Do Housing Submarkets Really Matter ?

BOURASSA, Steven C., HOESLI, Martin E., PENG, Vincent S.

Abstract

We maintain that the appropriate definition of submarkets depends on the use to which they will be put. For mass appraisal purposes, submarkets should be defined so that the accuracy of hedonic predictions will be optimized. Thus we test whether out-of-sample hedonic value predictions can be improved when a large urban housing market is divided into submarkets and we explore the effects of alternative definitions of submarkets on the accuracy of predictions. We compare a set of submarkets based on small geographical areas defined by real estate appraisers with a set of statistically generated submarkets consisting of dwellings that are similar but not necessarily contiguous. The empirical analysis uses a transactions database from Auckland, New Zealand. Price predictions are found to be most accurate when based on the housing market segmentation used by appraisers. We conclude that housing submarkets matter, and location plays the major role in explaining why they matter.

Do Housing Submarkets Really Matter?*

Steven C. Bourassa†

School of Urban and Public Affairs, University of Louisville, 426 W. Bloom Street, Louisville, Kentucky 40208

Martin Hoesli

HEC, University of Geneva, 40 boulevard du Pont-d'Arve, CH-1211 Geneva 4, Switzerland, and University of Aberdeen Business School, Edward Wright Building, Aberdeen AB24 3QY, Scotland

and

Vincent S. Peng

AMP Henderson Global Investors, 50 Bridge Street, Sydney, NSW 2000, Australia

This draft: November 8, 2002

* We thank Greg Schwann for providing the transactions data and Tom Thibodeau and an anonymous reviewer for helpful suggestions.

† To whom correspondence should be addressed.

Do Housing Submarkets Really Matter?

Abstract

We maintain that the appropriate definition of submarkets depends on the use to which they will be put. For mass appraisal purposes, submarkets should be defined so that the accuracy of hedonic predictions will be optimized. Thus we test whether out-of-sample hedonic value predictions can be improved when a large urban housing market is divided into submarkets and we explore the effects of alternative definitions of submarkets on the accuracy of predictions. We compare a set of submarkets based on small geographical areas defined by real estate appraisers with a set of statistically generated submarkets consisting of dwellings that are similar but not necessarily contiguous. The empirical analysis uses a transactions database from Auckland, New Zealand. Price predictions are found to be most accurate when based on the housing market segmentation used by appraisers. We conclude that housing submarkets matter, and location plays the major role in explaining why they matter.

1. INTRODUCTION

There is some agreement that housing markets are composed of a set of submarkets. Such submarkets are usually defined in terms of geographical areas or the physical characteristics of the dwellings. When spatial dimensions are used, housing market segmentation can rely on pre-existing geographic or political boundaries (Schnare and Struyk, 1976; Goodman and Kawai, 1982; Adair *et al*., 1996) or spatial partitions based on socio-economic or environmental characteristics (Schnare, 1980; Galster, 1987; Harsman and Quigley, 1995). Another way of delineating submarkets in spatial terms is offered by Palm (1978). She argues that information constraints and search costs may segment an urban housing market into different submarkets. Thus, Palm (1976) and Michaels and Smith (1990) investigate submarkets delineated by real estate agents. The use of physical characteristics of dwellings to define housing submarkets has been on the basis of the number of rooms (Schnare and Struyk, 1976), lot and floor area (Bajic, 1985), or the type of property, such as detached versus attached (Allen *et al*., 1995; Adair *et al*., 1996).

Some researchers have used statistical techniques to define housing submarkets. Dale-Johnson (1982) uses factor analysis on 13 variables, and extracts five factors that are used to define 10 submarkets. Maclennan and Tu (1996) investigate the structure of housing submarkets in Glasgow. They use principal component analysis to identify the individual variables that explain the highest proportion of the variation in the data. These variables are then used as the basis for cluster analysis, which in turn defines their submarkets. Goodman and Thibodeau (1998) use hierarchical methods to define submarkets in a study that focuses on the role of school districts in Dallas. Bourassa *et al*. (1999) use principal component analysis and cluster analysis to form housing submarkets for Sydney and Melbourne, Australia.

The structural stability across hedonic equations for assumed or statistically defined submarkets is tested in several studies (Dale-Johnson, 1982; Rothenberg *et al.*, 1991; Bourassa *et al.*, 1999), but the focus on the prices of characteristics (as is the case when hedonic models are used) rather than the characteristics themselves may not be the best way to test for submarkets. If the focus is on the practical applications of housing submarkets, however, then the hedonic approach is appropriate. This paper focuses on such applications. More specifically, we examine whether the recognition of submarkets can improve the reliability of out-of-sample hedonic predictions for a large housing market.

Automated valuation models, usually multiple regression or hedonic models, are increasingly being implemented for mass appraisal and mortgage underwriting (Mark and Goldberg, 1988; Hamilton, 1998; Pace and Gilley, 1989).[1] These methods permit a rapid, cost-effective and objective valuation of a property or a portfolio of properties and are being used in several countries (such as the U.S., the U.K., Canada, the Netherlands, and Switzerland). The multiple regression models that are used should provide price estimates that are as accurate as possible. In this context it is important to examine

---

[1] This is also gaining importance in commercial real estate markets (Crosson *et al*., 1996; Taylor *et al*., 2000).

whether a given housing market should be considered as a whole, or whether a market should be considered as a set of submarkets.

For the purpose of examining whether out-of-sample valuations are improved when submarkets are considered, we use data pertaining to dwelling sales in the city of Auckland, New Zealand, in 1996. Two sets of housing submarkets are considered. First, we use the "sales groups" defined by real estate appraisers. There are 34 sales groups consisting of small geographical areas. Second, we use principal component analysis and cluster analysis to define submarkets. Factors are extracted from the variables using principal component analysis. Factor scores are calculated and cluster analysis is applied to those scores to construct housing submarkets. Then we estimate hedonic equations for the city as a whole and for each submarket and we evaluate the accuracy of out-of-sample predictions. The analyses are performed both with and without controlling for spatial dependence.

The paper is organized as follows. We next present a conceptual framework for housing market segmentation (section 2). In sections 3 and 4, respectively, the data and method are discussed. Section 5 contains the results, while section 6 provides our conclusions.

2. A CONCEPTUAL FRAMEWORK FOR HOUSING MARKET SEGMENTATION

The typical approach to analyzing housing market segmentation involves estimating hedonic equations for various assumed or defined submarkets and then testing for structural stability across those equations (Rothenberg *et al.*, 1991). This method is based on a definition of submarket that relies on the concepts of substitutability and equilibrium. Substitutes are pairs of goods or services having the property that an increase in the price of one results in an increase in demand for the other. Pairs of goods or services with similar characteristics are more likely to be close substitutes than pairs with dissimilar characteristics. In equilibrium, prices of characteristics are assumed to equalize across substitutes. Prices are consistent within submarkets because submarkets contain close substitutes. However, it is possible that quite different dwellings could have similar hedonic functions yet not be substitutes. That is because the hedonic method focuses on the prices of characteristics rather than the existence or quantity of those characteristics.

A focus on hedonic prices is appropriate, however, if the aim is to segment housing markets for the purposes of automated valuation, either for property taxation or mortgage underwriting purposes. In this case, the market is divided into segments, and the sale prices and characteristics of the properties that did transact within a given segment are used to estimate hedonic equations, which in turn are used to estimate values for the properties that did not transact. The aim is not necessarily to define relatively homogeneous submarkets consisting of substitutable dwellings, but rather to segment the market in a way that allows for more accurate estimates of house values. These two goals may, in fact, be in conflict. For example, as a market is segmented into smaller and smaller (and more homogeneous) submarkets, the hedonic prices are estimated less precisely due to the inverse relationship between sample size and standard errors. Also, as a market is segmented into more homogeneous submarkets, variability in the hedonic characteristics will decrease and, consequently, some variables will drop out of the equation. However, if these equations are used to estimate prices for properties that did not transact, and there is more variability in the characteristics of those properties than in the characteristics of the properties that did transact, then the estimates will be inaccurate for some properties. This is a real problem because the number of properties that transact in a given time period is likely to be small relative to the number that do not transact, implying that the latter group will have more variability than the former. The general conclusion is that too much homogeneity may not be a good thing in practice.

In a recent paper, Bourassa *et al.* (1999) calculate weighted mean square errors from hedonic equations estimated for alternative definitions of submarkets. The weighted mean square errors are compared using an *F* test. This approach emphasizes the goodness-of-fit of the hedonic equations without attempting to assess how accurate they are for estimating values of properties that did not transact.[2] For various reasons, well-

---

[2] It has also been suggested that the *J* test could be used to compare alternative definitions of submarkets (Goodman and Dubin, 1990). Goodman and Dubin use the *J* test to compare a naïve north/south stratification with a more meaningful city/county

fitted hedonic equations may do a poor job of estimating values of properties that did not transact. For example, the properties that transacted may differ significantly from those that did not (Gatzlaff and Haurin, 1997, 1998).[3] Also, the definition of small, homogenous groups of properties may result in the problems described in the previous paragraph.

If the aim is to segment housing markets for the practical purpose of automated valuation, then the appropriate test should be the accuracy of out-of-sample estimates—that is, how precisely equations estimated for properties that did transact are able to value properties that did not. It is not possible to test this directly because the values of the properties that did not transact are unobservable. However, one way to do an approximate test would be to divide the sample of transactions into two parts using one set of properties to estimate hedonic equations that are then used to predict values for the remaining properties.

We hypothesize that a number of considerations will affect the predictive accuracy of hedonic equations estimated in this way. First, the method used to define submarkets will be important. We experiment with two definitions of submarkets, spatial and aspatial. If the well-known real estate dictum is correct and location is of prime importance, then geographically defined submarkets will perform best. Second, the definition of the sample used for estimation purposes will affect sample size and the variability of the data, both of which may affect the precision with which the hedonic equations are

---

stratification and find that the latter dominates the former. In most cases, however, neither of the alternatives is naïve and, consequently, neither classification is likely to dominate the other. As a simple test of this, we compared a detached versus attached dwelling stratification with two submarkets defined using principal component and cluster analysis. Neither stratification dominated the other. In this case, each contained significant information not contained in the other.

[3] Thus, in practice, it may be desirable to employ a sample selection bias correction technique such as that employed by Gatzlaff and Haurin. We do not need to use such a technique in the empirical analysis presented here because the sample of transactions is a random sample of our entire data set.

estimated. We experiment with two primary definitions of the sample: one includes transactions for all residential property types and the other includes transactions for detached dwellings only. Third, greater detail in the descriptive variables should improve the accuracy of predictions assuming that additional variables are not collinear with variables already included in the model. We test for this by expanding the list of variables for a subset of the detached dwelling sample.

3. DATA

The main source of data for this study is the official database of all real estate transactions in New Zealand. We use data pertaining to residential real estate, more specifically to "dwelling houses of a fully or semi-detached style situated on their own clearly defined piece of land" (hereafter referred to as *detached* dwellings) and "residential ownership home units which may be single or multi-story and which do not have the appearance of dwelling houses" (hereafter *attached* dwellings). We focus on sales for these two categories of residential properties in the city of Auckland in 1996. A total of 8,421 transactions were retained for the analysis: 5,716 sales of detached houses, and 2,705 sales of attached houses.[4] The database contains the date of sale, the sale price, and such information as: exact location, floor area, age, wall material and condition, and quality of the principal structure. The land area is provided for 65% of the detached dwellings, but none of the attached units. The detached units for which no land area is provided are generally "cross-leased", which means that the land is owned collectively by the owners of the dwellings on that site. The collective owners lease a fraction of the land to each individual owner for a "peppercorn", or nominal, rent.[5] Some attached units

---

[4] A sale was removed from the sample if it fell into one of the following categories: (a) the property had a land area larger than 0.25 hectares (this excluded properties that may have been sold primarily for redevelopment purposes); (b) the property had a floor area either less than 30 square meters (probably due to an error in data entry); or (c) the transaction was flagged as not being "arm's length".

[5] Cross-leasing is a simplified form of condominium ownership. There is no condominium association or dues or formal arrangement for maintenance of common

are also cross-leased, but most are "strata-titled"—that is, are condominiums. For all such cross-leased or strata-titled dwellings, we set land area equal to zero and set a dummy variable equal to one. For a large proportion of detached houses (4,880 properties), supplementary information used for mass appraisal purposes is available. These data include important characteristics such as water views, and the quality of landscaping and of the neighborhood. For the dependent variable in our hedonic models we use the sale price net of the value of any chattels (transformed as indicated below).

For the purposes of defining submarkets, we also use census data. The census is conducted in New Zealand every five years, and the data used in this study are from the census conducted in 1996, which is the year in which the transactions occurred. Each property in the database was assigned to a census "area unit" using a geographic information system (GIS). For each area unit, the following information was extracted and calculated: the densities of population and dwellings, homeownership rate, median household income, percentage of people receiving income support, average rooms per house, percentage unemployed, as well as ethnic composition.

The use of GIS also allowed the data set to be supplemented with the measure of the Euclidean distance between each property and the central business district (CBD). Dwellings have been geo-coded using street addresses that provide location with a general accuracy of plus or minus 20 meters.

Table 1 contains the means of the independent and dependent variables used in the regression analyses for the overall sample (Sample 1) of detached and attached properties (8,421 sales), for the full sample (Sample 2) of detached dwellings (5,716 sales), and for the sample (Sample 3) of detached dwellings for which supplementary mass appraisal data are available (4,880 sales).

[Insert Table 1 about here]

---

property. Instead, all owners participating in a cross-lease must abide by the ad hoc decisions of owners of a majority of the units. Peppercorn rents are on the order of 10 cents per year and are typically not collected.

4. METHOD

As noted above, it is not possible to directly test our ideas by using transacted properties as a sample of all properties due to the simple fact that market prices are not observed for properties that do not transact. Nevertheless, we are able to undertake a useful empirical study by using part of the transactions data to estimate values for the rest of the data.

We consider two sets of submarkets: a spatial classification that is used by government appraisers in New Zealand and a statistically generated aspatial classification. The first classification consists of 34 submarkets known as sales groups.[6] The sales groups are geographical areas considered by appraisers to be relatively homogeneous.

Our statistically derived submarkets are constructed as follows. We use principal component analysis to extract orthogonal factors from the characteristics of the properties, including the physical characteristics of the properties, the distance from the CBD, and demographic and socioeconomic characteristics of the areas in which the properties are located drawn from the census. The components that jointly account for at least 80% of the variance are retained, and for interpretation purposes, these components are rotated using a VARIMAX procedure. By VARIMAX rotation, the new principal components and the factor scores calculated on these components remain uncorrelated, which meets the requirement of using only non-collinear variables for cluster analysis. Factor scores are then used in cluster analysis to construct homogeneous submarkets. There is no requirement that submarkets consist of spatially contiguous dwellings.

The number of clusters is initially set to equal the number of spatial submarkets (34). As concluded by Afifi and Clark (1990), if the number of clusters to be grouped is known, a particularly appropriate method of clustering is the *K*-means method of MacQueen (1967). Therefore, a version of *K*-means clustering, for which the metric is squared Euclidean distance between cluster centroids, is used in this study. When a 34-

---

[6] Sales groups containing island properties were excluded from the analysis, and three central sales groups were combined because they had a relatively small number of transactions involving detached dwellings.

cluster solution is considered, however, some clusters contain very few observations, which would make it impossible to estimate hedonic equations for these submarkets. Therefore the cluster analysis was completed again with the constraint that the number of observations within each cluster should be equal to or greater than 50. This yielded 18 submarkets for the first sample, 15 for the second, and 14 for the third.

We note that the statistical method does not necessarily generate submarkets that are in some sense "superior" to the spatial ones defined by appraisers. Although our method explicitly takes into account numerous characteristics of the housing stock, the cluster analysis procedure implicitly gives equal weight to each of the factors defined by principal component analysis. However, it seems likely that some factors are more important than others, and an argument could be made that location captures many of the important factors. Thus geographical areas may well conform to the ideal sense of submarket more precisely than do aspatial groupings defined statistically with respect to multiple dimensions. By the same logic, excluding some non-locational factors from the cluster analysis may improve the results for the statistically generated submarkets.

As the purpose of this paper is to examine the usefulness of submarkets in an automated valuation context, we first consider a single market scenario, with one hedonic equation. We then estimate a single hedonic model with dummy variables for the sales groups. Such a procedure accounts for intercept differences across sales groups. To permit both the intercepts and slopes to vary, we then estimate hedonic models for each of the 34 sales groups. Finally, we perform hedonic regressions for the statistically defined housing submarkets. These analyses are conducted on each of the three samples of data.

To perform the hedonic regressions, 80% of the available transactions in each sample were used, the remaining observations being retained for out-of-sample testing. Transactions were divided into the two groups on the basis of random numbers. In reality, the sample of transacted properties would be a much smaller percentage of all properties and would not be a random sample. Consequently, the sample of transactions should be less representative than is the case with our experiment.

We employ both theoretical and empirical considerations in specifying the model. Both land area and floor area are expressed as natural logarithms due to the likelihood of diminishing returns as, *ceteris paribus*, the values of those variables increase in size. Distance to the CBD is also specified in logarithmic terms in view of the theory and empirical evidence that supports that relationship (see, *e.g.*, Mills and Hamilton, 1994). The dependent variable, house value, was also transformed into its natural logarithm as a means for making its distribution more normal, which in turn helps to normalize the distribution of the error term, a desirable characteristic for ordinary least squares estimators (see, *e.g.*, Kennedy, 1994). The relationship between the value and age of the dwelling is expected to follow a U-shaped curve, because very old houses can earn a premium due to their historic character and distinctive neighborhoods. Consequently, both the age of the dwelling and the square of age were included in the model. Quarterly time dummy variables were also added to the models. A number of other variables that consistently failed to contribute any explanatory power to the estimated equations were omitted.

For the citywide market and for each submarket, we estimate a hedonic regression and then use the estimated coefficients for out-of-sample prediction. For each out-of-sample prediction, we compute the absolute difference between the estimated value of the property and the actual sale price. We then calculate the percentage of differences that are within 10% and 20% of the sale prices. This is done for each of the three samples for the citywide equation with and without dummy variables for the sales groups, for the sales groups, and for the statistically generated submarkets. Then we repeat each set of predictions while controlling for spatial dependence. We adjust each prediction by the average residual for the neighborhood in which the property is located. For this purpose, we define the neighborhood to be the valuation roll, which is an area consisting of approximately 1,000 properties. Sales groups typically contain multiple valuation rolls.[7]

---

[7] Adjusting for spatial dependence using the average residuals for sales groups yielded similar results.

5. RESULTS

5.1. Hedonic Pricing Models

Table 2 provides the results for the citywide hedonic models without sales groups dummy variables for the three samples. The inclusion of additional variables in the third sample leads to an increase in the adjusted R^2 (from 0.69 to 0.72). Most variables selected in the models are significant at the 1% level. Table 3 gives results for the citywide models including the sales group dummy variables (the estimates for which are omitted from the table). The adjusted R^2 statistics increase about 8 to 10 percentage points when the sales groups dummies are included.

[Insert Tables 2 and 3 here]

The signs of the coefficients are as expected. For Sample 1, the dummy variable for detached dwellings is positive and significant at the 10% level or better, indicating that the sale price of detached units is higher than that of attached units. For housing physical characteristics, the logarithms of land and floor area are positively related to sale price, as is the square of the age of the property. Age itself is negatively related to the sale price, confirming the hypothesized U-shaped relationship between price and age. The quality and condition of the properties are also important. As expected, the logarithm of distance to the CBD is negatively related to sale price and is highly significant. The quarterly dummy variables suggest a seasonal decline in house prices during the winter, which is typical in Auckland.

The expanded variable set used with the third sample indicates that a water view is important in Auckland. The sale price is approximately 10% higher when there is such a view. Good landscaping, the number of attached garages, and to a lesser extent, a driveway, significantly affect dwelling prices. The quality of the neighborhood is very important, and higher quality levels are associated with higher prices. Without controlling for sales group membership (which allows the sales group variables to capture part of the variation in neighborhood quality), a property with high quality neighboring properties would be valued an average of 38% more than the same property with poor quality neighbors.

5.2. Out-of-Sample Percentage Prediction Errors

Table 4 reports the accuracy of predicted values in terms of the percentages of predictions that deviate less than 10% and 20% from the actual sales prices. For the Auckland-wide equation, just under two-thirds of the predictions deviate less than 20% from the sales prices (Samples 1 and 2). Regression equations with additional variables, such as view and quality of the neighborhood, improve the accuracy of predictions by a few percentage points (Sample 3).

[Insert Table 4 about here]

The inclusion of sales group dummy variables in the overall hedonic equations improves accuracy (at the 20% level) by about 10 to 15 percentage points. In the context of the use of submarkets in automated valuation, this result suggests that the sales group dummy variables capture important locational differences, such as socio-economic characteristics, that are important in predicting dwelling values. Not surprisingly, the addition of the sales group dummies benefits Sample 3 the least because additional locational variables, such as neighborhood quality, are already included in the estimations. However, the three samples have virtually the same level of accuracy for this specification.

When hedonic equations are estimated for individual sales groups, the results are about the same as for the citywide equations with sales group dummies. However, in this case, the accuracy of the predictions declines somewhat with sample size, consistent with our hypothesis.

The statistically defined submarkets do not perform as well as those based on the sales groups, with reduced accuracy of between five and seven percentage points. The superior performance of the models based on sales groups suggests that location is of prime importance when estimating hedonic equations for prediction purposes. This in turn suggests that the statistically defined submarkets might produce better results if they were formed by clustering locational factors rather than all factors. Details of the factors identified for Sample 1 are shown in Table 5. The factor patterns for Samples 2 and 3 are similar to that for Sample 1, except for the fact that additional variables are available for Sample 3. Experiments indicated that the best results are obtained when the cluster analysis is based on only the two most important factors for each model.[8] These factors load heavily on locational variables, namely distance from the CBD and characteristics of neighborhoods derived from the census. The results in the fifth line of Table 4 show the improvements in predictive accuracy obtained by defining submarkets in this manner. Although the results are generally not quite as good as when the sales groups are used, they are consistent with the idea that location is the single best criterion to use when defining submarkets.

[Insert Table 5 about here]

When results are adjusted for spatial dependence, the improvement in the proportion of predictions within a given percentage of sale price is greatest for equations which do not account for location or which account for location rather broadly. Improvements are in the seven to ten percentage point range for the citywide equations, while they are rather trivial for the citywide equations with sales group dummy variables and for the separate sales group equations. A similar pattern emerges for statistically defined submarkets: little improvement is obtained by adjusting for spatial dependence when submarkets are based on the two most important factors that are themselves locational in nature. These results underscore our conclusions regarding the importance of geography in delineating submarkets for mass appraisal purposes.

6. CONCLUSIONS

We start with the premise that the evaluation of alternative definitions of submarkets depends on the purpose for which the submarkets are constructed. If the purpose is to group close substitutes, then we argue that attention should be paid primarily to the characteristics of properties. On the other hand, if the aim is mass appraisal, then a focus on hedonic prices is warranted. In the latter case, the objective is to segment the market in a way that allows for accurate estimates of house values.

Using a sample of sales transactions from Auckland, New Zealand, we have demonstrated that housing submarkets defined as small geographical areas have more

[8] In this case, we constrained the number of submarkets to be the same as when they are defined based on all factors.

practical utility than submarkets defined using statistical techniques that disregard spatial contiguity. Adjusting for spatial dependence results in better predictions in most cases, although the degree of improvement depends on the level of spatial aggregation in the model. Not only do submarkets matter, but geography is what makes them matter. "Location, location, location" is not just a tired dictum. Moreover, our conclusions underscore the value of the practical knowledge of appraisers.

The broader implication of our results is that established neighborhood or other urban boundaries probably define suitable submarkets for mass appraisal purposes. In other words, it is probably not useful to employ elaborate statistical methods to define submarkets. However, such techniques may be useful in combining small geographical areas into larger areas for more basic research on the internal structure of cities. For example, such an approach could be employed to shed light on neighborhood patterns and dynamics.

REFERENCES

Adair, A. S., Berry, J. N. and McGreal, W. S. (1996). "Hedonic Modelling, Housing Submarkets and Residential Valuation," *J. Property Research* **13**, 67-83.

Afifi, A. A. and Clark, V. (1990). *Computer-Aided Multivariate Analysis*. London: Chapman and Hall.

Allen, M. T., Springer, T. M. and Waller, N. G. (1995). "Implicit Pricing Across Residential Submarkets," *J. Real Estate Finance Econ.* **11**, 137-151.

Bajic, V. (1985). "Housing-Market Segmentation and Demand for Housing Attributes: Some Empirical Findings," *AREUEA J.* **13**, 58-75.

Bourassa, S. C., Hamelink, F., Hoesli, M. and MacGregor, B. D. (1999). "Defining Housing Submarkets," *J. Housing Econ.* **8**, 160-183.

Crosson, S. T., Dannis, C. G. and Thibodeau, T. G. (1996). "Regression Analysis: A Cost-Effective Approach for the Valuation of Commercial Property Portfolios," *Real Estate Fin*. **13**(4), 20-28.

Dale-Johnson, D. (1982). "An Alternative Approach to Housing Market Segmentation Using Hedonic Pricing Data," *J. Urban Econ.* **11**, 311-332.

Galster, G. C. (1987). “Residential Segregation and Interracial Economic Disparities: A Simultaneous-Equations Approach,” *J. Urban Econ.* **21**, 22-44.

Gatzlaff, D. H. and Haurin, D. R. (1997) “Sample Selection Bias and Repeat-Sales Index Estimates,” *J. Real Estate Finance Econ.* **14**, 33-50.

Gatzlaff, D. H. and Haurin, D. R. (1998) “Sample Selection and Biases in Local House Value Indices,” *J. Urban Econ.* **43**, 199-222.

Goodman, A. C. and Dubin, R. A. (1990). “Sample Stratification with Non-Nested Alternatives: Theory and a Hedonic Example,” *Rev. Econ. Stat.* **72**, 168-173.

Goodman, A. C. and Kawai, M. (1982). “Permanent Income, Hedonic Prices, and Demand for Housing: New Evidence,” *J. Urban Econ.* **25**, 81-102.

Goodman, A. C. and Thibodeau, T. G. (1998). “Housing Market Segmentation,” *J. Housing Econ.* **7**, 121-143.

Hamilton, T. W. (1998). “Real Estate Market Segmentation,” *Assessment J.* **5**, 56-69.

Harsman, B. and Quigley, J. M. (1995). “The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco,” *J. Urban Econ.* **37**, 1-16.

Kennedy, P. (1994). *A Guide to Econometrics*, 3rd edition. Cambridge, MA: MIT Press.

Maclennan, D. and Tu, Y. (1996). “Economic Perspectives on the Structure of Local Housing Systems,” *Housing Stud.* **11**, 387-406.

MacQueen, J. B. (1967). “Some Methods for Classification and Analysis of Multivariate Observations,” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 281-297.

Mark, J. and Goldberg, M. A. (1988). “Multiple Regression Analysis and Mass Assessment: A Review of the Issues,” *Appraisal J.* **56**(1), 89-109.

Michaels, R. G. and Smith, V. K. (1990). “Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites,” *J. Urban Econ.* **28**, 223-242.

Mills, E. S. and Hamilton, B. W. (1994). *Urban Economics*, 5th edition. Reading, MA: Addison-Wesley.

Pace, R. K. and Gilley, O. W. (1989). "Appraisal Across Jurisdictions Using Bayesian Estimation with Bootstrapped Priors for Secondary Mortgage Market Applications," *Property Tax J.* **8**(1): 27-42.

Palm, R. (1976). "Urban Social Geography from the Perspective of the Real Estate Salesman," *Center for Real Estate and Urban Economics Research Report No. 38*, Berkeley: University of California.

Palm, R. (1978). "Spatial Segmentation of the Urban Housing Market," *Econ. Geography* **54**, 210-221.

Rothenberg, J., Galster G. C., Butler, R. V. and Pitkin, J. R. (1991). *The Maze of Urban Housing Markets*. Chicago: Univ. of Chicago Press.

Schnare, A. B. (1980). "Trends in Residential Segregation by Race: 1960-1970," *J. Urban Econ.* **7**, 293-301.

Schnare, A. B. and Struyk, R. J. (1976). "Segmentation in Urban Housing Markets," *J. Urban Econ.* **3**, 146-166.

Taylor, M., Rubin, G. and Lynford, L. (2000). "Submarkets Matter! Applying Market Information to Asset-Specific Decisions," *Real Estate Fin.* **17**(3), 7-26.

TABLE 1
Variable Means

Variables	Sample 1 Detached and attached	Sample 2 Detached only	Sample 3 Detached only (with expanded data)
Net sale price (NZ$)	294,996	323,955	328,398
Age of dwelling	37	45	46
Land area (square meters)	33	48	55
Cross-leased or strata-titled	0.54	0.35	
Floor area (square meters)	127	143	144
Detached houses (proportion)	0.45	1.00	1.00
Wall condition (proportion)			
Good	0.45	0.41	0.39
Average	0.53	0.56	0.58
Bad	0.02	0.03	0.03
Roof material (proportion)			
Tile	0.41	0.41	0.41
Metal	0.51	0.55	0.55
Other	0.07	0.04	0.04
Wall material (proportion)			
Wood	0.44	0.61	0.63
Brick	0.19	0.13	0.13
Fibrolite	0.06	0.07	0.06
Other	0.31	0.19	0.18
Quality of the principal structure (proportion)			
Superior	0.15	0.18	0.19
Average	0.82	0.77	0.76
Poor	0.03	0.05	0.05
Distance to CBD (kilometers)	6.2	6.8	6.8
Water view (proportion)			0.09
Modernization			0.26
Landscaping (proportion)			
Good			0.16
Average			0.79
Poor			0.05
Driveway			0.85
Quality of the neighborhood (proportion)			
Very good			0.03
Good			0.20
Average			0.68
Poor			0.09
Number of attached garages			0.75
Sample size	*8,421*	*5,716*	*4,880*

TABLE 2
Results for Citywide Hedonic Estimations without Sales Group Dummy Variables

Variables	Sample 1 Detached and attached	Sample 2 Detached only	Sample 3 Detached only (with expanded data)
Intercept	10.396***	12.370***	12.816***
Log of floor area	0.690***	0.642***	0.555***
Log of land area	1.898***	2.867***	2.651***
Cross-leased or strata-titled	0.099***	0.157***	0.173***
Detached dwelling	0.017*		
Age of dwelling	-0.002***	-0.002***	-0.003***
Age of dwelling squared	$4.720x10^{-5}$***	$3.366x10^{-5}$***	$4.720x10^{-5}$***
Walls in good condition	0.135***	0.135***	0.088***
Walls in average condition	0.092***	0.106***	0.063***
Dwelling with a tile roof	-0.031**	-0.049***	-0.030
Dwelling with a metal roof	-0.061***	-0.103***	-0.068***
Dwelling with wooden walls	-0.016	-0.018	-0.015
Dwelling with brick walls	-0.079***	-0.073***	-0.056***
Dwelling with fibrolite walls	-0.097***	-0.108***	-0.102***
Superior quality of the principal structure	0.231***	0.203***	0.148***
Average quality of the principal structure	0.094***	0.074***	0.070***
Log of distance to the CBD	-0.172***	-0.010***	-0.390***
Quarter 2	0.016*	0.011	0.008
Quarter 3	-0.011	-0.022**	-0.020**
Quarter 4	0.026***	0.011	0.015
Water view			0.103***
Modernization			0.034***
Average landscaping			0.026
Good landscaping			0.077***
Driveway			0.019
Average neighborhood			0.098***
Good neighborhood			0.231***
Very good neighborhood			0.323***
Number of attached garages			0.039***
Adjusted R^2	*0.704*	*0.688*	*0.722*

Note: The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

TABLE 3
Results for Citywide Hedonic Estimations with Sales Group Dummy Variables

Variables	Sample 1 Detached and attached	Sample 2 Detached only	Sample 3 Detached only (with expanded data)
Intercept	9.817***	10.850***	11.344***
Log of floor area	0.575***	0.502***	0.457***
Log of land area	2.535***	2.727***	2.560***
Cross-leased or strata-titled	0.100***	0.109***	0.128***
Detached dwelling	0.073***		
Age of dwelling	-0.004***	-0.004***	-0.004***
Age of dwelling squared	4.889×10^{-5}***	4.873×10^{-5}***	5.096×10^{-5}***
Walls in good condition	0.112***	0.115***	0.083***
Walls in average condition	0.069***	0.082***	0.052***
Dwelling with a tile roof	-0.040***	-0.051***	-0.028
Dwelling with a metal roof	-0.050***	-0.067***	-0.041**
Dwelling with wooden walls	-0.018**	7.577×10^{-4}	-0.006
Dwelling with brick walls	-0.023***	-0.026**	-0.019
Dwelling with fibrolite walls	-0.024**	-0.044***	-0.049***
Superior quality of the principal structure	0.146***	0.155***	0.124***
Average quality of the principal structure	0.047***	0.053***	0.050***
Log of distance to the CBD	-0.015**	-0.096***	-0.137***
Quarter 2	0.016**	0.013	0.014
Quarter 3	-0.013*	-0.023***	-0.019**
Quarter 4	0.023***	0.010	0.017
Water view			0.079***
Modernization			0.029***
Average landscaping			0.013
Good landscaping			0.060***
Driveway			0.010
Average neighborhood			0.021
Good neighborhood			0.067***
Very good neighborhood			0.205***
Number of attached garages			0.036***
Adjusted R^2	*0.806*	*0.785*	*0.798*

Note: The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.
Estimated coefficients for the sales group dummy variables are omitted from the table.

TABLE 4
Percentage of Predictions Within 10 and 20 Percent of Sale Price

Model	Sample 1 Detached and attached		Sample 2 Detached only		Sample 3 Detached only (with expanded data)	
	Within 10% of sale price	Within 20% of sale price	Within 10% of sale price	Within 20% of sale price	Within 10% of sale price	Within 20% of sale price
Not adjusted for spatial dependence						
Citywide equation	*35.0*	*63.3*	*37.2*	*62.5*	*39.2*	*66.6*
Citywide equation with sales group dummy variables	46.0	75.0	46.1	77.0	46.6	76.1
Separate sales group equations	48.2	76.8	45.5	74.0	44.9	72.0
Statistically defined submarkets based on all factors	41.0	69.7	40.7	67.1	39.9	67.5
Statistically defined submarkets based on two most important factors	44.2	73.8	41.2	71.1	43.1	73.5
Adjusted for spatial dependence						
Citywide equation	43.5	70.7	43.0	72.7	43.8	73.7
Citywide equation with sales group dummy variables	**48.3**	76.5	**47.5**	**77.2**	**48.0**	**77.5**
Separate sales group equations	47.4	**77.9**	44.6	75.1	44.1	72.4
Statistically defined submarkets based on all factors	46.3	74.1	44.2	73.1	43.1	70.0
Statistically defined submarkets based on two most important factors	46.8	75.9	46.6	74.4	46.5	74.1

Note: The highest percentage in each column is indicated in boldface. The lowest percentage is in italics.

TABLE 5
Principal Component Analysis for Sample 1 (Detached and Attached Dwellings)

A. Explanatory power of each factor

Factor	Percent variance explained	Cumulative	Eigenvalues
I	20.5	20.5	5.34
II	15.1	35.7	3.93
III	9.3	44.9	2.41
IV	7.0	51.9	1.81
V	5.4	57.3	1.40
VI	4.6	61.9	1.21
VII	4.5	66.4	1.17
VIII	4.3	70.7	1.11
IX	4.2	74.9	1.10
X	3.8	78.7	0.99
XI	3.7	82.3	0.95

B. The nature of the factors

Factor		Variable loadings (> |0.50|)	
I.	Socio-economic characteristics of neighborhoods	Percent receiving income support	0.92
		Percent Maori	0.89
		Percent Pacific Islander	0.88
		Percent unemployed	0.78
		Median household income	-0.84
II.	Housing tenure and location	Percent driving to work	0.90
		Homeownership rate	0.84
		Average number of bedrooms	0.79
		Distance from CBD	0.69
III.	Property type and age	Land area	0.88
		Detached dwelling	0.73
		Age	0.61
		Cross-leased or strata-titled	-0.90
IV.	Dwelling size and condition	Average wall condition	0.76
		Floor area	-0.64
V.	Roofing material	Other material	0.79
		Metal	-0.76
VI.	Asian neighborhood	Percent Asian	0.88
VII.	Structure quality	Poor	0.87
		Average	-0.77
VIII.	Wall material	Brick	0.90
		Other material	-0.61
IX.	Density	Dwelling density	0.81
		Population density	0.62
X.	Dwelling condition	Poor wall condition	0.94
XI.	Wall material	Fibrolite	0.95