significant impact on the total land quality. It is possible that large-scale urban development projects were completed in 2012, resulting in a higher anticipated land price for increased quality than the averaged transaction price.

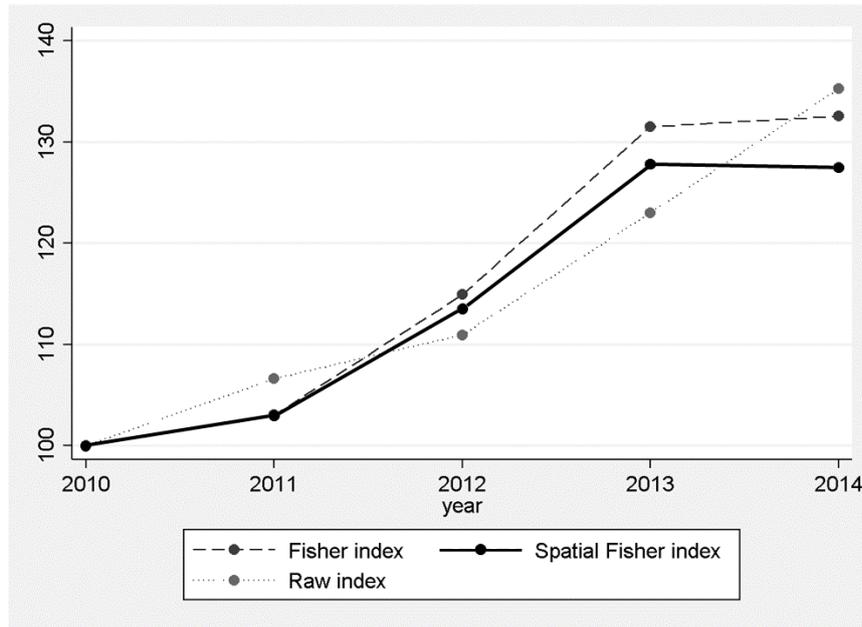


FIGURE 2 Fischer double imputed index based on a “semi-log” hedonic spatial Durbin error model (red) for urban land in Luxembourg

5.2 Robustness checks

Robustness checks were conducted in order to assess the performance of the underlying hedonic model and therefore the validity of the reported index. First, the model coefficient stability checks over time showed good results. The selected model is stable over time with regards to coefficient value, signs, and significance levels (Table 4). In addition, stability confirmation can also be concluded by comparing the similarity of the price trends between the non-spatial and spatial models (Figure 2). Second, the results are stable with less-strict outlier detection. Using only the first condition (section 4)–to have at least two observations in one year in one municipality–we obtained same tendency in the indices. Additionally, in order to confirm index stability over time, indices based on the models estimated on yearly data samples for five consecutive years were compared with indices based on the time-dummy hedonic model estimated on the joint five-year data pool (Figure 3). As there is an almost perfect match between the two types of indices, the authors conclude that the greatest drawback of an imputed index approach (i.e., instability over time) has been overcome.

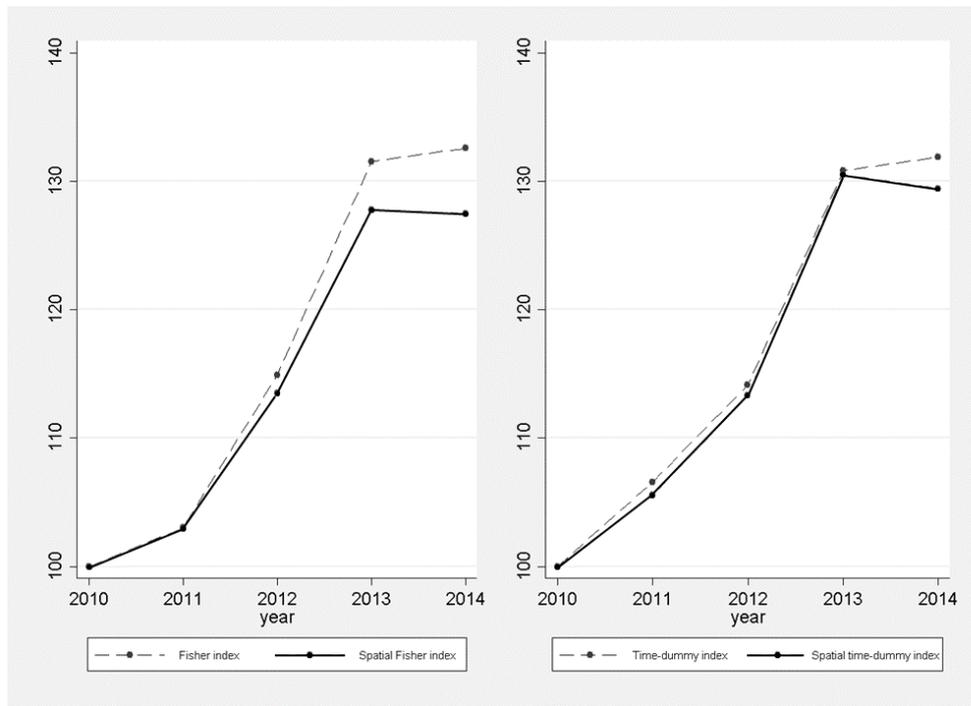


FIGURE 3 Double imputed and time-dummy index comparison

6. Conclusions

Many property-related indices have emerged to date. This article suggests the use of a residential and mixed-use land price index. It is based on the merged dataset of land transaction prices for Luxembourg reported by notarial deeds (AED) and cadastral data (PCN), together with geo-spatial characteristics (GIS).

The purpose of introducing the proposed urban land index is two-fold. First, it is assumed that a technical index based on urban land prices can capture very well spatial effects. As a result, this index can capture the value of large-scale urban development projects that entail an implicit quality improvement in neighboring areas. Second, a land price index can be of great use to provide a posteriori estimates of the impact of

changed land-use reflected in land price changes. Thus this insight might prove as necessary for policies designed to influence increased land transaction volume (e.g., land taxation) and aimed at improving land supply.

The index itself is based on a double imputed hedonic approach therefore this index that can capture individual differences between plots, and use this to illustrate differences in price while including the quality differences that have emerged over time. The initial problem associated with the difficulties in accounting for which land quality aspects should be included has been overcome by including the multitude of classes and referenced variables that respond to infrastructural upgrades over time, and policies that might influence environmental upgrades. In addition, using a spatial model to compute a land index, we are able to capture additional quality from observed and unobserved land characteristics. The index was tested for its robustness and it performs well. The proposed index is based on a spatial Durbin error model that performs well with regard to the estimated model parameters. The strong point of the proposed land price model is its ability to interpret the influence of different variables on land prices, expressed by the significance levels of chosen variables and the signs of the estimated coefficients of the total sample.

Furthermore, the index can be improved by including more variables that indicate hoped-for value and value-capturing potential, which could be a key to better describe the changes of quality over time. Of course, an updated dataset will bring new insights into price changes, but at the same time would help to improve the model valuation power and thus provide an even more reliable urban land price index.

References

- Andersson, E David. 2000. "Hypothesis Testing in Hedonic Price Estimation – On the Selection of Independent Variables." *The Annals of Regional Science* 34 (2): 293–304.
- Anselin, Luc, Anil K Bera, Raymond Florax, and Mann J Yoon. 1996. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26 (1): 77–104.
- Anselin, Luc, and Nancy Lozano-Gracia. 2008. "Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality." *Empirical Economics* 34 (1).

- Springer: 5–34.
- Anselin, Luc, Ibnu Syabri, and Youngihn Kho. 2006. “GeoDa: An Introduction to Spatial Data Analysis.” *Geographical Analysis* 38 (1). Wiley Online Library: 5–22.
- Bostic, Raphael W, Stanley D Longhofer, and Christian L Redfearn. 2007. “Land Leverage: Decomposing Home Price Dynamics.” *Real Estate Economics* 35 (2). Wiley Online Library: 183–208.
- Burrough, P A. 1989. “Fuzzy Mathematical Methods for Soil Survey and Land Evaluation.” *Journal of Soil Science* 40 (3):477–92.
- Clapp, John M. 1990. “A Methodology for Constructing Vacant Land Price Indices.” *Real Estate Economics* 18 (3). Wiley Online Library: 274–93.
- . 2003. “A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation.” *Journal of Real Estate Finance and Economics* 27 (3):303–20.
- Colwell, Peter F, and Gene Dilmore. 1999. “Who Was First? An Examination of an Early Hedonic Study.” *Land Economics* 75 (4). [Board of Regents of the University of Wisconsin System, University of Wisconsin Press]: 620–26.
- Costello, Greg. 2014. “Land Price Dynamics in a Large Australian Urban Housing Market.” *International Journal of Housing Markets and Analysis* 7 (1). Emerald:42–60.
- Davis, Morris A, and Jonathan Heathcote. 2007. “The Price and Quantity of Residential Land in the United States.” *Journal of Monetary Economics* 54 (8). Elsevier: 2595–2620.
- Demetriou, Demetris. 2016. “A Spatially Based Artificial Neural Network Mass Valuation Model for Land Consolidation.” *Environment and Planning B: Planning and Design* 44 (5): 864-883
- Demetriou, Demetris. 2018. “Automating the Land Valuation Process Carried out in Land Consolidation Schemes.” *Land Use Policy* 75 (June): 21–32.
- Diewert, W Erwin, Jan de Haan, and Rens Hendriks. 2015. “Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components.” *Econometric Reviews* 34 (1–2). Taylor & Francis: 106–26.

- Din, Allan, Martin Hoesli, and Andre Bender. 2001. "Environmental Variables and Real Estate Prices." *Urban Studies* 38 (11). SAGE Publications Ltd: 1989–2000.
- Elhorst, J Paul. 2010. "Applied Spatial Econometrics: Raising the Bar." *Spatial Economic Analysis* 5 (1). Routledge: 9–28.
- ESRI ArcGIS. 2011. "Release 10." Redlands, CA: Environmental Systems Research Institute.
- Eurostat. 2013. "Handbook on Residential Property Prices Indices." Edited by Jan de Han and W Erwin Diewert. Luxembourg: Publications Office of the European Union.
- Francke, Marc K, and Alex M van de Minne. 2016. "Land, Structure and Depreciation." *Real Estate Economics* 45 (2): 415–51.
- French, Nick. 2001. "Decision Theory and Real Estate Investment: An Analysis of the Decision-Making Processes of Real Estate Investment Fund Managers." *Managerial and Decision Economics* 22 (7). John Wiley & Sons: 399–410.
- . 2013. "The Discounted Cash Flow Model for Property Valuations: Quarterly Cash Flows." *Journal of Property Investment & Finance* 31 (2): 208–12.
- Garza, Nestor, and Colin Lizieri. 2016. "A Spatial-Temporal Assessment of the Land Value Development Tax." *Land Use Policy* 50: 449–60.
- Glaesener, Marie-Line, and Geoffrey Caruso. 2015. "Neighborhood Green and Services Diversity Effects on Land Prices: Evidence from a Multilevel Hedonic Analysis in Luxembourg." *Landscape and Urban Planning* 143:100–111.
- Glumac, B., Q. Han, W. Schaefer, and E. van der Krabben. 2015. "Negotiation Issues in Forming Public-Private Partnerships for Brownfield Redevelopment: Applying a Game Theoretical Experiment." *Land Use Policy* 47: 66–77.
- Goh, Yen Min, Greg Costello, and Greg Schwann. 2012. "Accuracy and Robustness of House Price Index Methods." *Housing Studies* 27 (5). Routledge: 643–66.
- Gravelius, Harry. 1914. *Flusskunde*. Vol. 1. GJ göschen.
- Grigsby, William. 1986. "The Dynamics of Neighborhood Change and Decline." Philadelphia, PA: University of Pennsylvania Press.

- Gwartney, Ted. 1970. "A Computerized Assessment Program." *The Assessment of Land Value*, 125–42.
- Haughwout, Andrew, James Orr, and David Bedoll. 2008. "The Price of Land in the New York Metropolitan Area."
- Hill, Robert J. 2013. "Hedonic Price Indexes For Residential Housing: A Survey, Evaluation and Taxonomy." *Journal of Economic Surveys* 27 (5): 879–914.
- Holland, D M. 1970. "The assessment of land value." In , xiii + 292 pp. Madison, Wis., USA: Univ. Wis. Press.
- Hui, E C M, V S M Ho, and D K H Ho. 2004. "Land Value Capture Mechanisms in Hong Kong and Singapore: A Comparative Analysis." *Journal of Property Investment and Finance* 22: 76–100.
- Hui, Eddie Chi-man, Barbara Yuk-ping Leung, and Ka-hung Yu. 2014. "The Impact of Different Land-Supplying Channels on the Supply of Housing." *Land Use Policy* 39: 244–53.
- IAAO. 2003. "Standard on Automated Valuation Models (AVMs)." Chicago, IL, USA: International Association of Assessing Officers.
- Jones, M, and R G Reed. 2018. "Open Space Amenities and Residential Land Use: An Australian Perspective." *Land Use Policy* 75 (June): 1–10.
- Lamboray, Claude. 2010. "Un Indice Des Prix Hedonique Des Appartements." *Economie et Statistiques*. Vol. 44. Luxembourg: STATEC.
- Lancaster, Kelvin J. 1966a. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2). University of Chicago Press: 132–57.
- . 1966b. "Change and Innovation in the Technology of Consumption." *The American Economic Review* 56 (1/2). American Economic Association: 14–23.
- LeSage, James P, and R Kelley Pace. 2009. *Introduction to Spatial Econometrics (Statistics, Textbooks and Monographs)*. CRC Press.
- Liu, Zheng, Pengfei Wang, and Tao Zha. 2013. "Land-price Dynamics and Macroeconomic

- Fluctuations.” *Econometrica* 81 (3). Wiley Online Library:1147–84.
- Matsuo, Toshiyuki, Shu Wakasaki, and 若崎周. 1999. Land evaluation system and method, and medium storing land evaluation program, issued 1999.
- Moore, J Wayne. 2009. “A History of Appraisal Theory and Practice Looking Back from IAAO’s 75th Year.” *Journal of Property Tax Assessment & Administration* 6 (3). International Association of Assessing Officers: 23–49.
- Mulley, Corinne. 2013. “Accessibility and Residential Land Value Uplift: Identifying Spatial Variations in the Accessibility Impacts of a Bus Transitway.” *Urban Studies* 51 (8). SAGE Publications Ltd: 1707–24.
- Nichols, Joseph B, Stephen D Oliner, and Michael R Mulhall. 2013. “Swings in Commercial and Residential Land Prices in the United States.” *Journal of Urban Economics* 73 (1): 57–76.
- Northam, Ray M. 1971. “Vacant Urban Land in the American City.” *Land Economics* 47 (4). [Board of Regents of the University of Wisconsin System, University of Wisconsin Press]: 345–55.
- Osland, Liv. 2010. “Spatial Variation in Job Accessibility and Gender: An Intraregional Analysis Using Hedonic House-Price Estimation.” *Environment and Planning A: Economy and Space* 42 (9). SAGE Publications Ltd: 2220–37.
- Preclin, V. 2015. “L’indice Des Prix Des Maisons Anciennes.” *Economie et Statistiques*. Luxembourg: STATEC.
- Rosen, Sherwin. 1974. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” *Journal of Political Economy* 82 (1). University of Chicago Press: 34–55.
- Rosiers, François Des, and Marius Thériault. 2008. “Property Appraisal and the Measurement of Urban Externalities – Some Methodological Issues.” In *Mass Appraisal Methods – An International Perspective for Property Valuers*, edited by Tom Kauko and Maurizio d’Amato, 111–47. Wiley-Blackwell (UK).
- Rosiers, François Des, Marius Thériault, and Université Laval. *Faculté des sciences de*

- l'administration. Direction de la Recherche. 1999. "House Prices and Spatial Dependence: Towards an Integrated Procedure to Model Neighborhood Dynamics." Québec: Faculté des sciences de l'administration de l'Université Laval, Direction de la recherche.
- Samsura, D A A, E van der Krabben, and A M A van Deemen. 2010. "A Game Theory Approach to the Analysis of Land and Property Development Processes." *Land Use Policy* 27 (2): 564–78.
- Shimizu, Chihiro, and Kiyohiko Nishimura. 2006. "Biases in Appraisal Land Price Information: The Case of Japan." *Journal of Property Investment & Finance* 24 (2). Emerald Group Publishing Limited: 150–75.
- Soguel, Nils, Alexandre Tangerini, and Jacques Pictet. 2007. "How to Measure Scope Variables when no Metrics Exist : Application to Landscape Quality Measurement and Hedonic Price Evaluation." *Revue d'économie politique* 117 (5). Paris: Dalloz: 827–41.
- Spinney, Jamie, Pavlos Kanaroglou, and Darren Scott. 2011. "Exploring Spatial Dynamics with Land Price Indexes." *Urban Studies* 48 (4): 719–35.
- Wade Brorsen, B, Damona Doye, and Kalyn B Neal. 2015. "Agricultural Land and the Small Parcel Size Premium Puzzle." *Land Economics* 91 (3):572–85.
- Wu, Yuzhe, Xiaoling Zhang, Martin Skitmore, Yan Song, and Eddie C M Hui. 2014. "Industrial Land Price and Its Impact on Urban Growth: A Chinese Case Study." *Land Use Policy* 36 (0):199–209.
- Yamazaki, Ritsuko. 2001. "Empirical Testing of Real Option Pricing Models Using Land Price Index in Japan." *Journal of Property Investment & Finance* 19 (1). MCB UP Ltd: 53–72.

