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 140,931 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

May 2026

Version 1.2

Abstract

We estimate a large-scale hedonic price model for the Canadian residential real-estate market using a cross-section of 140,931 active MLS listings drawn from the nine provinces present in the data. A semi-logarithmic specification decomposes dwelling prices into structural attributes, dwelling type and ownership form, and 1,153 neighbourhood (Forward Sortation Area) fixed effects estimated by absorbing least squares. Moving from a purely structural model to the full specification raises the explained share of price variation from 46% to 77%, establishing that location is the dominant price determinant in Canada: neighbourhood effects alone account for roughly thirty percentage points of explanatory power. The living-area elasticity is 0.51–0.62, each full bathroom commands a premium of about 11–15%, and—conditional on floor space—bedroom counts are economically negligible. The implicit prices are stable across alternative samples and trimming rules, and the model delivers strong out-of-sample valuation accuracy (held-out $R^2 = 0.76$; median absolute error of 16%). The neighbourhood effects absorb the bulk of the spatial dependence in prices—Moran's I falls from 0.46 to 0.08—and we further document diminishing returns to floor space, an urban price gradient that decays with distance to major metros, and quantile and cross-province heterogeneity in the implicit prices. We map the resulting neighbourhood premia, which span a factor of roughly nine between the most and least expensive areas.

Keywords: Hedonic pricing, Housing markets, Canada, Fixed effects, Spatial heterogeneity, Automated valuation

JEL Classification: R31, R21, C21, C55

1. Introduction

Housing is the largest asset class on household balance sheets and the collateral behind most household debt, and its price is the single most consequential relative price most families ever face. Yet a dwelling is the archetypal heterogeneous good: no two houses are identical, and the most important attribute—*where the dwelling stands*—is not a scalar that can be read off a listing sheet. The hedonic approach pioneered by Lancaster (1966) and Rosen (1974) resolves this heterogeneity by treating the dwelling as a bundle of attributes, each commanding an implicit price determined in equilibrium by the interaction of buyers' marginal willingness to pay and sellers' marginal cost of supply. Regressing the (log) price of a dwelling on its measurable characteristics recovers these implicit prices, and the resulting estimates provide the empirical backbone for constant-quality house-price indices (Bailey et al., 1963; Hill, 2013), property assessment and mass appraisal (McCluskey et al., 2013), mortgage valuation, and the welfare analysis of local amenities (Black, 1999; Chay and Greenstone, 2005).

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 far beyond folklore. Assessment authorities must apportion value between land and improvements; property-tax reform hinges on the incidence of the land component; the macro literature has shown that land, not structure, drives both the level and the growth of aggregate housing wealth (Davis and Heathcote, 2007; Knoll et al., 2017); and the dispersion of location values across cities is the fingerprint of supply constraints and agglomeration (Saiz, 2010; Gyourko et al., 2013; Combes and Gobillon, 2015). Credible answers require holding the structural bundle fixed while letting location vary—precisely what a hedonic model 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 **140,931** active MLS listings spanning the nine provinces present in the data—from British Columbia to Newfoundland and Labrador—we decompose dwelling prices into structural attributes, dwelling type and ownership form, and a rich set of **1,153 neighbourhood fixed effects** defined at the Forward Sortation Area (FSA) level. To our knowledge this is among the most geographically comprehensive single-equation hedonic exercises assembled for Canada, where the hedonic tradition has remained largely metropolitan in scope (Des Rosiers et al., 1996; Haider and Miller, 2000). It 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 thousands of dummy variables (Guimarães and Portugal, 2010; Correia, 2017).

Our main findings can be summarised as follows. The purely structural model explains 46% of the variation in log prices; adding province fixed effects raises this to 57%, and replacing them with FSA fixed effects lifts it to **77%**. Neighbourhood location *alone* therefore accounts for roughly thirty percentage points of explanatory power—more than all structural attributes combined, and a listing-level counterpart to the large land shares found in aggregate data (Davis and Heathcote, 2007; Knoll et al., 2017). Conditional on location, the living-area elasticity is estimated at 0.51–0.62: a 10% larger dwelling sells for about 5–6% more. Each additional full bathroom commands a premium of 11–15%, whereas the number of bedrooms is economically negligible once floor space is held fixed—a recurrent finding in the hedonic literature (Sirmans et al., 2005). The implicit prices are remarkably stable across alternative samples, trimming rules and dwelling types, and the model attains an out-of-sample R^2 of 0.76 with a median absolute valuation error of 16%, a level of accuracy competitive with the mass-appraisal benchmarks reported in the valuation literature (McCluskey et al., 2013; Mullainathan and Spiess, 2017). We map the estimated neighbourhood premia and show that the highest-valued FSAs—concentrated in the City of Vancouver and the Greater Toronto Area—trade at more than triple the national-median level net of structure. A battery of additional analyses sharpens the picture: the neighbourhood effects absorb 82% of the spatial autocorrelation in raw residuals (Moran's I falls from 0.46 to 0.08); floor space exhibits clear diminishing returns; the location component of price decays with distance to the nearest major metropolis, as the monocentric tradition predicts (Alonso, 1964; Muth, 1969; Mills, 1967); and the implicit prices vary sensibly across the price distribution and across provinces while transferring well out-of-region.

The contribution is threefold. *First*, we provide the first (to our knowledge) listing-level structure-versus-location decomposition spanning the entire Canadian market, quantifying the neighbourhood as its single largest priced component. *Second*, we bring the high-dimensional fixed-effects technology of modern applied microeconomics (Abowd et al., 1999; Correia, 2017) to a national housing cross-section and validate the design directly, showing that more than a thousand absorbed neighbourhood intercepts eliminate the bulk of spatial dependence that parametric spatial models are built to address (Dubin, 1988; Basu and Thibodeau, 1998; Gibbons and Overman, 2012). *Third*, we supply a transparent accuracy benchmark for the automated-valuation debate: every coefficient of the model is inspectable, yet its held-out performance is competitive with the mass-appraisal standards reported in the literature (Clapp, 2003; McCluskey et al., 2013).

The remainder of the paper is organised as follows. Section 2 positions the paper in six related literatures. Section 3 describes the data and variable construction. Section 4 presents the empirical strategy. Section 5 reports the main results. Section 6 provides robustness checks, heterogeneity analyses, and out-of-sample validation. Section 7 interprets the findings against the literature and discusses implications, limitations and future work. Section 8 concludes.

2. Related Literature

This paper sits at the intersection of six literatures: the hedonic framework and its identification; functional form and distributional heterogeneity in housing hedonics; the treatment of space—submarkets, spatial econometrics and fixed effects; the theory of amenity capitalization; urban spatial structure and the land share of housing value; and property valuation, from price indices to automated valuation models. We review each in turn and state 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 rather than the goods themselves, 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 traces out the envelope of buyers' bid functions and sellers' offer functions; its gradient with respect to a characteristic identifies, at the margin, the implicit price of that characteristic. A subsequent literature clarified exactly what the second stage—recovering preference parameters from the implicit prices—can and cannot deliver. Epple (1987) formalised the simultaneity problem that plagues second-stage demand estimation; Ekeland et al. (2004) show that nonlinearity of the hedonic price function aids identification of preferences; Bajari and Benkard (2005) develop a tractable demand estimator with unobserved product characteristics; and Kuminoff et al. (2010) document how sensitive welfare estimates are to specification. We deliberately remain at the first stage and target the implicit-price schedule itself, which is the relevant object for valuation, assessment and index construction, and which is identified under much weaker assumptions than the underlying preferences.

2.2. Functional form and heterogeneity in the implicit prices

A large applied literature studies which attributes belong in a housing hedonic and what 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 meta-tabulation also shows that bedroom coefficients are frequently small or negative once floor area is controlled—the pattern we recover for Canada. Malpezzi (2003) surveys functional-form choices and argues that the semi-logarithmic specification—log price on linear (or log) characteristics—is a robust default: it accommodates the right-skew of prices, yields coefficients interpretable as approximate percentage effects (Halvorsen and Palmquist, 1980), and mitigates heteroskedasticity. We adopt this specification throughout. A complementary literature relaxes its restrictions in two directions. McMillen and Redfearn (2010) estimate fully nonparametric hedonic surfaces and show that parametric functional forms can mask economically meaningful curvature—a concern we address directly with a quadratic-in-log-area specification that reveals diminishing returns to 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 single conditional-mean coefficient can be an incomplete summary; we document this for the Canadian market in Section 6.2. *Gap:* these strands are almost exclusively metropolitan or regional in scope; national-scale 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, location is intrinsic to its value, and unobserved local amenities induce strong spatial dependence in prices. The empirical importance of this dependence is long established: Dubin (1988) showed that ignoring spatially autocorrelated errors distorts hedonic inference, and Basu and Thibodeau (1998) documented pervasive residual autocorrelation in Dallas transactions. Can (1992) and Anselin (1988) formalised spatial autocorrelation in hedonic errors, while Pace et al. (1998) and, in the Canadian context, Dubé and Legros (2013) developed spatio-temporal weight-matrix approaches. A parallel strand argues that housing markets are segmented into *submarkets* within which prices are set: Goodman and Thibodeau (1998) formalise market segmentation tests, and Bourassa et al. (2003) show that accounting for submarkets materially improves prediction accuracy—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 head-to-head and conclude that geographic submarket controls capture most of what parametric spatial models capture, at a fraction of the complexity.

Our approach takes this conclusion to its logical end: rather than imposing a spatial weight matrix, we absorb a fixed effect for each of 1,153 FSA neighbourhoods, allowing the data to assign an arbitrary price level to each area and identifying 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) and now standard in applied microeconomics, with computation following Guimarães and Portugal (2010) and Correia (2017) and inference clustered at the neighbourhood level (Cameron and Miller, 2015). The choice also speaks to a methodological debate: Gibbons and Overman (2012) caution that parametric spatial-lag models are often hard to interpret causally 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 directly, testing for residual spatial autocorrelation with Moran's I (Moran, 1950) before and after absorbing the neighbourhood effects. *Gap:* 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 capitalization

Why should a neighbourhood intercept capture economic value at all? 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 capitalization of taxes and school spending into house prices, to Roback (1982), who embedded amenity capitalization in a general spatial equilibrium of wages and rents. Quasi-experimental hedonic studies have since measured the capitalization 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 such capitalized amenities into a single FSA-level premium rather than attempting to disentangle them; the estimated premium is therefore an upper envelope on the value of location 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) on uniquely sharp Berlin data. Our distance-to-metro gradient is the reduced-form counterpart of this prediction at national scale, and the agglomeration literature reviewed by Combes and Gobillon (2015) supplies the economic mechanism. A related macro-oriented literature measures how much of housing value is land rather than structure: Davis and Heathcote (2007) put the U.S. residential land share near 46% in aggregate, Knoll et al. (2017) show that land appreciation explains about 80% of the global house-price increase since World War II, and Albouy (2016) document enormous cross-city dispersion in land values. On the supply side, Saiz (2010) shows that geographic constraints generate exactly the kind of persistent cross-location price dispersion that our fixed effects absorb, and Gyourko et al. (2013) formalise the resulting "superstar city" dynamics—with Vancouver and Toronto as textbook candidates (Glaeser et al., 2005; Mok and Chan, 2018). *Gap:* this literature measures the land/location share with aggregate or city-level data; a listing-level decomposition for an entire national market, in which the location share is estimated directly 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 and taxonomises 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 "location, location, location" decomposition of house-price determinants, 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 terms. *Gap:* few studies report honestly held-out accuracy for a fully transparent hedonic model at national scale; our out-of-sample results provide exactly that benchmark.

2.7. Canadian evidence

Canada has a productive hedonic tradition, largely centred on single metropolitan areas: the Quebec City program of Des Rosiers and coauthors measures the capitalization of shopping centres (Des Rosiers et al., 1996) and landscaping quality (Des Rosiers et al., 2002), and Haider and Miller (2000) apply spatial autoregressive techniques to Toronto transactions. Policy work documents the spatial dispersion of Canadian prices (Mok and Chan, 2018), with Vancouver and Toronto at the expensive extreme (Glaeser et al., 2005). *Gap:* to our knowledge no previous study estimates a single hedonic system spanning the Canadian market coast to coast at the listing level; the national scope, the 1,153 absorbed neighbourhood effects, and the validated out-of-sample accuracy are, jointly, the contribution of this paper.

3. Data

3.1. Source and coverage

The analysis draws on a de-duplicated snapshot of the Canadian Multiple Listing Service (MLS), `realtor_mls_unique.duckdb`, containing 172,019 unique active "for-sale" listings. Each record carries the list price, a structured set of building and lot attributes, ownership and dwelling-type fields, and—crucially for this study—a geocoded location (latitude, longitude and postal code). Geographic coverage is essentially complete: coordinates and postal codes are present for more than 99.9% of records, so every listing can be mapped to its Forward Sortation Area (FSA), the first three characters of the Canadian postal code and our unit of neighbourhood.

3.2. Sample construction and variable parsing

The raw fields are semi-structured and require parsing. Bedroom counts reported as "3 + 1" (main plus lower level) are summed. Living areas, reported inconsistently in square feet or square metres, are harmonised to square metres (preferring the explicit floor-area measurement and converting from square feet at 1 $\text{ft}^2 = 0.0929\ \text{m}^2$). Lot sizes are parsed from free-text fields into square metres where a numeric value can be recovered, with a missingness indicator retained. We restrict the sample to residential dwellings (houses, condominiums, plexes, townhouses and apartments) with a strictly positive list price and non-missing core structural fields, and we 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 in hedonic work with listing data (Malpezzi, 2003). The resulting estimation sample contains **140,931 dwellings**, of which 82,334 are houses and 57,857 are condominiums.

The dependent variable is the natural logarithm of the list price. Structural regressors are: ln 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/condo, row/townhouse, duplex, triplex, fourplex, manufactured home, other), the ownership form (freehold, condominium/strata, leasehold, etc.), and the broad listing category. The neighbourhood fixed effect is the FSA; FSAs with fewer than 25 listings are pooled into a province-level residual category so that each absorbed effect is estimated from a reasonable number of observations.

3.3. Descriptive statistics

Table 1 reports the descriptive statistics. The median dwelling lists for roughly \$565,000 and offers about 125 m^2 of living space, two full bathrooms and three bedrooms. Both price and price per square metre are strongly right-skewed—means exceed medians throughout—motivating the log transformation illustrated in Figure 1, whose right panel is approximately symmetric and underpins the semi-logarithmic hedonic specification.

The provincial composition is dominated by Ontario, Quebec, British Columbia and Alberta, which together account for the large majority of listings; the Atlantic provinces and the Prairies are present but thinner, and Prince Edward Island and the territories contain no listings in this snapshot. Median price per square metre ranges from roughly \$2,000 in Newfoundland and Labrador to over \$6,100 in British Columbia (Figure 2), foreshadowing the dominant role of location documented below.

Table 1. Descriptive statistics

Variable	Mean	SD	P25	Median	P75
List price (CAD)	763,592	475,478	449,000	639,888	928,000
Price per m^2 (CAD)	5,382	2,682	3,580	4,781	6,593
Living area (m^2)	151.7	79.1	93.8	134.6	185.8
Bedrooms	3.11	1.32	2.00	3.00	4.00
Full bathrooms	2.32	1.11	2.00	2.00	3.00
Half bathrooms	0.35	0.52	0.00	0.00	1.00
Parking spaces	2.72	11.20	0.00	2.00	4.00
Storeys	2.44	6.01	1.50	2.00	2.00
Lot area (m^2)	2,621	10,818	0	0	656

Notes: Estimation sample of 140,931 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.

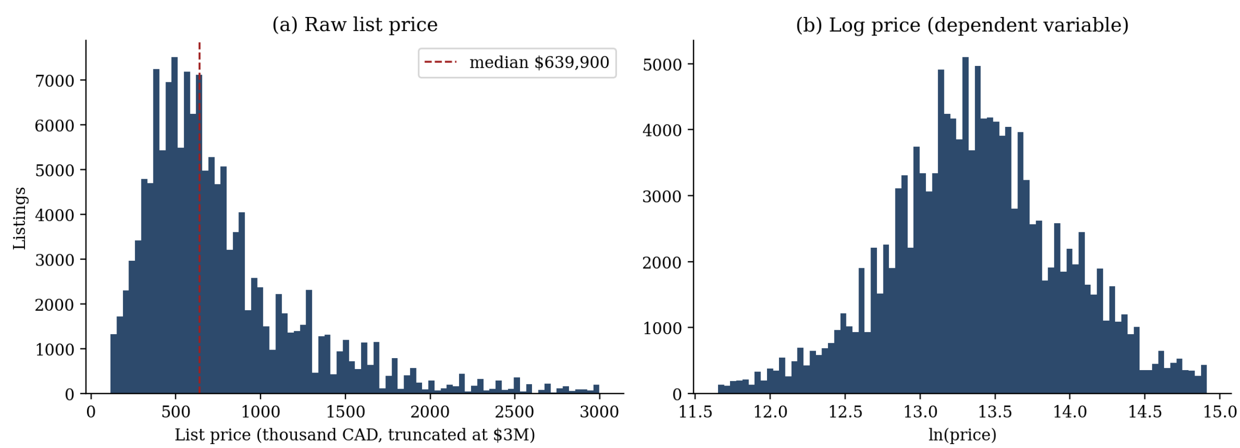


Figure 1. Distribution of list prices. Panel (a) shows the raw price (truncated at $3M for readability) with the sample median marked; panel (b) shows log price, the dependent variable, which is close to symmetric.

4. Empirical Strategy

4.1. The hedonic equation

We estimate the semi-logarithmic hedonic price equation

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

where P_i is the list price of dwelling i; Area_i is living area in square metres; $\mathbf{x}_i$ collects the remaining structural attributes (bedrooms, full and half bathrooms, parking, storeys, the lot indicator and ln lot area); $\mathbf{d}_i$ is a vector of dwelling-type and ownership dummies; $\mu_{f(i)}$ is a fixed effect for the FSA

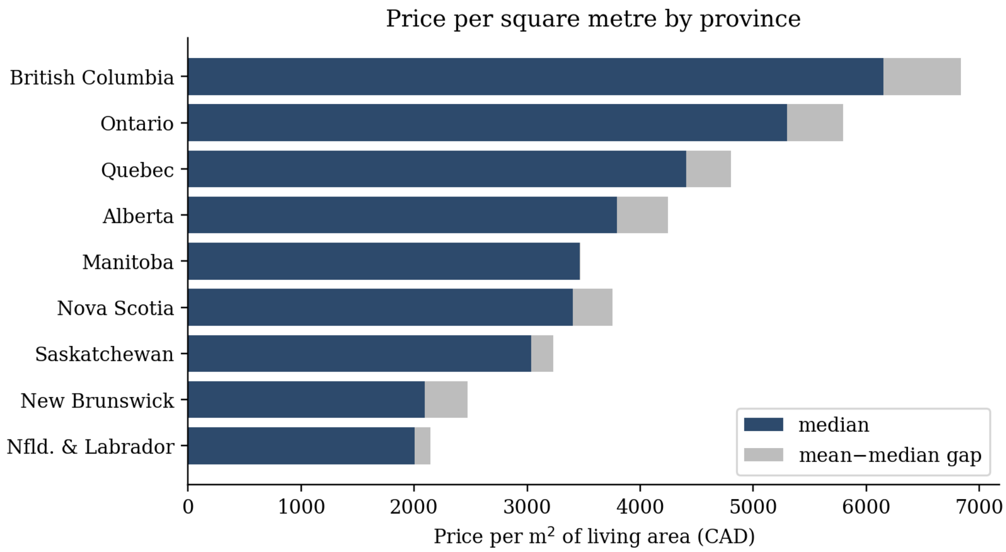


Figure 2. Median price per square metre of living area by province. Grey extensions show the mean–median gap, reflecting right-skew from high-value listings.

neighbourhood f to which dwelling i belongs; and ε_i is an idiosyncratic error.

Because β multiplies a logged regressor, it is the *elasticity* of price with respect to floor space. The elements of γ attached to count variables are semi-elasticities: following Halvorsen and Palmquist (1980), a coefficient γ implies an approximate proportional price change of $100\,(e^{\gamma}-1)\%$ for a one-unit increase, which we report whenever the distinction from the raw coefficient is material.

4.2. High-dimensional location effects

The neighbourhood fixed effects μ_f are the heart of the design. By including a separate intercept for each of the 1,153 FSAs, we allow 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 in β and γ are then identified purely from variation *within* neighbourhoods, comparing dwellings that differ in their physical attributes but share a location. This is the housing counterpart of the within estimator, 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).

Including more than a thousand dummies 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 FSA effects are partialled out of both the dependent variable and the regressors before the structural coefficients are estimated (Guimarães and Portugal, 2010; Correia, 2017), and the slope estimates are numerically identical to full-dummy OLS. All standard errors are clustered at the FSA level to allow for arbitrary within-neighbourhood correlation and heteroskedasticity (Cameron and Miller, 2015).

4.3. Specification ladder

We report a ladder of nested specifications that isolates the marginal contribution of each block of controls:

- **M1 – Structural:** structural attributes only.
- **M2 – + Type/Ownership:** adds dwelling-type and ownership dummies.
- **M3 – + Province:** adds province fixed effects.
- **M4 – Houses + FSA:** replaces province with FSA fixed effects, houses only.
- **M5 – Grand model:** FSA fixed effects over all residential dwellings.

Columns M1–M3 are estimated on the house subsample to keep the structural interpretation clean; M5 is the preferred grand specification estimated over the full residential sample. The gap in R^2 between M3 and M5 measures the explanatory value of resolving location at the neighbourhood rather than the provincial scale.

4.4. Retransformation and out-of-sample evaluation

Because the model is estimated in logs, predicted price levels require a retransformation correction. We use Duan's smearing estimator (Duan, 1983), multiplying $\exp(\widehat{\ln P})$ by the sample mean of $\exp(\widehat{\varepsilon})$, which is consistent without assuming log-normal errors. For predictive validation we randomly split the data 80/20, estimate the model on the training fold (restricting evaluation to neighbourhoods observed in training, since out-of-support FSA effects are not identified), and report held-out fit and percentage-error metrics on the test fold.

5. Results

5.1. The value of location

Figure 3 summarises the explanatory power of the specification ladder. The structural-only model (M1) accounts for 46.4% of the variation in log prices. Adding dwelling-type and ownership controls (M2) barely moves the fit, but introducing province fixed effects (M3) raises R^2 to 56.8%, and resolving location at the FSA scale lifts it to **76.2%** for houses (M4) and **76.7%** for the grand model (M5). The implication is stark: moving from province to neighbourhood resolution adds about twenty percentage points of explained variance, and *location as a whole accounts for roughly thirty percentage points*—more than the entire structural bundle. This is the central result of the paper and a quantitative statement of the realtor's adage that what matters is "location, location, location."

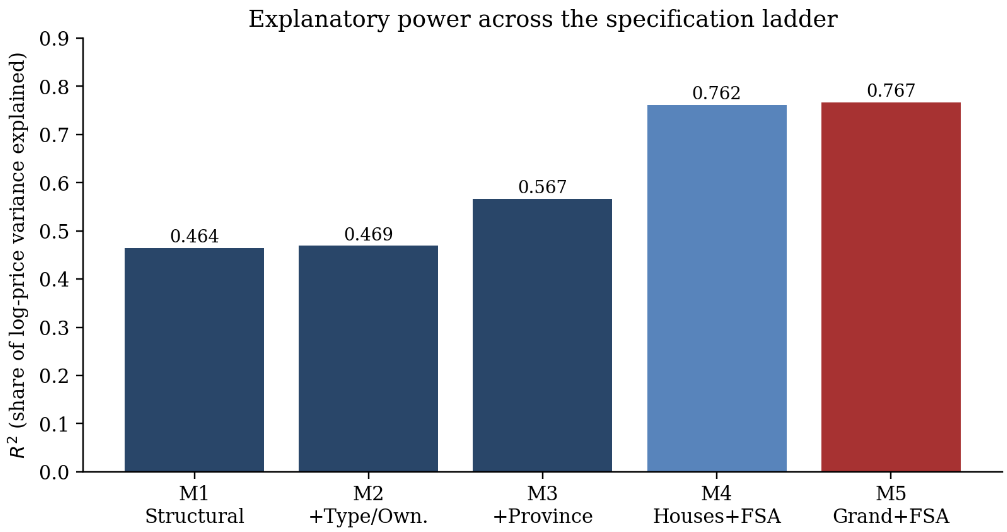


Figure 3. Share of log-price variation explained (R^2) across the five nested specifications. The jump from M3 (province) to M4/M5 (FSA) quantifies the value of resolving location at the neighbourhood scale.

5.2. Implicit prices of structural attributes

Table 2 reports the regression estimates. The living-area elasticity is remarkably stable across specifications, at 0.66 in the raw structural model and settling near 0.55 once location is controlled for: a 10% larger dwelling commands roughly a 5.5% higher price. The slight decline as controls are added is consistent with larger homes being located in more expensive areas—a confound the FSA effects remove.

Full bathrooms carry one of the strongest structural premia: about 0.11 log points in the grand model, or roughly an **11% price increase** per additional bathroom, rising to nearly 15% in the province-FE model. Half bathrooms attract a small *negative* conditional coefficient, which we read not as a disamenity but as a proxy for older or more compartmentalised floor plans once total area and full baths are held fixed. Bedrooms are economically negligible conditional on living area: holding floor space constant, subdividing it into more bedrooms does not raise value—a textbook hedonic finding (Sirmans et al., 2005) that recurs across our samples. Lot information enters positively (ln lot elasticity around 0.03) but modestly, reflecting both the noisiness of the parsed lot field and the fact that, within a neighbourhood, lot variation is compressed.

Figure 4 presents the structural implicit prices 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.

The price–size gradient by dwelling type (Figure 5) confirms the log–log structure: median price rises concavely with living area for both dwelling types. Unconditionally, condominiums list *above* houses of the same size—a compositional effect of their concentration in the expensive

Table 2. Hedonic regression estimates

	(1) Structural	(2) +Type/Own.	(3) +Province	(4) Grand+FSA
ln living area	0.657***	0.650***	0.585***	0.547***
	(0.022)	(0.022)	(0.016)	(0.009)
Full bathrooms	0.136***	0.141***	0.138***	0.109***
	(0.010)	(0.010)	(0.006)	(0.004)
Half bathrooms	-0.035***	-0.039***	-0.026***	-0.036***
	(0.009)	(0.009)	(0.009)	(0.009)
Bedrooms	0.005	0.003	-0.003	-0.002
	(0.004)	(0.004)	(0.003)	(0.003)
Parking spaces	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Storeys	0.001	0.000	-0.010	0.001***
	(0.004)	(0.004)	(0.007)	(0.000)
Has lot info	-0.212***	-0.221***	-0.034	-0.196***
	(0.038)	(0.038)	(0.039)	(0.028)
$\ln(1+\text{lot m}^2)$	0.026***	0.027***	0.008*	0.030***
	(0.004)	(0.004)	(0.005)	(0.004)
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	0.464	0.469	0.567	0.767
Observations	82,334	82,334	82,334	140,931

Notes: Dependent variable is ln(price). Columns (1)–(3) use the house subsample; column (4) is the grand model over all residential dwellings with 1,153 absorbed FSA fixed effects. Cluster-robust standard errors (by FSA) in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

metropolitan markets—whereas conditional on location and ownership the estimated dwelling-type effects show that houses command the premium, in line with Table 2.

5.3. The geography of housing value

Figure 6 maps the spatial structure of value directly. The familiar outline of populated Canada emerges from the listing coordinates. The Vancouver corridor and the Greater 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.

To translate location into a clean dollar statement we recover each FSA's fixed effect from the

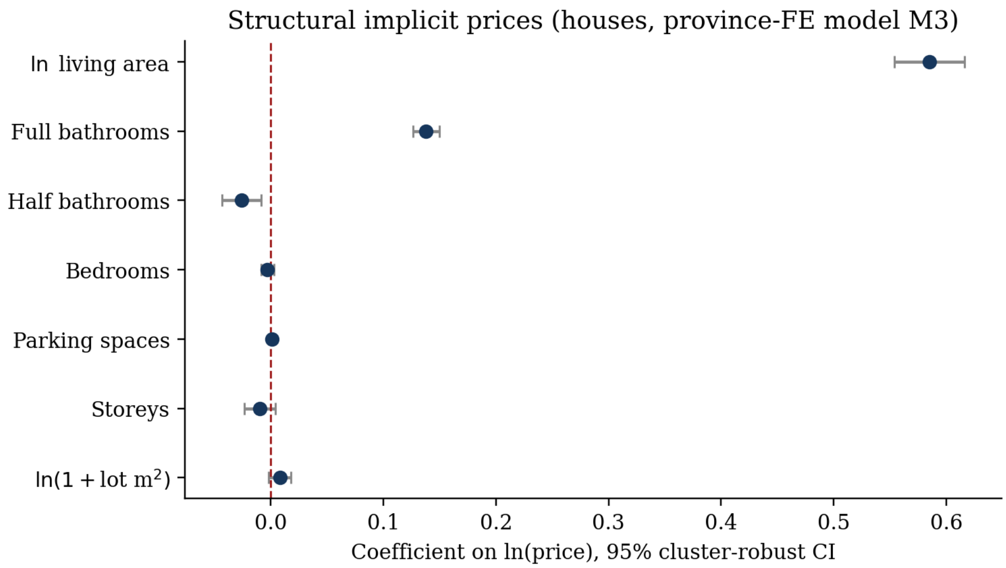


Figure 4. Marginal implicit prices of structural attributes (houses, province-FE model M3) with 95% cluster-robust confidence intervals, on the log-price scale.

grand model—its price premium net of structure, dwelling type and ownership—and express it relative to the national median (Figure 7). The highest-valued neighbourhoods, all in the City of Vancouver, trade at **150–200% above** the national-median neighbourhood for an otherwise identical dwelling; the lowest, in rural Saskatchewan, Manitoba and Newfoundland, sit **60–67% below**. The full premium distribution thus spans a factor of roughly nine between the most and least expensive neighbourhoods, dwarfing the price range attributable to any single structural attribute.

5.4. Decomposing the variance of prices

Figure 8 casts the specification ladder as a decomposition of the variance of log prices into the share explained by each successive block of controls. Physical structure accounts for 46% of the variance; dwelling type and ownership add a further 0.4 points; province adds about 10 points; and resolving location to the neighbourhood adds a further 20 points, for a total location contribution near 30 points. Just under a quarter of the variance remains unexplained and is attributable to idiosyncratic pricing and unobserved dwelling quality. The visual makes the headline unmistakable: the single largest identified block of housing value in Canada is the neighbourhood.

5.5. Nonlinearity: diminishing returns to floor space

The constant-elasticity assumption is convenient but restrictive. Re-estimating the grand model with a quadratic in log living area yields a positive linear term and a significantly negative quadratic term ($\widehat{\beta}_1 = 1.06$, $\widehat{\beta}_2 = -0.053$), implying that the marginal elasticity of price with respect to floor space

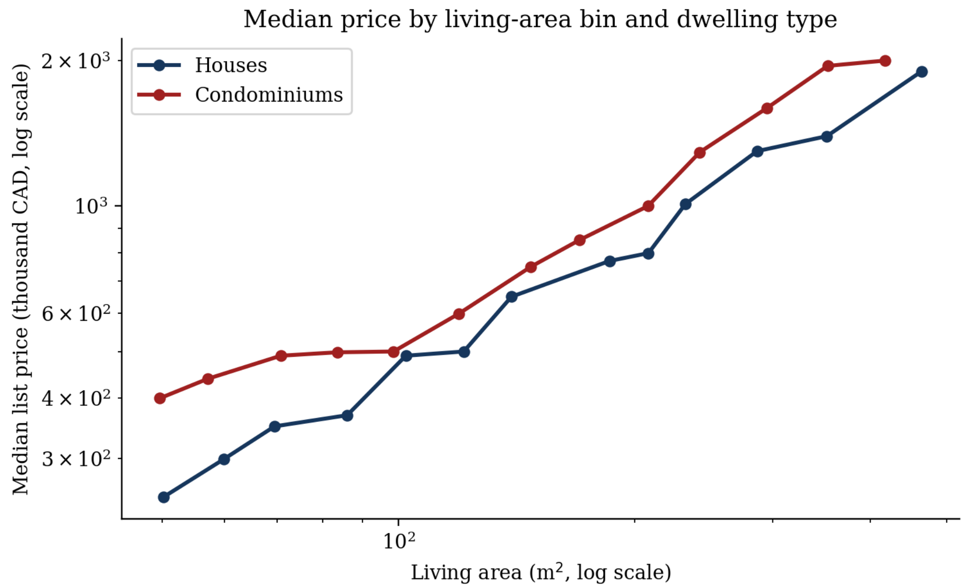
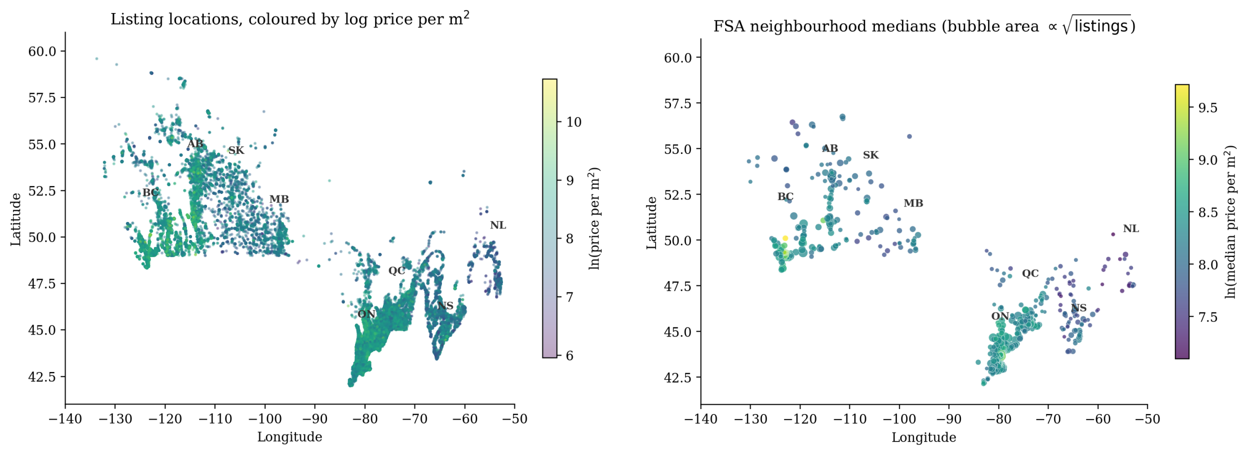


Figure 5. Median price by living-area bin and dwelling type. The concave gradient is linearised by the semi-log specification.

declines with dwelling size. Figure 9 traces the implied marginal elasticity: it falls from roughly 0.65 for a compact 60 m^2 dwelling to about 0.45 for a large 350 m^2 home. Economically, the first square metres of living space are valued most highly and additional space is subject to diminishing returns—consistent with the nonparametric hedonic surfaces of McMillen and Redfearn (2010). The constant-elasticity estimate of 0.55 is best read as an average over the size distribution.

5.6. The urban price gradient

Classic urban theory predicts that, holding structure fixed, value declines with distance from employment centres (Alonso, 1964; Muth, 1969; Mills, 1967; Ahlfeldt et al., 2015). We compute the great-circle distance from each listing to the nearest of nine major Canadian metropolitan centres and relate it to the location component of price (the residual from a structure-only model). Figure 10 confirms a pronounced gradient: dwellings within the metropolitan core carry location premia of tens of percent over the structure-only benchmark, and the premium decays steadily with distance, turning negative beyond roughly 80 km. A log-linear fit implies a semi-elasticity of -0.085 ($t > 100$): a doubling of distance to the nearest metro is associated with an 8.5% lower location premium. This agglomeration gradient is precisely the force that the FSA fixed effects absorb non-parametrically in the grand model (Combes and Gobillon, 2015).



(a) All listings

(b) FSA neighbourhood medians

Figure 6. Spatial distribution of housing value. Colour encodes log price per m^2; bubble area in panel (b) is proportional to $\sqrt{\text{listings}}$. Labels mark the provinces with substantial samples; the data cover nine provinces, while the territories and Prince Edward Island contain no listings.

6. Robustness, Heterogeneity, and Validation

6.1. Stability of the implicit prices

Table 3 re-estimates the grand FSA-fixed-effects model on six alternative samples and specifications. The size elasticity stays within the narrow band 0.51–0.62 and the bathroom premium within 0.10–0.12 log points across every cut: restricting to houses, restricting to condominiums, tightening the price trim to the 0.5/99.5 percentiles, dropping FSAs with fewer than fifty listings, and limiting the sample to the three largest provinces. The condominium subsample exhibits both the highest fit ($R^2 = 0.81$) and the largest size elasticity, consistent with condominium prices being more tightly pinned down by floor area and location and less by idiosyncratic lot and structure features. The overall picture is one of striking parameter stability: the headline implicit prices are not artefacts of a particular sample definition.

6.2. Implicit prices along the price distribution

OLS recovers the implicit price at the conditional mean, but buyers at the bottom and top of the market may value attributes differently (Zietz et al., 2008). We estimate quantile hedonic regressions (Koenker and Bassett, 1978) at the 10th through 90th percentiles of price (Table 4, Figure 11). Two patterns stand out. The living-area elasticity is roughly flat-to-rising, climbing from 0.56 at the bottom to 0.60 at the top, indicating that floor space is valued slightly more in expensive segments. The lot elasticity rises more steeply across the distribution, consistent with land being a luxury component of value. The bathroom premium is stable around 0.10–0.13 throughout. The

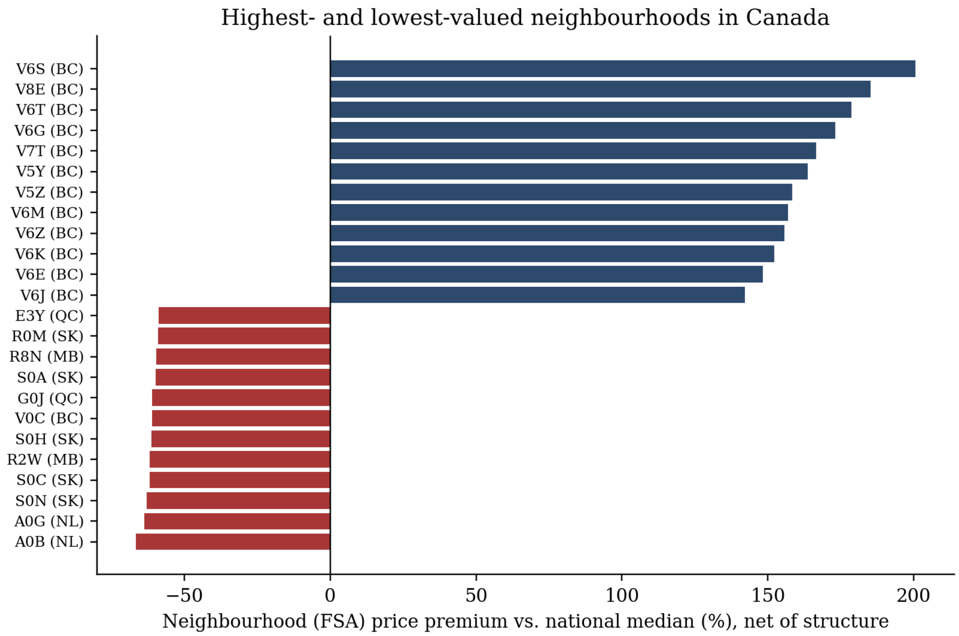


Figure 7. Highest- and lowest-valued neighbourhoods (FSAs) in Canada, measured as the estimated location premium relative to the national-median neighbourhood, net of structure, dwelling type and ownership. Only FSAs with at least 50 listings are shown.

mean-based estimates in Table 2 are thus representative, but they mask economically sensible distributional variation.

6.3. Residual spatial autocorrelation

A central justification for the neighbourhood fixed effects is that they should absorb the spatial dependence that pervades raw housing residuals (Anselin, 1988; Dubin, 1988; Basu and Thibodeau, 1998). We test this directly by computing Moran's I (Moran, 1950) on the residuals, using row-standardised k-nearest-neighbour spatial weights ($k = 10$) on a random sample of 15,000 listings. The structure-only model leaves enormous spatial autocorrelation in its residuals, $I = 0.46$ ($z = 145$, $p < 0.01$): nearby dwellings are mispriced in the same direction—the signature of omitted location. The grand model with FSA fixed effects cuts this to $I = 0.08$ ($z = 23$), an 82% reduction, confirming that the neighbourhood effects absorb the overwhelming majority of the spatial signal. Figure 12 contrasts the two Moran scatterplots. The small residual autocorrelation that remains is within-neighbourhood and could be addressed by finer geographies or an explicit spatial model (LeSage and Pace, 2009), but it is an order of magnitude smaller than the dependence the fixed effects remove, vindicating the design over a parametric spatial-lag alternative (Gibbons and Overman, 2012).

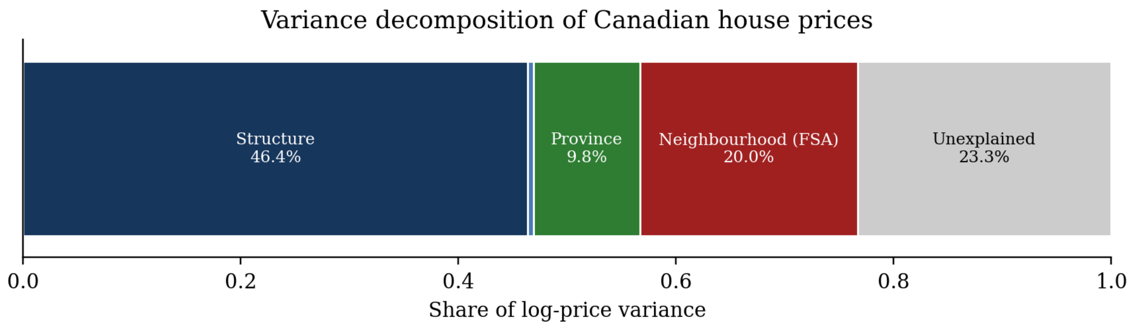


Figure 8. Variance decomposition of Canadian log house prices into structure, dwelling type/ownership, province, neighbourhood (FSA) and the unexplained residual, from the nested specification ladder.

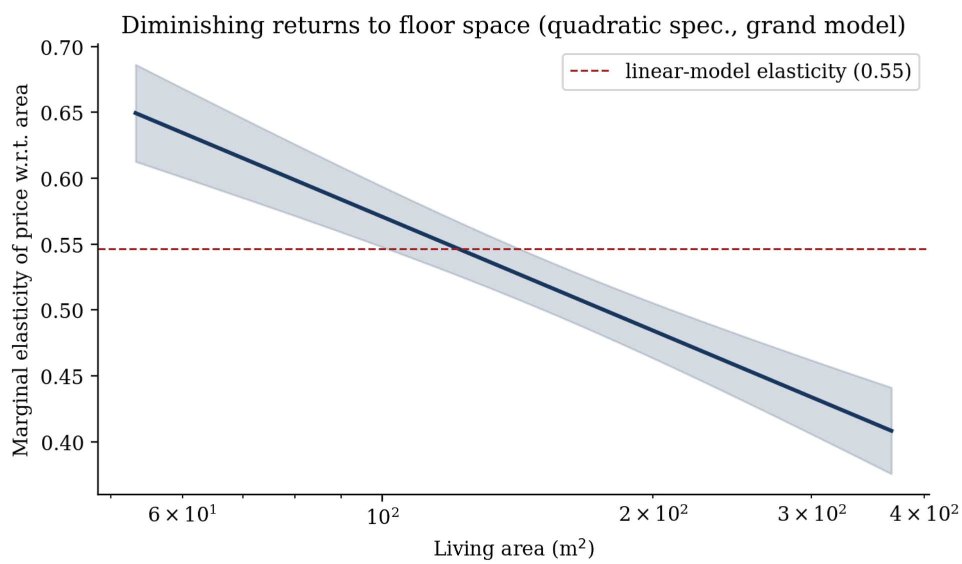


Figure 9. Marginal elasticity of price with respect to living area as a function of dwelling size, from a grand model with a quadratic in log area (95% cluster-robust band). The dashed line is the constant-elasticity estimate.

6.4. Do structural prices transfer across space?

As a demanding test of external validity we perform leave-one-province-out cross-validation: the structural model is estimated on all provinces but one and used to predict the held-out province, allowing only a province-specific intercept (the price *level* is not identified out of region). Table 5 reports the within-province R^2. The structural implicit prices transfer well to most of the country, with a mean held-out R^2 of 0.36 and values above 0.40 for the large central and western markets; transfer is weaker for the small Atlantic samples, where idiosyncratic stock and thin data dominate. That structural prices generalise across provinces—even as price *levels* differ by a factor of nine—reinforces the paper's central decomposition: structure is broadly priced the same everywhere, and it is location that varies.

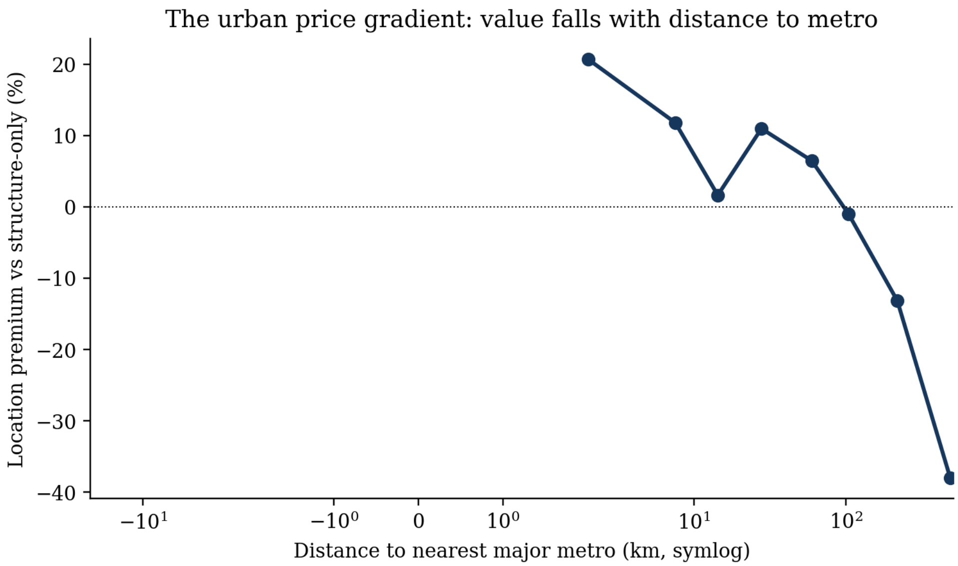


Figure 10. Urban price gradient: the location component of price (relative to a structure-only benchmark) against distance to the nearest of nine major Canadian metros. The horizontal axis is on a symmetric-log scale.

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 the high-price coastal markets—about 0.49 in British Columbia and 0.50 in Ontario—and highest in the Prairies, reaching 0.65–0.66 in Saskatchewan and Manitoba. The pattern is intuitive: where land and location dominate value (Vancouver, Toronto), 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 bathroom premium shows the mirror pattern, larger in the Prairies and Atlantic provinces than in the coastal metros.

6.6. Out-of-sample valuation accuracy

A hedonic model that fits in-sample need not predict well. Table 6 and Figure 14 report performance on a randomly held-out 20% of listings. The model attains an out-of-sample R^2 of **0.764** on log price—essentially identical to its in-sample fit, indicating negligible over-fitting despite the thousand-plus location effects. In price levels, after Duan smearing, the **median absolute valuation error is 15.8%**, the mean is 22.5%, and **59% of held-out dwellings are priced within ±20%** of their actual list price (34% within ±10%). These figures are within the accuracy bands reported in the mass-appraisal and automated-valuation literature (McCluskey et al., 2013; Clapp, 2003) and establish the transparent hedonic specification as a credible valuation benchmark (Mullainathan and

Table 3. Robustness of key implicit prices across samples

Sample / specification	N	R^2	ln area	Full bath	ln lot
Baseline (all residential)	140,931	0.767	0.547 (0.009)	0.109 (0.004)	0.030 (0.004)
Houses only	82,334	0.762	0.513 (0.010)	0.103 (0.004)	0.055 (0.003)
Condominiums only	57,857	0.813	0.619 (0.017)	0.121 (0.005)	0.005 (0.003)
Price trimmed 0.5/99.5%	139,552	0.765	0.536 (0.009)	0.107 (0.003)	0.028 (0.003)
FSAs with $\geq$50 listings	130,590	0.760	0.542 (0.010)	0.109 (0.004)	0.030 (0.004)
ON/QC/BC only	109,313	0.762	0.532 (0.010)	0.096 (0.004)	0.030 (0.003)

Notes: Each row re-estimates the grand FSA-fixed-effects model on a different sample. Cluster-robust (FSA) standard errors in parentheses. The size elasticity and bathroom premium are stable across all cuts.

Spiess, 2017).

6.7. Model fit and residual behaviour

Panel (a) of Figure 15 plots predicted against actual log prices for the grand model; the cloud hugs the 45-degree line. The residuals (panel (b)) are approximately Gaussian and centred on zero, with only 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, renovations, finish) that the data do not record.

7. Discussion

7.1. Magnitudes in the light of the literature

The location share. Our headline decomposition—structure explains 46% of log-price variance, location a further 30 points—is a cross-sectional, listing-level statement, but its magnitude aligns closely 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, with much higher shares in coastal metros, and Knoll et al. (2017) attribute about 80% of the global house-price run-up since 1950 to land. Land and "location" are not identical objects—our FSA premia capitalize amenities and access as well as physical land scarcity (Cheshire and Sheppard,

Table 4. Quantile hedonic estimates across the price distribution

Quantile	ln area	Full bath	ln lot
$\tau = 0.1$	0.557 (0.008)	0.181 (0.004)	-0.014 (0.001)
$\tau = 0.25$	0.534 (0.005)	0.146 (0.002)	-0.011 (0.001)
$\tau = 0.5$	0.552 (0.005)	0.114 (0.002)	-0.002 (0.001)
$\tau = 0.75$	0.601 (0.006)	0.087 (0.003)	0.009 (0.001)
$\tau = 0.9$	0.598 (0.010)	0.071 (0.004)	0.023 (0.002)
OLS	0.584	0.126	0.007

Notes: House subsample, province fixed effects, structural controls. Analytical standard errors in parentheses. The size elasticity rises and the lot elasticity rises with price.

1995; Albouy, 2016)—but both measurements say the same thing: the dwelling itself is the minority component of housing value. The ×9 span of our neighbourhood premia likewise echoes the enormous cross-city dispersion of land values documented by Albouy (2016) and is exactly the pattern predicted where supply constraints bind heterogeneously across locations (Saiz, 2010; Gyourko et al., 2013): Vancouver—sea, mountains and an agricultural land reserve—anchors the top of our premium distribution, as superstar-city dynamics would predict.

Structural implicit prices. The living-area elasticity of 0.55 sits inside the range catalogued by Sirmans et al. (2005) across decades of published hedonics, and its decline from 0.66 to 0.55 as location controls tighten is the classic signature of omitted-locational-quality bias 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 bathroom premium (11–15%) is at the upper end of published estimates, which we attribute to bathrooms proxying for unobserved renovation status and finish quality in listing data that lack an age variable. The diminishing marginal elasticity of floor space (from $\sim$0.65 at 60 m^2 to $\sim$0.45 at 350 m^2) gives parametric confirmation of the curvature that McMillen and Redfearn (2010) detect nonparametrically.

Spatial dependence. The 82% reduction of Moran's I (0.46 to 0.08) speaks directly to the oldest empirical worry in housing hedonics (Dubin, 1988; Basu and Thibodeau, 1998). It quantifies,

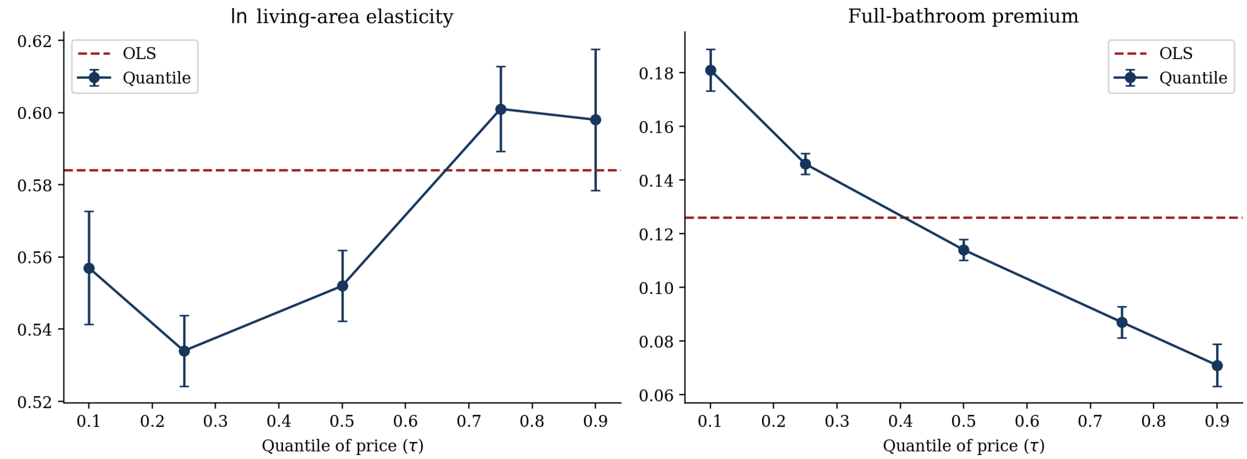


Figure 11. Implicit prices across the conditional price distribution. Quantile estimates (with 95% confidence intervals) of the living-area elasticity (left) and the full-bathroom premium (right); the dashed line is the OLS estimate.

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 $I = 0.08$ is within-FSA dependence that an explicit spatio-temporal model (Pace et al., 1998; Dubé and Legros, 2013) could exploit, particularly for prediction. It does, however, shift the burden of proof toward parsimony: a fixed effect per neighbourhood is transparent, imposes no weight-matrix assumptions, and removes five-sixths of the spatial signal.

Quantile and provincial heterogeneity. The rising size elasticity across the price distribution (0.56 at $\tau = 0.1$ to 0.60 at $\tau = 0.9$) and the rising lot elasticity mirror the distributional patterns of Zietz et al. (2008), and the coastal-versus-Prairies gradient in the size elasticity (0.49 in BC to 0.66 in MB) is 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 those slopes move within an economically meaningful band. Fully interacted (submarket-specific) coefficient systems are the natural next step, at a substantial cost in transparency.

Valuation accuracy. A held-out R^2 of 0.764 and a median absolute error of 15.8% place the transparent hedonic model within the accuracy bands reported in the mass-appraisal comparison literature (McCluskey et al., 2013) and close to the performance that machine-learning methods deliver on comparable tasks (Mullainathan and Spiess, 2017). Two caveats temper the comparison:

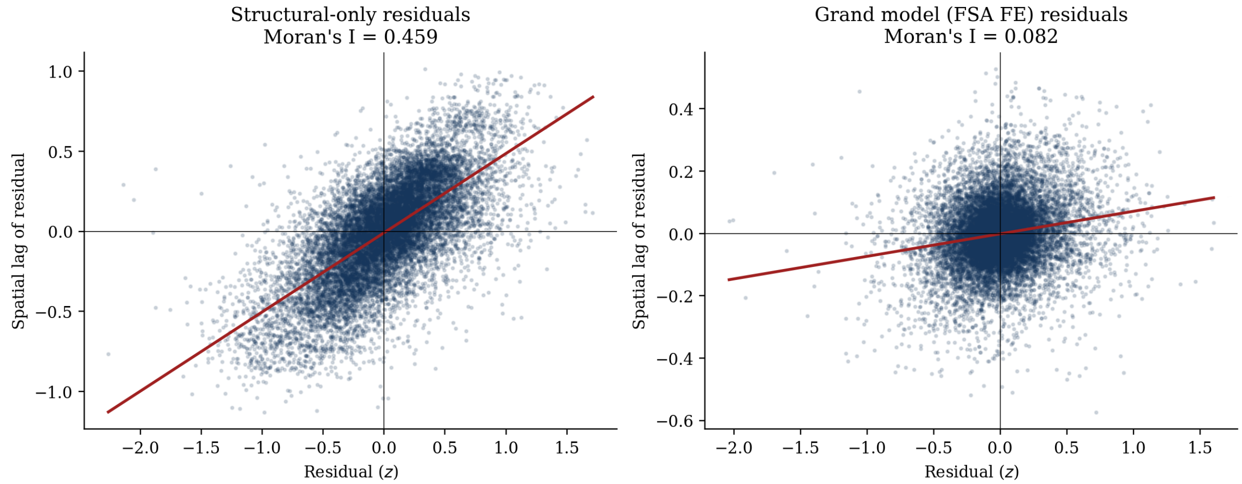


Figure 12. Moran scatterplots of model residuals against their spatial lag ($k = 10$ nearest neighbours, 15,000-listing sample). Left: structure-only model. Right: grand model with FSA fixed effects. The slope is Moran's I; it collapses from 0.46 to 0.08.

our target is the *list* price rather than the transaction price, and our evaluation is restricted to neighbourhoods observed in training. 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 rather than implicitly.

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 artifacts—are surely not it. The 0.08 residual Moran's I suggests finer geography would still add value. Second, Kuminoff et al. (2010) warn that hedonic estimates can be fragile to specification; our robustness battery (six samples, quantiles, quadratic terms) addresses the first-stage version of this concern, but any second-stage welfare use of our premia would inherit the well-known identification problems (Epple, 1987). Third, our LOPO exercise shows structural prices transfer imperfectly (mean held-out R^2 of 0.36, negative for Alberta), 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—not structure—is where most of the assessable base lives, so assessment uniformity depends first

Table 5. Leave-one-province-out spatial cross-validation

Held-out province	N	Within-province R^2
ON	48,781	0.549
MB	3,015	0.527
SK	3,400	0.475
NS	1,913	0.450
NL	1,016	0.378
BC	29,976	0.317
QC	30,556	0.303
AB	22,236	-0.101
Mean	—	0.362

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.

on getting location surfaces right; the FSA premia we publish are directly usable as such a surface. For *index construction*, the stability of structural implicit prices across provinces supports pooled hedonic indices with local intercepts (Hill, 2013), while the quantile results caution that constant-quality adjustment differs across market segments. For *automated valuation*, the results quantify the price of transparency: a fully inspectable model concedes little accuracy on within-support predictions, echoing Mullainathan and Spiess (2017)'s point that prediction tasks discipline, rather than replace, economic structure. For *housing policy*, a ×9 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 and Gobillon, 2015); policies that expand supply where premia are highest attack the decomposition's dominant term.

7.4. Limitations

Five limitations bound the interpretation. (i) Prices are *list* prices: sellers set them strategically (Genesove and Mayer, 2001) and they interact with search and bargaining (Han and Strange, 2015), so implicit prices measured on listings can differ from transaction-based ones, especially in tight markets. (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 quality sorting. (iii) Lot information is sparse and noisily parsed from free text, limiting the precision of the land-versus-structure margin at the listing level. (iv) The analysis is a single cross-section: it characterises the spatial and structural *level* of prices, not their dynamics, and cannot speak to the efficiency

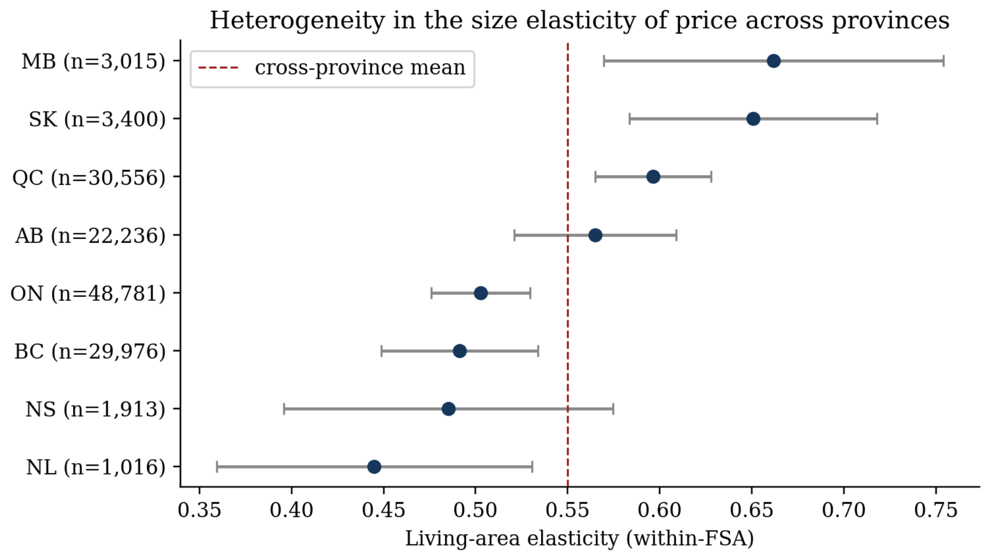


Figure 13. Living-area elasticity of price by province, estimated within FSAs, with 95% cluster-robust confidence intervals. The elasticity is smallest in the expensive coastal markets and largest in the Prairies.

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.

7.5. Future work

The natural agenda follows the limitations: linking listings to closing prices and time-on-market to estimate the list-to-sale wedge (Genesove and Mayer, 2001; Han and Strange, 2015); enriching the attribute set with age and quality; decomposing the neighbourhood premia into capitalized 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 with spatio-temporal structure (Dubé and Legros, 2013); extending the cross-section to a panel 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 140,931 MLS listings and 1,153 absorbed neighbourhood fixed effects, we estimate a grand hedonic model of the Canadian residential housing market. The exercise yields three robust conclusions. First, location dominates: resolving geography from the provincial to the neighbourhood scale raises explained price variation from 57% to 77%, and location as a

Table 6. Out-of-sample valuation performance (80/20 split)

Metric	Value
Training listings	112,732
Test listings	28,187
Out-of-sample R^2 (log price)	0.764
RMSE (log points)	0.282
Median absolute % error	15.8%
Mean absolute % error	22.5%
Share priced within ±10%	33.8%
Share priced within ±20%	59.5%

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.

whole accounts for roughly thirty percentage points of R^2—more than every structural attribute combined, and a listing-level counterpart to the large land shares documented in aggregate data (Davis and Heathcote, 2007; Knoll et al., 2017). Second, the structural implicit prices behave as hedonic theory predicts and are strikingly stable across samples: living area carries an elasticity near 0.55, each full bathroom adds about 11–15%, and bedroom counts are economically negligible once floor space is held fixed—magnitudes squarely within the range catalogued by decades of hedonic work (Sirmans et al., 2005). Third, the neighbourhood premia are vast—a factor of roughly nine separates the most and least expensive FSAs—and the model translates into a credible valuation tool, predicting held-out prices with an out-of-sample R^2 of 0.76 and a median error of 16%. These conclusions survive an extensive battery of checks: the implicit prices are stable across subsamples and across the price distribution, the neighbourhood effects absorb 82% of the spatial autocorrelation in residuals, floor space displays diminishing returns, value decays with distance to major metros as the monocentric tradition predicts, and the structural prices transfer across provinces in leave-one-province-out cross-validation.

The reading we propose is deliberately modest and, we believe, well supported: in Canada, the single largest priced component of a home is its neighbourhood. That fact is simultaneously a measurement result (for assessment, taxation and index construction), a validation result (a transparent fixed-effects design removes the spatial dependence that motivates far more elaborate machinery), and a policy result (the premium dispersion that supply constraints and agglomeration generate is the dominant term in the decomposition). Section 7 details the limitations—list prices, unobserved quality, noisy lot data, a single cross-section—and the research agenda they imply. Even within those bounds, the model provides a transparent, reproducible benchmark for automated

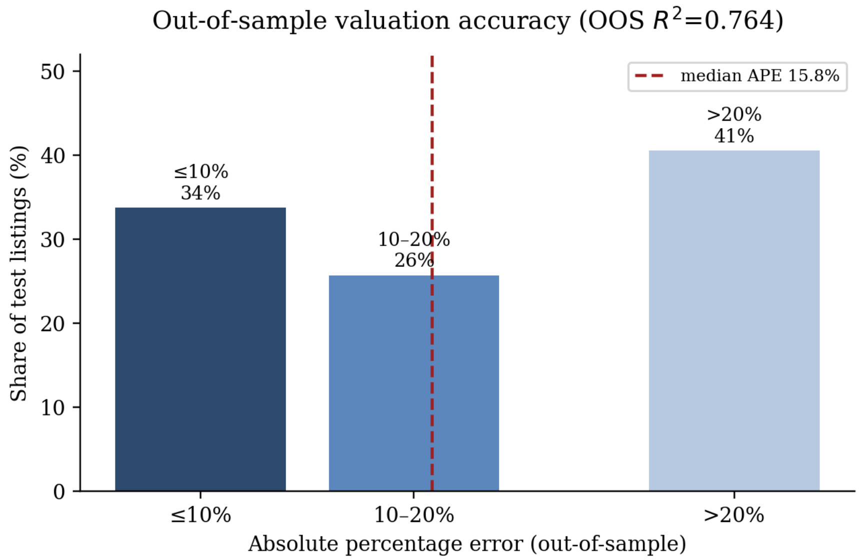
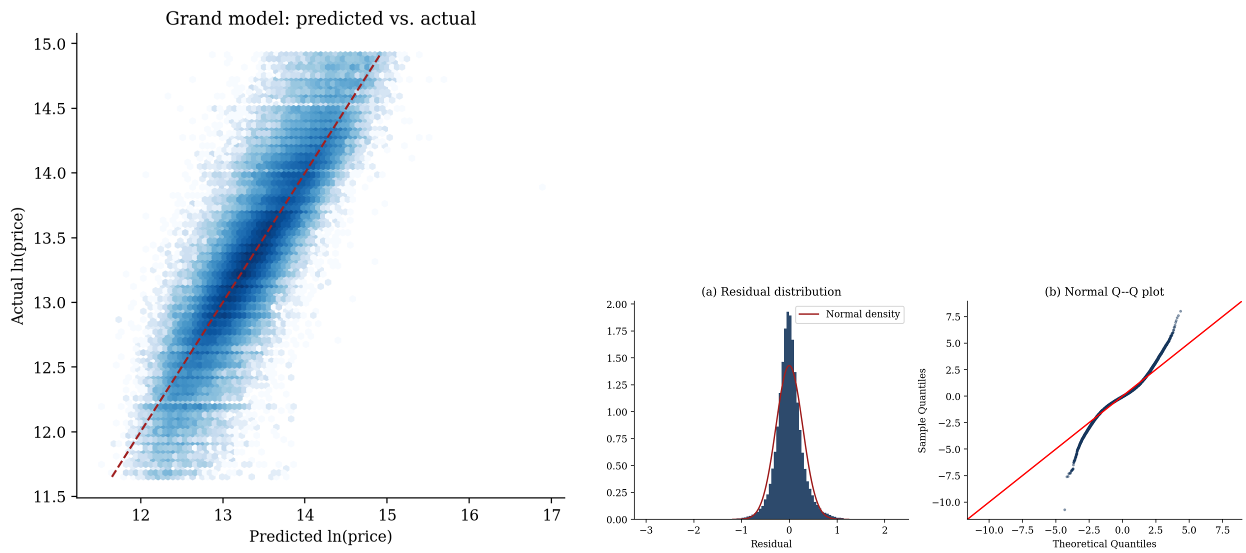


Figure 14. Out-of-sample valuation accuracy on the held-out test fold: the share of listings priced within $\pm10\%$, between 10 and 20%, and beyond 20% of the actual list price. The out-of-sample R^2 is reported in the title.

valuation, market monitoring and the welfare analysis of local amenities across the Canadian residential market.

Reproducibility. Data engineering was performed in `DuckDB` and `pandas`; estimation used `statsmodels` (cluster-robust OLS) and `linearmodels` (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.



(a) Predicted vs. actual

(b) Residual diagnostics

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

References

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.

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

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.

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

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," *Handbook of Regional and Urban Economics*, 2015, *5*, 247–348.

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

Court, Andrew T., "Hedonic Price Indexes with Automotive Examples," *The Dynamics of Automobile Demand, General Motors*, 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.

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.

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.

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," *The Price Statistics of the Federal Government, NBER*, 1961, pp. 173–196.

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.

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

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.

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.

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

Mok, Diana and Sheryl Chan, "House Prices and the Geography of the Canadian Housing Market," *Canadian Journal of Urban Research*, 2018, *27* (1), 1–18.

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.

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.

Sirmans, G. Stacy, David A. Macpherson, and Emily N. Zietz, "The Composition of Hedonic Pricing Models," *Journal of Real Estate Literature*, 2005, *13* (1), 3–43.

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.