UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 9

A Grand Hedonic Model of the Canadian Housing Market

Decomposing the Value of Structure and Location across 142,121 MLS Listings with High-Dimensional Neighbourhood Fixed Effects

Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

October 2026

Version 1.3

Abstract

We estimate a national hedonic price function for the Canadian residential market on a cross-section of 142,121 active MLS listings from nine provinces. A semi-logarithmic specification relates the log list price to structural attributes, dwelling type and ownership form, and 1,155 absorbed neighbourhood (Forward Sortation Area) fixed effects estimated by absorbing least squares with neighbourhood-clustered inference. Moving from a structural model to the full specification raises the explained share of log-price variation from 44% to 78%. Because structure and location are correlated, the attribution of this gain depends on the order in which blocks enter: a Shapley–Owen decomposition assigns about 34 percentage points to structure and 36 to location, so that the two are of comparable importance and the neighbourhood is the single largest block once province and neighbourhood are grouped. Conditional on location, the living-area elasticity is 0.53 (0.49–0.62 across samples), each full bathroom adds about 12%, and bedroom counts are economically negligible. The implicit prices of area and bathrooms survive province-level wild-cluster bootstrap inference, whereas the small half-bathroom and bedroom effects do not. Neighbourhood effects remove 87% of the spatial autocorrelation in residuals (Moran's I 0.50 to 0.06). The model values held-out listings in neighbourhoods seen in training with an out-of-sample R^2 of 0.77 and a median absolute percentage error of 15.6%; for neighbourhoods never seen in training, spatial interpolation of the neighbourhood effect still yields $R^2 = 0.74$ and a median error of 17%. Floor space displays mild diminishing returns, the location premium falls by about 5.5% per doubling of distance to the nearest metropolitan centre, and implicit prices vary across the price distribution and across provinces. Neighbourhood premia span a factor of about ten between the most and least expensive areas. Because the dependent variable is the asking price rather than the transaction price, all estimates describe the pricing of listings, not of sales.

Keywords: hedonic pricing, housing markets, Canada, high-dimensional fixed effects, neighbourhood effects, spatial autocorrelation, automated valuation model, list prices

JEL Classification: R31, R21, C21, C55

1. Introduction

Housing is the largest asset on most household balance sheets and the collateral behind most household debt, and its price is the most consequential relative price that most families ever face. A dwelling is also the archetypal heterogeneous good: no two houses are identical, and its most important attribute—*where it stands*—cannot be read off a listing sheet as a scalar. The hedonic approach of Lancaster (1966) and Rosen (1974) resolves this heterogeneity by treating the dwelling as a bundle of attributes, each of which commands an implicit price determined in equilibrium by the interaction of buyers' bid functions and sellers' offer functions. Regressing the log price of a dwelling on its measurable characteristics recovers these implicit prices, and the resulting hedonic price function is the empirical backbone of constant-quality house-price indices (Bailey et al., 1963; Hill, 2013), mass appraisal (McCluskey et al., 2013), mortgage collateral valuation, and the welfare analysis of local amenities (Black, 1999; Chay and Greenstone, 2005; Bishop et al., 2020).

How much of a home's value is the structure, and how much is the location? The question is old—it is the empirical content of the realtor's adage "location, location, location" (Kiel and Zabel, 2008)—but its answer matters well beyond folklore. Assessment authorities must apportion value between land and improvements; the incidence of property taxation depends on the land component; the macro literature attributes most of the level and growth of aggregate housing wealth to land rather than structure (Davis and Heathcote, 2007; Knoll et al., 2017); and the dispersion of location values across cities is the imprint of supply constraints and agglomeration (Saiz, 2010; Gyourko et al., 2013; Combes et al., 2019). A credible answer requires holding the structural bundle fixed while letting location vary, which is what a hedonic price function with fine geographic fixed effects delivers.

This paper estimates such a model for the Canadian residential market at national scale. Using a cross-section of 142,121 active MLS listings from the nine provinces present in the data, we relate the log list price to structural attributes, dwelling type and ownership form, and 1,155 neighbourhood fixed effects defined at the Forward Sortation Area (FSA) level, the first three characters of the postal code. The Canadian hedonic tradition has remained largely metropolitan in scope (Des Rosiers et al., 1996; Haider and Miller, 2000; Thériault et al., 2003; Dubé et al., 2013), and national evidence has been confined to city-level price indices (Allen et al., 2009; Hossain and Latif, 2009); to our knowledge no previous study estimates a single listing-level hedonic price function spanning the Canadian market. The exercise is made feasible by an absorbing least-squares estimator in the lineage of Abowd et al. (1999) that sweeps out the high-dimensional location effects without materialising more than a thousand dummy variables (Guimarães and Portugal, 2010; Correia, 2017).

Two features of the design deserve emphasis at the outset, because they govern the interpretation

of every number that follows. First, the dependent variable is the *asking* price of an active listing, not a transaction price. List prices are set strategically by sellers and interact with search and bargaining (Horowitz, 1992; Genesove and Mayer, 2001; Merlo and Ortalo-Magné, 2004; Han and Strange, 2016), so the implicit prices we estimate describe how sellers and their agents price attributes when they bring a dwelling to market. We refer to them as implicit *list* prices wherever the distinction matters and discuss the resulting wedge in Section 7. Second, the neighbourhood fixed effects are a device for measurement, not for causal identification: they absorb every location-specific amenity, observed or not, so that structural implicit prices are identified from within-neighbourhood variation, and the fixed effects themselves summarise the capitalised value of location without disentangling its components.

Our findings can be summarised as follows. A model with structural attributes alone explains 44% of the variation in log list prices across all residential dwellings; adding dwelling type, ownership and 1,155 neighbourhood effects raises the share to 78%. Because larger and better-equipped dwellings are not randomly located, the attribution of the explained variance to structure and location depends on the order in which the blocks are entered. We therefore complement the sequential specification ladder with a Shapley–Owen decomposition, which splits the shared variance symmetrically (Israeli, 2007; Huettner and Sunder, 2012): structure receives about 34 percentage points and location about 36, so that the two are of comparable importance and the neighbourhood is the single largest block once province and neighbourhood are grouped. The sequential ladder, which credits the overlap to whichever block enters first, assigns 44 points to structure and 32 to location. Conditional on location, the living-area elasticity is 0.53, so that a 10% larger dwelling is listed at about 5% more; each full bathroom adds about 12%; and the number of bedrooms is economically negligible once floor space is held fixed, a recurrent finding of the hedonic literature (Sirmans et al., 2005, 2006). These implicit prices are stable across eight alternative samples and specifications, and the area and bathroom effects survive a province-level wild-cluster bootstrap, although the smaller half-bathroom and bedroom effects do not. The neighbourhood effects remove 87% of the spatial autocorrelation in residuals (Moran's I falls from 0.50 to 0.06). Out of sample, the model values held-out listings in neighbourhoods observed in training with an R^2 of 0.77 and a median absolute percentage error (MdAPE) of 15.6%; when entire neighbourhoods are held out and their effects are interpolated from neighbouring FSAs, the R^2 is 0.74 and the MdAPE 17%. Floor space exhibits mild diminishing returns, the location premium declines by about 5.5% per doubling of distance to the nearest metropolitan centre, and implicit prices vary across the price distribution and across provinces. The neighbourhood premia, net of structure, span a factor of about ten between the most expensive FSAs—on Vancouver's west side, in Whistler and in West Vancouver—and the least expensive, in rural Newfoundland, Saskatchewan and Manitoba.

The contribution is threefold. First, we provide a listing-level decomposition of Canadian

housing value into structure and location that spans the entire national market and whose attribution rule is made explicit; we show that the familiar "location dominates" statement holds under a symmetric attribution but not under a structure-first sequential one, and we quantify the overlap. Second, we bring the high-dimensional fixed-effects technology of applied microeconomics to a national housing cross-section and validate it directly, documenting how much spatial dependence more than a thousand absorbed neighbourhood intercepts remove and how sensitive inference is to the level of clustering. Third, we supply a transparent accuracy benchmark for the automated-valuation debate, distinguishing predictive accuracy within the spatial support of the training data from accuracy in neighbourhoods never observed, a distinction that random holdouts conceal (Roberts et al., 2017).

The remainder of the paper is organised as follows. Section 2 positions the paper in the related literatures. Section 3 describes the data and the construction of the estimation sample. Section 4 presents the empirical strategy, including the decomposition and inference procedures. Section 5 reports the main results and Section 6 the robustness, heterogeneity and validation analyses. Section 7 interprets the findings against the literature and discusses limitations; Section 8 concludes. The appendix details the variable construction and reconciles the present estimates with the May 2026 version of this paper.

2. Related Literature

The paper draws on, and speaks to, six literatures: the hedonic framework and its identification; functional form and distributional heterogeneity in housing hedonics; the treatment of space through submarkets, spatial econometrics and fixed effects; the theory of amenity capitalisation; urban spatial structure and the land share of housing value; and valuation, from price indices to automated valuation models. A final subsection reviews the Canadian evidence. Each subsection closes by stating where the present study enters.

2.1. The hedonic framework and its identification

The hedonic method has two intellectual roots. Lancaster (1966) recast consumer theory in terms of the characteristics embodied in goods, while Court (1939) and Griliches (1961) developed the empirical machinery of quality-adjusted price indices. Rosen (1974) unified these ideas in an equilibrium model in which the observed price schedule is the envelope of buyers' bid functions and sellers' offer functions, so that its gradient with respect to a characteristic identifies the implicit price of that characteristic at the margin. Sheppard (1999) surveys the subsequent housing applications. A later literature clarified what the second stage—recovering preference parameters from implicit prices—can and cannot deliver. Epple (1987) formalised the simultaneity problem

of second-stage demand estimation; Ekeland et al. (2004) showed that nonlinearity of the price function aids identification; Bajari and Benkard (2005) developed a tractable demand estimator with unobserved product characteristics; Bishop and Timmins (2019) proposed a second-stage estimator that dispenses with instruments; and Kuminoff et al. (2010) and Bishop et al. (2020) documented how sensitive welfare estimates are to the specification of the first stage and codified best practice, including the use of fine spatial fixed effects. We remain deliberately at the first stage and target the implicit-price schedule itself, which is the relevant object for valuation, assessment and index construction and is identified under much weaker assumptions than the underlying preferences.

2.2. Functional form and heterogeneity in the implicit prices

A large applied literature asks which attributes belong in a housing hedonic and which functional form to impose. Sirmans et al. (2005) catalogue the regressors used in decades of published models and document the central role of living area, bathrooms, lot size, age and location; their tabulation also shows that bedroom coefficients are often small or negative once floor area is controlled, and the meta-analysis of Sirmans et al. (2006) confirms that the implicit prices of area and bathrooms are among the most stable in the literature. Malpezzi (2003) argues that the semi-logarithmic specification is a robust default: it accommodates the right skew of prices, yields coefficients that are approximately proportional effects, and mitigates heteroskedasticity. The interpretation of its coefficients requires care, however: for indicator variables the proportional effect is $e^{\gamma} - 1$ rather than γ (Halvorsen and Palmquist, 1980), and the sampling variance of $\hat{\gamma}$ biases even that transformation, which Kennedy (1981) corrects. We adopt the semi-log form and report corrected percentage effects. A complementary literature relaxes the functional form. McMillen and Redfearn (2010) estimate nonparametric hedonic surfaces and show that parametric forms can mask economically meaningful curvature, a concern we address with quadratic and spline specifications in floor space. Quantile-regression approaches (Koenker and Bassett, 1978; Zietz et al., 2008) show that implicit prices differ systematically along the price distribution, so that a conditional-mean coefficient is an incomplete summary; we document this for Canada in Section 6.3. These strands are almost exclusively metropolitan or regional in scope; national evidence on functional form and distributional heterogeneity estimated on a single consistent dataset is rare.

2.3. Space in housing models: submarkets, spatial econometrics, fixed effects

Because housing is immobile, unobserved local amenities induce strong spatial dependence in prices. Dubin (1988) showed that spatially autocorrelated errors distort hedonic inference and Basu and Thibodeau (1998) documented pervasive residual autocorrelation in Dallas transactions. Can (1992) and Anselin (1988) formalised spatial dependence in hedonic models, Kelejian and Prucha

(1998) provided feasible estimators for spatial autoregressive specifications, and Kelejian and Prucha (2001) derived the asymptotic distribution of the Moran statistic that we use as a diagnostic. Pace et al. (1998) and, in the Canadian context, Dubé and Legros (2013) developed spatio-temporal weight matrices. A parallel strand argues that housing markets are segmented into submarkets: Goodman and Thibodeau (1998) formalise segmentation tests, and Bourassa et al. (2003) show that accounting for submarkets improves prediction, with the practical finding that simple geographic definitions perform as well as statistically derived ones. Bourassa et al. (2007) and Case et al. (2004) compare these approaches and conclude that geographic submarket controls capture most of what parametric spatial models capture at a fraction of the complexity. More recently, Hill and Scholz (2018) show that geospatial splines improve hedonic imputation indices, and Bogin et al. (2019) construct price indices at the ZIP code level, a geography comparable to the FSA.

Our approach takes the submarket conclusion to its logical end: rather than imposing a spatial weight matrix, we absorb a fixed effect for each of 1,155 neighbourhoods and identify structural implicit prices from within-neighbourhood variation. This is the housing analogue of the high-dimensional fixed-effects designs pioneered in labour economics by Abowd et al. (1999), with computation following Guimarães and Portugal (2010) and Correia (2017). The choice speaks to a methodological debate: Gibbons and Overman (2012) argue that parametric spatial-lag models are often hard to interpret and that flexible fixed effects are frequently preferable, while LeSage and Pace (2009) develop the spatial-autoregressive alternative. We side with the fixed-effects approach but validate it, testing for residual spatial autocorrelation with Moran's I (Moran, 1950) before and after absorbing the neighbourhood effects. Inference in such designs raises its own questions: Cameron and Miller (2015) and Abadie et al. (2023) discuss the choice of clustering level, and MacKinnon and Webb (2017) the wild-cluster bootstrap when clusters are few; we report alternative variance estimators and a province-level bootstrap. Submarket and spatial-fixed-effect studies are overwhelmingly single-city; we are not aware of a prior implementation at the scale of an entire national market with more than a thousand absorbed neighbourhoods.

2.4. Amenity capitalisation

Why should a neighbourhood intercept measure economic value? The theoretical answer runs from Tiebout (1956), in which households sort across jurisdictions according to local public goods, through Oates (1969), who first measured the capitalisation of taxes and school spending into house prices, to Roback (1982), who embedded amenity capitalisation in a spatial equilibrium of wages and rents. Quasi-experimental hedonic studies have since measured the capitalisation of specific amenities: school quality (Black, 1999), air quality (Chay and Greenstone, 2005) and local crime risk (Linden and Rockoff, 2008), while Cheshire and Sheppard (1995) and Albouy (2016) quantify the total land-value content of local amenities. Our neighbourhood fixed effects deliberately bundle

all capitalised amenities into a single FSA-level premium rather than attempting to disentangle them; the estimated premium is therefore a composite that future work can decompose amenity by amenity.

2.5. Urban spatial structure and the land share of housing value

The monocentric tradition of Alonso (1964), Muth (1969) and Mills (1967) predicts that land rents—and hence house prices, holding structure constant—decline with distance from employment centres, a prediction whose modern quantitative form is estimated by Ahlfeldt et al. (2015). Our distance-to-metro gradient is a reduced-form counterpart of this prediction at national scale, and the agglomeration literature reviewed by Combes and Gobillon (2015) supplies the economic mechanism; Combes et al. (2019) estimate the corresponding house- and land-price gradients across French cities. A related literature measures how much of housing value is land rather than structure: Davis and Heathcote (2007) put the aggregate U.S. residential land share near 46%, Davis and Palumbo (2008) and Albouy et al. (2018) document its enormous dispersion across metropolitan areas, Knoll et al. (2017) attribute about 80% of the global house-price increase since World War II to land, and Diewert et al. (2015) show how a hedonic regression can itself be used to split a price index into land and structure components. On the supply side, Saiz (2010) shows that geographic constraints generate persistent cross-location price dispersion, and Gyourko et al. (2013) formalise the resulting "superstar city" dynamics, of which Vancouver and Toronto are natural candidates (Glaeser et al., 2005). This literature measures the land or location share with aggregate or city-level data; a listing-level decomposition for an entire national market, estimated from within-neighbourhood variation, complements it from the bottom up.

2.6. Valuation: price indices, mass appraisal and machine learning

The hedonic regression is one of two workhorses of constant-quality house-price measurement, alongside the repeat-sales method of Bailey et al. (1963) that underlies the Case and Shiller (1989) index tradition; Hill (2013) surveys the hedonic index literature. The same estimated surface doubles as a valuation engine: Clapp (2003) develops a location-value surface for automated valuation, Kiel and Zabel (2008) systematise the structure-versus-location decomposition of house prices, and the mass-appraisal literature benchmarks prediction accuracy across model classes (McCluskey et al., 2013). The rise of machine-learning valuation (Mullainathan and Spiess, 2017) has sharpened the question of what transparency costs in accuracy. A methodological point from the predictive literature is directly relevant: when observations are spatially clustered, random holdouts overstate the accuracy that can be expected in locations not represented in the training data, and spatially blocked holdouts are required (Roberts et al., 2017). Few studies report held-out accuracy for

a transparent hedonic model at national scale, and fewer still distinguish within-support from out-of-support accuracy; our validation provides both.

2.7. List prices and the Canadian evidence

Because our dependent variable is the asking price, the literature on listing behaviour bounds the interpretation of our estimates. Horowitz (1992) models the list price as a ceiling that shapes the arrival of offers; Genesove and Mayer (2001) show that sellers facing nominal losses set higher asking prices; Anglin et al. (2003) document the trade-off between the asking price and time on the market; Merlo and Ortalo-Magné (2004) observe the full bargaining sequence from listing to sale; Genesove and Han (2012) and Han and Strange (2015) review the search-and-matching microstructure; and Han and Strange (2016) show that the asking price acts both as a commitment and as a signal. The wedge between list and transaction prices is therefore systematic rather than random noise, and we return to it in Section 7.

Canada has a productive hedonic tradition that is largely metropolitan: the Quebec City programme of Des Rosiers, Thériault and coauthors measures the capitalisation of shopping centres (Des Rosiers et al., 1996), landscaping (Des Rosiers et al., 2002) and the interaction of location with attribute values (Thériault et al., 2003); Haider and Miller (2000) apply spatial autoregressive techniques to Toronto; and Dubé et al. (2013) estimate the capitalisation of commuter rail on Montreal's South Shore. At the national level, Allen et al. (2009) test for segmentation across Canadian city price indices and Hossain and Latif (2009) study the determinants of aggregate price volatility. To our knowledge no previous study estimates a single hedonic price function spanning the Canadian market at the listing level; the national scope, the absorbed neighbourhood effects, the explicit attribution rule for the structure–location decomposition and the within- versus out-of-support validation are, jointly, the contribution of this paper.

3. Data

3.1. Source and coverage

The analysis draws on a de-duplicated snapshot of active "for-sale" listings from the Canadian Multiple Listing Service (MLS), stored as `realtor_mls_unique.duckdb` and containing 172,019 unique listings. Each record carries the asking price, a semi-structured set of building and lot attributes, ownership and dwelling-type fields, and a geocoded location (latitude, longitude and postal code). Coordinates are present for every record and postal codes for 99.9% of them, so that each listing can be assigned to its Forward Sortation Area (FSA), the first three characters of the Canadian postal code. Canada Post defines FSAs as the geographic units from which mail delivery

originates; the first letter identifies the province or territory, or a sector within Ontario and Quebec, and the 2021 Census recognised 1,643 FSAs nationally (Statistics Canada, 2022). FSAs are postal conveniences rather than planning units, but they are the finest geography that can be assigned to every listing without a boundary file, and their size—a few city blocks in dense urban cores, a rural county elsewhere—is of the order of what the submarket literature uses (Bourassa et al., 2003; Bogin et al., 2019).

Two properties of the snapshot bear on everything that follows. First, the dependent variable is the asking price of an active listing. Asking prices are set by sellers and their agents, are typically above the eventual transaction price, and respond to market tightness and listing strategy (Horowitz, 1992; Genesove and Mayer, 2001; Anglin et al., 2003; Han and Strange, 2016); we therefore interpret every coefficient as an implicit *list* price. Second, the snapshot is a single cross-section of the stock of listings at one point in time, not a flow of sales over a period, so that no time dimension is available and the composition of the sample reflects which dwellings were on the market rather than which dwellings sold.

3.2. Sample construction and variable parsing

The raw fields are semi-structured and require parsing; Appendix A documents every rule. Bedroom counts reported as "3 + 1" (main plus lower level) are summed. Living areas, reported inconsistently in square feet or square metres and sometimes as a band, are harmonised to square metres (preferring the explicit floor-area measurement, mapping banded entries to the upper bound of the band, and converting at 1 $\text{ft}^2 = 0.0929\ \text{m}^2$). Lot sizes are parsed from free text into square metres where a numeric value with a recognisable unit can be recovered; the largest 1% of parsed lots, which are mostly multi-acre acreage strings, are treated as unparsed, and a missingness indicator is retained. Storey and parking counts are taken as reported and set to zero when absent (32% and 27% of listings respectively), so that their coefficients should be read as conditional on a count being reported; both fields contain implausible extreme values (up to 593 storeys and 3,620 parking spaces) that we address in the robustness analysis.

The province of each listing is derived from the first letter of its FSA rather than from the province field of the scraper, which records the regional seed of the extraction and mislabels 11.7% of listings near provincial borders (for instance, New Brunswick and Ottawa-area listings seeded as Quebec). Listings whose FSA letter does not correspond to a province (27 records in the territories or with invalid codes) are dropped. We restrict the sample to residential dwellings (houses, condominiums, plexes, townhouses and apartments) with a strictly positive asking price, non-missing living area, bedroom and bathroom counts and a valid FSA, which leaves 147,236 listings, and we then trim the extreme 1% tails of price and living area to limit the influence of data-entry errors and ultra-luxury outliers, a standard precaution with listing data (Malpezzi,

2003). The estimation sample contains 142,121 dwellings, of which 83,524 are houses, 57,853 condominiums and 744 plexes, townhouses and apartments. Asking prices range from $114,900 to $2,998,800 and living areas from 46 to 465 m^2.

The dependent variable is the natural logarithm of the asking price. Structural regressors are the log of living area; counts of bedrooms, full bathrooms and half bathrooms; parking spaces; storeys; and an indicator for the availability of lot information together with $\ln(1 + \text{lot area})$. Categorical controls comprise the dwelling type (house, apartment, row or townhouse, duplex, triplex, fourplex, manufactured home, other), the ownership form (freehold, condominium or strata, leasehold, co-operative, other, not reported) and the broad listing category. The neighbourhood is the FSA. The sample contains 1,667 distinct FSAs; the 521 FSAs with fewer than 25 listings, which together hold 5,385 listings (3.8% of the sample), are pooled into one residual category per province, so that the model absorbs 1,155 location effects: 1,146 individual FSAs and nine pooled categories. The median absorbed FSA contains 86 listings and the mean 123. The pooled categories are not neighbourhoods and we verify in Section 6.1 that dropping them leaves the estimates unchanged.

3.3. Descriptive statistics

Table 1 reports the descriptive statistics. The median dwelling is listed at $639,900 and offers 135 m^2 of living space, two full bathrooms and three bedrooms. Price and price per square metre are strongly right-skewed—means exceed medians throughout—which motivates the log transformation illustrated in Figure 1; the distribution of the log price is close to symmetric and underpins the semi-logarithmic specification. The lot-area column reflects the parsing rules: 38% of listings, mostly condominiums, carry no usable lot information and are coded as zero, and the median parsed lot among the remainder is 690 m^2.

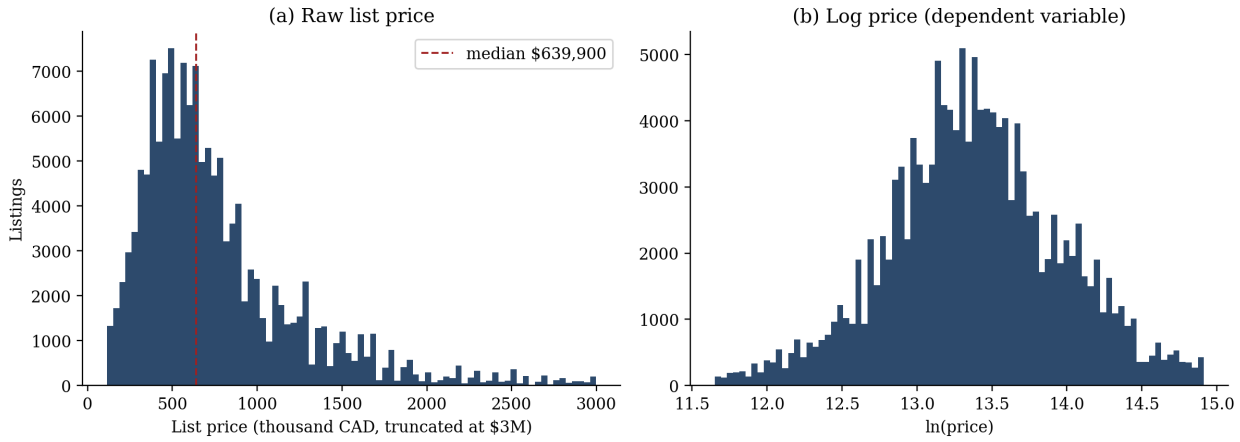


Figure 1. Distribution of asking prices. Panel (a) shows the raw price (truncated at $3M for readability) with the sample median marked; panel (b) shows the log price, the dependent variable.

Table 1. Descriptive statistics

Variable	Mean	SD	P25	Median	P75
List price (CAD)	766,862	478,647	449,000	639,900	929,900
Price per m^2 (CAD)	5,388	2,690	3,581	4,785	6,600
Living area (m^2)	152.1	79.3	94.1	135.3	185.8
Bedrooms	3.12	1.33	2.00	3.00	4.00
Full bathrooms	2.32	1.12	2.00	2.00	3.00
Half bathrooms	0.35	0.53	0.00	0.00	1.00
Parking spaces	2.76	11.49	0.00	2.00	4.00
Storeys	1.80	6.10	0.00	1.00	2.00
Lot area (m^2)	2,531	11,820	0	367	906

Notes: Estimation sample of 142,121 residential dwellings (houses, condominiums, plexes, townhouses and apartments) from the Canadian MLS, after parsing and trimming the extreme 1% tails of price and living area.

Ontario (55,897 listings), British Columbia (32,994), Quebec (19,949) and Alberta (19,463) account for 90% of the sample; Saskatchewan, Manitoba, Nova Scotia and New Brunswick each contribute about 3,000 listings and Newfoundland and Labrador 1,274. Prince Edward Island and the territories are absent. Median asking prices per square metre range from about $2,000 in Newfoundland and Labrador and $2,300 in New Brunswick to $5,200 in Ontario and $6,000 in British Columbia (Figure 2), foreshadowing the role of location documented below.

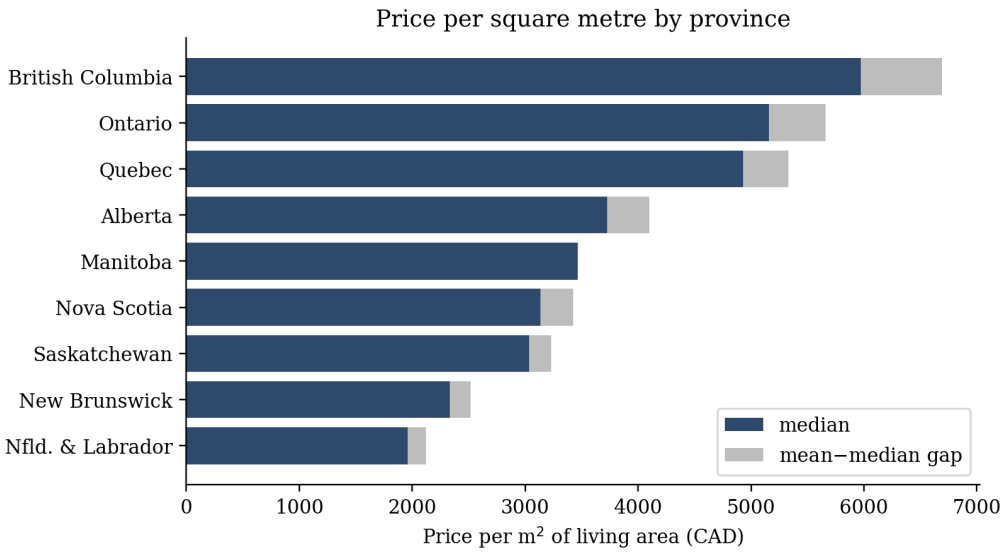


Figure 2. Median asking price per square metre of living area by province. Grey extensions show the mean–median gap, reflecting right skew from high-value listings. Provinces are assigned from the FSA.

4. Empirical Strategy

4.1. The hedonic price function

We estimate the semi-logarithmic hedonic price function

$$\ln P_i = \alpha + \beta \ln(\text{Area}_i) + \mathbf{x}_i'\gamma + \mathbf{d}_i'\delta + \mu_{f(i)} + \varepsilon_i, \tag{1}$$

where P_i is the asking price of dwelling i; Area_i is its living area in square metres; $\mathbf{x}_i$ collects the remaining structural attributes (bedrooms, full and half bathrooms, parking spaces, storeys, the lot-information indicator and $\ln(1+\text{lot area})$); $\mathbf{d}_i$ is a vector of dwelling-type, ownership and listing-category indicators; $\mu_{f(i)}$ is a fixed effect for the FSA $f(i)$ in which the dwelling is located; and ε_i is an idiosyncratic error with $\mathbb{E}[\varepsilon_i \mid \text{Area}_i, \mathbf{x}_i, \mathbf{d}_i, f(i)] = 0$. Equation (1) is a first-stage hedonic regression in the sense of Rosen (1974): its coefficients are implicit prices along the equilibrium price schedule of listings, not preference parameters.

Because β multiplies a logged regressor, it is the elasticity of the asking price with respect to floor space. A coefficient γ_k on a count variable implies a proportional change of $e^{\gamma_k} - 1$ per additional unit, which for the magnitudes encountered here differs from γ_k by less than one percentage point; we report both where the difference is material. For the indicator variables in $\mathbf{d}_i$ the proportional effect relative to the omitted category is $e^{\delta_k} - 1$ (Halvorsen and Palmquist, 1980), and because $\hat{\delta}_k$ is estimated with error the unbiased version is the Kennedy (1981) estimator

$$\hat{g}_k = \exp\!\left(\hat{\delta}_k - \tfrac{1}{2}\widehat{V}(\hat{\delta}_k)\right) - 1, \tag{2}$$

which we use throughout for percentage effects of indicators.

4.2. High-dimensional location effects and identification

The neighbourhood fixed effects μ_f are the heart of the design. A separate intercept for each of the 1,155 FSA groups allows every neighbourhood an arbitrary price level that absorbs all location-specific amenities—school quality, transit access, coastline, employment density—whether or not they are observed. The structural implicit prices β and γ are then identified from variation *within* neighbourhoods, by comparing dwellings that differ in physical attributes but share a location. This is the within estimator of panel econometrics applied to a spatial grouping, and it addresses the most pernicious source of omitted-variable bias in hedonic work: the correlation between structural quality and unobserved locational quality (Can, 1992; Bourassa et al., 2007; Bishop et al., 2020). It does not address within-neighbourhood sorting of unobserved dwelling quality (age, renovation, finish) on observed attributes, which remains the main threat to the structural coefficients and which

we discuss in Section 7.

Including more than a thousand indicator variables directly is numerically wasteful. We instead estimate Equation (1) by absorbing least squares, in the high-dimensional fixed-effects tradition initiated by Abowd et al. (1999): the dependent variable and every regressor are demeaned within FSA groups, $\tilde{z}_i = z_i - \bar{z}_{f(i)}$, and the slopes are obtained by ordinary least squares on the demeaned data (Guimarães and Portugal, 2010; Correia, 2017). With a single fixed-effect dimension the transformation is exact, so the slope estimates are numerically identical to those of the full-dummy regression, and the fixed effects are recovered as $\hat{\mu}_f = \overline{(\ln P - \hat{\beta}\ln \text{Area} - \mathbf{x}'\hat{\gamma} - \mathbf{d}'\hat{\delta})}_f$. The R^2 we report is that of the full model including the absorbed effects.

4.3. Inference

Listings in the same neighbourhood share unobserved shocks, so our baseline standard errors are clustered at the FSA level ($G = 1{,}155$ clusters), which allows arbitrary within-neighbourhood correlation and heteroskedasticity (Cameron and Miller, 2015). Whether this is the right level is a substantive question (Abadie et al., 2023): if implicit prices are set by provincial markets, or if the sampling of listings is correlated at the provincial scale, clustering by province would be appropriate, but with nine provinces the asymptotic cluster-robust variance is unreliable. We therefore also report heteroskedasticity-robust (HC1), province-clustered and two-way (FSA and province) standard errors, and for the province level we implement the restricted wild-cluster bootstrap-t of Cameron and Miller (2015) with Rademacher weights. With $G = 9$ clusters the $2^9 = 512$ distinct sign vectors are enumerated exhaustively, as recommended by MacKinnon and Webb (2017), so that the bootstrap p-value is exact up to its granularity of $2/512$.

4.4. Specification ladder and variance decomposition

We report a ladder of nested specifications that isolates the marginal explanatory contribution of each block of controls. Model M1 contains the structural attributes only; M2 adds the dwelling-type, ownership and category indicators; M3 adds province fixed effects; M4 replaces the province effects with FSA fixed effects; and M5—the grand model—absorbs FSA effects over the full residential sample. Columns M1–M4 are estimated on the house subsample to keep the structural interpretation clean; M5 is the preferred specification. The gap in R^2 between M3 and M4 measures the explanatory value of resolving location at the neighbourhood rather than the provincial scale.

A sequential ladder attributes the variance explained jointly by two blocks to whichever block enters first. Because larger dwellings are not randomly distributed across neighbourhoods, this attribution rule matters for the structure-versus-location question, and we therefore complement it with an order-free decomposition. Let $R^2(S)$ denote the coefficient of determination of the model

containing the blocks in a set $S \subseteq \mathscr{B} = \{\text{structure}, \text{type/ownership}, \text{province}, \text{FSA}\}$. The Shapley value of block b is

$$\phi_b = \sum_{S \subseteq \mathscr{B} \setminus \{b\}} \frac{|S|!\,(|\mathscr{B}| - |S| - 1)!}{|\mathscr{B}|!} \left[R^2(S \cup \{b\}) - R^2(S)\right], \tag{3}$$

the average marginal contribution of b over all $|\mathscr{B}|!$ orderings (Israeli, 2007; Grömping, 2007); the ϕ_b sum to the R^2 of the full model. Because province effects are nested in FSA effects, $R^2(S \cup \{\text{province}\}) = R^2(S)$ whenever S contains the FSA block, so the province block only receives credit in orderings where it precedes the FSA block. The Owen value treats province and FSA as a single location group and applies Equation (3) to the three groups $\{\text{structure}\}$, $\{\text{type/ownership}\}$ and $\{\text{location}\}$ (Huettner and Sunder, 2012); it is the natural statistic for the structure-versus-location question because it does not split location credit across two nested geographies. Each $R^2(S)$ is computed by absorbing least squares when S contains the FSA block and by OLS otherwise, so the decomposition requires $2^4 = 16$ regressions.

4.5. Functional form, distributional heterogeneity and the urban gradient

To test the constant-elasticity restriction we re-estimate Equation (1) with a quadratic in log area, $\beta_1 \ln \text{Area}_i + \beta_2 (\ln \text{Area}_i)^2$, whose marginal elasticity $\beta_1 + 2\beta_2 \ln \text{Area}$ varies with size, and with a piecewise-linear spline, $\beta_0 \ln \text{Area}_i + \sum_{k=1}^{4} \beta_k \max(\ln \text{Area}_i - c_k, 0)$, with knots c_k at the quintiles of living area, whose segment elasticities $\sum_{j \le k} \beta_j$ do not impose monotonicity. Quantile hedonic regressions (Koenker and Bassett, 1978) estimate, for $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$,

$$\hat{\theta}(\tau) = \arg\min_{\theta} \sum_i \rho_\tau\left(\ln P_i - \mathbf{z}_i'\theta\right), \qquad \rho_\tau(u) = u\left(\tau - \mathbf{1}\{u < 0\}\right), \tag{4}$$

where $\mathbf{z}_i$ stacks the structural attributes, ownership indicators and province effects on the house subsample; standard errors use the Koenker–Bassett kernel sandwich with the Hall–Sheather bandwidth. The urban gradient relates the location component of the price, defined as the residual u_i of the model with structural, type and ownership controls but no location effects, to the great-circle distance d_i from the listing to the nearest of nine metropolitan centres (Toronto, Montreal, Vancouver, Calgary, Ottawa–Gatineau, Edmonton, Winnipeg, Quebec City and Halifax) through $u_i = \kappa + \lambda \ln d_i + \eta_i$, so that a doubling of distance changes the location premium by $100\,(2^{\lambda} - 1)$ per cent; we also allow a quadratic in $\ln d_i$ and metro-specific intercepts. Finally, spatial dependence in the residuals is measured by Moran's I (Moran, 1950),

$$I = \frac{\sum_i \sum_j w_{ij}\, \hat{e}_i \hat{e}_j}{\sum_i \hat{e}_i^2}, \tag{5}$$

with row-standardised k-nearest-neighbour weights w_{ij} based on great-circle distance, computed on random subsamples of 15,000 listings for $k \in \{5, 10, 20\}$, and with significance assessed by 199 random permutations of the residuals; the asymptotic normal approximation of Kelejian and Prucha (2001) yields the same conclusion.

4.6. Retransformation and out-of-sample evaluation

Because the model is estimated in logs, predicted price levels require a retransformation. We use the smearing estimator of Duan (1983),

$$\hat{P}_i = \exp\left(\widehat{\ln P_i}\right) \cdot \hat{\psi}, \qquad \hat{\psi} = \frac{1}{N_{\text{train}}} \sum_{j \in \text{train}} \exp(\hat{\varepsilon}_j), \tag{6}$$

which is consistent without assuming log-normal errors. Predictive accuracy is measured on held-out listings by the out-of-sample R^2 of the log price, $1 - \sum_i (\ln P_i - \widehat{\ln P_i})^2 / \sum_i (\ln P_i - \overline{\ln P})^2$, by the root mean squared log error, and in levels by the median absolute percentage error (MdAPE) $\text{median}_i \, |\hat{P}_i - P_i| / P_i$ and the shares of listings valued within $\pm 10\%$ and $\pm 20\%$ of their asking price. Three validation designs are used. The first is a random 80/20 split of listings, repeated over five seeds, with evaluation restricted to FSAs observed in training since out-of-support fixed effects are not identified; this is the design most valuation studies report and it measures accuracy within the spatial support of the data. The second holds out 20% of the FSAs entirely and values their listings with the training coefficients and an imputed neighbourhood effect, either the mean effect of the training FSAs in the same province or an inverse-distance weighted mean of the five nearest training FSAs; this spatially blocked design measures accuracy in neighbourhoods never observed (Roberts et al., 2017). The third is leave-one-province-out cross-validation of the structural model, which asks whether the structural implicit prices transfer across provincial markets when only a province-specific intercept is allowed.

5. Results

5.1. The explanatory value of location

Figure 3 summarises the explanatory power of the specification ladder on the house subsample. Structural attributes alone (M1) account for 46.7% of the variation in log asking prices. Adding dwelling-type and ownership controls (M2) barely moves the fit, introducing province fixed effects (M3) raises the R^2 to 60.5%, and resolving location at the FSA scale (M4) lifts it to 76.9%; the grand model over all residential dwellings (M5) reaches 77.6%. Moving from provincial to neighbourhood resolution therefore adds about sixteen percentage points of explained variance, and the two location

blocks together add thirty points to a model that already contains structure and dwelling type. Just under a quarter of the variance remains unexplained and is attributable to idiosyncratic pricing, listing strategy and unobserved dwelling quality.

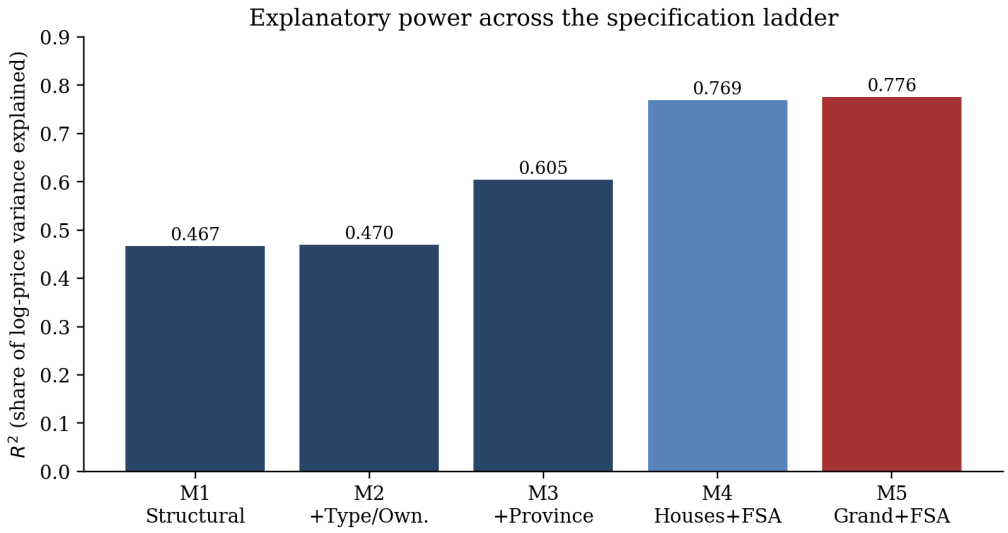


Figure 3. Share of log asking-price variation explained (R^2) across the five nested specifications. M1–M4 are estimated on houses, M5 on all residential dwellings. The jump from M3 (province) to M4 (FSA) quantifies the value of resolving location at the neighbourhood scale.

5.2. Decomposing the variance: structure versus location

The sequential increments of the ladder are a legitimate but order-dependent statement. Table 2 and Figure 4 report, for the full residential sample and for houses, the sequential attribution alongside the Shapley and Owen values of Equation (3). In the sequential ordering that enters structure first, structure receives 43.6 points of the 77.6% explained in the full sample, dwelling type 1.6, province 14.3 and neighbourhood 18.1. When the shared variance is split symmetrically over all orderings, structure falls to 33.0 points and the neighbourhood rises to 28.2, with 7.5 points for dwelling type and ownership and 8.9 for province; grouping the two nested geographies into a single location block, the Owen value assigns 34.0 points to structure, 7.3 to dwelling type and ownership and 36.3 to location. The ten-point gap between the sequential and the Shapley attribution of structure is the variance that structure and location explain jointly, and it has an economic reading: large dwellings are disproportionately located in inexpensive neighbourhoods and condominiums in expensive ones, so that each block "borrows" explanatory power from the other when it enters first.

Three conclusions follow. First, location and structure are of comparable importance in explaining the cross-section of Canadian asking prices: roughly one third of the variance each under symmetric attribution, with the remaining variance split between dwelling type and the residual.

Table 2. Decomposition of the explained variance of log list prices (percentage points of R^2)

Block	Sequential	Shapley	Owen (grouped)
All residential ($N = 142,121$, $R^2 = 0.776$)			
Structure	43.6	33.0	34.0
Dwelling type and ownership	1.6	7.5	7.3
Province	14.3	8.9	—
Neighbourhood (FSA)	18.1	28.2	36.3[a]
Houses ($N = 83,524$, $R^2 = 0.769$)			
Structure	46.7	31.8	33.8
Dwelling type and ownership	0.3	0.9	1.0
Province	13.5	10.8	—
Neighbourhood (FSA)	16.5	33.5	42.2[a]

Notes: *Sequential* is the increment in R^2 when the block is added in the order structure → type/ownership → province → FSA. *Shapley* averages the increment over all 4! orderings of the four blocks (Israeli, 2007; Grömping, 2007). *Owen* treats province and FSA as a single location group and averages over the 3! orderings of the three groups (Huettner and Sunder, 2012). Province dummies are nested in the FSA effects, so the province block adds nothing once FSA effects are present.

[a] Location group (province + FSA).

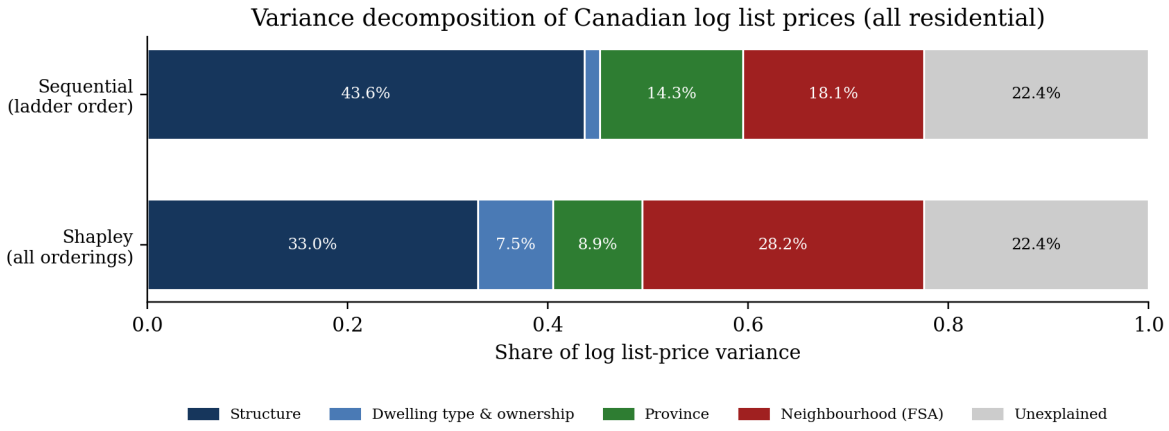


Figure 4. Variance decomposition of log asking prices (all residential dwellings) into structure, dwelling type and ownership, province, neighbourhood (FSA) and the unexplained residual. The upper bar attributes the shared variance to the block that enters first in the ladder order; the lower bar averages over all orderings (Shapley values, Table 2).

Second, the statement that location is the single largest priced component holds when province and neighbourhood are treated as one block (36.3 against 34.0 points), is reversed under the structure-first sequential ordering (32.4 against 43.6), and is reversed again within the four-block Shapley decomposition, where structure (33.0) exceeds the neighbourhood alone (28.2). The question "how much is location worth?" thus has no answer that is independent of the attribution rule, and papers that report only a sequential R^2 ladder overstate the block that enters first. Third, the house subsample behaves similarly but with a larger location share under symmetric attribution (Owen value 42.2 against 33.8 for structure), consistent with houses being more heterogeneous in lot and land content than condominiums.

5.3. Implicit prices of structural attributes

Table 3 reports the regression estimates. The living-area elasticity is 0.68 in the structural model, 0.55 once province effects are included and 0.53 in the grand model: a 10% larger dwelling is listed at roughly 5.3% more, holding the neighbourhood and the other attributes fixed. The decline as location controls tighten is the signature of omitted locational quality: within a province, and a fortiori within a neighbourhood, large dwellings are more often found in cheaper places, so that the unconditional elasticity overstates the implicit price of space. The estimate is precisely determined, with an FSA-clustered standard error of 0.009.

Full bathrooms carry the strongest structural premium after floor space. The grand-model coefficient of 0.112 implies an increase in the asking price of 11.9% ($e^{0.112}-1$) per additional full bathroom at constant living area, and the province-effects model on houses yields 0.148, or 16%. Half bathrooms attract a small negative conditional coefficient (-0.034, about -3.3%); Section 6.2 shows that this effect, unlike those of area and full bathrooms, is not statistically distinguishable from zero once inference allows for province-level dependence, and we do not interpret it further. Bedrooms are economically negligible conditional on living area: the grand-model coefficient of -0.006 per bedroom is statistically different from zero under FSA clustering but implies that subdividing a given floor area into one more room lowers the asking price by about half a percent, and the coefficient is indistinguishable from zero in the house-only columns and under province-level inference. This is the textbook hedonic result that floor space, not room count, carries value (Sirmans et al., 2005, 2006). Parking spaces and storeys have coefficients that are essentially zero; the storeys coefficient in column (4) is statistically significant only because of a handful of implausible counts, and it disappears when the variable is capped (Section 6.1).

The two lot variables must be read jointly. The indicator for available lot information has a coefficient of -0.459 and the log lot area a coefficient of 0.067, so that a listing reporting a median-sized lot of 690 m^2 ($\ln 691 = 6.5$) is priced $-0.459 + 0.067 \times 6.5 = -0.02$ log points relative to a listing with no lot information, while the lot elasticity among listings with a parsed

Table 3. Hedonic regression estimates

	(1) Structural	(2) +Type/Own.	(3) +Province	(4) Grand+FSA
ln living area	0.680***	0.677***	0.545***	0.529***
	(0.022)	(0.022)	(0.016)	(0.009)
Full bathrooms	0.130***	0.136***	0.148***	0.112***
	(0.008)	(0.009)	(0.005)	(0.004)
Half bathrooms	-0.011	-0.017	-0.016	-0.034***
	(0.013)	(0.013)	(0.011)	(0.007)
Bedrooms	0.001	-0.001	-0.001	-0.006***
	(0.004)	(0.004)	(0.003)	(0.002)
Parking spaces	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Storeys	-0.036*	-0.037*	-0.034*	0.001***
	(0.021)	(0.022)	(0.020)	(0.000)
Has lot info	0.049	0.047	-0.228***	-0.459***
	(0.050)	(0.051)	(0.030)	(0.018)
ln(1+lot m^2)	0.015**	0.014**	0.034***	0.067***
	(0.007)	(0.007)	(0.004)	(0.003)
Dwelling-type FE	No	Yes	Yes	Yes
Ownership FE	No	Yes	Yes	Yes
Province FE	No	No	Yes	—
FSA fixed effects	No	No	No	Yes
R^2	0.467	0.470	0.605	0.776
Observations	83,524	83,524	83,524	142,121

Notes: Dependent variable is ln(list price). Columns (1)–(3) use the house subsample; column (4) is the grand model over all residential dwellings with 1,155 absorbed FSA fixed effects (incl. nine province-level residual categories pooling FSAs with fewer than 25 listings). Columns (2)–(4) also include dwelling-type, ownership and listing-category dummies (Table 4). Cluster-robust standard errors (by FSA) in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

lot is 0.067. Because 84% of condominiums and 7% of houses report no lot, the indicator is close to a dwelling-type marker and its coefficient has no interpretation on its own. The lot elasticity is modest for two reasons: the field is noisily parsed from free text, and within a neighbourhood lot variation is compressed.

Figure 5 presents the structural implicit prices of the province-effects model with 95% cluster-robust confidence intervals, making visually plain the dominance of living area and bathrooms and the near-zero conditional effects of bedrooms, parking and storeys.

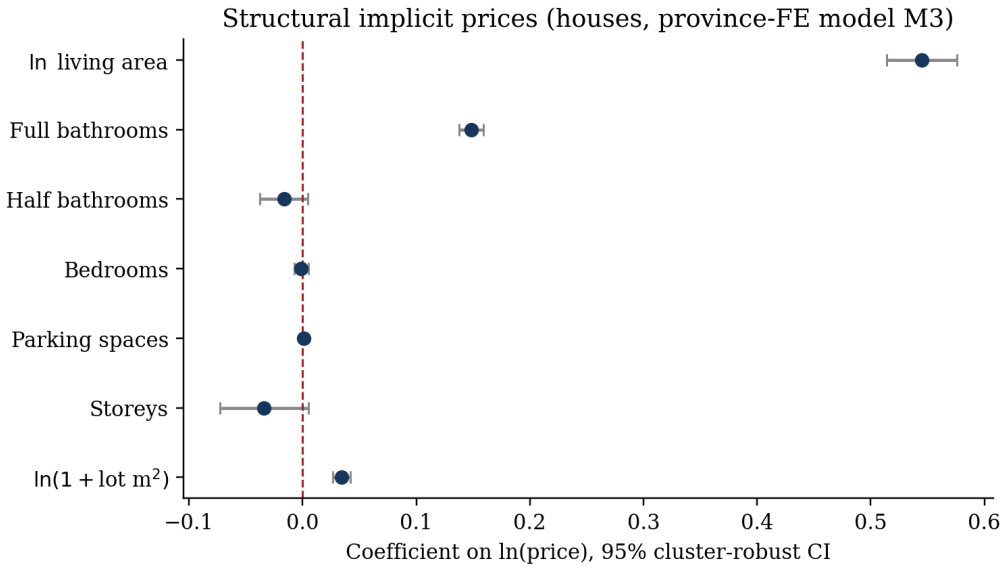


Figure 5. Marginal implicit list prices of structural attributes (houses, province-effects model M3) with 95% cluster-robust confidence intervals, on the log-price scale.

5.4. Dwelling type and ownership

Table 4 reports the dwelling-type and ownership effects of the grand model with the percentage effects of Equation (2). Conditional on living area, the other structural attributes and the neighbourhood, a house is listed 46% above an apartment of the same size, a triplex 43% and a fourplex 39% above, a duplex 16% above, and a row or townhouse at the same price as an apartment; manufactured homes are not significantly different from apartments once their neighbourhoods and sizes are controlled. The house premium is large and deserves a precise reading: it is the price of the land and the exclusive use that a detached dwelling embodies over and above the parsed lot variables, and it is estimated from neighbourhoods that contain both houses and apartments, that is, from mixed urban FSAs. Relative to co-operative ownership, freehold and condominium or strata titles both command a premium of about 10%, and leasehold a discount of 12%, in line with the restrictions that co-operative and leasehold tenure place on resale.

Figure 6 shows the unconditional price–size gradient by dwelling type: median asking price rises concavely with living area for both houses and condominiums, and the semi-log specification linearises this curvature. Unconditionally, condominiums list *above* houses of the same size, a compositional effect of their concentration in expensive metropolitan neighbourhoods, whereas conditional on neighbourhood and ownership houses command the premium reported in Table 4. The contrast illustrates the identifying role of the neighbourhood effects.

Table 4. Dwelling-type and ownership effects in the grand model

	Coefficient	SE	% effect
Dwelling type (ref.: apartment)			
House	0.377***	(0.041)	+45.7
Row / townhouse	0.002	(0.008)	+0.1
Duplex	0.146***	(0.016)	+15.8
Triplex	0.359***	(0.056)	+43.0
Fourplex	0.328***	(0.048)	+38.7
Manufactured home	-0.040	(0.050)	-4.0
Other	0.100**	(0.040)	+10.4
Ownership (ref.: co-op / co-ownership)			
Freehold	0.102***	(0.032)	+10.7
Condominium / strata	0.097***	(0.031)	+10.2
Leasehold	-0.129***	(0.049)	-12.2
Not reported	0.101	(0.062)	+10.4
Other	-0.142**	(0.055)	-13.3

Notes: Grand model (M5), all residential dwellings, FSA fixed effects absorbed, FSA-clustered standard errors. The percentage effect is the Kennedy (1981) estimator $100\,[\exp(\hat{\gamma} - \frac{1}{2}\widehat{V}(\hat{\gamma})) - 1]$, which corrects the Halvorsen and Palmquist (1980) transformation for estimation error. Effects are relative to the omitted category and conditional on living area, the other structural attributes and the neighbourhood. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

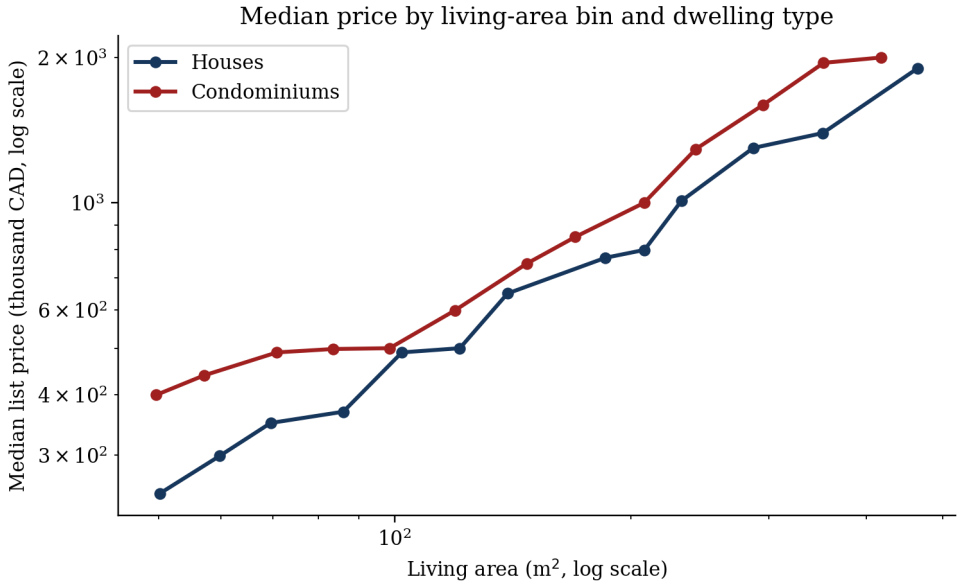
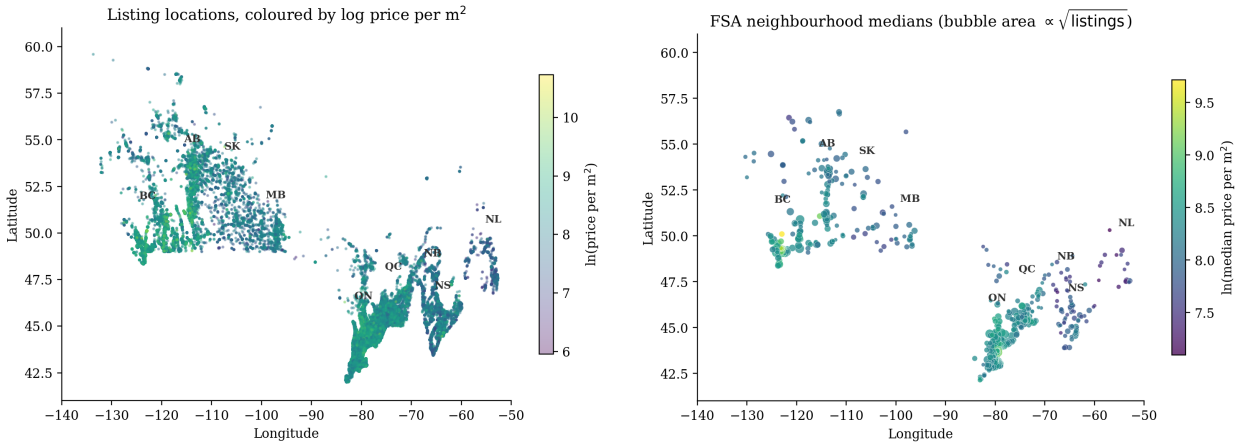


Figure 6. Median asking price by living-area bin and dwelling type (unconditional). The concave gradient is linearised by the semi-log specification; the condominium premium at given size is compositional and reverses once location is controlled.

5.5. The geography of housing value

Figure 7 maps the spatial structure of value directly. The outline of populated Canada emerges from the listing coordinates; the Vancouver corridor and the Toronto–Golden Horseshoe area sit at the top of the price-per-square-metre distribution, while the Prairies and Atlantic Canada anchor the bottom. The right panel aggregates to FSA medians, the geographic unit absorbed in the grand model.



(a) All listings **(b)** FSA neighbourhood medians

Figure 7. Spatial distribution of housing value. Colour encodes the log asking price per m^2; bubble area in panel (b) is proportional to $\sqrt{\text{listings}}$. Labels mark provinces with substantial samples; the data cover nine provinces, while the territories and Prince Edward Island contain no listings.

To translate location into a dollar statement we recover each FSA's fixed effect from the grand model—its price premium net of structure, dwelling type and ownership—and express it relative to the median FSA among the 870 neighbourhoods with at least 50 listings (Figure 8). The highest-valued neighbourhoods are all in British Columbia's Lower Mainland and Sea-to-Sky corridor: V6S (Dunbar–Point Grey, +226%) and V6T (University Endowment Lands, +201%) on Vancouver's west side, V8E (Whistler, +197%), V6G (Vancouver West End, +185%), V7T (West Vancouver, +179%) and a further seven Vancouver FSAs between +150% and +175%; the first Ontario neighbourhood, M5R (the Annex, Toronto), ranks fourteenth at +142%. The lowest-valued neighbourhoods are in rural Newfoundland (A0B and A0G, −69% and −64%), rural Saskatchewan (S0N, S0C, S0H, S0A, around −62%), northern and rural Manitoba (R8N Thompson, R0M), New Brunswick (E3Y Grand Falls, E2A Bathurst) and the north-east of British Columbia (V0C), with one urban exception, R2W in Winnipeg's North End (−62%). An otherwise identical dwelling is therefore listed at about 3.3 times the national-median neighbourhood level in Dunbar–Point Grey and at 0.31 times that level in rural Newfoundland, a ratio of about ten between the extremes and of 6.7 between the first and ninety-ninth percentiles of the premium distribution. No single structural attribute moves the asking price by anything approaching this range.

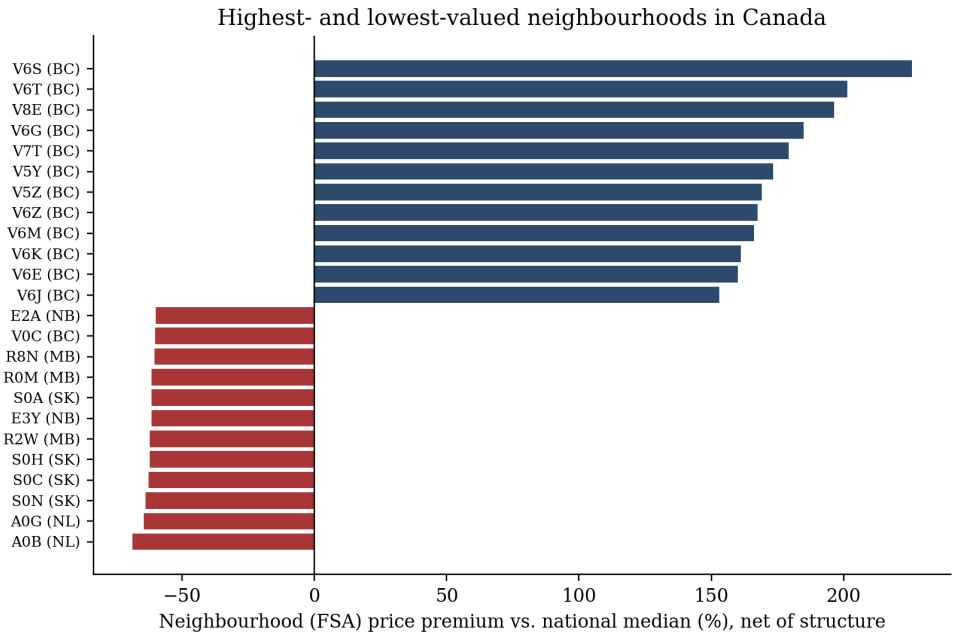


Figure 8. Highest- and lowest-valued neighbourhoods (FSAs) in Canada, measured as the estimated location premium relative to the median neighbourhood, net of structure, dwelling type and ownership. Only FSAs with at least 50 listings are shown; provinces are assigned from the FSA.

5.6. Nonlinearity in floor space

The constant-elasticity assumption is convenient but restrictive. Re-estimating the grand model with a quadratic in log living area yields $\hat{\beta}_1 = 1.147$ and $\hat{\beta}_2 = -0.063$ (FSA-clustered t-statistic of the quadratic term above 10), so that the marginal elasticity declines from about 0.63 for a 60 m^2 dwelling to about 0.41 for a 350 m^2 home (Figure 9). The piecewise-linear spline of Table 5, which does not impose monotonicity, qualifies this picture: the elasticity is 0.57 below 87 m^2, peaks at 0.65 between 87 and 113 m^2, is 0.58 up to 149 m^2, and falls to 0.41–0.48 above 149 m^2. Diminishing returns are thus concentrated in the upper half of the size distribution rather than smooth, and their economic magnitude is modest: the spline raises the R^2 from 0.7756 to 0.7767, and the constant elasticity of 0.53 is an adequate average for most purposes. The pattern is consistent with the curvature that McMillen and Redfearn (2010) detect nonparametrically and with the lower value of marginal space in large dwellings, where the land component and the fixed cost of the structure dominate.

5.7. The urban price gradient

Urban theory predicts that, holding structure fixed, value declines with distance from employment centres (Alonso, 1964; Muth, 1969; Mills, 1967; Ahlfeldt et al., 2015). Table 6 and Figure 10 relate the location component of the asking price—the residual of the model without location effects—to

Table 5. Living-area elasticity by size segment (piecewise-linear spline, grand model)

Segment	Living area (m^2)	Elasticity
1	46–87	0.571 (0.021)
2	87–113	0.650 (0.021)
3	113–149	0.584 (0.022)
4	149–209	0.409 (0.017)
5	209–465	0.477 (0.013)
R^2: linear / quadratic / spline	0.7756 / 0.7765 / 0.7767	

Notes: Linear spline in ln(area) with knots at the quintiles of living area, all other grand-model controls and FSA fixed effects included. Each entry is the elasticity of price with respect to area within the segment; FSA-clustered standard errors in parentheses.

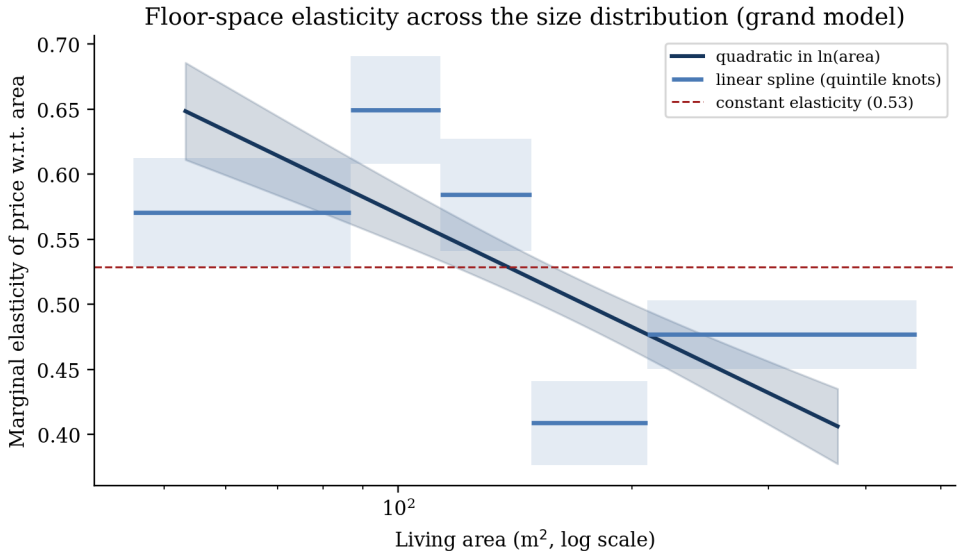


Figure 9. Marginal elasticity of the asking price with respect to living area as a function of dwelling size, from the grand model with a quadratic in log area (dark line, 95% FSA-clustered band) and with a piecewise-linear spline with knots at the size quintiles (horizontal segments, 95% bands). The dashed line is the constant-elasticity estimate.

the great-circle distance to the nearest of nine metropolitan centres. Listings within 5 km of a centre carry a mean location premium of 21% over the structure-only benchmark; the premium is 12% at 5–10 km, close to zero at 10–20 km, 11% at 20–40 km and 6% at 40–80 km, turns negative beyond about 80 km, and reaches -13% at 160–320 km and -38% beyond 320 km. The log-linear fit implies $\hat{\lambda} = -0.082$ (FSA-clustered standard error 0.007), so that a doubling of the distance to the nearest metro is associated with a 5.5% lower location premium ($2^{-0.082} - 1$) and a tenfold increase with a 24% lower premium. The slope is essentially unchanged when nearest-metro intercepts are added (-0.079), which shows that it is not driven by differences in price level across metropolitan systems, and the quadratic specification indicates that the gradient steepens with distance: it is flat within the first few kilometres and reaches about -0.10 per log unit at 40 km and -0.21 at 500 km.

Table 6. The urban price gradient: location premium and distance to the nearest metro

Panel A: mean location premium by distance band			
Distance band	Listings	Median km	Premium (%)
0–5 km	12,979	2	+20.7
5–10 km	11,358	8	+11.8
10–20 km	20,373	14	+1.6
20–40 km	21,594	28	+11.0
40–80 km	20,927	60	+6.5
80–160 km	25,565	104	-1.0
160–320 km	19,856	218	-13.2
>320 km	9,469	485	-38.0
Panel B: regressions of the location premium on $\ln(\textit{distance})$			
	Slope	SE	R^2
Pooled	-0.0821	(0.0074)	0.083
Nearest-metro intercepts	-0.0793	(0.0045)	0.336
Quadratic: $\ln d$	0.0733	(0.0275)	0.106
$(\ln d)^2$	-0.0229	(0.0037)	

Notes: The location premium is the residual of the model with structural, type and ownership controls but no location effects; the premium in Panel A is $100\,(e^{\bar{u}} - 1)$ for the band mean $\bar{u}$. Distance is the great-circle distance to the nearest of nine metropolitan centres (Toronto, Montreal, Vancouver, Calgary, Ottawa–Gatineau, Edmonton, Winnipeg, Quebec City, Halifax); listings within 0.5 km of a centre are excluded from Panel B. FSA-clustered standard errors.

Two caveats apply. The non-monotonicity between 10 and 40 km reflects the heterogeneity of Canadian suburbs—inner suburbs of Montreal and Toronto with modest premia sit at the same distance as high-value suburbs of Vancouver—and the fact that a single point per metropolitan area is a crude representation of polycentric employment. And the gradient explains only 8% of the variance of the location component (34% with metro intercepts): distance to the centre is one

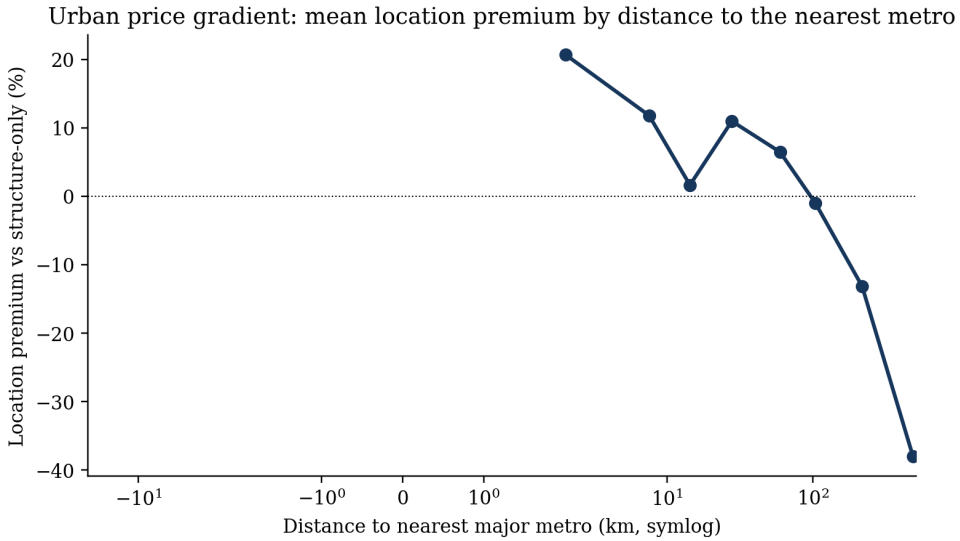


Figure 10. Urban price gradient: mean location premium of the asking price (relative to a structure-only benchmark) by distance band to the nearest of nine major Canadian metropolitan centres. The horizontal axis is on a symmetric-log scale.

determinant of location value among many, which is precisely why the grand model absorbs location non-parametrically rather than imposing a gradient (Combes and Gobillon, 2015).

6. Robustness, Inference, Heterogeneity and Validation

6.1. Stability of the implicit prices across samples

Table 7 re-estimates the grand model on eight alternative samples and specifications. The living-area elasticity ranges from 0.49 (houses only) to 0.62 (condominiums only) and lies between 0.51 and 0.53 in every cut that keeps both dwelling types; the full-bathroom premium ranges from 0.096 (Ontario, Quebec and British Columbia only) to 0.121 (condominiums). Tightening the price trim to the 0.5 and 99.5 percentiles, dropping FSAs with fewer than fifty listings, capping storey and parking counts at their 99th percentiles and excluding the nine pooled residual categories all leave the headline coefficients within one or two standard errors of the baseline. The condominium subsample exhibits both the highest fit ($R^2 = 0.81$) and the largest size elasticity, consistent with condominium prices being more tightly determined by floor area and location and less by lot and structure features that the data do not record. Two cuts deserve comment. Capping storeys and parking lowers the elasticity to 0.51 and raises the R^2 to 0.78, confirming that the extreme counts are data errors rather than information; and excluding the pooled categories raises the R^2 to 0.785, as expected when heterogeneous residual groups are removed, without moving the coefficients. The implicit list prices are therefore not artefacts of a particular sample definition.

Table 7. Robustness of key implicit prices across samples

Sample / specification	N	R^2	ln area	Full bath	ln lot
Baseline (all residential)	142,121	0.776	0.529	0.112	0.067
			(0.009)	(0.004)	(0.003)
Houses only	83,524	0.769	0.487	0.107	0.084
			(0.010)	(0.004)	(0.003)
Condominiums only	57,853	0.812	0.620	0.121	0.005
			(0.017)	(0.005)	(0.004)
Price trimmed 0.5/99.5%	140,714	0.773	0.519	0.110	0.064
			(0.009)	(0.004)	(0.002)
FSAs with $\geq$50 listings	131,780	0.769	0.525	0.113	0.066
			(0.010)	(0.004)	(0.003)
ON/QC/BC only	108,840	0.753	0.521	0.096	0.063
			(0.011)	(0.004)	(0.003)
Storeys/parking capped at P99	142,121	0.779	0.513	0.110	0.059
			(0.010)	(0.004)	(0.003)
Excl. pooled residual FSAs	136,736	0.785	0.522	0.110	0.069
			(0.009)	(0.004)	(0.002)

Notes: Each row re-estimates the grand FSA-fixed-effects model on a different sample. Cluster-robust (FSA) standard errors in parentheses. The last two rows cap storeys and parking counts at their 99th percentiles and drop the province-level categories that pool FSAs with fewer than 25 listings.

6.2. Sensitivity of inference to the level of clustering

Table 8 reports the grand-model coefficients with heteroskedasticity-robust, FSA-clustered, province-clustered and two-way clustered standard errors, together with the p-value of the province-level wild-cluster bootstrap. Clustering by FSA inflates the HC1 standard errors by a factor of 2.5 to 3, and clustering by province inflates them by a further factor of four, so that the living-area elasticity has a standard error of 0.004 under HC1, 0.009 under FSA clustering and 0.037 under province clustering. The ranking of the coefficients is unaffected: living area, full bathrooms and lot area have province-level t-statistics of 14, 9 and 11 and bootstrap p-values at the minimum attainable value of 0.004, so that their implicit prices are statistically robust even under the most conservative design. The half-bathroom and bedroom coefficients, by contrast, have province-level t-statistics of -1.1 and -0.7 and bootstrap p-values of 0.52 and 0.58: their statistical significance in Table 3 depends on treating neighbourhoods as independent, and we accordingly describe them as economically and statistically negligible. The two-way estimator is indistinguishable from the province-level one, which indicates that the province-level dependence dominates.

Table 8. Grand-model coefficients under alternative variance estimators

		Standard error				Wild-cluster
	Coef.	HC1	FSA ($G = 1,155$)	Province ($G = 9$)	Two-way	p-value
ln living area	0.529	0.0037	0.0094	0.0373	0.0375	0.004
Full bathrooms	0.112	0.0016	0.0038	0.0120	0.0121	0.004
Half bathrooms	-0.034	0.0021	0.0066	0.0315	0.0316	0.523
Bedrooms	-0.006	0.0011	0.0021	0.0084	0.0085	0.582
ln(1+lot m^2)	0.067	0.0011	0.0025	0.0060	0.0061	0.004

Notes: Grand model (M5) estimated on the FSA-demeaned data. HC1 is the heteroskedasticity-robust estimator; FSA and province cluster by neighbourhood and by province; two-way clusters by both. The last column is the p-value of a restricted wild-cluster bootstrap-t test at the province level with Rademacher weights and full enumeration of the $2^9 = 512$ sign vectors (Cameron and Miller, 2015; MacKinnon and Webb, 2017); its smallest attainable value is $2/512 = 0.004$.

6.3. Implicit prices along the price distribution

OLS recovers the implicit price at the conditional mean, but buyers and sellers at the bottom and top of the market may value attributes differently (Zietz et al., 2008). Table 9 and Figure 11 report quantile hedonic regressions of Equation (4) at the 10th to 90th percentiles of the asking price on the house subsample with province effects. The living-area elasticity is essentially flat across the distribution, between 0.57 and 0.59, so that floor space is valued proportionally in cheap and expensive segments alike. The full-bathroom premium declines monotonically from 0.19 log points at the first decile to 0.10 at the ninth: an additional bathroom matters most, in proportional terms, in the lower segment of the market, where it is most likely to signal a renovated or larger-than-typical dwelling. The lot elasticity rises from 0.006 to 0.053 across the distribution, consistent with land being a luxury component of value whose implicit price rises with the price of the dwelling. The conditional-mean estimates of Table 3 are thus representative for floor space but mask economically sensible distributional variation for bathrooms and land.

6.4. Residual spatial autocorrelation

A central justification for the neighbourhood fixed effects is that they should absorb the spatial dependence that pervades hedonic residuals (Anselin, 1988; Dubin, 1988; Basu and Thibodeau, 1998). We test this directly with Moran's I of Equation (5). On a random subsample of 15,000 listings with $k = 10$ nearest neighbours, the residuals of the model without location effects display strong positive spatial autocorrelation, $I = 0.50$ (permutation z-score 141, $p = 0.005$, the smallest value attainable with 199 permutations): nearby dwellings are mispriced in the same direction, the signature of omitted location. The grand model with FSA fixed effects cuts this to $I = 0.06$ ($z = 16$),

Table 9. Quantile hedonic estimates across the price distribution

Quantile	ln area	Full bath	ln lot
$\tau = 0.1$	0.566	0.188	0.006
	(0.009)	(0.004)	(0.002)
$\tau = 0.25$	0.569	0.160	0.012
	(0.005)	(0.002)	(0.001)
$\tau = 0.5$	0.570	0.142	0.023
	(0.005)	(0.002)	(0.001)
$\tau = 0.75$	0.588	0.122	0.039
	(0.006)	(0.002)	(0.001)
$\tau = 0.9$	0.567	0.101	0.053
	(0.010)	(0.004)	(0.002)
OLS	0.547	0.148	0.035

Notes: House subsample, province fixed effects (province derived from the FSA), structural controls and ownership dummies. Kernel-based heteroskedasticity-robust standard errors (Epanechnikov kernel, Hall–Sheather bandwidth) in parentheses; the OLS row reports HC1 standard errors.

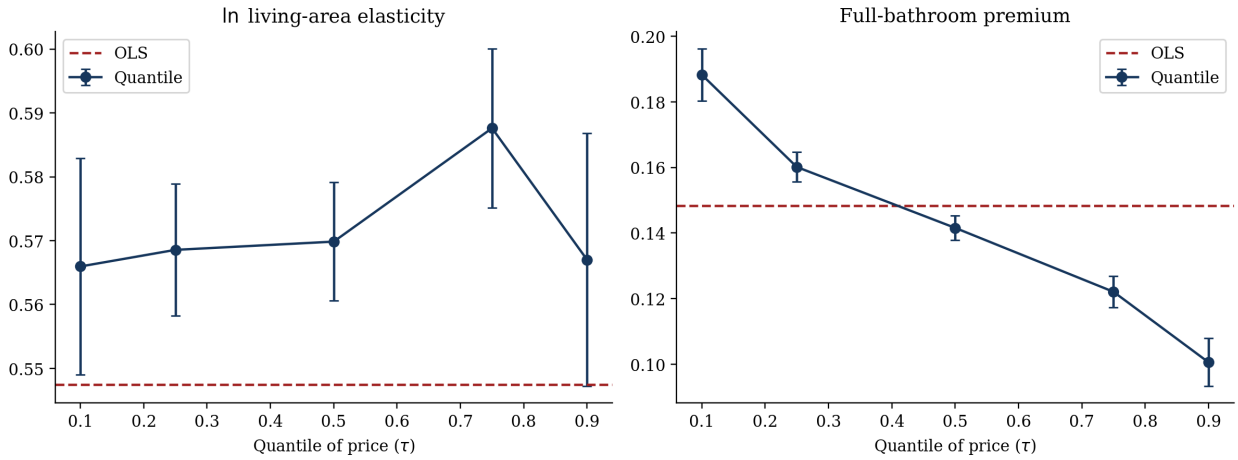


Figure 11. Implicit list prices across the conditional price distribution. Quantile estimates (with 95% confidence intervals) of the living-area elasticity (left) and the full-bathroom premium (right); the dashed line is the OLS estimate.

a reduction of 87%. Figure 12 contrasts the two Moran scatterplots, and Table A1 in the appendix shows that the conclusion is insensitive to the weight matrix and the subsample: for $k = 5$ the statistic falls from 0.52 to 0.11 and for $k = 20$ from 0.47 to 0.03, and a second subsample gives the same figures to the second decimal. The residual autocorrelation that remains is within-neighbourhood dependence that the FSA grid cannot capture and that could be addressed by finer geographies or an explicit spatial error model (Kelejian and Prucha, 1998; LeSage and Pace, 2009); it is an order of magnitude smaller than the dependence that the fixed effects remove, which supports the design over a parametric spatial-lag alternative (Gibbons and Overman, 2012). The permutation test rejects the null of no residual autocorrelation in every case, so the FSA effects reduce but do not eliminate spatial dependence, and the clustered inference of Section 4.3 is required.

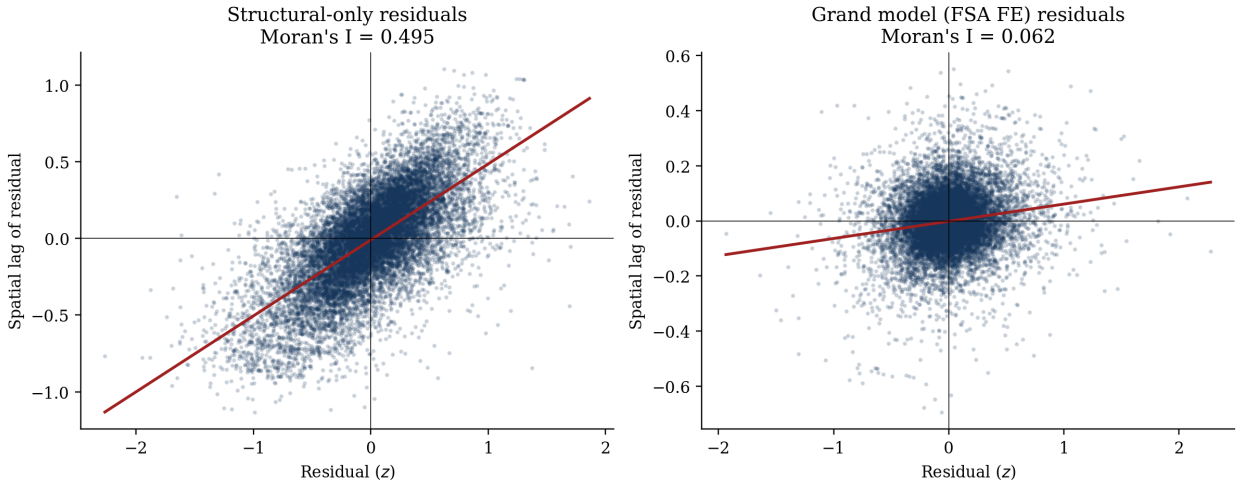


Figure 12. Moran scatterplots of model residuals against their spatial lag ($k = 10$ nearest neighbours, 15,000-listing subsample). Left: model without location effects. Right: grand model with FSA fixed effects. The slope is Moran's I.

6.5. Heterogeneity across provinces

Estimating the within-FSA model province by province reveals economically meaningful heterogeneity in the size elasticity (Figure 13). The elasticity is lowest in Newfoundland and Labrador (0.41), Ontario (0.47), British Columbia (0.49) and Nova Scotia (0.50), intermediate in Quebec (0.57), Alberta (0.59) and New Brunswick (0.61), and highest in Manitoba (0.66) and Saskatchewan (0.73). For the large markets the pattern is intuitive: where land and location dominate value, as in Toronto and Vancouver, an extra square metre of structure adds proportionally less, whereas in lower-priced markets the building itself is a larger share of value and floor space carries more weight. The low elasticity in Newfoundland and Labrador, a cheap market, does not fit this reading and is estimated on 1,274 listings with a wide confidence interval; it may reflect the prevalence of

large, inexpensive rural dwellings whose size is weakly priced. The bathroom premium is largest in the Atlantic provinces and Manitoba (0.20–0.26) and smallest in British Columbia (0.09) and Ontario (0.12), mirroring the elasticity pattern.

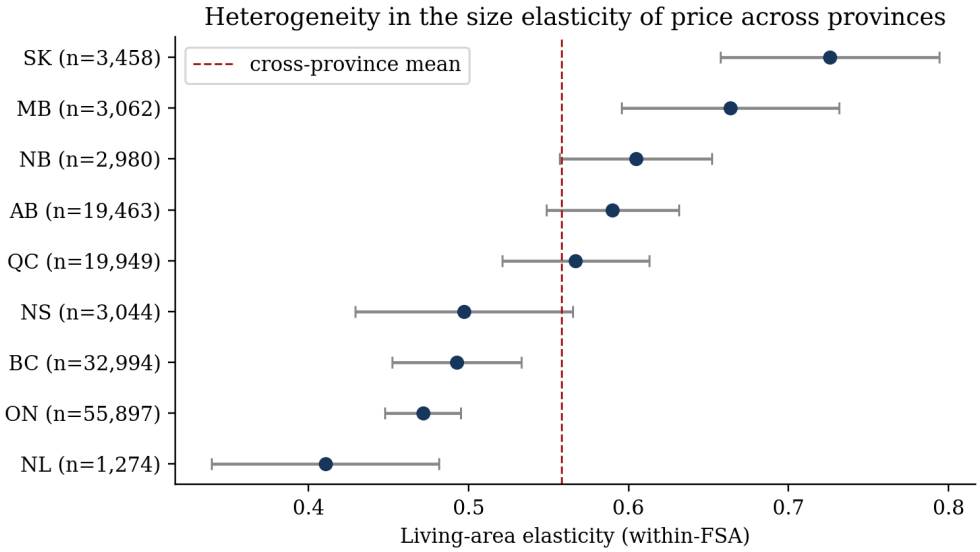


Figure 13. Living-area elasticity of the asking price by province, estimated within FSAs, with 95% cluster-robust confidence intervals. Provinces are assigned from the FSA.

6.6. Do structural prices transfer across provinces?

As a demanding test of transferability we perform leave-one-province-out cross-validation: the model with structural, type and ownership controls is estimated on all provinces but one and used to predict the held-out province, allowing only a province-specific intercept. Table 10 reports the within-province R^2. The structural implicit prices transfer reasonably to Manitoba, Alberta, New Brunswick, Ontario, Saskatchewan and Nova Scotia (within-province R^2 of 0.46–0.52) and to Newfoundland and Labrador (0.40), but poorly to Quebec (0.19) and not at all to British Columbia (-0.49): a model estimated on the rest of Canada predicts British Columbia's asking prices worse than the provincial mean does. The mean held-out R^2 is 0.34. The British Columbia result is the mirror image of the heterogeneity just documented—the province has the second-lowest size elasticity and the lowest bathroom premium, so that national coefficients systematically overprice large, well-equipped dwellings there—and it is reinforced by the extreme dispersion of location value within the province, which a single intercept cannot absorb. The lesson is twofold: structural implicit prices are broadly similar across most of the country, but they are not universal constants, and province-specific (or, better, neighbourhood-specific) intercepts are not sufficient to transfer a hedonic model into the most expensive market.

Table 10. Leave-one-province-out spatial cross-validation

Held-out province	N	Within-province R^2
MB	3,062	0.521
AB	19,463	0.519
NB	2,980	0.515
ON	55,897	0.483
SK	3,458	0.474
NS	3,044	0.464
NL	1,274	0.404
QC	19,949	0.185
BC	32,994	-0.494
Mean	—	0.341

Notes: The structural hedonic model is estimated on all provinces except one and used to predict the held-out province; a province-specific intercept is allowed (the price *level* is not identified out of sample), so the metric captures whether the *structural* implicit prices transfer across space.

6.7. Out-of-sample valuation accuracy

A hedonic model that fits in sample need not value well out of sample. Table 11 and Figure 14 report performance on a randomly held-out 20% of listings whose neighbourhoods are observed in training. The model attains an out-of-sample R^2 of 0.772 on the log asking price, essentially identical to its in-sample fit, which indicates negligible over-fitting despite the thousand-plus location effects. In levels, after Duan smearing, the median absolute percentage error is 15.6%, the mean 21.9%, and 60% of held-out dwellings are valued within $\pm 20\%$ of their asking price (34% within $\pm 10\%$). Table 12 shows that these figures are stable across five random splits (MdAPE $15.5 \pm 0.2\%$, R^2 0.769 ± 0.010). They are within the accuracy bands reported for mass-appraisal models (McCluskey et al., 2013) and establish the transparent specification as a credible benchmark, with the caveat that the target is the asking price, which is smoother and more predictable than the transaction price.

Random holdouts measure accuracy within the spatial support of the training data. The right-hand columns of Table 12 answer a different question: how well does the model value listings in neighbourhoods it has never seen? Holding out 20% of the FSAs (229 neighbourhoods and about 27,000 listings per replication) and imputing their effect as the mean effect of the training FSAs in the same province yields an R^2 of 0.52 and a MdAPE of 25%, with only 41% of listings within $\pm 20\%$; imputing it instead as the inverse-distance weighted mean of the five nearest training FSAs recovers most of the lost accuracy, with an R^2 of 0.74, a MdAPE of 17.3% and 56% of listings within $\pm 20\%$. The gap between the two imputations is a direct measure of the spatial autocorrelation of the neighbourhood effects themselves: location value is smooth enough across adjacent FSAs

Table 11. Out-of-sample valuation performance (80/20 split)

Metric	Value
Training listings	113,696
Test listings	28,425
Out-of-sample R^2 (log price)	0.772
RMSE (log points)	0.279
Median absolute % error	15.6%
Mean absolute % error	21.9%
Share priced within ±10%	34.1%
Share priced within ±20%	60.1%

Notes: Model trained on a random 80% of listings and scored on the held-out 20% (restricted to neighbourhoods observed in training). Prices back-transformed with Duan's smearing estimator. Table 12 reports repeated splits and spatial (FSA-level) holdouts.

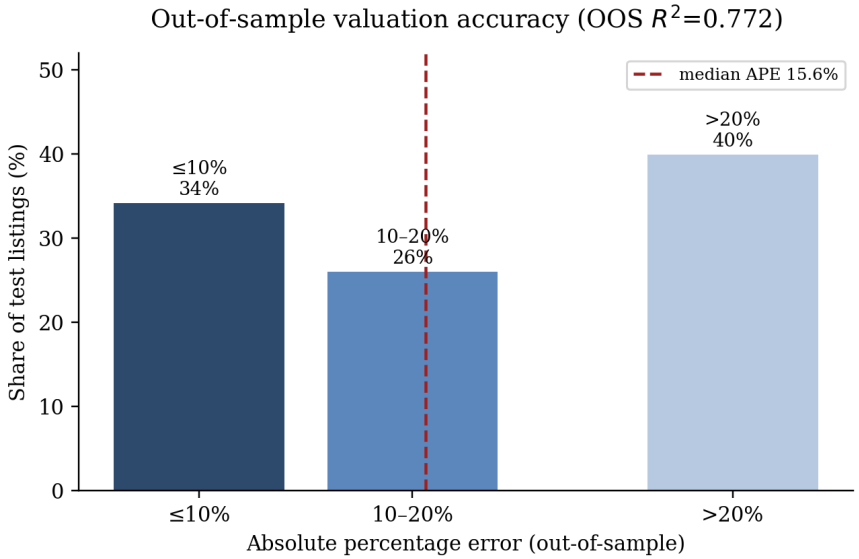


Figure 14. Out-of-sample valuation accuracy on the held-out test fold: the share of listings valued within ±10%, between 10 and 20%, and beyond 20% of the asking price.

that a nearest-neighbour interpolation of the fixed effects loses only about two points of MdAPE relative to observing the neighbourhood, whereas a province-level prior loses ten. For valuation practice this means that the model generalises to new neighbourhoods provided their location effect is interpolated spatially rather than set to a regional mean, and it illustrates the general point of Roberts et al. (2017) that random cross-validation overstates the accuracy to be expected in locations outside the training support.

Table 12. Random versus spatial holdouts: valuation accuracy

Metric	Random 80/20 (within support)	Held-out FSAs (20%)	
		Province-mean effect	IDW of 5 nearest FSAs
Out-of-sample R^2 (log price)	0.769 (0.010)	0.520 (0.034)	0.743 (0.014)
RMSE (log points)	0.280 (0.005)	0.402 (0.013)	0.294 (0.008)
Median absolute % error	15.5 (0.2)	25.0 (1.0)	17.3 (0.6)
Mean absolute % error (top 1% trimmed)	20.4 (0.1)	32.3 (1.0)	22.5 (0.9)
Share within $\pm$10%	34.3 (0.4)	21.2 (1.0)	31.2 (1.2)
Share within $\pm$20%	60.3 (0.4)	40.9 (1.7)	56.1 (1.6)

Notes: Means over five replications, standard deviations across replications in parentheses. Column 1: random 80/20 splits of listings, evaluation restricted to FSAs observed in training. Columns 2–3: 20% of the FSAs (229 of the non-pooled neighbourhoods) are held out entirely; their listings are valued with the structural coefficients from the training FSAs and a neighbourhood effect imputed either as the mean effect of the training FSAs in the same province or as the inverse-distance-weighted mean of the five nearest training FSAs (centroid distance). Prices back-transformed with Duan's smearing estimator. The trimmed mean excludes the largest 1% of errors in each replication: a few listings with implausible parking or storey counts (up to 3,620 spaces) otherwise dominate the untrimmed mean.

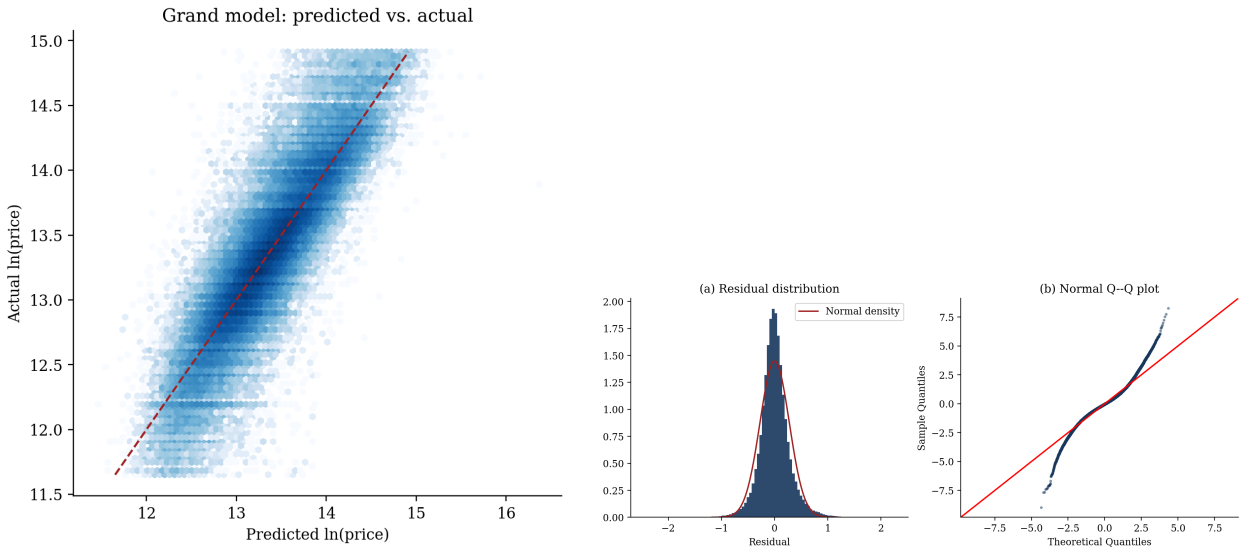
6.8. Model fit and residual behaviour

Panel (a) of Figure 15 plots predicted against actual log asking prices for the grand model; the cloud hugs the 45-degree line. The residuals (panel (b)) are approximately Gaussian and centred on zero with mild heavy tails, typical of housing data and accommodated by the cluster-robust inference. We interpret the remaining dispersion as a combination of genuine idiosyncratic pricing, listing strategy, and dwelling-level quality (age, renovation, finish) that the data do not record.

7. Discussion

7.1. Magnitudes in the light of the literature

The location share. Our decomposition—structure and location each explaining roughly one third of the variance of log asking prices under symmetric attribution, with location slightly ahead



(a) Predicted vs. actual **(b)** Residual diagnostics

Figure 15. Grand-model (M5) goodness of fit and residual behaviour.

when province and neighbourhood are grouped—is a cross-sectional, listing-level statement, but its order of magnitude is consistent with what aggregate approaches find from the opposite direction. Davis and Heathcote (2007) estimate that land accounts for roughly 46% of the value of the U.S. housing stock, Davis and Palumbo (2008) and Albouy et al. (2018) document that the share ranges from under 20% in inland cities to over 70% in coastal metropolitan areas, and Knoll et al. (2017) attribute about 80% of the global house-price increase since 1950 to land. Land and "location" are not identical objects—our FSA premia capitalise amenities and access as well as physical land scarcity (Cheshire and Sheppard, 1995; Albouy, 2016), and a variance share is not a value share—but both measurements say that the dwelling itself is at most half of housing value. The tenfold span of our neighbourhood premia echoes the dispersion of metropolitan land values in Albouy et al. (2018) and the French gradients of Combes et al. (2019), and it is the pattern predicted where supply constraints bind heterogeneously across locations (Saiz, 2010; Gyourko et al., 2013): Vancouver, hemmed in by sea, mountains and an agricultural land reserve, anchors the top of the distribution, as superstar-city dynamics would predict.

The decomposition also carries a methodological message. The sequential R^2 ladder that hedonic papers routinely report credits the variance shared by structure and location to whichever block enters first, and in our data the two blocks share about ten percentage points. The Shapley–Owen decomposition makes the attribution rule explicit; it is cheap to compute with absorbing least squares and we recommend it whenever the relative importance of regressor blocks is itself a finding.

Structural implicit prices. The living-area elasticity of 0.53 sits inside the range catalogued by Sirmans et al. (2005) and the meta-analysis of Sirmans et al. (2006), and its decline from 0.68 to 0.53 as location controls tighten is the classic signature of omitted locational quality that within-neighbourhood designs are built to remove (Kiel and Zabel, 2008). The near-zero conditional bedroom coefficient reproduces one of the most robust findings of that meta-literature: floor space, not room count, carries value. Our full-bathroom premium (12% in the grand model, 16% on houses with province effects) is at the upper end of published estimates; we attribute this to bathrooms proxying for unobserved renovation status and finish quality in listing data that lack an age variable, and the declining quantile profile—0.19 log points at the first decile against 0.10 at the ninth—is consistent with that proxy role being strongest in the lower segment of the market. The mild diminishing marginal elasticity of floor space (from about 0.63 at 60 m^2 to 0.41 at 350 m^2 under the quadratic, concentrated above 150 m^2 under the spline) gives parametric confirmation of the curvature that McMillen and Redfearn (2010) detect nonparametrically, while its small contribution to fit supports the semi-log default of Malpezzi (2003).

Inference. The comparison of variance estimators speaks to the question posed by Abadie et al. (2023): clustering should reflect the level at which the sampling or the assignment of regressors is correlated. Clustering by FSA, our baseline, triples the heteroskedasticity-robust standard errors; clustering by province multiplies them by a further factor of four. The economically important implicit prices—area, full bathrooms, lot—are robust to this choice, but the small half-bathroom and bedroom effects are not, and we have revised their interpretation accordingly. Studies that report only heteroskedasticity-robust standard errors on spatially clustered listing data will overstate the precision of small coefficients by a wide margin.

Spatial dependence. The 87% reduction of Moran's I speaks to the oldest empirical worry in housing hedonics (Dubin, 1988; Basu and Thibodeau, 1998). It quantifies, for a national market, the claim of Bourassa et al. (2007) and Gibbons and Overman (2012) that geographic controls—here, three-character postal geography—do most of the work that parametric spatial-lag structures are designed to do. Our result does not make spatial econometrics redundant: the residual I of 0.03–0.11 depending on k is within-FSA dependence that an explicit spatial error model (Kelejian and Prucha, 1998) or spatio-temporal weights (Pace et al., 1998; Dubé and Legros, 2013) could exploit, particularly for prediction, and the success of the inverse-distance interpolation of neighbourhood effects in Table 12 shows that the fixed effects themselves are spatially smooth. It does, however, shift the burden of proof toward parsimony: a fixed effect per neighbourhood is transparent, imposes no weight-matrix assumptions, and removes most of the spatial signal.

Quantile and provincial heterogeneity. The flat size elasticity and the declining bathroom premium across the price distribution are a different pattern from the rising elasticities that Zietz et al. (2008) report for Orem, Utah, which illustrates that distributional heterogeneity is market-specific. The provincial gradient in the size elasticity (0.47 in Ontario and 0.49 in British Columbia against 0.66 in Manitoba and 0.73 in Saskatchewan) and the failure of the national structural coefficients to transfer to British Columbia are consistent with the submarket literature's core claim that implicit prices, not just price levels, vary across segments (Goodman and Thibodeau, 1998; Bourassa et al., 2003). This heterogeneity qualifies our own grand model: the absorbed intercepts allow every neighbourhood its own price level, but the structural slopes are pooled, and Section 6.5 shows that those slopes move within an economically meaningful band. Fully interacted, submarket-specific coefficient systems are the natural next step, at a cost in transparency.

Valuation accuracy. A within-support MdAPE of 15.6% places the transparent hedonic model within the accuracy bands reported in the mass-appraisal comparison literature (McCluskey et al., 2013) and close to what machine-learning methods deliver on comparable tasks (Mullainathan and Spiess, 2017). Two caveats temper the comparison. First, our target is the asking price, which sellers set with reference to the same observable attributes and neighbourhood comparables that the model uses, so that the accuracy of a list-price model is an upper bound on what the same specification would achieve on transaction prices. Second, the spatial holdout shows that accuracy in neighbourhoods absent from the training data is materially lower (MdAPE of 17% with spatial interpolation, 25% with a provincial prior); valuation studies that report only random holdouts on spatially clustered data overstate out-of-support performance (Roberts et al., 2017). Within those bounds, the result supports the position of Clapp (2003): most of the predictive content of "black-box" valuation lies in the location surface, which a fixed-effect design captures explicitly and auditably.

7.2. Points of tension with prior findings

Three tensions deserve explicit statement. First, the submarket literature sometimes finds that *how* submarkets are defined matters little for prediction (Bourassa et al., 2003); our single-country evidence cannot adjudicate the optimal geography, and FSAs—postal artefacts—are surely not it. The residual Moran's I and the smoothness of the fixed effects across adjacent FSAs both suggest that a finer or better-shaped geography would add value. Second, Kuminoff et al. (2010) and Bishop et al. (2020) warn that hedonic estimates can be fragile to specification; our robustness battery (eight samples, quantiles, quadratic and spline terms, four variance estimators) addresses the first-stage version of this concern, but any second-stage welfare use of our premia would inherit the identification problems of Epple (1987). Third, the leave-one-province-out exercise shows that

structural prices transfer imperfectly (mean held-out R^2 of 0.34, negative for British Columbia), which cuts against reading the grand model's pooled slopes as universal constants and aligns with the heterogeneity that segmented-market models predict (Goodman and Thibodeau, 1998).

7.3. Implications

For *assessment and property taxation*, the decomposition implies that neighbourhood-level value is at least as large a component of the assessable base as structure, so that assessment uniformity depends first on getting location surfaces right; the FSA premia we publish, and their spatial interpolability, are directly usable as such a surface. For *index construction*, the broad stability of structural implicit prices across most provinces supports pooled hedonic indices with local intercepts (Hill, 2013; Hill and Scholz, 2018), while the British Columbia result and the quantile profile of the bathroom premium caution that constant-quality adjustment differs across markets and segments. For *automated valuation*, the results quantify the price of transparency: a fully inspectable model concedes little accuracy on within-support predictions, and its out-of-support accuracy depends on how the location effect is extrapolated rather than on the structural coefficients. For *housing policy*, a tenfold neighbourhood premium span within one country, concentrated in two metropolitan systems, is the observable imprint of restricted supply against agglomeration demand (Saiz, 2010; Gyourko et al., 2013; Combes et al., 2019); policies that expand supply where premia are highest attack the term of the decomposition that is largest in the most expensive markets.

7.4. Limitations

Six limitations bound the interpretation. (i) Prices are asking prices. Sellers set them strategically (Horowitz, 1992; Genesove and Mayer, 2001), they trade off against time on the market (Anglin et al., 2003), and the eventual transaction price emerges from search and bargaining (Merlo and Ortalo-Magné, 2004; Genesove and Han, 2012; Han and Strange, 2016). Implicit list prices can therefore differ from implicit transaction prices, especially where the list-to-sale ratio varies systematically with attributes or with market tightness—for instance if overpricing is more common for large dwellings or in slow markets—and the composition of active listings over-represents dwellings that have been slow to sell. (ii) The data lack year of construction, renovation status and interior quality; the neighbourhood effects absorb the cross-neighbourhood component of this unobserved quality, but the within-FSA structural estimates remain exposed to within-neighbourhood sorting, and the bathroom premium in particular is likely to carry some of it. (iii) Several fields are noisy: lot information is sparse and parsed from free text, storey and parking counts contain implausible values and are zero when unreported, and the province variable had to be reconstructed from the postal code. (iv) The analysis is a single cross-section: it characterises the spatial and structural level of

asking prices at one moment, not their dynamics, and cannot speak to the efficiency questions of the repeat-sales tradition (Case and Shiller, 1989). (v) FSA boundaries are postal conveniences; premia estimated at this resolution average over genuinely finer neighbourhood variation, as the residual spatial autocorrelation confirms, and the nine pooled residual categories are not neighbourhoods at all. (vi) The sample is a snapshot of one listing platform; coverage of rural areas, of private sales and of new construction sold off-plan is unknown.

7.5. Future work

The agenda follows the limitations: linking listings to closing prices and time on the market to estimate the list-to-sale wedge attribute by attribute (Anglin et al., 2003; Han and Strange, 2016); enriching the attribute set with age and quality; decomposing the neighbourhood premia into capitalised amenities—schools, transit, environmental quality—in the quasi-experimental tradition (Black, 1999; Chay and Greenstone, 2005; Linden and Rockoff, 2008); modelling the residual within-FSA dependence and the spatial smoothness of the fixed effects with an explicit spatial structure (Kelejian and Prucha, 1998; Dubé and Legros, 2013); extending the cross-section to a panel of snapshots to study dynamics and policy incidence; and estimating submarket-specific coefficient systems to map the heterogeneity that our provincial results only sketch (Goodman and Thibodeau, 1998; Bourassa et al., 2003).

8. Conclusion

Using a nationwide cross-section of 142,121 MLS listings and 1,155 absorbed neighbourhood fixed effects, we estimate a hedonic price function for the Canadian residential market and decompose the variance of log asking prices into structure and location. The exercise yields four conclusions. First, location and structure are of comparable importance: resolving geography from the provincial to the neighbourhood scale raises explained variation from 60% to 77%, and a Shapley–Owen decomposition assigns about 34 percentage points of R^2 to structure and 36 to location, so that the neighbourhood is the single largest block once province and neighbourhood are grouped, while the sequential ladder that most hedonic papers report would credit structure with 44 points and location with 32. The answer to "how much is location worth?" is therefore inseparable from the attribution rule, and we have made that rule explicit. Second, the structural implicit list prices behave as hedonic theory predicts and are stable across eight samples: living area carries an elasticity of 0.53, each full bathroom adds about 12%, and bedroom counts are economically negligible once floor space is held fixed. The area and bathroom effects survive province-level wild-cluster bootstrap inference; the small half-bathroom and bedroom effects do not. Third, the neighbourhood premia are vast—a factor of about ten separates Vancouver's west side from rural Newfoundland for an

identical dwelling—and the neighbourhood effects remove 87% of the spatial autocorrelation in residuals. Fourth, the model is a credible valuation tool whose accuracy depends on spatial support: it values held-out listings in known neighbourhoods with an out-of-sample R^2 of 0.77 and a median absolute error of 15.6%, and listings in neighbourhoods never observed with an R^2 of 0.74 and a median error of 17% when the neighbourhood effect is interpolated from adjacent FSAs. Floor space displays mild diminishing returns concentrated above 150 m^2, the location premium declines by about 5.5% per doubling of distance to the nearest metropolitan centre, implicit prices vary sensibly across the price distribution and across provinces, and the structural prices transfer across most provinces but not to British Columbia.

The reading we propose is deliberately modest. In Canada, the neighbourhood is at least as large a component of a dwelling's asking price as the dwelling itself, and it is the component whose dispersion is greatest. That fact is simultaneously a measurement result (for assessment, taxation and index construction), a validation result (a transparent fixed-effects design removes most of the spatial dependence that motivates far more elaborate machinery, and its fixed effects are smooth enough to be interpolated), and a policy result (the premium dispersion that supply constraints and agglomeration generate is largest precisely in the most expensive markets). Section 7 details the limitations—asking rather than transaction prices, unobserved quality, noisy fields, a single cross-section, postal rather than planning geography—and the research agenda they imply. Within those bounds, the model provides a transparent and reproducible benchmark for automated valuation, market monitoring and the analysis of location value across the Canadian residential market.

Reproducibility. Data engineering was performed in `DuckDB` and `pandas`; estimation used `statsmodels` for cluster-robust OLS and quantile regression and `linearmodels` for absorbing least squares; figures were produced in `matplotlib`. All tables and figures are generated programmatically from the source database by the numbered scripts in the accompanying repository, and Appendix C reconciles the present estimates with those of the May 2026 version of this paper.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *Quarterly Journal of Economics*, 2023, *138* (1), 1–35.

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Albouy, David, "What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities," *Review of Economics and Statistics*, 2016, *98* (3), 477–487.

_ , Gabriel Ehrlich, and Minchul Shin, "Metropolitan Land Values," *Review of Economics and Statistics*, 2018, *100* (3), 454–466.

Allen, Jason, Robert Amano, David P. Byrne, and Allan W. Gregory, "Canadian City Housing Prices and Urban Market Segmentation," *Canadian Journal of Economics*, 2009, *42* (3), 1132–1149.

Alonso, William, *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, MA: Harvard University Press, 1964.

Anglin, Paul M., Ronald Rutherford, and Thomas M. Springer, "The Trade-off Between the Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting," *Journal of Real Estate Finance and Economics*, 2003, *26* (1), 95–111.

Anselin, Luc, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, 1988.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 1963, *58* (304), 933–942.

Bajari, Patrick and C. Lanier Benkard, "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 2005, *113* (6), 1239–1276.

Basu, Sabyasachi and Thomas G. Thibodeau, "Analysis of Spatial Autocorrelation in House Prices," *Journal of Real Estate Finance and Economics*, 1998, *17* (1), 61–85.

Bishop, Kelly C. and Christopher Timmins, "Estimating the Marginal Willingness to Pay Function Without Instrumental Variables," *Journal of Urban Economics*, 2019, *109*, 66–83.

_ , Nicolai V. Kuminoff, H. Spencer Banzhaf, Kevin J. Boyle, Kathrine von Gravenitz, Jaren C. Pope, V. Kerry Smith, and Christopher D. Timmins, "Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality," *Review of Environmental Economics and Policy*, 2020, *14* (2), 260–281.

Black, Sandra E., "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 1999, *114* (2), 577–599.

Bogin, Alexander, William Doerner, and William Larson, "Local House Price Dynamics: New Indices and Stylized Facts," *Real Estate Economics*, 2019, *47* (2), 365–398.

Bourassa, Steven C., Eva Cantoni, and Martin Hoesli, "Spatial Dependence, Housing Submarkets, and House Price Prediction," *Journal of Real Estate Finance and Economics*, 2007, *35* (2), 143–160.

_ , Martin Hoesli, and Vincent S. Peng, "Do Housing Submarkets Really Matter?," *Journal of Housing Economics*, 2003, *12* (1), 12–28.

Cameron, A. Colin and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Can, Ayşe, "Specification and Estimation of Hedonic Housing Price Models," *Regional Science and Urban Economics*, 1992, *22* (3), 453–474.

Case, Bradford, John Clapp, Robin Dubin, and Mauricio Rodriguez, "Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models," *Journal of Real Estate Finance and Economics*, 2004, *29* (2), 167–191.

Case, Karl E. and Robert J. Shiller, "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 1989, *79* (1), 125–137.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Cheshire, Paul and Stephen Sheppard, "On the Price of Land and the Value of Amenities," *Economica*, 1995, *62* (246), 247–267.

Clapp, John M., "A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation," *Journal of Real Estate Finance and Economics*, 2003, *27* (3), 303–320.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 247–348.

_ **, Gilles Duranton, and Laurent Gobillon**, "The Costs of Agglomeration: House and Land Prices in French Cities," *Review of Economic Studies*, 2019, *86* (4), 1556–1589.

Correia, Sergio, "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator," 2017. Working paper, Duke University.

Court, Andrew T., "Hedonic Price Indexes with Automotive Examples," in "The Dynamics of Automobile Demand," New York: General Motors Corporation, 1939, pp. 99–117.

Davis, Morris A. and Jonathan Heathcote, "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, 2007, *54* (8), 2595–2620.

_ **and Michael G. Palumbo**, "The Price of Residential Land in Large US Cities," *Journal of Urban Economics*, 2008, *63* (1), 352–384.

Diewert, W. Erwin, Jan de Haan, and Rens Hendriks, "Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components," *Econometric Reviews*, 2015, *34* (1–2), 106–126.

Duan, Naihua, "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, 1983, *78* (383), 605–610.

Dubé, Jean and Diego Legros, "A Spatio-temporal Measure of Spatial Dependence: An Example Using Real Estate Data," *Papers in Regional Science*, 2013, *92* (1), 19–30.

_ **, Marius Thériault, and François Des Rosiers**, "Commuter Rail Accessibility and House Values: The Case of the Montreal South Shore, Canada, 1992–2009," *Transportation Research Part A: Policy and Practice*, 2013, *54*, 49–66.

Dubin, Robin A., "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms," *Review of Economics and Statistics*, 1988, *70* (3), 466–474.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim, "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 2004, *112* (S1), S60–S109.

Epple, Dennis, "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 1987, *95* (1), 59–80.

Genesove, David and Christopher Mayer, "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 2001, *116* (4), 1233–1260.

_ **and Lu Han**, "Search and Matching in the Housing Market," *Journal of Urban Economics*, 2012, *72* (1), 31–45.

Gibbons, Stephen and Henry G. Overman, "Mostly Pointless Spatial Econometrics?," *Journal of Regional Science*, 2012, *52* (2), 172–191.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks, "Why Have Housing Prices Gone Up?," *American Economic Review*, 2005, *95* (2), 329–333.

Goodman, Allen C. and Thomas G. Thibodeau, "Housing Market Segmentation," *Journal of Housing Economics*, 1998, *7* (2), 121–143.

Griliches, Zvi, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in "The Price Statistics of the Federal Government," New York: National Bureau of Economic Research, 1961, pp. 173–196.

Grömping, Ulrike, "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician*, 2007, *61* (2), 139–147.

Guimarães, Paulo and Pedro Portugal, "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects," *Stata Journal*, 2010, *10* (4), 628–649.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai, "Superstar Cities," *American Economic Journal: Economic Policy*, 2013, *5* (4), 167–199.

Haider, Murtaza and Eric J. Miller, "Effects of Transportation Infrastructure and Location on Residential Real Estate Values: Application of Spatial Autoregressive Techniques," *Transportation Research Record*, 2000, *1722* (1), 1–8.

Halvorsen, Robert and Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 1980, *70* (3), 474–475.

Han, Lu and William C. Strange, "The Microstructure of Housing Markets: Search, Bargaining, and Brokerage," in "Handbook of Regional and Urban Economics," Vol. 5, Elsevier, 2015, pp. 813–886.

_ **and** _ , "What Is the Role of the Asking Price for a House?," *Journal of Urban Economics*, 2016, *93*, 115–130.

Hill, Robert J., "Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy," *Journal of Economic Surveys*, 2013, *27* (5), 879–914.

_ **and Michael Scholz**, "Can Geospatial Data Improve House Price Indexes? A Hedonic Imputation Approach with Splines," *Review of Income and Wealth*, 2018, *64* (4), 737–756.

Horowitz, Joel L., "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 1992, *7* (2), 115–129.

Hossain, Belayet and Ehsan Latif, "Determinants of Housing Price Volatility in Canada: A Dynamic Analysis," *Applied Economics*, 2009, *41* (27), 3521–3531.

Huettner, Frank and Marco Sunder, "Axiomatic Arguments for Decomposing Goodness of Fit According to Shapley and Owen Values," *Electronic Journal of Statistics*, 2012, *6*, 1239–1250.

Israeli, Osnat, "A Shapley-Based Decomposition of the R-Square of a Linear Regression," *Journal of Economic Inequality*, 2007, *5* (2), 199–212.

Kelejian, Harry H. and Ingmar R. Prucha, "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 1998, *17* (1), 99–121.

_ **and** _ , "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics*, 2001, *104* (2), 219–257.

Kennedy, Peter E., "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 1981, *71* (4), 801.

Kiel, Katherine A. and Jeffrey E. Zabel, "Location, Location, Location: The 3L Approach to House Price Determination," *Journal of Housing Economics*, 2008, *17* (2), 175–190.

Knoll, Katharina, Moritz Schularick, and Thomas Steger, "No Price Like Home: Global House Prices, 1870–2012," *American Economic Review*, 2017, *107* (2), 331–353.

Koenker, Roger and Gilbert Bassett, "Regression Quantiles," *Econometrica*, 1978, *46* (1), 33–50.

Kuminoff, Nicolai V., Christopher F. Parmeter, and Jaren C. Pope, "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities?," *Journal of Environmental Economics and Management*, 2010, *60* (3), 145–160.

Lancaster, Kelvin J., "A New Approach to Consumer Theory," *Journal of Political Economy*, 1966, *74* (2), 132–157.

LeSage, James and R. Kelley Pace, *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, 2009.

Linden, Leigh and Jonah E. Rockoff, "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 2008, *98* (3), 1103–1127.

MacKinnon, James G. and Matthew D. Webb, "Wild Bootstrap Inference for Wildly Different Cluster Sizes," *Journal of Applied Econometrics*, 2017, *32* (2), 233–254.

Malpezzi, Stephen, "Hedonic Pricing Models: A Selective and Applied Review," in Tony O'Sullivan and Kenneth Gibb, eds., *Housing Economics and Public Policy*, Blackwell, 2003, pp. 67–89.

McCluskey, William J., Michael McCord, Peadar T. Davis, Martin Haran, and David McIlhatton, "Prediction Accuracy in Mass Appraisal: A Comparison of Modern Approaches," *Journal of Property Research*, 2013, *30* (4), 239–265.

McMillen, Daniel P. and Christian L. Redfearn, "Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions," *Journal of Regional Science*, 2010, *50* (3), 712–733.

Merlo, Antonio and François Ortalo-Magné, "Bargaining over Residential Real Estate: Evidence from England," *Journal of Urban Economics*, 2004, *56* (2), 192–216.

Mills, Edwin S., "An Aggregative Model of Resource Allocation in a Metropolitan Area," *American Economic Review*, 1967, *57* (2), 197–210.

Moran, Patrick A. P., "Notes on Continuous Stochastic Phenomena," *Biometrika*, 1950, *37* (1/2), 17–23.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

Muth, Richard F., *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*, Chicago: University of Chicago Press, 1969.

Oates, Wallace E., "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *Journal of Political Economy*, 1969, *77* (6), 957–971.

Pace, R. Kelley, Ronald Barry, John M. Clapp, and Mauricio Rodriquez, "Spatiotemporal Autoregressive Models of Neighborhood Effects," *Journal of Real Estate Finance and Economics*, 1998, *17* (1), 15–33.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Roberts, David R., Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann, "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure," *Ecography*, 2017, *40* (8), 913–929.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1974, *82* (1), 34–55.

Rosiers, François Des, Antonio Lagana, Marius Thériault, and Marcel Beaudoin, "Shopping Centres and House Values: An Empirical Investigation," *Journal of Property Valuation and Investment*, 1996, *14* (4), 41–62.

_ , Marius Thériault, Yan Kestens, and Paul Villeneuve, "Landscaping and House Values: An Empirical Investigation," *Journal of Real Estate Research*, 2002, *23* (1–2), 139–162.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Sheppard, Stephen, "Hedonic Analysis of Housing Markets," in Paul Cheshire and Edwin S. Mills, eds., *Handbook of Regional and Urban Economics*, Vol. 3, Elsevier, 1999, pp. 1595–1635.

Sirmans, G. Stacy, David A. Macpherson, and Emily N. Zietz, "The Composition of Hedonic Pricing Models," *Journal of Real Estate Literature*, 2005, *13* (1), 1–44.

_ , Lynn MacDonald, David A. Macpherson, and Emily Norman Zietz, "The Value of Housing Characteristics: A Meta Analysis," *Journal of Real Estate Finance and Economics*, 2006, *33* (3), 215–240.

Statistics Canada, "Census Forward Sortation Area Boundary File, Reference Guide, Census Year 2021," Catalogue no. 92-179-G2021001, Statistics Canada, Ottawa 2022.

Thériault, Marius, François Des Rosiers, Paul Villeneuve, and Yan Kestens, "Modelling Interactions of Location with Specific Value of Housing Attributes," *Property Management*, 2003, *21* (1), 25–62.

Tiebout, Charles M., "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

Zietz, Joachim, Emily N. Zietz, and G. Stacy Sirmans, "Determinants of House Prices: A Quantile Regression Approach," *Journal of Real Estate Finance and Economics*, 2008, *37* (4), 317–333.

Appendices

A. Variable construction

The raw `listings` table stores every attribute as text. The estimation sample is built by the first script of the pipeline (`01_build_sample.py`) from the rules below, which were reconstructed from the description of the original (May 2026) pipeline and validated against its stored outputs; the original cleaning code was not preserved.

Bedrooms. The integer components of the raw string are summed, so that "3 + 1" (three bedrooms on the main level and one below grade) is coded as four.

Living area. The explicit floor-area measurement is preferred. Exact entries are used as reported; banded entries such as "1100–1500 sqft" are mapped to the upper bound of the band, the convention that reproduces the published living-area distribution. When the field is absent the interior-size field is used. Square feet are converted at $1\ \text{ft}^2 = 0.0929\ \text{m}^2$; entries labelled m^2 are used as such.

Lot area. The free-text lot field is parsed when it contains a number with a recognisable unit (square feet, square metres, acres or hectares), a phrase of the form "under x acre", an acreage range (mapped to its upper bound), or a frontage-by-depth pair in feet. Parsed lots above the 99th percentile of positive values are treated as unparsed, because they are dominated by multi-acre acreage strings whose relationship to the dwelling is unclear. The variable entering the regressions is $\ln(1 + \text{lot area})$ together with an indicator equal to one when a positive lot area was parsed.

Counts. Full bathrooms, half bathrooms, parking spaces and storeys are taken as reported; half bathrooms, parking and storeys are set to zero when absent. No cap is applied in the baseline; Table 7 caps storeys and parking at their 99th percentiles.

Dwelling type and ownership. The raw building type is consolidated into eight groups (house; apartment; row or townhouse; duplex; triplex; fourplex; manufactured home, which pools manufactured, mobile and park-model homes; other), and the ownership type into six (freehold; condominium or strata; leasehold; co-operative or co-ownership; other; not reported). The broad listing category (house, condominium, plex, townhouse, apartment) is also retained as a control.

Neighbourhood and province. The FSA is the first three characters of the postal code, retained when it matches the pattern letter–digit–letter. The province is the Canada Post region of the

first letter: A, Newfoundland and Labrador; B, Nova Scotia; C, Prince Edward Island; E, New Brunswick; G, H and J, Quebec; K, L, M, N and P, Ontario; R, Manitoba; S, Saskatchewan; T, Alberta; V, British Columbia (Statistics Canada, 2022). The letters X and Y (territories) and the unassigned letters O and W (27 records in total) are dropped. The scraper's own province field, which records the regional seed of the extraction, is retained for reference only; it disagrees with the FSA-based province for 11.7% of the estimation sample.

Filters. Listings are kept when the asking price is strictly positive, living area, bedrooms and bathrooms are non-missing and the FSA is valid (147,236 of 172,019 raw records); the 1st and 99th percentiles of price and living area are then applied as inclusive bounds, leaving 142,121 listings. FSAs with fewer than 25 listings are pooled into one residual category per province.

B. Additional results

Table A1. Moran's I of residuals: sensitivity to the weight matrix and subsample

	Subsample A		Subsample B	
Neighbours	Structure only	Grand model	Structure only	Grand model
$k = 5$	0.525 ($z = 107$)	0.111 ($z = 21$)	0.517 ($z = 108$)	0.100 ($z = 20$)
$k = 10$	0.496 ($z = 140$)	0.063 ($z = 19$)	0.491 ($z = 151$)	0.058 ($z = 17$)
$k = 20$	0.469 ($z = 199$)	0.034 ($z = 15$)	0.470 ($z = 191$)	0.031 ($z = 14$)

Notes: Moran's I with row-standardised k-nearest-neighbour weights (great-circle distance) on two independent random subsamples of 15,000 listings. z-scores from 199 random permutations; every permutation p-value equals its minimum attainable value of $1/200 = 0.005$. "Structure only" is the residual of the model with structural, type and ownership controls but no location effects.

C. Reconciliation with the May 2026 version

The May 2026 version of this paper (v1.2) reported estimates from an upstream pipeline that was not preserved. The present version re-estimates every result from the reconstructed pipeline described in Appendix A, with two substantive differences: the province of each listing is derived from its FSA rather than from the scraper's regional seed, and the reconstructed trimming rule retains 142,121 rather than 140,931 listings (a difference of 0.8%, concentrated at the upper living-area bound). All headline magnitudes are preserved (Table A2); the estimates that changed materially are those that use the province variable—the province-effects R^2, the leave-one-province-out ranking and the provincial heterogeneity—and the two lot coefficients, whose free-text parser could not be

recovered exactly. The variance decomposition, the inference comparison, the spline, the spatial holdout and the Moran sensitivity analysis are new to this version.

Table A2. Headline estimates: May 2026 version (v1.2) and present version (v1.3)

Quantity	v1.2 (May 2026)	v1.3 (this version)
Estimation sample	140,931	142,121
Absorbed location effects	1,153	1,155
R^2: M1 / M2 / M3 (houses)	0.464 / 0.469 / 0.567	0.467 / 0.470 / 0.605
R^2: M4 (houses, FSA) / M5 (grand)	0.762 / 0.767	0.769 / 0.776
Living-area elasticity, M5 (SE)	0.547 (0.009)	0.529 (0.009)
Full-bathroom coefficient, M5 (SE)	0.109 (0.004)	0.112 (0.004)
Bedrooms, M5	-0.002	-0.006
ln lot / lot indicator, M5	0.030 / -0.196	0.067 / -0.459
Out-of-sample R^2 / MdAPE	0.764 / 15.8%	0.772 / 15.6%
Moran's I, structure $\rightarrow$ grand ($k = 10$)	0.46 $\rightarrow$ 0.08	0.50 $\rightarrow$ 0.06
Gradient slope on ln distance	-0.085	-0.082
Quadratic in ln area: β_1 / β_2	1.065 / -0.053	1.147 / -0.063
LOPO mean within-province R^2	0.362 (AB negative)	0.341 (BC negative)
Size elasticity range across provinces	0.44 (NL)–0.66 (MB)	0.41 (NL)–0.73 (SK)
Premium ratio, most / least expensive FSA	≈ 9	≈ 10

Notes: v1.2 values are the published numbers of the May 2026 version, archived in `results/reference/`; v1.3 values are produced by the current pipeline (`results/reproduced/`). The v1.2 province variable was the scraper's regional seed, which mislabelled about 12% of listings; the v1.2 statement that the lowest-premium FSAs included "E2A (QC)" and "R0M (SK)" reflected this error (E is New Brunswick and R is Manitoba). In v1.2 a doubling of distance was described as lowering the location premium by 8.5%; the correct reading of a slope of -0.085 on ln distance is -5.7% per doubling.