

Using Hedonics to Create Land and Structure Price Indexes for the Ottawa Condominium Market

Kate Burnett-Isaacs
Statistics Canada
May 21, 2015

Abstract: Statistics Canada is developing a New Condominium Apartment Price Index (NCAPI) to measure the price change of new condo units over time. This work is important because the existing house price index (New Housing Price Index) does not include condominium apartment buildings. One of the main requirements for this work is that separate land and structure price indexes be produced. An emerging field in hedonics for residential property price indexes is to use a non-linear regression model to create a breakdown of the land and structure components of a residential property price index. The methods for creating a land-structure split are just developing in the literature and apply only to single family homes. These techniques have not been adequately tested on condos. This paper investigates the suitability and feasibility of using one hedonic model to create a land, structure and total price index for the Ottawa, Canada condo apartment market between 2005 and 2009. There are a few practical difficulties in carrying out this study, namely associating the land of a condo building to the value of an individual condo unit and separating the interrelationship between land and structure components on the value of a condo property as a whole. The results of the hedonic model are volatile, raising concerns on the use of a non-linear model and potential multicollinearity between land and structure variables. Using a construction cost index as a variable in the proposed model does improve the stability of the resulting indexes, suggesting that multicollinearity does play a role in causing the volatility in the original results. However, the volatility attributed to the non-linear nature of the model and the skewness of the total index towards the land component brings into question the practical uses of the proposed hedonic model. This paper raises the benefits and limitations of using one hedonic model to create separate land and structure price indexes and the assumptions that have to be made in order to turn this theory into a reality.

1. Introduction

Currently, Statistics Canada produces the New Housing Price Index (NHPI), which measures changes in the selling prices of new houses where detailed specifications pertaining to each house remain the same between two consecutive periods. To calculate the index, price and quantity data are captured for specific house models until those models cease to exist. When a model is discontinued, a new model with similar characteristics is used as a replacement and then quality adjusted to match the discontinued model. This method is called the matched model method and is used to ensure that only pure price changes are measured, not price changes resulting from a change in quality characteristics. This method works well when house models are homogeneous, as is the case in Canada, where many new subdivisions are

developed outside of the city centre. The NHPI does not contain condominium apartments (condo) and so a New Condominium Apartment Price Index (NCAPI) is being developed by Statistics Canada to fill this gap in the coverage of new housing prices in Canada. Like the NHPI, the NCAPI will attempt to use a matched model approach to ensure that condo prices are of constant quality over time. However, the condo market is more heterogeneous than newly built homes, so it can be more difficult to keep a condo model constant over consecutive periods. Therefore, the match-model method may not be feasible. Instead, the use of a hedonic regression model could be an alternative approach to calculate an index using data on the price and characteristics of the condo unit.

One of the main user requirements for both the NHPI and the NCAPI is the breakdown of the indexes into separate land and structure price indexes. The Consumer Price Index (CPI) requires this breakdown for their Shelter index: the structure component is used as a proxy for replacement costs of owner-occupied dwellings, and the total index is used for mortgage interest costs and commission fees. Residential property land and structure indexes are also required for the System of National Accounts. The National Economic Accounts Division (NEAD) Investment Accounts requires a structure only index for the deflation of residential assets; and NEAD Financial and Wealth Accounts uses both the structure and land components for the national balance sheet. The NCAPI and its land and structure price indexes would be the only source of such information for the residential housing market.

This study focuses on the Ottawa condominium apartment market to determine if suitable and feasible land and structure price indexes can be produced using one hedonic model. The suitability of the land and structure indexes is based primarily on the stability of the indexes and the comparison with the benchmark time dummy hedonic index. The feasibility of the land and structure indexes are based on the ability to systematically generate suitable price indexes on a regular basis. This paper is broken down in the following manner: section 2 covers the background of hedonics and its use in residential property price indexes; section 3 explains the data used in this study; section 4 explains the proposed methodology to create a land-structure split using one hedonic model; the results are summarized in section 5 and discussed in section 6.

2. Background

Ideally, it would be easy to decompose a house price into a land component and a structure component. The latter component can be viewed as the cost to build the structure itself. It has been suggested that a construction cost index can be used as a proxy for the structure component. The land component measures the impact that location and amenities, in addition to land size, have on the total price of a house (Davis and Palumbo, 2008). This breakdown becomes more complex when creating a land structure split for condos. Condo units not only share structural space with the condo building, such as a lobby and hallways, but they also share the land the building is built upon. To what extent these common areas can be attributed to the price of a condo unit is unclear. In addition, it is unclear if land size for a condo unit is a

two dimensional space – the plot of land the condo building sits on – or a three dimensional space – the plot of land and the view from a condo unit.

There is an extensive literature on calculating a total index for housing using hedonics – mostly for single family homes and to a lesser extent condos. An emerging field in hedonics for residential property price indexes is the method proposed for creating a breakdown of the land and structure components of a housing index. A hedonic model can be set up so it has a separate land and structure component. Potential multicollinearity between land and structure variables can make it difficult to truly separate the two components in a hedonic regression. A possible solution can be to use a construction cost index, like Statistics Canada's Apartment Building Construction Price Index (ABCPI), to remove that multicollinearity, resulting in a true land and structure value (Diewert and Shimizu, 2013). The methods to create this breakdown are only just developing in the literature and apply only to single family homes. These techniques have not been adequately tested on condos. A difficulty in creating this breakdown in a condo index is determining the relationship between the association of land for a condo building and the value of a condo unit. Previous studies on hedonic models for condos do not include land size as a variable in their regressions, for example Song and Wilhelmsson (2009). Therefore certain assumptions on the distribution of land per condo unit are made in this study.

3. Data

The source of data for condo prices and characteristics is a subset from a Canadian real estate board database. A research dataset was developed for new and resale condos in Ottawa, Canada for the 2005 – 2009 period. High rise condos, which are defined as those condo buildings with over four floors, are the focus of this study because of the potential use of the ABCPI as a proxy for high rise condo building structure costs. The dataset contains the condo characteristic and postal code location variables necessary for hedonics. However, the dataset does not contain a variable for land size. This complicates the conceptual discussion regarding building a land index. However, given the absence of land size as a variable in other condo hedonic models, it is assumed that land size is not a critical factor in modeling a condo price. For the purposes of this study, it is assumed that the main driver of condo land prices is the location, captured by 3 digit postal code dummy variables. The data is provided monthly, but since the ABCPI data is only available quarterly, the index is calculated on a quarterly basis with 99 to 308 observations per quarter. The ABCPI is produced internally to Statistics Canada and final index values, rebased to quarter one 2005, are used in this study.

4. Methodology

This study uses the methodology proposed by Diewert and Shimizu (2013) and Diewert, de Haan and Hendriks (2011), which suggests that a hedonic regression can be constructed with separate land and structure components.

4.1 Basic Model

The basic model is a non-linear regression equation estimated for 20 quarters between 2005 and 2009. The base period for the index is quarter one, 2005. The model includes the variables selling price of the condo (P_i^t), age of the building (A_i^t), condo living area (S_i^t), number of bedrooms (R_i^t), and 3 digit postal code dummy variables (D_{ij}^t). This model is estimated by Ordinary Least Squares.

The hedonic model is as follows:

$$(1) \quad P_i^t = \alpha^t (\sum_{j=1}^J \omega_j^t D_{ij}^t) + \beta^t (1 - \delta A_i^t) (1 + \gamma R_i^t) S_i^t + \varepsilon_i^t$$

The land component is:

$$(2) \quad \alpha^t (\sum_{j=1}^J \omega_j^t D_{ij}^t)$$

Where α^t is the price for the land component.

The structure component is:

$$(3) \quad \beta^t (1 - \delta A_i^t) (1 + \gamma R_i^t) S_i^t$$

Where β^t is the price per square foot of living area.

The land price index is defined as

$$(4) \quad P_L^t = \frac{\hat{\alpha}^t}{\hat{\alpha}^0}$$

The land quantity is defined as

$$(5) \quad Q_L^t = \alpha^t \sum_{i=1}^N (\sum_{j=1}^J \omega_j^t D_{ij}^t)$$

and the structure price index is defined as

$$(6) \quad P_S^t = \frac{\hat{\beta}^t}{\hat{\beta}^0}$$

The structure quantity is defined as

$$(7) \quad Q_S^t = \beta^t (\sum_{i=1}^N (1 - \delta A_i^t) (1 + \gamma R_i^t) S_i^t)$$

A total price index is created using the Laspeyres formula. The Laspeyres formula is used because it is the same method used to calculate NHPI and the matched model NCAPI.

$$(8) \quad I^t = \frac{P_L^t Q_L^0 + P_S^t Q_S^0}{P_L^0 Q_L^0 + P_S^0 Q_S^0} \times 100$$

4.2 Model Using the ABCPI

In past studies of the hedonic land-structure split, multicollinearity between the land and structure variables has warranted the use of a construction cost index as a variable in the hedonic model (Eurostat, 2013). Multicollinearity occurs when two or more variables are correlated with each other. This can cause unstable estimates and produce incorrect signs or magnitudes for the estimates (Greene, 2003). In this study of condos, the ABCPI is used based on the assumption that the movement of condo unit construction costs are the same as those for apartment buildings. This assumption is based on the grounds that increasingly, apartment buildings are being constructed with similar finishes as condos.

The model using the ABCPI is as follows:

$$(9) \quad P_i^t = \alpha^t (\sum_{j=1}^J \omega_j^t D_{ij}^t) + \beta^0 ABCPI^t (1 - \delta A_i^t) (1 + \gamma R_i^t) S_i^t + \varepsilon_i^t$$

The land price index and quantity will remain the same as in the original model, however the structure price index is now defined as

$$(10) \quad P_S^t = \frac{\beta^0 ABCPI^t}{\beta^0 ABCPI^0} = \frac{ABCPI^t}{ABCPI^0}$$

And the structure quantity is defined as

$$(11) \quad Q_S^t = \beta^0 ABCPI^t (\sum_{i=1}^N (1 - \delta A_i^t) (1 + \gamma R_i^t) S_i^t)$$

The total price index is again calculated using a Laspeyres formula.

5. Results

5.1 The Benchmark Model

The time dummy model is the staple method in the literature to create a hedonic index. The time dummy model is a fixed effects model that allows the characteristic variables to be constant over time. A dummy variable is attributed to each time period, where the base group is the starting period. The index then becomes the exponential of the time dummy coefficient estimates, multiplied by 100. In this case, the time dummy model is as follows:

$$(12) \quad \ln P_i = \alpha + \sum_{i=1}^{26} \omega_i PostalCode + \beta_1 Livarea_i + \beta_2 Bedroom_i + \sum_{t=1}^{20} \delta_t Time$$

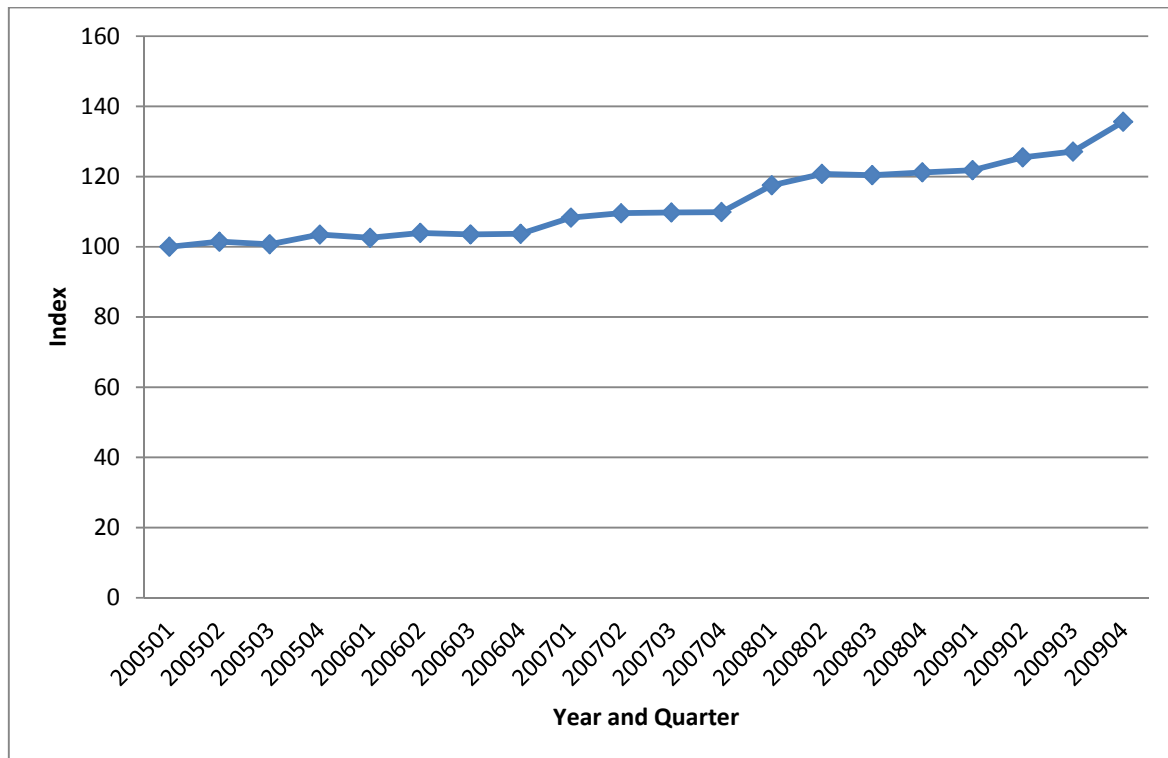
The index is calculated by:

$$(13) \quad I^t = \exp(\hat{\delta}_t) \times 100$$

The downside of the time dummy model is that it cannot be broken down into land and structure components. However, the time dummy model gives a good benchmark for analyzing the suitability of the proposed basic and ABCPI models. In addition, the time dummy model is also useful to analyze the quality of the data. The dataset used in the time dummy model

contains 3820 observations and as shown in Figure 1, the time dummy index does not present any indication of significant outliers. However, the time dummy model has an R^2 of 0.6211, suggesting that improvements can be made in explaining the data.

Figure 1: Total Condo Time Dummy Variable Index



5.2 The Basic Model

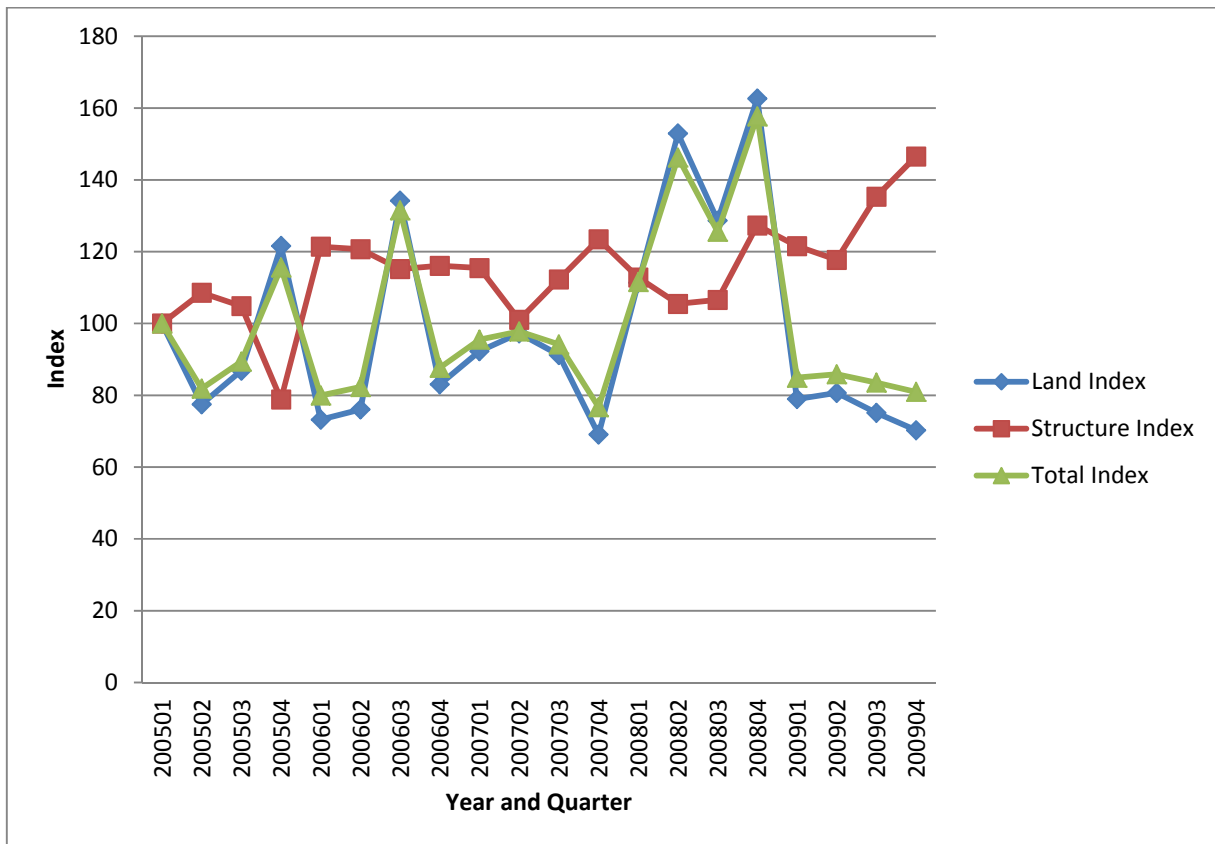
The basic model is the first step in creating a land-structure split from one hedonic model. This model contains a constant depreciation rate (δ) and a constant number of bedrooms (γ) coefficient, set at 0.006 and -0.025, respectively. These coefficients were determined by running the basic model using the entire five years of data. The model now appears as follows:

$$(14) \quad P_i^t = \alpha^t (\sum_{j=1}^J \omega_j^t D_{ij}^t) + \beta^t (1 - 0.006A_i^t)(1 - 0.025R_i^t)S_i^t + \varepsilon_i^t$$

The model is estimated for each quarter between 2005 and 2009, resulting in land and structure coefficient estimates for each of the 20 time periods. It is important to note that the only coefficient deemed significant and unbiased in the estimation of this model is β^t . Analysis will be pursued even if results are biased.

From the coefficient estimates of alpha and beta, the land, structure and total indexes are created. The coefficient estimates are summarized in Appendix Table 2 and the indexes are illustrated in Figure 2 and summarized in Table 3.

Figure 2: The Basic Model Land, Structure and Total Indexes



The resulting indexes of the basic model are volatile, particularly the land index. There are two possible explanations for this volatility. Firstly, non-linear models are often unstable and the estimates of alpha swing dramatically depending on outliers present in the data and the postal code dummy variables included in the model. The estimates of beta are less volatile but can still have large movements between time periods. Secondly, multicollinearity can be a cause for the volatility in the coefficient estimates over the 20 quarters. Multicollinearity can cause large movements in the results given even small changes in the data and as shown in Figure 2, the results of the basic model vary significantly from one time period to another. Since the land component of this model is expressed as postal code dummy variables, which have a binary distribution and not a normal, continuous, distribution, it is not possible to test the multicollinearity between land and structure variables. Therefore the extent to which multicollinearity is a cause for the volatility seen in the results of the basic model can only be speculated.

As illustrated in Figure 2, the total index is driven by the land index because of the large weight of the land component. The weighting is skewed due to the large difference in the magnitudes of the coefficients alpha and beta (see Appendix Table 2). The weights of the land and structure components are a product of their coefficient estimates and the sum of their respective variables, resulting in large differences, as shown in Table 4.

5.2 The ABCPI Model

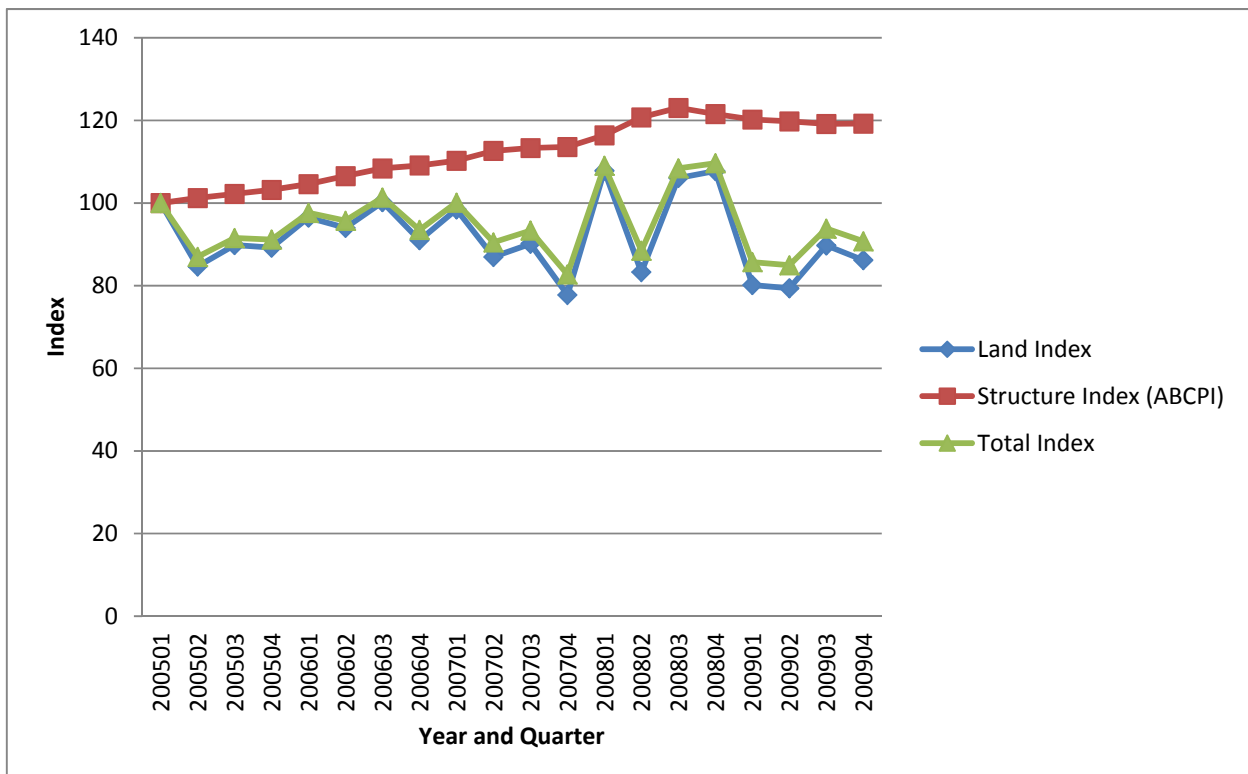
As seen in the basic model, the resulting indexes are volatile, possibly caused by multicollinearity between the land and structure variables. In order to eliminate any multicollinearity, the ABCPI is included in the model as a function of the structure component. The beta coefficient is now held constant at β^0 , more precisely 0.241, and the model appears as follows:

$$(15) \quad P_i^t = \alpha^t (\sum_{j=1}^J \omega_j^t D_{ij}^t) + 0.241(ABCPI^t)(1 - 0.006A_i^t)(1 - 0.025R_i^t)S_i^t + \varepsilon_i^t$$

In this model, the construction cost index is now assumed to be the structure index. The assumption has to be made that the price movements of condo unit structures are the same as the price movements of apartment building structures. The use of a construction cost index as a proxy for a structure index brings into question the relationship between supply side costs and demand side, market driven prices. The assumption in this study is that structure prices are driven by supply side forces and land prices are driven by demand side forces (Davis and Heathcote, 2007).

The ABCPI model is re-estimated for each quarter and the estimates of alpha are summarized in Appendix Table 2 and the index results are illustrated in Figure 3 and summarized in Table 3.

Figure 3: The ABCPI Model Land, Structure and Total Index

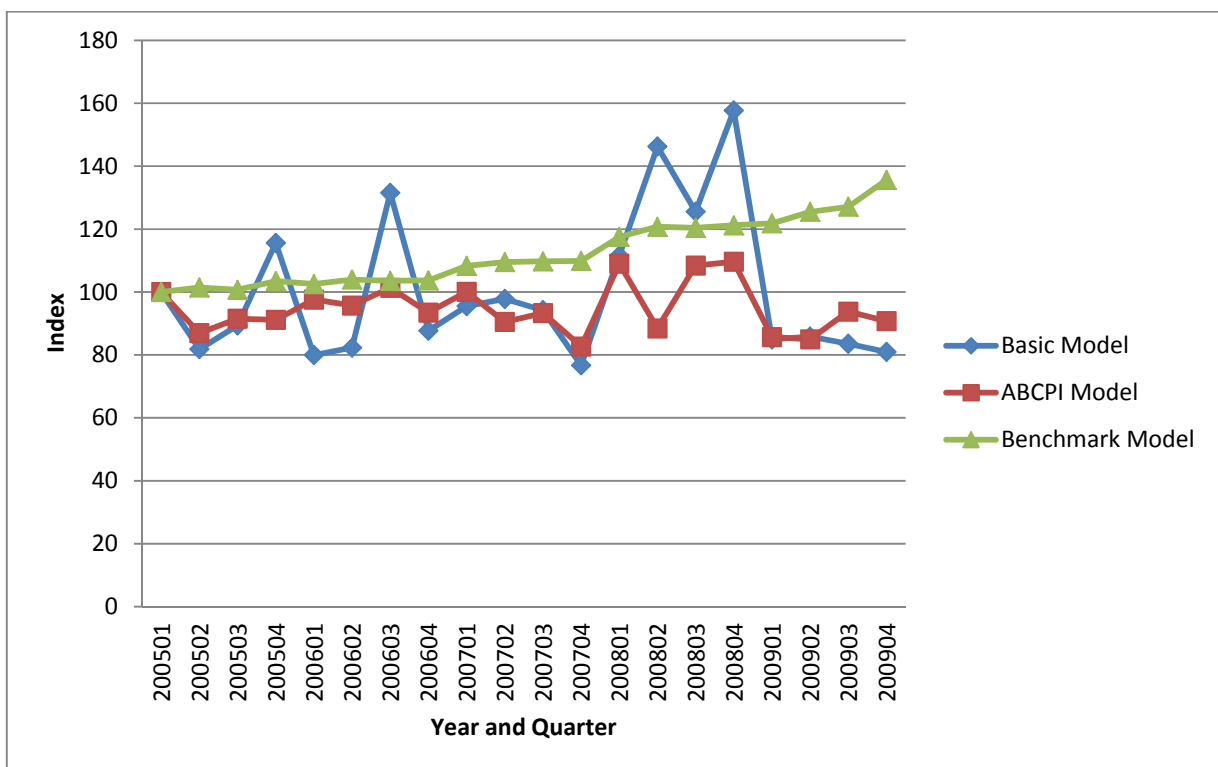


As shown in Figure 3, the ABCPI model results are more stable. Though, given the lack of a land size variable, it is not possible to test the correlation between land and structure variables, the more stable ABCPI model index results suggest that multicollinearity is an attributing factor to the instability in the basic model. Like in the case of the basic model, the total index is driven by its land component due to the relatively large weighting of the land values.

6. Discussion

The purpose of this study is to determine if using one hedonic model to create a land-structure split is a suitable and feasible mechanism for use in the NCAPI. The suitability of the land, structure and total condo price indexes is based on the appropriateness of the results in representing accurate price movements. In order to determine what is an accurate price movement, the total indexes of the two models, the basic model and the ABCPI model, are compared to the benchmark, time dummy model, shown in Figure 4.

Figure 4: Comparison with Benchmark Model



Both the basic and the ABCPI model indexes are more volatile than the benchmark index. Given the unstable nature of non-linear regression models, these results are not surprising. However, the fact that the total basic and ABCPI model indexes are driven by their respective land indexes may be skewing the true results. Namely, these three indexes do not follow the same overall trends. The benchmark model has a clear positive trend. The basic model, though less clear, has a positive trend. However, the ABCPI model has a slight negative trend. Given the

positive slope of the construction cost index itself, if the land and structure components were more proportionally weighted, the ABCPI model would likely match the positive trends seen in the other two models. However, this discrepancy does raise concerns on which trend is correct and which model best represents the data.

The significant difference in the magnitudes of the coefficient estimates for alpha and beta are the primary drivers of the weight differences between the land and structure components. Not including land size as a variable in either model may be a main cause of this difference. Using land size in the hedonic model may balance the magnitudes of the coefficient estimates for land and structure, creating more appropriate total indexes, by removing any omitted variable bias attributed to land size. Having land size as a normally distributed, continuous, variable included in the model will also help pin point multicollinearity problems between the land and structure components and help validate the use of the ABCPI model.

Investigations have been made into attaining land size data for condos. However, land size data for condos is seldom reported. Even for those cases where land size is reported, to include them in a hedonic model would entail assumptions about the distribution of land for a condo unit. As previously mentioned, consideration needs to be made when attributing land size to one condo unit, particularly concerning whether land for a condo unit is a two dimensional or three dimensional space. Whether land size is included in the model or not, assumptions will have to be made when using the proposed models and users of the condo index need to be confident in those assumptions.

The model using the ABCPI is a more stable model and produces more realistic price movements. However, the use of this model depends on the assumption that the ABCPI is an appropriate proxy for the structure component of condos. Traditionally, condos tended to have higher finishes and better quality inputs than apartment buildings; however there is an increasing trend to build apartment buildings with the same type of finishes as a condo building. The proposed methodology uses the ABCPI movement, not the absolute value of prices, making it a reasonable proxy for condo building construction cost movements. Another concern is that the ABCPI measures the construction costs of an entire building and not of a single unit. Again, assumptions will have to be made upon the allocation of construction costs between a residential building and a single unit.

The feasibility of using one of the proposed hedonic models in the production of a condo index must also be assessed. Producing an index requires reliability in processing inputs and in analysing the results. The hedonic model must produce replicable results and the price movements must be explainable. The direction of some index movements between the basic and ABCPI models are opposing, for unexplainable reasons, such as in the cases of quarter 1, 2006 and quarter 2, 2008 (see Figure 4). The process must also be replicable across modellers. The instability of both proposed non-linear models poses problems in the consistent production of land, structure and total indexes. The results can change dramatically depending on the postal code dummy variables included. In some cases, negative or extremely high values are attained with the same variable make up as other time periods. This brings subjectivity into the process. Though as much consistency as possible is kept across the model time periods,

adjustments in the postal code dummy variables included are made, in this study, based on the results. Not only are judgements made on what postal code dummy variables to include, but judgement is also made on what constitutes a reasonable result for alpha and beta. A basis for using hedonics is that it reduces subjectivity from the quality adjustment process. However, the process to stabilize either the basic model or the ABCPI model requires judgement from the econometric modeller. Therefore, the proposed methods to calculate land and structure indexes will be difficult to systematically implement and provide consistent results over time.

The volatility in the results and the subjectivity of the hedonic model limit the suitability and feasibility of using these hedonic models in practice. However, this work has brought to light useful tools and appropriate assumptions that can be made in creating condo land and structure price indexes. The matched model approach can produce a sound total condo price index, however there are limitations to its use in deconstructing a land and structure price index. The notion used in this study that a total condo price index can be decomposed into land and structure components can be carried over to implicitly derive the land price index. The next steps to this investigation include collecting data on common space (land and structure) characteristics and prices to use in land and structure index calculation. Consultations with builders have indicated that reporting land cost or land value can be difficult. Therefore, research continues to find suitable data and variables to derive a land price index for condos.

7. Conclusion

Given the heterogeneity of the condo market, the match model approach may not be feasible in creating a condo price index of constant quality. Hedonics is an attractive method to calculate this component of a residential property price index. Land and structure price indexes are a requirement for the NCAPI, however, the methods of calculating these indexes using hedonics has not been fully explored in the literature. This study investigates the suitability and feasibility of using one hedonic model to create a land, structure and total price index. In practice, the volatility of the basic model raises concerns on the use of a non-linear model and potential multicollinearity between land and structure variables. Using the ABCPI as a variable in the proposed model does improve the stability of the resulting indexes, suggesting that multicollinearity does play a role in the volatility of the basic model indexes. The lack of a land size variable does cause concerns in the results. The total index is driven by its land component due to the large discrepancies in the magnitudes of alpha and beta. The skewed total indexes do not have consistent trends compared with each other and the benchmark model, raising concerns about the ability for the ABCPI model to represent the data. Overall, this investigation has raised the benefits and concerns of using one hedonic model to create an index and the assumptions that have to be made in order to turn this theory into a reality. Using the lessons

learned in this investigation, research continues to find appropriate data in order to calculate land and structure indexes.

8. Resources

Davis, M., and M. Palumbo. (2008). 'The price of residential land in large US cities'. *Journal of Urban Economics* 63, 352-384.

Davis, M., and J. Heathcote. (2007). 'The Price and Quantity of Residential Land in the United States'. *Journal of Monetary Economics*, 52:8, 2595-2620.

Diewert, W.E., J. de Haan, and R. Hendriks. (2011). 'The Decomposition of a House Price Index into Land and Structure Components: A hedonic Regression Approach'. *The Valuation Journal*, 6, 58-106.

Diewert, W.E., and C. Shimizu, (2013). 'Residential Property Price Indexes for Tokyo'. *Working Copy*.

Eurostat. *Handbook on Residential Property Prices Indices (RPPIs)*. Luxembourg: Publications Office of the European Union.

Greene, W. H. (2003). *Econometric Analysis*, 5th ed. Upper Saddle River: Pearson Education, Inc.

Song, H.S., and M. Wilhelmsson. (2010). 'Improved price index for condominiums'. *Journal of Property Research*, 27:1, 39-60.

9. Appendix

Table 1: Summary statistics for refined data

Year	Quarter	Mean Price (\$000s)	Standard Deviation	Sample Size
2005	1	193.71	45.42	165
	2	198.26	52.33	218
	3	189.07	51.04	180
	4	191.31	50.29	112
2006	1	189.15	57.87	172
	2	200.08	58.53	255
	3	200.62	59.94	198
	4	201.94	58.03	150
2007	1	219.43	67.53	193
	2	210.49	62.24	290
	3	202.30	56.74	231
	4	213.26	66.41	147
2008	1	223.48	63.94	184
	2	230.21	66.15	263
	3	221.39	62.78	183
	4	240.88	75.91	104
2009	1	237.34	73.66	157
	2	247.09	74.02	308
	3	221.97	63.64	186
	4	262.62	77.31	99

Table 2: Beta and alpha coefficient estimates per model

Year and Quarter	Basic Model		ABCPI Model
	α^t	β^t	α^t
200501	5.95	0.241	11.90
200502	4.61	0.262	10.07
200503	5.17	0.253	10.69
200504	7.23	0.190	10.62
200601	4.36	0.293	11.48
200602	4.53	0.291	11.18
200603	7.98	0.278	11.93
200604	4.94	0.280	10.83
200701	5.49	0.278	11.72
200702	5.79	0.244	10.35
200703	5.43	0.271	10.72
200704	4.11	0.298	9.25
200801	6.62	0.272	12.83
200802	9.10	0.254	9.91
200803	7.66	0.257	12.62
200804	9.68	0.307	12.82
200901	4.70	0.293	9.54
200902	4.80	0.284	9.44
200903	4.47	0.326	10.68
200904	4.18	0.353	10.26

Table 3: Model land, structure and total index values

Year and Quarter	Basic Model			ABCPI Model		
	Land Index	Structure Index	Total Index	Land Index	Structure Index	Total Index
200501	100.0	100.0	100.0	100.0	100.0	100.0
200502	77.5	108.6	81.9	84.7	101.2	87.0
200503	86.9	104.9	89.4	89.8	102.2	91.5
200504	121.6	78.8	115.6	89.2	103.2	91.1
200601	73.2	121.4	80.0	96.5	104.6	97.6
200602	76.1	120.7	82.3	94.0	106.5	95.7
200603	134.2	115.2	131.5	100.2	108.4	101.3
200604	83.0	116.1	87.7	91.0	109.1	93.5
200701	92.3	115.4	95.5	98.5	110.2	100.1
200702	97.3	101.0	97.8	86.9	112.6	90.5
200703	91.3	112.3	94.2	90.1	113.3	93.3
200704	69.1	123.5	76.7	77.8	113.6	82.7
200801	111.3	112.8	111.6	107.8	116.4	109.0
200802	152.9	105.5	146.3	83.3	120.7	88.5
200803	128.7	106.6	125.6	106.1	123.0	108.4
200804	162.7	127.3	157.7	107.7	121.5	109.6
200901	79.0	121.6	84.9	80.1	120.2	85.7
200902	80.7	117.7	85.9	79.4	119.8	85.0
200903	75.1	135.3	83.6	89.7	119.2	93.8
200904	70.3	146.5	81.0	86.2	119.2	90.8

Table 4: Land and structure weights

Basic Model		ABCPI Model	
Land Quantity	Structure Quantity	Land Quantity	Structure Quantity
141431.31	23065.58	143676.79	23065.58