Université du Québec en Outaouais

Département des sciences administratives

Working Paper No. 1

Heterogeneous Amenity Capitalization in the Quebec Residential Market: A Spatial Quantile Regression Approach

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 1.1

Abstract

Hedonic regressions estimated at the conditional mean impose a single implicit price for each housing attribute across the price distribution. This paper asks whether the capitalisation of neighbourhood amenities into Quebec residential asking prices varies along the conditional price distribution, and whether it survives an adjustment for spatial dependence. Using a January 2026 snapshot of 42,481 Centris listings in four segments—single-family houses, condominiums, multiplexes and vacant land—we estimate quantile regressions at $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and a two-stage spatial quantile regression in which a k-nearest-neighbour spatial lag of neighbouring log prices is instrumented by spatially lagged covariates. For houses, the coefficients of the highway, transit and park indicators decline monotonically along the distribution (highway from 0.205 at $\tau = 0.10$ to 0.034 at $\tau = 0.90$; transit from 0.178 to 0.053; park from 0.122 to 0.014) and the hospital indicator changes sign (from −0.085 to +0.041); the equality of the tail coefficients is rejected for all four, while the school discount of about 11 log points is constant. Condominiums show no detectable heterogeneity; multiplexes and vacant land show it for two amenities each. Moran's I on OLS residuals lies between 0.51 and 0.58 in every segment; the median spatial-lag coefficient ranges from 0.13 (multiplexes) to 0.32 (land), and its introduction attenuates the median highway, transit and school differentials for houses by a third to a half. The estimates are quantile-specific capitalisation gradients in asking prices, not willingness-to-pay measures; inference relies on a paired bootstrap that ignores spatial dependence, and the spatial-lag coefficients carry no standard errors. All reported figures are drawn from the preserved log of the original estimation run, and the robustness analyses that a re-estimation should provide are stated explicitly.

Keywords: hedonic price function, quantile regression, spatial lag, amenity capitalisation, implicit prices, housing submarkets, asking prices, Quebec

JEL Classification: R31, C21, C31, Q51

1. Introduction

The hedonic price function of Rosen (1974) remains the standard instrument for recovering the implicit prices of housing attributes. Regressing prices on structural characteristics, location and neighbourhood amenities decomposes market values into attribute-specific gradients that are used in appraisal, mass assessment, cost–benefit analysis and the evaluation of urban policy (Malpezzi, 2003; Sirmans et al., 2005; Bishop et al., 2020). Almost all of this work estimates the conditional mean. A least-squares hedonic regression therefore imposes a single implicit price for each attribute, whatever the position of the property in the price distribution: a buyer of a $275,000 house is assumed to capitalise park or transit proximity in the same proportion as the buyer of a $1.5 million house. Nothing in hedonic theory requires this. The hedonic price function is an envelope of bid functions of heterogeneous households, and the gradient of that envelope can legitimately differ across segments of the market when preferences, budget constraints and substitution possibilities differ (Rosen, 1974; Epple, 1987; Ekeland et al., 2004). Empirically, quantile regression has repeatedly rejected coefficient constancy across the conditional price distribution (Zietz et al., 2008; McMillen, 2008; Mak et al., 2010; Liao and Wang, 2012).

A second well-documented feature of housing data is spatial dependence. Neighbouring properties share unobserved locational attributes and are priced with reference to one another, so that regression residuals are spatially autocorrelated and the price of a property is correlated with the prices of its neighbours (Can, 1992; Basu and Thibodeau, 1998; Anselin, 1988; LeSage and Pace, 2009). Combining the two concerns leads to spatial quantile regression. The combination is not straightforward, because a spatial lag of the dependent variable is endogenous and must be instrumented inside a quantile objective; the available estimators are the two-stage "fitted-value" quantile regressions of Amemiya (1982), Powell (1983), Chen and Portnoy (1996) and Kim and Muller (2004), the instrumental-variable quantile regression of Chernozhukov and Hansen (2005, 2006, 2008) adapted to spatial autoregressive models by Su and Yang (2011), and the conditionally parametric approaches surveyed by McMillen (2013). Applications to housing remain few (Kostov, 2009; Liao and Wang, 2012; Wen et al., 2017; Gu et al., 2024; Squires et al., 2026).

This paper applies a two-stage spatial quantile regression to residential asking prices in the province of Quebec. The data are a snapshot of 42,481 Centris listings, the province-wide multiple listing service, covering four product types: single-family houses, condominiums, small multi-unit buildings ("plexes") and vacant land. For each segment we estimate the conditional quantile function of the log asking price at $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$, first without and then with a spatial lag of neighbouring log prices instrumented by the spatially lagged covariates, and we test whether the coefficients of six broker-reported amenity indicators (highway, hospital, park, school, public transit and golf course) differ between the tails of the distribution. The contribution is empirical

and modest in scope. We know of no spatial quantile hedonic study of the Quebec market, where the hedonic tradition of Des Rosiers et al. (2000), Thériault et al. (2003), Kestens et al. (2006) and Dubé et al. (2011, 2013, 2014) has concentrated on conditional-mean models of Québec City and Montreal; and we know of no study that compares amenity capitalisation across quantiles for several product types drawn from the same listing platform.

The results, all of which are obtained from asking prices in a single cross-section and must be read as conditional capitalisation differentials rather than as welfare measures (Kuminoff and Pope, 2014), are as follows. For single-family houses, the implicit prices of highway, transit and park proximity decline monotonically along the conditional distribution: the highway coefficient falls from 0.205 at the 10th percentile to 0.034 at the 90th, transit from 0.178 to 0.053 and park from 0.122 to 0.014, and the equality of the coefficients at $\tau = 0.10$ and $\tau = 0.90$ is rejected at the 1 percent level for each of them. The hospital indicator changes sign, from −0.085 at the 10th percentile to +0.041 at the 90th, and the difference is again significant. School proximity carries a stable discount of about 11 log points at every quantile. For condominiums no amenity coefficient differs significantly between the tails; for multiplexes the highway and hospital coefficients do; for vacant land the school and golf coefficients do. Residual spatial autocorrelation is strong in every segment (Moran's I between 0.51 and 0.58 on OLS residuals), and the instrumented spatial lag estimated at the median lies between 0.13 (multiplexes) and 0.32 (vacant land). Introducing the lag attenuates the median highway, transit and school coefficients for houses by roughly a third to a half, which indicates that part of what a non-spatial model attributes to an amenity indicator is the price level of the surrounding listings.

Three qualifications shape the presentation. First, the amenity indicators are binary broker classifications without distance or quality information, and the asking price is not a transaction price. Second, the inference relies on a paired bootstrap that resamples listings independently; it ignores spatial dependence and the generated-regressor nature of the fitted spatial lag, and no standard errors are available for the spatial-lag coefficient itself. Third, the numerical record of the study consists of the console log of the original estimation run; the raw Centris snapshot is not redistributable and could not be re-run for this revision. Every number in the paper is taken from that log, the tables and figures are generated from it by a versioned script, and the limitations this imposes are stated explicitly in Section 6 and Appendix A.

The remainder of the paper is organised as follows. Section 2 reviews the hedonic, spatial and quantile literatures and situates the Quebec evidence. Section 3 describes the Centris data and the construction of the four estimation samples. Section 4 sets out the estimators, the spatial weights, the inference procedures and the interpretation of the estimates. Section 5 reports the results, Section 6 discusses robustness and the limits of the available evidence, Section 7 interprets the findings, and Section 8 concludes.

2. Literature

2.1. Hedonic prices and what a cross-section identifies

In Rosen (1974), a differentiated good is a bundle of attributes $\mathbf{z}$ and the market price function $P(\mathbf{z})$ is the locus of tangencies between the bid functions of heterogeneous buyers and the offer functions of heterogeneous sellers. Its gradient $\partial P/\partial z_k$ is the implicit (marginal) price of attribute k: at the equilibrium, it equals the marginal bid of the buyer who selects that bundle. Recovering marginal willingness to pay *functions* from the gradient is the second stage of Rosen's procedure, and it requires exclusion restrictions, multiple markets or functional-form assumptions that a single cross-section does not supply (Epple, 1987; Ekeland et al., 2004; Bajari and Benkard, 2005). The distinction matters for the interpretation of quantile estimates. The coefficient of an amenity indicator at the τ-th conditional quantile is a feature of the equilibrium price schedule in the part of the market occupied by properties whose price, given their characteristics, lies at that quantile; it is not the preference parameter of a particular type of household. Kuminoff and Pope (2014) show, moreover, that even a well-identified capitalisation effect of a change in a public good equals marginal willingness to pay only under restrictive conditions on sorting and on the shape of the price function. We therefore use the term *implicit price* throughout and avoid "willingness to pay" except when referring to the literature.

The log-linear specification is the dominant functional form in applied work (Halvorsen and Pollakowski, 1981; Cropper et al., 1988; Malpezzi, 2003; Sirmans et al., 2005). Cropper et al. (1988) find in simulations that simple forms, including the semi-log, are the most robust to omitted variables and mismeasured attributes, and Cassel and Mendelsohn (1985) argue that flexible forms buy little when the true specification is unknown. In a semi-log model the coefficient of a binary attribute is converted to a proportional price differential as $\exp(\beta) - 1$ (Halvorsen and Palmquist, 1980). Kuminoff et al. (2010), Parmeter and Pope (2013) and Bishop et al. (2020) review the specification choices that most affect the credibility of hedonic estimates—spatial fixed effects, functional form, the treatment of unobserved neighbourhood quality—and recommend reporting the sensitivity of implicit prices to them.

2.2. Spatial dependence in hedonic models

Residuals of hedonic regressions are spatially autocorrelated because neighbouring properties share unmeasured locational attributes and because prices are formed with reference to comparable nearby sales (Can, 1992; Basu and Thibodeau, 1998). Two canonical models address this. The spatial autoregressive (SAR) model adds a spatial lag of the dependent variable, $\ln \mathbf{P} = \rho \mathbf{W} \ln \mathbf{P} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{W}$ is a weights matrix; the spatial error model places the dependence in the disturbance; and

the spatial Durbin model adds spatially lagged covariates $\mathbf{WX}$ (Anselin, 1988; LeSage and Pace, 2009; Elhorst, 2010; Anselin, 2010). Because $\mathbf{W}\ln\mathbf{P}$ is endogenous, the SAR model is estimated by maximum likelihood or by spatial two-stage least squares with $\mathbf{WX}$ (and higher-order lags) as instruments (Kelejian and Prucha, 1998; Lee, 2003; Kelejian and Prucha, 2010). In a SAR model the marginal effect of a covariate is not its coefficient but a matrix expression involving $(\mathbf{I}-\rho\mathbf{W})^{-1}$, which generates direct, indirect and total effects (LeSage and Pace, 2009, 2014); Small and Steimetz (2012) show that for an amenity which is itself spatially distributed, the willingness-to-pay interpretation of these effects depends on whether the spillover reflects preferences or the mechanics of price formation. Gibbons and Overman (2012) and Gibbons et al. (2015) caution that the exclusion restriction on $\mathbf{WX}$ is rarely credible and that spatial lag models are better regarded as descriptive adjustments than as structural models; we adopt that reading.

Applications to hedonic pricing include Kim et al. (2003) and Anselin and Lozano-Gracia (2008) for air quality, Bourassa et al. (2007, 2010) for prediction, and the choice of $\mathbf{W}$ has been studied by Getis (2009) and Stakhovych and Bijmolt (2009), who find that misspecifying $\mathbf{W}$ affects ρ more than the remaining coefficients. Spatially varying coefficients are an alternative to a spatial lag (Bitter et al., 2007), and recent work combines spatial structure with machine-learning predictors (Gyger et al., 2025; Chen et al., 2025); the implicit price of location itself, in a multinodal metropolitan area, is the subject of Waddell et al. (1993). In the Canadian context, Zhang et al. (2019) document spatial dependence in provincial house price dynamics; the Québec City and Montreal studies of Des Rosiers et al. (2000, 2002), Thériault et al. (2003), Dubé et al. (2011, 2013, 2014) and Dubé and Legros (2013) use spatial and spatio-temporal weights built from transaction microdata, and Dubé and Legros (2014) provide a textbook treatment of these methods on microdata.

2.3. Quantile regression in housing

Quantile regression (Koenker and Bassett, 1978; Koenker, 2005; Koenker and Hallock, 2001) estimates the conditional quantile function $Q_\tau(\ln P \mid \mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}(\tau)$. Because $\boldsymbol{\beta}(\tau)$ is allowed to vary with τ, the method characterises the whole conditional distribution instead of its mean, does not require distributional assumptions on the error term and is robust to outliers in the dependent variable (Buchinsky, 1998; Hao and Naiman, 2007). Inference can be based on the asymptotic sandwich, which requires an estimate of the conditional density of the error at zero, or on the bootstrap (Buchinsky, 1995; Hahn, 1995; Horowitz, 1998). For dependent data, cluster and block versions of the bootstrap are available (Lahiri, 2003; Parente and Santos Silva, 2016; Hagemann, 2017), and the joint behaviour of estimates at several quantiles is treated by Koenker and Xiao (2002).

Zietz et al. (2008) is the reference application to house prices: using Utah transactions, they find that most structural coefficients increase with the quantile and that the implicit prices of several

attributes are not constant. McMillen (2008) uses quantile regression to decompose the change in the distribution of Chicago house prices into changes in characteristics and changes in coefficients, and McMillen (2013) develops the spatial extensions. Mak et al. (2010) report, for Hong Kong, that the floor-area coefficient at the 90th percentile is several times that at the 10th; Farmer and Lipscomb (2010) use the quantile pattern of coefficients to identify submarket competition in a United States metropolitan area. Spatial quantile applications include Kostov (2009) for agricultural land, Liao and Wang (2012) for Shanghai apartments, Wen et al. (2017) for the Hangzhou canal landscape, Gu et al. (2024) for hospital accessibility in a Chinese city and Squires et al. (2026) for airport proximity in New Zealand. The semiparametric M-quantile approach of Schirripa Spagnolo et al. (2024) and the geographically weighted quantile regression of Chen et al. (2012) relax linearity or spatial stationarity of the coefficients. We are not aware of a published quantile hedonic study of the Quebec market.

2.4. Estimating a spatial lag inside a quantile regression

Three families of estimators exist. The first is the fitted-value two-stage approach: the endogenous regressor is replaced by its projection on the instruments, obtained by least squares (Amemiya, 1982; Powell, 1983) or by a quantile regression at the same τ (Chen and Portnoy, 1996; Kim and Muller, 2004), and the quantile regression of interest is run on the fitted value. Kim and Muller (2004) show that the quantile-based first stage yields a consistent estimator of the structural quantile coefficients under weaker conditions than the least-squares first stage, which is consistent for the slopes only when the composite first-stage error is independent of the instruments. The second family is the inverse quantile regression of Chernozhukov and Hansen (2005, 2006, 2008), which does not replace the endogenous regressor but searches over its coefficient for the value at which the instruments are irrelevant in the quantile regression of $y - \rho \mathbf{W}y$; Su and Yang (2011) adapt this estimator to spatial autoregressive models with dependent, non-identically distributed errors. The third is the conditionally parametric or locally weighted approach of McMillen (2013), which lets the coefficients vary smoothly over space instead of modelling the spatial lag. The estimator used in this paper belongs to the first family with a least-squares first stage; Section 4.4 defines it precisely and states what it does and does not deliver.

2.5. Submarkets and product types

That a single hedonic function should not be imposed on an entire metropolitan market is an old theme (Palm, 1978; Goodman and Thibodeau, 1998; Watkins, 2001). Segmented models predict better (Bourassa et al., 2003, 2007), submarkets form a lattice of imperfect substitutes (Pryce, 2013), and data-driven segmentations are now common (Rey-Blanco et al., 2024). Most of this literature

segments by location or by price; segmentation by product type is less studied, although the attribute bundles of a single-family house, a condominium unit, a small rental building and a vacant parcel are so different that pooling them is rarely attempted. Kestens et al. (2006) show, on Québec City transactions matched to buyer surveys, that implicit prices vary with household profile—a direct motivation for allowing them to vary across the price distribution.

2.6. Amenity capitalisation and the Quebec evidence

Parks and open space generally command premiums that decline with distance and depend on the type of open space (Crompton, 2001; Anderson and West, 2006; Troy and Grove, 2008); Dubé et al. (2025) find, for Québec City, a premium for houses with a direct view of a park beyond the proximity effect. School quality is capitalised (Black, 1999; Gibbons and Machin, 2008), but physical proximity to a school building is a different attribute from school quality and can carry a discount through traffic and noise. Rail transit premiums are positive on average but heterogeneous (Debrezion et al., 2007; Mohammad et al., 2013); for Montreal, Dubé et al. (2011) and Dubé et al. (2013) find positive commuter-rail effects that depend on distance and on the period, and Dubé et al. (2014) propose a spatial difference-in-differences estimator for such changes. Highways and traffic generate competing accessibility and noise effects (Theebe, 2004; Nelson, 2004); hospital accessibility has been found to capitalise heterogeneously across price quantiles (Gu et al., 2024); golf courses carry premiums for adjacent lots (Do and Grudnitski, 1995). Glaeser et al. (2008) document that lower-income households are more dependent on public transit, which suggests a larger transit premium at the lower end of the price distribution, and the sorting of heterogeneous households over neighbourhoods that underlies any such pattern goes back to Tiebout (1956) and is formalised by Bayer et al. (2007). Social-housing externalities in Québec City are studied by Dubé et al. (2023). None of these Quebec studies examines the distribution of the capitalisation effect across price quantiles, which is the gap addressed here.

3. Data

3.1. Source

The data are a snapshot of active residential listings on Centris, the listing platform of the Québec Professional Association of Real Estate Brokers, which aggregates brokered listings for the whole province. The snapshot files are dated 8 January 2026 (Unix timestamps of the extraction) and contain 42,481 listings in four product categories: 18,920 single-family houses, 9,347 condominium units, 4,619 multiplexes (two to six residential units) and 9,595 vacant parcels. Each record carries the asking price, structural attributes, lot and floor dimensions, geographic coordinates, the

municipality, the type of parking and a free-text "amenities nearby" field completed by the listing broker. The snapshot is not redistributed; the repository documents its structure and the scripts that process it.

Two features of the source condition the analysis. First, the dependent variable is an asking price. Asking prices exceed eventual transaction prices on average and the gap varies with market conditions, with the seller's pricing strategy and with time on market (Horowitz, 1992; Knight, 2002; Anglin et al., 2003; Haurin et al., 2010). The hedonic coefficients estimated on asking prices are therefore implicit prices of the listing schedule; they equal those of the transaction schedule only if the listing premium is, conditionally on the covariates, unrelated to the attributes of interest. Section 4.10 returns to this point. Second, the sample is a stock of listings at one date, not a flow of sales over a period. Properties that sell quickly are under-represented relative to properties that linger, and the listing stock in January is thinner than in spring; both features may differ across segments and price levels.

3.2. Variables

Table 1 defines the variables. Structural attributes are the number of bedrooms, the number of bathrooms (full bathrooms plus one half for each powder room), the number of storeys, the log of interior floor area in square feet, the log of lot area in square feet, the number of parking spaces and an indicator for a garage; the multiplex specification adds the number of units. Interior area is parsed from the floor-area measurement field (for condominiums, from the interior-size field when available) and lot area from the lot-size field, with conversion of square metres, acres, hectares and arpents to square feet. Location is represented by the log of the straight-line distance to downtown Montreal, computed by a flat-earth approximation from the coordinates and floored at 0.5 km, and by three regional indicators derived from the municipality name: the Montreal metropolitan area (the island, Laval and the North and South Shore suburbs), the Québec City area (including Lévis) and the remaining regions, with Gatineau (including Hull and Aylmer) as the omitted category in every segment.

Table 1. Variable definitions

Variable	Type	Definition	Prior sign
Dependent variable			
$\ln P$	Continuous	Natural log of asking price (CAD)	—
Structural characteristics			
Bedrooms	Count	Number of bedrooms	+
Bathrooms	Count	Full baths + 0.5 × half baths	+
Stories	Count	Number of stories	+/−
ln(Floor area)	Continuous	Log of interior area (sq.ft.)	+
ln(Land area)	Continuous	Log of lot size (sq.ft.)	+
Parking	Count	Number of parking spaces	+
Garage	Binary	= 1 if "Garage" in parking-type field	+
Unit count	Count	Number of units (plex only)	+
Locational variables			
ln(Dist. MTL)	Continuous	Log straight-line distance to downtown Montreal (km, floored at 0.5)	−
Montreal	Binary	= 1 if Montreal metro area	+
Québec City	Binary	= 1 if Québec City metro area	+/−
Regions	Binary	= 1 if outside the three metropolitan areas	−
Gatineau	Binary	omitted reference category	—
Amenity indicators			
Highway	Binary	= 1 if "Highway" in amenities-nearby field	+/−
Hospital	Binary	= 1 if "Hospital" in amenities-nearby field	+/−
Park	Binary	= 1 if "Park" in amenities-nearby field	+
School	Binary	= 1 if "School" in amenities-nearby field	+/−
Transit	Binary	= 1 if "Transit" in amenities-nearby field	+
Golf	Binary	= 1 if "Golf" in amenities-nearby field	+
Spatial variable (2SQR)			
$\mathbf{W}\ln\mathbf{P}$	Continuous	Inverse-distance weighted mean of the ten nearest log asking prices	+

Notes: Continuous variables are centred at segment means. Amenity indicators are keyword matches in the broker-completed Centris "amenities nearby" field and carry no distance or quality information. Prior signs summarise the literature of Section 2; +/− denotes an ambiguous prior.

The six amenity indicators equal one when the words “Highway”, “Hospital”, “Park”, “School”, “Transit” or “Golf” appear in the broker’s amenities-nearby field. They are therefore binary, broker-reported and undistanced: they do not say how far the amenity is, how large or of what quality it is, nor whether “transit” refers to a metro station, a commuter-rail station or a bus stop. A broker who regards an amenity as relevant to the marketing of a property is more likely to mention it, so the indicators partly measure the broker’s judgement about the market positioning of the listing. These limitations are discussed in Section 6.3; they argue for reading the estimates as capitalisation differentials associated with the broker’s classification rather than with physical proximity.

3.3. Sample construction

Listings with a non-positive or missing asking price or without coordinates are dropped. Houses, condominiums and multiplexes require non-missing bedrooms and bathrooms; floor area is required when it is available for at least 30 percent of the segment (houses and condominiums, not multiplexes), and must then exceed 100 sq. ft.; lot area is required when it is available for at least 30 percent of the segment (all three) and is winsorised at its 99th percentile; storeys are imputed at the segment median and parking spaces at zero when missing. Vacant parcels require a lot area above 10 sq. ft. The log asking price and, where present, the log floor area and log lot area are then trimmed at their 0.5th and 99.5th percentiles. Table 2 reports the resulting estimation samples: 6,250 houses, 1,031 condominiums, 3,755 multiplexes and 9,239 parcels, a total of 20,275 listings. The retention rate is lowest for condominiums (11 percent), because interior area is missing for most condominium listings in the snapshot. This selection is not innocuous: a condominium listing that reports its interior area may differ systematically from one that does not, and the condominium estimates should be read as conditional on that reporting.

Table 2. Sample construction by market segment

	Houses	Condominiums	Multiplexes	Vacant land
Raw listings	18,920	9,347	4,619	9,595
Estimation sample	6,250	1,031	3,755	9,239
Retention rate (%)	33.0	11.0	81.3	96.3
Number of regressors	17	17	17	11
Minimum asking price ($)	99,000	177,700	130,000	16,000
Median asking price ($)	579,000	425,000	799,000	150,000
Maximum asking price ($)	5,295,000	2,932,760	4,699,999	8,900,000

Notes: Centris active listings, Quebec, snapshot of May 2026. The estimation sample retains listings with a positive asking price, valid coordinates, non-missing structural variables required by the segment specification (Table 4), floor area above 100 sq. ft. where applicable, land area above zero (above 10 sq. ft. for land), land area winsorised at its 99th percentile, and asking price, floor area and land area trimmed at the 0.5th and 99.5th percentiles of their logarithms. The number of regressors excludes the constant. Minimum and maximum prices are those of the estimation sample. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

3.4. Descriptive statistics

Table 3 reports the prevalence of each amenity indicator and the regional composition of the four samples; the median asking prices are $579,000 for houses, $425,000 for condominiums, $799,000 for multiplexes and $150,000 for parcels. Amenity mentions are most frequent for condominiums (transit 73 percent, park 81 percent, school 83 percent) and multiplexes, and least frequent for vacant land (transit 8 percent), which reflects the urban location of the former and the peripheral location of the latter. The regional composition differs markedly across segments: 70 percent of the houses, 70 percent of the condominiums, 44 percent of the multiplexes and 89 percent of the parcels lie outside the three metropolitan areas, and the Montreal area accounts for 47 percent of the multiplexes but only 19 percent of the houses. Median asking prices in the Montreal area are higher than elsewhere in every segment ($868,942 against $520,000 for houses in the other regions; $1,049,000 against $525,000 for multiplexes). The amenity indicators are thus correlated with the regional and distance variables, and their coefficients must be read conditionally on those controls. The preserved log of the estimation run does not record the moments of the structural variables (bedrooms, bathrooms, areas); they are therefore not tabulated here, and Appendix A lists them among the outputs to be

restored when the pipeline is re-run.

Table 3. Amenity prevalence, regional composition and median asking prices of the estimation samples

	Houses	Condominiums	Multiplexes	Vacant land
Amenity indicators (share of listings, %)				
Highway	48.5	72.5	54.5	24.3
Hospital	30.5	51.5	45.2	13.6
Park	60.3	81.2	66.9	30.5
School	69.2	82.9	73.3	34.2
Transit	34.8	73.5	54.4	8.3
Golf course	39.7	40.0	24.5	26.6
Garage	56.9	37.1	31.6	—
Regional composition (share of listings, %)				
Montreal area	18.9	25.9	47.5	8.6
Québec City area	3.8	3.0	4.8	1.8
Gatineau	7.3	1.5	3.4	0.7
Other regions	70.0	69.6	44.3	88.9
Median asking price by region ($)				
Montreal area	868,942	582,000	1,049,000	499,000
Québec City area	536,384	450,000	649,000	327,000
Gatineau	534,856	319,000	659,000	314,900
Other regions	520,000	375,000	525,000	147,900
All listings	579,000	425,000	799,000	150,000
Estimation sample (N)	6,250	1,031	3,755	9,239

Notes: Amenity indicators equal one when the corresponding keyword appears in the Centris "amenities nearby" field; the garage indicator is built from the parking-type field and is not defined for vacant land. Regional shares are computed from the listing counts recorded in the log (e.g. 1,180 Montreal-area houses). Moments of the structural variables (bedrooms, bathrooms, areas) were not written to the preserved log and are therefore not reported; see Section 3. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

4. Methodology

4.1. Hedonic specification

For each segment the baseline is the semi-log hedonic equation

$$\ln P_i = \alpha + \mathbf{X}_i'\boldsymbol{\beta} + \mathbf{A}_i'\boldsymbol{\gamma} + \mathbf{R}_i'\boldsymbol{\delta} + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{1}$$

where P_i is the asking price of listing i, $\mathbf{X}_i$ collects the structural and locational variables of the segment (Table 4), $\mathbf{A}_i$ the amenity indicators and $\mathbf{R}_i$ the three regional indicators. Continuous regressors are centred at their segment means, which leaves the slopes unchanged and makes the intercept the conditional log price of a mean-characteristic property in the reference region (Marquardt, 1980). Equation (1) is estimated by OLS with HC3 standard errors (MacKinnon and White, 1985; Long and Ervin, 2000; Cribari-Neto, 2004) and serves as the conditional-mean benchmark. For a binary attribute with coefficient γ_k, the proportional price differential is $\exp(\gamma_k) - 1$ (Halvorsen and Palmquist, 1980); for small coefficients the two are close and we quote coefficients in log points.

Table 4. Regressors included in each segment specification

Variable	Houses	Condominiums	Multiplexes	Vacant land
Bedrooms	✓	✓	✓	—
Bathrooms	✓	✓	✓	—
Stories	✓	✓	✓	—
ln(Floor area)	✓	✓	—	—
ln(Land area)	✓	✓	✓	✓
Parking spaces	✓	✓	✓	—
Garage	✓	✓	✓	—
Unit count	—	—	✓	—
ln(Dist. Montreal)	✓	✓	✓	✓
Regional dummies	✓	✓	✓	✓
Amenity indicators (6)	✓	✓	✓	✓
Spatial lag $\widehat{\mathbf{W}\ln\mathbf{P}}$ (2SQR only)	✓	✓	✓	✓
Number of regressors (excl. constant)	17	17	17	11

Notes: Floor area is unavailable for most multiplex listings and is excluded from that segment; lot area is available for condominiums (lot share of the building) and is retained. Gatineau is the omitted regional category in every segment. The amenity indicators are highway, hospital, park, school, transit and golf course; the garage indicator is counted with the structural variables and is undefined for land.

4.2. Quantile regression

The τ-th conditional quantile of the log asking price is modelled as linear in the same regressors,

$$Q_\tau(\ln P_i \mid \mathbf{X}_i, \mathbf{A}_i, \mathbf{R}_i) = \alpha(\tau) + \mathbf{X}_i'\boldsymbol{\beta}(\tau) + \mathbf{A}_i'\boldsymbol{\gamma}(\tau) + \mathbf{R}_i'\boldsymbol{\delta}(\tau), \tag{2}$$

and the coefficient vector $\boldsymbol{\theta}(\tau) = (\alpha(\tau), \boldsymbol{\beta}(\tau)', \boldsymbol{\gamma}(\tau)', \boldsymbol{\delta}(\tau)')'$ is estimated by minimising the check-function objective of Koenker and Bassett (1978),

$$\hat{\boldsymbol{\theta}}(\tau) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \rho_\tau\left(\ln P_i - \mathbf{Z}_i'\boldsymbol{\theta}\right), \qquad \rho_\tau(u) = u\left[\tau - \mathbf{1}(u < 0)\right], \tag{3}$$

with $\mathbf{Z}_i = (1, \mathbf{X}_i', \mathbf{A}_i', \mathbf{R}_i')'$. We estimate $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$, the grid used by Zietz et al. (2008) and Liao and Wang (2012), which is fine enough to distinguish monotone, U-shaped and sign-changing coefficient profiles while keeping the tail estimates away from the extremes where the conditional density is thin. The linear program is solved by the interior-point routine of

`statsmodels` with a ceiling of 5,000 iterations. Fit is summarised by the pseudo-R^2 of Koenker and Machado (1999),

$$R^1(\tau) = 1 - \frac{\sum_i \rho_\tau\left(\ln P_i - \mathbf{Z}_i'\hat{\boldsymbol{\theta}}(\tau)\right)}{\sum_i \rho_\tau\left(\ln P_i - \hat{Q}_\tau\right)}, \tag{4}$$

where $\hat{Q}_\tau$ is the unconditional sample quantile. $R^1(\tau)$ compares the weighted absolute deviations of the fitted model to those of the constant-only model at the same τ; it is a local measure and is not comparable to the OLS R^2.

It is worth stating what the quantile index means here. τ indexes the conditional distribution of the log asking price given the covariates, not the unconditional price distribution: a listing at $\tau = 0.10$ is one whose price is low *relative to what its characteristics and location predict*, not necessarily a cheap listing. The conditional and unconditional rankings are correlated, so it is convenient to speak of the "lower" and "upper" segments of the market, but the coefficients describe how the price schedule is shaped at different conditional ranks.

4.3. Spatial weights

Spatial relations are represented by a k-nearest-neighbour matrix with $k = 10$ and inverse-distance weights, built within each segment from the listing coordinates with a k-d tree (Bentley, 1975):

$$w_{ij} = \frac{d_{ij}^{-1}}{\sum_{l \in \mathcal{N}_i} d_{il}^{-1}} \ \text{if } j \in \mathcal{N}_i, \qquad w_{ij} = 0 \ \text{ otherwise}, \tag{5}$$

where $\mathcal{N}_i$ is the set of the ten listings nearest to i and d_{ij} their distance. The matrix is row-standardised by construction, so $[\mathbf{W} \ln \mathbf{P}]_i$ is a weighted average of the ten neighbouring log prices and the lag coefficient is read as the elasticity of a listing's price to the average price of its neighbours. Nearest-neighbour weights avoid the islands produced by distance bands in a province where listing density varies by several orders of magnitude (Getis, 2009); their cost is that the ten neighbours of a rural parcel may lie tens of kilometres away while those of a Montreal condominium lie in the same building. Two implementation details should be recorded. The neighbour search and the distances d_{ij} are computed in unprojected latitude–longitude degrees; at the latitude of Quebec one degree of longitude is about 70 percent of a degree of latitude, so the neighbourhoods are slightly elongated in the north–south direction. And listings with identical coordinates (several units in the same building) receive a distance floored at 10^{-10}, so that such a neighbour receives almost the entire weight of row i. Neither feature affects the row-standardisation, but both are candidates for the sensitivity analysis described in Section 6.

4.4. Two-stage spatial quantile regression

The spatial quantile model adds a spatial lag of the dependent variable to equation (2):

$$Q_\tau(\ln P_i \mid \mathbf{Z}_i, [\mathbf{W}\ln\mathbf{P}]_i) = \mathbf{Z}_i'\boldsymbol{\theta}(\tau) + \rho(\tau)\,[\mathbf{W}\ln\mathbf{P}]_i. \tag{6}$$

The lag is endogenous because the neighbours' prices depend on the price of i through the same mechanism. Following the spatial two-stage least squares logic of Kelejian and Prucha (1998), the lag is instrumented by the spatially lagged covariates, and following the fitted-value two-stage quantile regression of Amemiya (1982) and Powell (1983), generalised by Chen and Portnoy (1996) and Kim and Muller (2004), the endogenous regressor is replaced in the quantile objective by its first-stage projection. The first stage is the least-squares regression

$$\mathbf{W}\ln\mathbf{P} = \boldsymbol{\Pi}_0 + \mathbf{Z}\boldsymbol{\Pi}_1 + \mathbf{W}\mathbf{Z}\boldsymbol{\Pi}_2 + \mathbf{v}, \qquad \widehat{\mathbf{W}\ln\mathbf{P}} = \hat{\boldsymbol{\Pi}}_0 + \mathbf{Z}\hat{\boldsymbol{\Pi}}_1 + \mathbf{W}\mathbf{Z}\hat{\boldsymbol{\Pi}}_2, \tag{7}$$

where $\mathbf{Z}$ excludes the constant and $\mathbf{WZ}$ contains the spatial lags of all covariates, including the amenity and regional indicators. The second stage is the quantile regression

$$\hat{\boldsymbol{\theta}}(\tau), \hat{\rho}(\tau) = \arg\min \sum_{i=1}^{N} \rho_\tau\left(\ln P_i - \mathbf{Z}_i'\boldsymbol{\theta} - \rho\,[\widehat{\mathbf{W}\ln\mathbf{P}}]_i\right). \tag{8}$$

We call the resulting estimator 2SQR. Three properties should be kept in mind. First, because the first stage is least squares rather than a quantile regression at τ, the estimator is the two-stage least absolute deviations estimator of Amemiya (1982) generalised to an arbitrary τ, not the estimator recommended by Kim and Muller (2004); it is consistent for the slopes of equation (6) when the composite error $u_i(\tau) + \rho(\tau)v_i$ has a conditional τ-quantile that does not depend on the instruments, which holds under independence of the errors from the covariates and their spatial lags but not under general heteroskedasticity. Second, it is not the inverse quantile regression of Chernozhukov and Hansen (2006) and Su and Yang (2011), which would be the natural estimator if the spatial lag were to be treated as a structural endogenous regressor under weaker conditions; we did not implement it. Third, the second-stage residuals are computed with the *fitted* lag. Since $\widehat{\mathbf{W}\ln\mathbf{P}}$ is a linear combination of $\mathbf{Z}$ and $\mathbf{WZ}$, the second stage is, from the point of view of residual diagnostics, a quantile regression on $\mathbf{Z}$ and one linear combination of $\mathbf{WZ}$; its residuals should not be expected to be less spatially autocorrelated than those of equation (2), and they are not (Section 5.5). The structural residuals $\ln P_i - \mathbf{Z}_i'\hat{\boldsymbol{\theta}}(\tau) - \hat{\rho}(\tau)[\mathbf{W}\ln\mathbf{P}]_i$, which use the observed lag, are the appropriate object for that diagnostic and were not computed in the original run.

The exclusion restriction is that neighbours' characteristics affect a listing's price only through neighbours' prices. As Gibbons and Overman (2012) emphasise, this is an assumption of convenience:

the lot sizes or amenity classifications of neighbouring listings may describe the neighbourhood directly. We therefore regard $\hat{\rho}(\tau)$ as a descriptive measure of how strongly a listing's conditional quantile co-moves with the neighbouring price level, and the 2SQR amenity coefficients as the part of the amenity differential that is not accounted for by that co-movement. Instrument relevance is assessed with the first-stage R^2; the partial F-statistic on the excluded instruments $\mathbf{WZ}$ (Stock and Yogo, 2005) was not recorded by the original run and the overall first-stage F reported in Table 13 is only an upper bound for it.

4.5. Effect decomposition under the spatial lag

In a spatial autoregressive model the response of the vector of log prices to a change in regressor k is $(\mathbf{I} - \rho\mathbf{W})^{-1}\beta_k$, whose diagonal gives direct effects, off-diagonal elements indirect effects and row sums total effects (LeSage and Pace, 2009). Computing these matrices for samples of up to 9,239 listings, and bootstrapping them, was not attempted. We report instead the scalar approximation that treats $\mathbf{W}$ as if every row sum of $(\mathbf{I} - \rho\mathbf{W})^{-1}$ were $1/(1-\rho)$:

$$\text{Direct}_k(\tau) \approx \beta_k(\tau), \qquad \text{Indirect}_k(\tau) \approx \frac{\rho(\tau)\,\beta_k(\tau)}{1-\rho(\tau)}, \qquad \text{Total}_k(\tau) \approx \frac{\beta_k(\tau)}{1-\rho(\tau)}. \tag{9}$$

The approximation is exact for the total effect when $\mathbf{W}$ is row-standardised and understates the direct effect slightly, because the diagonal of $(\mathbf{I} - \rho\mathbf{W})^{-1}$ exceeds one by the feedback term (LeSage and Pace, 2014). The indirect share of the total is $\rho(\tau)$ by construction, so the decomposition adds no information beyond $\hat{\rho}(\tau)$ and $\hat{\beta}_k(\tau)$; it is reported because dollar-equivalent magnitudes at the median are easier to read than log points, and it is evaluated at $\tau = 0.50$ only, the single quantile for which the original run recorded the second-stage coefficients.

4.6. Inference

The asymptotic covariance of a quantile regression estimator depends on the conditional density of the error at zero, which is awkward to estimate in the tails (Koenker, 2005). Standard errors for the coefficients of equation (2) are therefore obtained by the paired (non-parametric) bootstrap (Freedman, 1981; Efron and Tibshirani, 1993; Buchinsky, 1995; Hahn, 1995): $B = 200$ samples of N listings are drawn with replacement, the quantile regression is re-estimated on each, and the standard error of $\hat{\theta}_k(\tau)$ is

$$\widehat{\text{SE}}_{\text{boot}}[\hat{\theta}_k(\tau)] = \left[\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_k^{(b)}(\tau) - \bar{\theta}_k(\tau)\right)^2\right]^{1/2}, \tag{10}$$

with $\bar{\theta}_k(\tau)$ the bootstrap mean. A replication that fails to converge is discarded and the standard error is computed from the successful replications, provided at least 100 succeed; all twenty segment–quantile combinations met this condition. Significance is assessed by the normal approximation $\hat{\theta}_k / \widehat{\mathrm{SE}}_{\mathrm{boot}}$. Two caveats apply. The paired bootstrap resamples listings as if they were independent, which they are not; with Moran's I above 0.5 on the residuals, the bootstrap standard errors are likely to understate the sampling variability, and a block or cluster bootstrap (Lahiri, 2003; Cameron et al., 2008; Hagemann, 2017) or a bootstrap that respects the spatial structure of Moran's test (Jin and Lee, 2015) would be preferable. And the bootstrap was applied to the quantile regressions of equation (2) only; the 2SQR second stage was not bootstrapped, so no standard errors are available for $\hat{\rho}(\tau)$ or for the 2SQR coefficients, and the generated-regressor problem of the fitted lag (Powell, 1983) is not addressed.

4.7. Tests of inter-quantile equality

Whether the implicit price of amenity k differs between two quantiles is tested with

$$W_k(\tau_1, \tau_2) = \frac{[\hat{\gamma}_k(\tau_1) - \hat{\gamma}_k(\tau_2)]^2}{\widehat{\mathrm{Var}}_{\mathrm{boot}}[\hat{\gamma}_k(\tau_1)] + \widehat{\mathrm{Var}}_{\mathrm{boot}}[\hat{\gamma}_k(\tau_2)]} \overset{H_0}{\sim} \chi^2(1), \tag{11}$$

for the pairs $(0.10, 0.90)$, $(0.25, 0.75)$ and $(0.10, 0.50)$. The denominator omits the covariance between the two estimators, which is generally positive because both are estimated on the same sample (Koenker and Xiao, 2002); the statistic therefore tends to understate the evidence against H_0, and the rejections reported below are conservative in that respect, although not with respect to the spatial dependence ignored by the bootstrap. The test is in the spirit of the quantile-based heteroskedasticity test of Koenker and Bassett (1982), which asks whether slopes differ across quantiles. A joint test across all amenities would require the full bootstrap covariance matrix of the stacked coefficient vector at the two quantiles, which was not saved; we do not report an omnibus statistic and instead count the amenities for which H_0 is rejected.

4.8. Residual spatial autocorrelation

Spatial dependence in the residuals e_i of each model is measured by Moran's I,

$$I = \frac{N}{S_0} \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}}, \qquad S_0 = \sum_i \sum_j w_{ij}, \tag{12}$$

with $\mathbf{e}$ demeaned, and is standardised with the expectation $-1/(N-1)$ and the randomisation variance of Cliff and Ord (1981), which adjusts for the kurtosis of the residuals. The statistic is computed on OLS residuals, on the residuals of each quantile regression and on the second-stage

residuals of each 2SQR, using the same **W**.

4.9. Dollar-equivalent implicit prices

To express the quantile coefficients in dollars, the log-point differential at quantile τ is applied to the unconditional sample quantile of the asking price $P(\tau) = \exp(\hat{Q}_\tau)$:

$$\pi_k(\tau) = P(\tau)\left[\exp\big(\hat{\gamma}_k(\tau)\big) - 1\right]. \tag{13}$$

This is a presentational device, not an additional estimate: it scales a conditional differential by a price level that is representative of the τ-th part of the market, and it inherits all the qualifications of the coefficient.

4.10. Identification and interpretation

The estimates are conditional associations in a single cross-section of asking prices and support no causal claim about the effect of building a park or opening a transit station. Four issues delimit their interpretation. Households sort across locations on unobserved preferences that are correlated with amenity access (Bayer et al., 2007), so an amenity indicator may proxy for the type of neighbourhood and of household that selects it; the regional indicators, the distance gradient and the spatial lag absorb part of this, not all of it. The amenity indicators are broker classifications; measurement error in a binary regressor attenuates its coefficient, and a broker's propensity to mention an amenity may itself vary with the price segment, which would produce quantile heterogeneity in the coefficient without any heterogeneity in capitalisation. Asking prices differ from transaction prices by a premium that depends on market conditions and seller strategy; the coefficients are unbiased for the transaction-price schedule only if that premium is, conditionally on the covariates, unrelated to the amenity indicators at each quantile. And the quantile coefficients are gradients of an equilibrium price schedule, not willingness-to-pay parameters of a household type (Kuminoff and Pope, 2014); a declining profile of the transit coefficient across τ is consistent with stronger transit dependence among buyers of lower-priced homes (Glaeser et al., 2008), but equally with differences in the supply of transit-adjacent housing or in broker reporting across segments. We use the expression *quantile-specific capitalisation gradient* for the estimates and reserve mechanisms for Section 7, where they are presented as interpretations rather than findings.

5. Results

All estimates in this section are those of the original estimation run, whose console log is the only preserved numerical record of the study; the tables and figures are generated from that log by a

versioned script (Appendix A). Coefficients are in log points of the asking price.

5.1. Conditional-mean benchmark and spatial autocorrelation

Table 5 reports the fit of the OLS specification (1) and Moran's I on its residuals. The adjusted R^2 is 0.619 for houses, 0.669 for condominiums, 0.576 for multiplexes and 0.154 for vacant land; the last figure reflects a specification that contains only lot area, distance, region and six amenity indicators, when zoning, servicing and development potential are the main determinants of parcel prices (Colwell and Munneke, 1997). Moran's I on the OLS residuals is 0.506 for houses, 0.551 for condominiums, 0.528 for multiplexes and 0.580 for land, with z-statistics between 22 and 67. These values are high by the standards of urban hedonic studies, which typically report residual Moran statistics of 0.2–0.4 on neighbourhood-level weights (Basu and Thibodeau, 1998; Bourassa et al., 2007); two features of the design raise them. The samples span the whole province, so that the regional indicators and a single distance gradient leave considerable between-municipality variation in the residuals, and the ten-nearest-neighbour weights are very local, so that unmodelled street- and building-level price levels are strongly reflected in **We**. Whatever its source, residual dependence of this magnitude means that the OLS and quantile standard errors computed under independence should be treated as lower bounds, and it motivates the spatial lag of Section 5.4.

Table 5. OLS fit and Moran's I test on OLS residuals, by segment

Segment	N	Adjusted R^2	Moran's I	z-statistic	p-value
Single-family houses	6,250	0.619	0.506	51.9	< 0.001
Condominiums	1,031	0.669	0.551	22.0	< 0.001
Multiplexes	3,755	0.576	0.528	40.2	< 0.001
Vacant land	9,239	0.154	0.580	67.2	< 0.001

Notes: OLS estimates of equation (1) with HC3 standard errors. Moran's I uses the row-standardised inverse-distance k-nearest-neighbour matrix ($k = 10$) of equation (5); the z-statistic uses the randomisation (kurtosis-adjusted) variance of Cliff and Ord (1981); p-values are two-sided normal. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

5.2. Quantile regressions: single-family houses

Table 6 reports the amenity coefficients of equation (2) at the five quantiles, with the OLS coefficient for comparison, and Figure 1 plots the profiles for all four segments; the structural, locational and

regional coefficients for houses are in Appendix Table 17. Four profiles stand out.

The highway indicator has the steepest decline. Its coefficient is 0.205 at $\tau = 0.10$, 0.156 at 0.25, 0.113 at the median, 0.071 at 0.75 and 0.034, no longer significant, at 0.90; the OLS coefficient of 0.133 lies between the median and the first quartile. Transit proximity follows the same pattern from 0.178 to 0.053, with the OLS value of 0.113 again near the median. Park proximity declines from 0.122 at the 10th percentile to 0.038 at the median (significant at the 5 percent level) and to 0.014 at the top, where it is indistinguishable from zero. For these three indicators the conditional-mean coefficient is a poor summary of the schedule: it overstates the differential in the upper part of the conditional distribution by a factor of two to four and understates it in the lower part by a third to a half.

The hospital indicator changes sign. It is -0.085 at $\tau = 0.10$, -0.061 at 0.25 and -0.027 at the median, all significant, and then +0.001 at 0.75 and +0.041 at 0.90, the last marginally significant at the 10 percent level. The OLS coefficient of -0.049 describes neither tail. The school indicator, by contrast, is uniformly negative, between -0.102 and -0.136, and the garage and golf-course indicators are positive and comparatively stable (garage 0.047–0.100; golf 0.015–0.063), with the smallest values at the 90th percentile.

Table 6. Amenity coefficients by quantile — single-family houses

	Quantile regression, τ					OLS
Amenity indicator	0.10	0.25	0.50	0.75	0.90	
Highway	0.205***	0.156***	0.113***	0.071***	0.034	0.133***
	(0.033)	(0.018)	(0.015)	(0.014)	(0.025)	
Hospital	−0.085***	−0.061***	−0.027**	0.001	0.041*	−0.049***
	(0.028)	(0.015)	(0.013)	(0.015)	(0.024)	
Park	0.122***	0.070***	0.038**	0.015	0.014	0.042**
	(0.034)	(0.019)	(0.019)	(0.021)	(0.025)	
School	−0.102***	−0.136***	−0.110***	−0.110***	−0.111***	−0.122***
	(0.033)	(0.022)	(0.021)	(0.019)	(0.024)	
Transit	0.178***	0.162***	0.094***	0.074***	0.053**	0.113***
	(0.027)	(0.020)	(0.017)	(0.016)	(0.025)	
Golf course	0.041	0.063***	0.038***	0.036***	0.015	0.052***
	(0.026)	(0.016)	(0.013)	(0.012)	(0.020)	
Garage	0.061***	0.100***	0.096***	0.082***	0.047**	0.100***
	(0.020)	(0.015)	(0.012)	(0.014)	(0.022)	
Pseudo-R^2 / adj. R^2	0.355	0.364	0.400	0.436	0.444	0.619
N	6,250					

Notes: Dependent variable: log asking price. Each column is a separate estimate of equation (2) (quantile regression) or equation (1) (OLS) including all structural, locational and regional controls listed in Table 4. Bootstrap standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels (two-sided normal test using bootstrap standard errors, $B = 200$ paired resamples). OLS stars use HC3 standard errors and are reported as printed in the log (the log applies the thresholds 0.001, 0.01 and 0.05 to the OLS column, so OLS stars are conservative relative to the quantile columns). The pseudo-R^2 is that of Koenker and Machado (1999). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

Table 7 tests the equality of each amenity coefficient across three quantile pairs. For the $(0.10, 0.90)$ pair the statistic is 17.1 for highway, 12.1 for hospital, 11.6 for transit and 6.7 for park, each rejecting equality at the 1 percent level, against 0.05 for school, 0.6 for golf and 0.2 for garage. The $(0.25, 0.75)$ pair yields the same ranking with somewhat smaller statistics (13.8, 8.2, 11.7 and 3.7), and the $(0.10, 0.50)$ pair rejects equality for highway, transit and park at the 5 percent level but not for hospital. Four of the seven indicators therefore display significant heterogeneity between

the tails, and the heterogeneity is concentrated in the lower half of the conditional distribution: the adjacent-quantile statistics in Figure 2 are largest between $\tau = 0.25$ and $\tau = 0.50$ for transit (6.8) and between 0.25 and 0.75 for highway, while the upper pairs are small.

Table 7. Wald tests of inter-quantile equality for three quantile pairs — single-family houses

Amenity indicator	$\tau = 0.10$ vs. 0.90	$\tau = 0.25$ vs. 0.75	$\tau = 0.10$ vs. 0.50
Highway	17.12***	13.78***	6.42**
Hospital	12.11***	8.22***	3.63*
Park	6.66***	3.66*	4.66**
School	0.05	0.83	0.04
Transit	11.64***	11.73***	6.99***
Golf course	0.63	1.84	0.01
Garage	0.22	0.75	2.26

Notes: See Table 12. ***, ** and * denote rejection at the 1, 5 and 10 percent levels. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

The pseudo-R^2 rises from 0.355 at $\tau = 0.10$ to 0.444 at $\tau = 0.90$ (Table 8, Figure 5). The covariates thus account for a larger share of the check-function loss in the upper part of the conditional distribution. One reading is that low-priced listings carry more idiosyncratic information—condition, deferred maintenance, motivated sellers—that is not in the listing fields; the structural coefficients in Appendix Table 17 are consistent with this, the floor-area elasticity rising from 0.51 at $\tau = 0.10$ to 0.69 at $\tau = 0.90$ and the Montreal premium from 0.28 to 0.44. The same appendix table also shows that the log-distance coefficient is implausibly large (1.6 to 2.5 with bootstrap standard errors of 0.4 to 0.8) and positive once the regional indicators are included; the coefficients on this control are not interpretable as a distance gradient, and Appendix A flags the scaling of the distance variable in the original run as a point to be verified when the pipeline is re-executed.

Table 8. Pseudo-R^2 of the quantile regressions and adjusted R^2 of OLS, by segment

	Quantile regression, τ					OLS
Segment	0.10	0.25	0.50	0.75	0.90	adj. R^2
Single-family houses	0.355	0.364	0.400	0.436	0.444	0.619
Condominiums	0.377	0.414	0.449	0.468	0.456	0.669
Multiplexes	0.417	0.420	0.383	0.357	0.341	0.576
Vacant land	0.086	0.107	0.112	0.115	0.084	0.154

Notes: Pseudo-R^2 of Koenker and Machado (1999), equation (4); it is not comparable to the OLS R^2 and only loosely comparable across quantiles. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

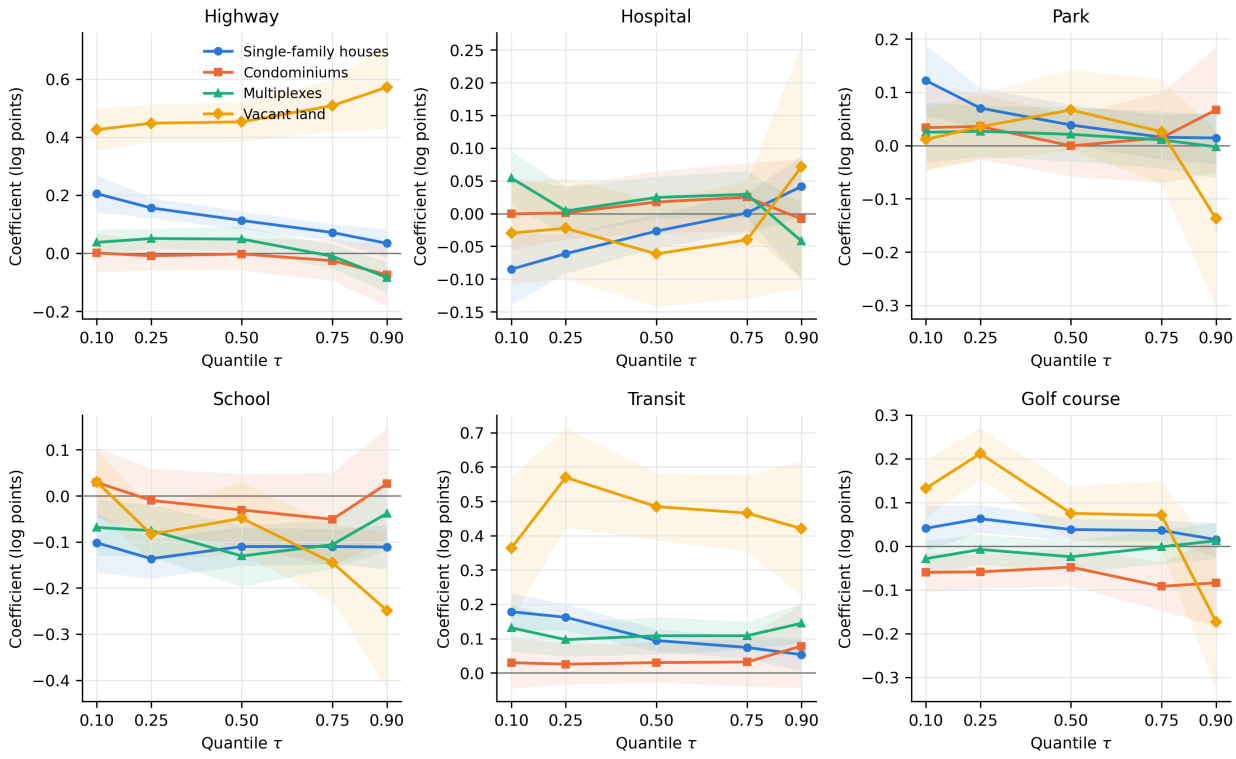


Figure 1. Amenity coefficients by quantile and segment

Notes: Quantile regression coefficients of equation (2) with 95 percent bootstrap confidence bands (±1.96 bootstrap standard errors, $B = 200$). The garage indicator, absent from the land specification, is omitted. Rebuilt from the preserved log by `scripts/06_tables_figures_from_log.py`.

5.3. Quantile regressions: condominiums, multiplexes and vacant land

Tables 9 to 11 report the corresponding estimates for the other segments, and Table 12 the tail-equality tests for all four. For condominiums ($N = 1{,}031$) no amenity coefficient is significant at any quantile except the golf-course indicator, which is negative (-0.048 to -0.092) and significant at four of the five quantiles, and the garage indicator, which is large and significant throughout (0.105 to 0.164). The tail-equality statistics are all below 1.4. The bootstrap standard errors are two to three times those of the house sample, so the absence of heterogeneity is as much a statement about power as about the market; it is also consistent with a product whose urban location makes most amenity mentions nearly universal (Table 3) and whose price is dominated by interior area, bathrooms and location, the OLS adjusted R^2 being the highest of the four segments.

Table 9. Amenity coefficients by quantile — condominiums

	Quantile regression, τ					OLS
Amenity indicator	0.10	0.25	0.50	0.75	0.90	
Highway	0.001	−0.009	−0.003	−0.026	−0.074	−0.023
	(0.034)	(0.025)	(0.029)	(0.035)	(0.055)	
Hospital	−0.001	0.001	0.018	0.025	−0.008	0.008
	(0.028)	(0.021)	(0.024)	(0.026)	(0.047)	
Park	0.034	0.036	−0.001	0.015	0.067	0.027
	(0.040)	(0.032)	(0.029)	(0.043)	(0.060)	
School	0.029	−0.010	−0.031	−0.051	0.026	−0.030
	(0.039)	(0.035)	(0.040)	(0.051)	(0.062)	
Transit	0.030	0.025	0.030	0.032	0.078	0.064**
	(0.038)	(0.030)	(0.029)	(0.037)	(0.062)	
Golf course	−0.060**	−0.059***	−0.048**	−0.092***	−0.084	−0.078***
	(0.023)	(0.020)	(0.022)	(0.028)	(0.051)	
Garage	0.105***	0.159***	0.164***	0.116***	0.122***	0.134***
	(0.038)	(0.030)	(0.027)	(0.026)	(0.039)	
Pseudo-R^2 / adj. R^2	0.377	0.414	0.449	0.468	0.456	0.669
N	1,031					

Notes: Dependent variable: log asking price. Each column is a separate estimate of equation (2) (quantile regression) or equation (1) (OLS) including all structural, locational and regional controls listed in Table 4. Bootstrap standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels (two-sided normal test using bootstrap standard errors, $B = 200$ paired resamples). OLS stars use HC3 standard errors and are reported as printed in the log (the log applies the thresholds 0.001, 0.01 and 0.05 to the OLS column, so OLS stars are conservative relative to the quantile columns). The pseudo-R^2 is that of Koenker and Machado (1999). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

For multiplexes ($N = 3{,}755$) two indicators display significant tail heterogeneity. The highway coefficient is positive at the first quartile and the median (0.051 and 0.049) and negative at the 90th percentile (−0.084), with $W = 11.6$; the hospital coefficient is 0.054 at $\tau = 0.10$ and −0.042 at $\tau = 0.90$, with $W = 7.1$. Transit proximity is the strongest and most stable amenity in this segment, at 0.10 to 0.14 at every quantile, and the school discount is again negative everywhere but

concentrated in the middle of the distribution (-0.131 at the median against -0.038 at the top). The garage indicator declines from 0.104 to 0.004 between $\tau = 0.10$ and $\tau = 0.75$ ($W = 4.4$ for the tails, 10.4 for the interquartile pair), which is natural for a product bought partly for rental income, where a garage adds less to the revenue of an expensive building than to the value of an owner-occupied one. The pseudo-R^2 declines with the quantile for this segment (0.417 to 0.341), the opposite of the house pattern.

Table 10. Amenity coefficients by quantile — multiplexes

	Quantile regression, τ					OLS
Amenity indicator	0.10	0.25	0.50	0.75	0.90	
Highway	0.037	0.051***	0.049**	−0.010	−0.084***	0.040*
	(0.023)	(0.017)	(0.022)	(0.021)	(0.027)	
Hospital	0.054***	0.004	0.025	0.029	−0.042	0.004
	(0.021)	(0.019)	(0.016)	(0.018)	(0.030)	
Park	0.025	0.027	0.021	0.011	−0.002	0.012
	(0.029)	(0.024)	(0.026)	(0.027)	(0.030)	
School	−0.069**	−0.075***	−0.131***	−0.106***	−0.038	−0.094***
	(0.031)	(0.028)	(0.034)	(0.027)	(0.041)	
Transit	0.131***	0.097***	0.108***	0.108***	0.144***	0.129***
	(0.035)	(0.025)	(0.028)	(0.021)	(0.030)	
Golf course	−0.029	−0.007	−0.024	−0.001	0.012	0.012
	(0.021)	(0.017)	(0.019)	(0.020)	(0.021)	
Garage	0.104***	0.086***	0.039***	0.004	0.037	0.062***
	(0.020)	(0.018)	(0.015)	(0.018)	(0.025)	
Pseudo-R^2 / adj. R^2	0.417	0.420	0.383	0.357	0.341	0.576
N	3,755					

Notes: Dependent variable: log asking price. Each column is a separate estimate of equation (2) (quantile regression) or equation (1) (OLS) including all structural, locational and regional controls listed in Table 4. Bootstrap standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels (two-sided normal test using bootstrap standard errors, $B = 200$ paired resamples). OLS stars use HC3 standard errors and are reported as printed in the log (the log applies the thresholds 0.001, 0.01 and 0.05 to the OLS column, so OLS stars are conservative relative to the quantile columns). The pseudo-R^2 is that of Koenker and Machado (1999). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

For vacant land ($N = 9{,}239$) the highway and transit indicators dominate, with coefficients of 0.43 to 0.57 and 0.36 to 0.57 respectively, that is, differentials of 50 to 80 percent in the asking price of a parcel; both are stable across quantiles. The tail tests reject equality for school ($W = 9.0$) and golf ($W = 13.1$): the school indicator is close to zero at the 10th percentile and reaches −0.249 at the 90th, and the golf indicator is positive at the lower quantiles (0.132 and 0.212) and negative at

the top (−0.173). The pseudo-R^2 of the land regressions never exceeds 0.12.

Table 11. Amenity coefficients by quantile — vacant land

	Quantile regression, τ					OLS
Amenity indicator	0.10	0.25	0.50	0.75	0.90	
Highway	0.426***	0.448***	0.453***	0.509***	0.572***	0.475***
	(0.037)	(0.033)	(0.035)	(0.045)	(0.073)	
Hospital	−0.030	−0.022	−0.062	−0.040	0.071	−0.036
	(0.039)	(0.039)	(0.041)	(0.046)	(0.094)	
Park	0.011	0.035	0.067*	0.026	−0.137	−0.005
	(0.030)	(0.032)	(0.039)	(0.050)	(0.084)	
School	0.030	−0.083***	−0.049	−0.145***	−0.249***	−0.113***
	(0.035)	(0.028)	(0.040)	(0.043)	(0.086)	
Transit	0.364***	0.570***	0.485***	0.466***	0.421***	0.426***
	(0.106)	(0.073)	(0.050)	(0.057)	(0.100)	
Golf course	0.132***	0.212***	0.075**	0.071*	−0.173**	0.086**
	(0.033)	(0.030)	(0.032)	(0.040)	(0.078)	
Pseudo-R^2 / adj. R^2	0.086	0.107	0.112	0.115	0.084	0.154
N	9,239					

Notes: Dependent variable: log asking price. Each column is a separate estimate of equation (2) (quantile regression) or equation (1) (OLS) including all structural, locational and regional controls listed in Table 4. Bootstrap standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels (two-sided normal test using bootstrap standard errors, $B = 200$ paired resamples). OLS stars use HC3 standard errors and are reported as printed in the log (the log applies the thresholds 0.001, 0.01 and 0.05 to the OLS column, so OLS stars are conservative relative to the quantile columns). The pseudo-R^2 is that of Koenker and Machado (1999). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

Table 12. Wald tests of equal amenity coefficients at $\tau = 0.10$ and $\tau = 0.90$, by segment

Amenity indicator	Houses	Condominiums	Multiplexes	Vacant land
Highway	17.12***	1.34	11.58***	3.18*
Hospital	12.11***	0.02	7.09***	0.98
Park	6.66***	0.22	0.43	2.79*
School	0.05	0.00	0.35	9.00***
Transit	11.64***	0.44	0.08	0.15
Golf course	0.63	0.18	1.88	13.11***
Garage	0.22	0.10	4.44**	—
Rejections at 5%	4/7	0/7	3/7	2/6

Notes: Entries are the statistic of equation (11), $W = [\hat{\beta}_k(0.10) - \hat{\beta}_k(0.90)]^2 / [\widehat{\mathrm{Var}}(\hat{\beta}_k(0.10)) + \widehat{\mathrm{Var}}(\hat{\beta}_k(0.90))]$, with bootstrap variances, referred to a $\chi^2(1)$ distribution. The denominator omits the (typically positive) covariance between the two quantile estimators, which makes the test conservative. ***, ** and * denote rejection at the 1, 5 and 10 percent levels. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

Across the four segments, significant tail heterogeneity is found for four of six amenities for houses, none for condominiums, two for multiplexes and two for land. The pattern that is common to the three built segments is a transit differential that is positive at every quantile, and a school differential that is negative at most quantiles; the pattern that is specific to houses is the monotone decline of the highway, transit and park differentials with the quantile and the sign change of the hospital differential.

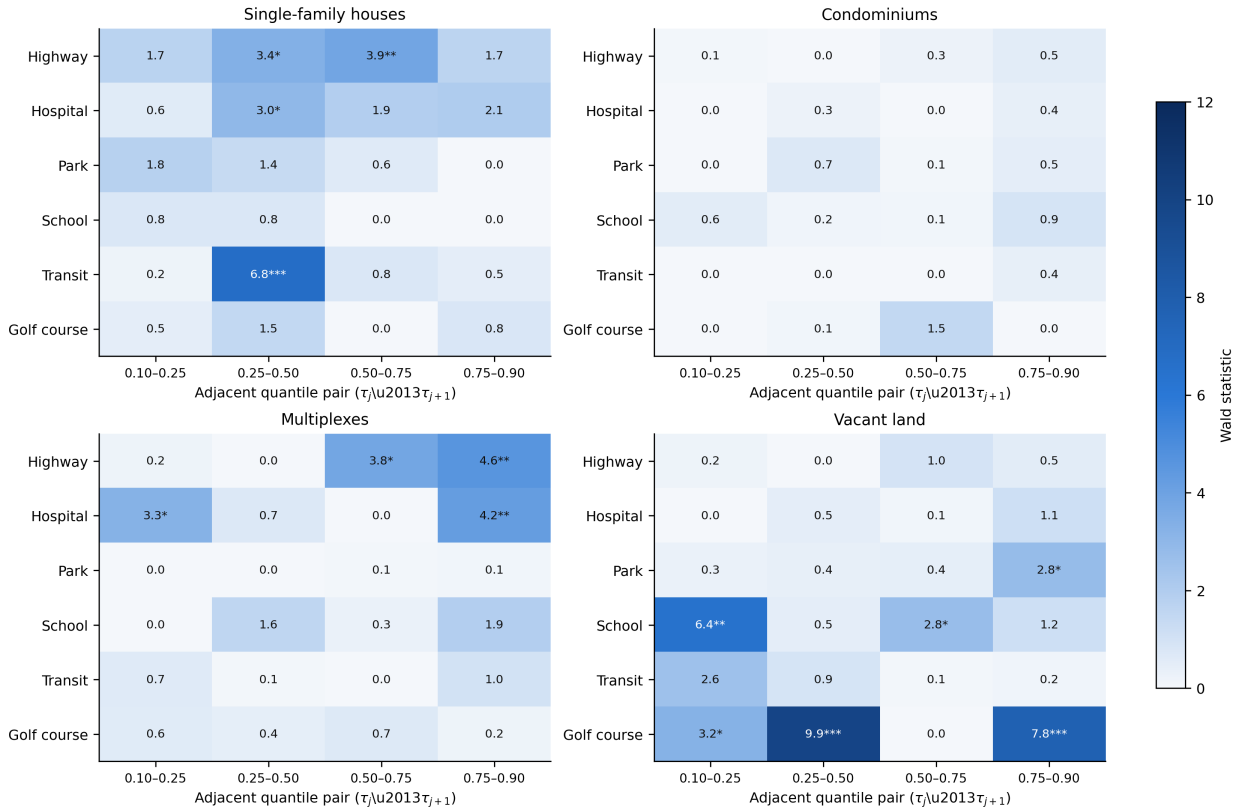


Figure 2. Wald statistics for equality of amenity coefficients at adjacent quantiles

Notes: Each cell is the statistic of equation (11) for the pair of adjacent quantiles in the column, computed from the coefficients and bootstrap standard errors recorded in the log. ***, **, *: rejection at the 1, 5 and 10 percent levels. Rebuilt from the preserved log.

5.4. Two-stage spatial quantile regression

Table 13 reports, for each segment, the fit of the first stage (7), the spatial-lag coefficient of the median second stage (8), and the median amenity coefficients with and without the lag; Figure 3 displays the three sets of coefficients side by side. The first-stage R^2 is 0.672 for houses, 0.719 for condominiums, 0.650 for multiplexes and 0.204 for land, so that the lagged covariates explain a large share of the neighbouring price level in the built segments and a small share for parcels; the overall first-stage F implied by these values is in the hundreds, but the partial F on the excluded instruments, which is the relevant weak-instrument statistic, was not recorded (Section 6.3). The median spatial-lag coefficient is 0.291 for houses, 0.203 for condominiums, 0.131 for multiplexes and 0.322 for land. The ordering is sensible: the undifferentiated product, land, is the one whose asking price co-moves most with neighbouring asking prices, and the rental building, priced partly on income, the one that co-moves least. Without standard errors for $\hat{\rho}$—the second stage was not bootstrapped—we do not claim significance, and we note that the lag coefficients at the other four quantiles were not recorded, so that nothing can be said here about the variation of ρ across the

conditional distribution.

Table 13. Two-stage spatial quantile regression at the median: first-stage fit, spatial lag and amenity coefficients

	OLS	QR, $\tau = 0.50$	2SQR, $\tau = 0.50$
Single-family houses ($N = 6,250$)			
Highway	0.133	0.113	0.057
Hospital	−0.049	−0.027	−0.006
Park	0.042	0.038	0.042
School	−0.122	−0.110	−0.072
Transit	0.113	0.094	0.071
Golf course	0.052	0.038	0.023
Garage	0.100	0.096	0.091
First-stage R^2	—	—	0.672
Implied overall first-stage $F(34,\ 6,215)$	—	—	374
Spatial lag $\hat{\rho}(0.50)$	—	—	0.291
Condominiums ($N = 1,031$)			
Highway	−0.023	−0.003	−0.002
Hospital	0.008	0.018	0.024
Park	0.027	−0.001	0.005
School	−0.030	−0.031	−0.017
Transit	0.064	0.030	0.024
Golf course	−0.078	−0.048	−0.041
Garage	0.134	0.164	0.130
First-stage R^2	—	—	0.719
Implied overall first-stage $F(34,\ 996)$	—	—	75
Spatial lag $\hat{\rho}(0.50)$	—	—	0.203
Multiplexes ($N = 3,755$)			
Highway	0.040	0.049	0.043
Hospital	0.004	0.025	0.024
Park	0.012	0.021	0.029
School	−0.094	−0.131	−0.129
Transit	0.129	0.108	0.098
Golf course	0.012	−0.024	−0.023
Garage	0.062	0.039	0.030
First-stage R^2	—	—	0.650
Implied overall first-stage $F(34,\ 3,720)$	—	—	203
Spatial lag $\hat{\rho}(0.50)$	—	—	0.131
Vacant land ($N = 9,239$)			
Highway	0.475	0.453	0.324
Hospital	−0.036	−0.062	−0.011
Park	−0.005	0.067	0.066
School	−0.113	−0.049	−0.018

Introducing the lag attenuates the median amenity coefficients for houses: highway falls from 0.113 to 0.057, transit from 0.094 to 0.071, school from −0.110 to −0.072 and hospital from −0.027 to −0.006, while park (0.038 to 0.042) and garage (0.096 to 0.091) are essentially unchanged. The attenuation is largest for the indicators that are themselves spatially clustered—highways, transit lines and schools serve whole neighbourhoods, whereas a garage is a property attribute—which is what a spatial lag that absorbs neighbourhood price levels should produce (Kim et al., 2003; Small and Steimetz, 2012). The same qualitative pattern appears for multiplexes (garage 0.039 to 0.030, school unchanged) and for land, where the highway coefficient falls from 0.453 to 0.324 and transit from 0.485 to 0.349. Table 14 converts the median 2SQR coefficients into the scalar total effects of equation (9): with $\hat{\rho} = 0.291$ the multiplier for houses is 1.41, so that the total highway differential is 0.081 log points, or about $48,800 at the median asking price, the transit differential 0.100 ($60,900) and the school differential −0.101 (−$55,600). Because the indirect share equals $\hat{\rho}$ by construction, these figures restate the second-stage coefficients at a different scale rather than add to them.

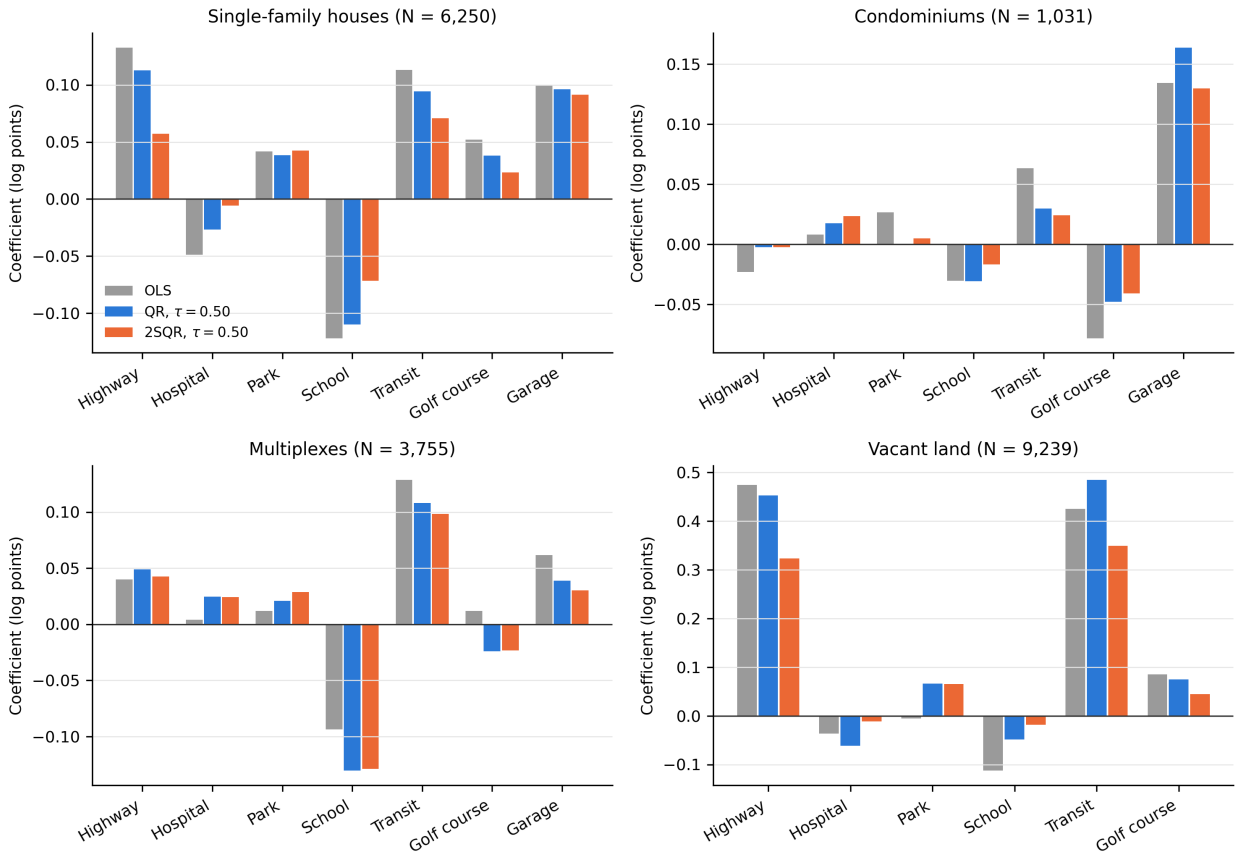


Figure 3. Amenity coefficients at the median: OLS, quantile regression and 2SQR

Notes: OLS coefficients of equation (1), median quantile regression coefficients of equation (2), and median second-stage coefficients of the 2SQR, equation (8). No standard errors are available for the 2SQR coefficients. Rebuilt from the preserved log.

Table 14. Scalar approximation of direct, indirect and total amenity effects in the median 2SQR

Amenity indicator	Direct $\approx \hat{\beta}_k$	Indirect $\approx \hat{\rho}\hat{\beta}_k/(1-\hat{\rho})$	Total $\approx \hat{\beta}_k/(1-\hat{\rho})$	Total, $ at median
Single-family houses ($\hat{\rho}(0.50) = 0.291$, $1/(1-\hat{\rho}) = 1.410$, median price $579,000)				
Highway	0.057	0.024	0.081	+48,767
Hospital	−0.006	−0.002	−0.008	−4,635
Park	0.042	0.017	0.060	+35,626
School	−0.072	−0.029	−0.101	−55,569
Transit	0.071	0.029	0.100	+60,914
Golf course	0.023	0.010	0.033	+19,356
Garage	0.091	0.037	0.129	+79,543
Condominiums ($\hat{\rho}(0.50) = 0.203$, $1/(1-\hat{\rho}) = 1.254$, median price $425,000)				
Highway	−0.002	−0.001	−0.003	−1,251
Hospital	0.024	0.006	0.029	+12,726
Park	0.005	0.001	0.006	+2,620
School	−0.017	−0.004	−0.021	−8,898
Transit	0.024	0.006	0.030	+13,121
Golf course	−0.041	−0.010	−0.051	−21,205
Garage	0.130	0.033	0.163	+75,116
Multiplexes ($\hat{\rho}(0.50) = 0.131$, $1/(1-\hat{\rho}) = 1.151$, median price $799,000)				
Highway	0.043	0.006	0.049	+40,228
Hospital	0.024	0.004	0.028	+22,517
Park	0.029	0.004	0.033	+26,804
School	−0.129	−0.019	−0.148	−110,219
Transit	0.098	0.015	0.113	+95,646
Golf course	−0.023	−0.004	−0.027	−21,111
Garage	0.030	0.004	0.035	+28,140
Vacant land ($\hat{\rho}(0.50) = 0.322$, $1/(1-\hat{\rho}) = 1.475$, median price $150,000)				
Highway	0.324	0.154	0.477	+91,709
Hospital	−0.011	−0.005	−0.016	−2,441
Park	0.066	0.031	0.097	+15,322
School	−0.018	−0.009	−0.027	−4,003
Transit	0.349	0.166	0.515	+101,101
Golf course	0.045	0.021	0.067	+10,331

Notes: Scalar approximations of equation (9) evaluated at $\tau = 0.50$; the dollar column is $\tilde{P}_{0.50}$ [exp(Total) − 1] with $\tilde{P}_{0.50}$ the median asking price of the segment. The log labels $\hat{\beta}_k/(1-\hat{\rho})$ "direct"; it is the total effect of the scalar approximation, and the columns are relabelled accordingly (direct + indirect = total). No standard errors are available (Section 6.3). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

5.5. Residual autocorrelation after the spatial lag

Table 15 reports Moran's I on the residuals of the quantile regressions and of the 2SQR second stages. For houses the QR residuals have I between 0.515 and 0.524, slightly above the OLS value, and the second-stage 2SQR residuals between 0.518 and 0.545, slightly above the QR values; in the other three segments the two sets of statistics are of the same magnitude (0.53–0.61) and the 2SQR values are not systematically lower. The spatial lag therefore does not reduce measured residual autocorrelation. This is the expected outcome of the diagnostic as it was computed, not a failure of the model: the residuals of the second stage are formed with the fitted lag, which is a linear combination of the covariates and their spatial lags, so that the component of $\ln P_i$ that co-moves with the *observed* neighbouring prices is still in the residual. Residual autocorrelation after a spatial lag must be assessed on the structural residuals, which use $\hat{\rho}(\tau)[\mathbf{W}\ln\mathbf{P}]_i$ rather than its projection; those were not computed in the original run. Until they are, the evidence on whether a first-order spatial lag is sufficient, or whether a spatial error component or finer spatial fixed effects are also needed, is simply unavailable, and we refrain from the statement that the 2SQR "reduces but does not eliminate" spatial dependence.

Table 15. Moran's I on residuals: OLS, quantile regression and 2SQR second stage, by quantile and segment

	OLS	QR residuals, τ					2SQR second-stage residuals, τ				
Segment		0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90
Single-family houses	0.506	0.514	0.515	0.524	0.519	0.515	0.518	0.519	0.536	0.543	0.545
Condominiums	0.551	0.574	0.561	0.552	0.529	0.555	0.554	0.541	0.555	0.534	0.552
Multiplexes	0.528	0.559	0.533	0.516	0.524	0.516	0.562	0.536	0.522	0.539	0.542
Vacant land	0.580	0.597	0.585	0.585	0.585	0.589	0.607	0.599	0.595	0.594	0.588

Notes: All statistics are significant at any conventional level (z-statistics between 21 and 70). The 2SQR residuals are those of the second stage, $\ln P_i - \hat{\alpha}(\tau) - \mathbf{X}_i'\hat{\boldsymbol{\beta}}(\tau) - \hat{\rho}(\tau)[\widehat{\mathbf{W}\ln\mathbf{P}}]_i$, computed with the *fitted* spatial lag; because the fitted lag is a linear combination of $\mathbf{X}$ and $\mathbf{WX}$, these residuals cannot be expected to display less spatial autocorrelation than the QR residuals (Section 5.5). Structural residuals using the observed lag were not computed in the original run. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

5.6. Dollar-equivalent implicit prices

Table 16 scales the house coefficients by the unconditional price quantiles ($275,000, $400,000, $579,000, $894,000 and $1,485,000), and Figure 4 plots the corresponding profiles for all segments. Two mechanisms interact in these figures. A declining coefficient applied to a rising price base can yield a flat or hump-shaped dollar profile: the highway differential is $62,700 at the 10th

percentile, peaks at $69,200 at the median and is $51,700 at the 90th, and the park differential falls from $35,600 to $21,300. A stable coefficient applied to a rising base yields a steeply increasing dollar profile: the school discount grows from $26,600 to $156,200 and the garage differential from $17,300 to $72,000 between the tails. The hospital differential turns from −$22,400 to +$62,800. These are differentials in asking prices between otherwise similar listings with and without the broker's mention of the amenity; they are reported to give a sense of scale and should not be read as values of the amenity to a household.

Table 16. Dollar-equivalent implicit prices of amenity indicators by quantile — single-family houses (CAD)

Amenity indicator	$\tau = 0.10$ ($275K)	$\tau = 0.25$ ($400K)	$\tau = 0.50$ ($579K)	$\tau = 0.75$ ($894K)	$\tau = 0.90$ ($1,485K)
Highway	+62,663***	+67,602***	+69,207***	+66,200***	+51,655
Hospital	−22,403***	−23,779***	−15,372**	+869	+62,776*
Park	+35,648***	+29,106***	+22,701**	+13,965	+21,292
School	−26,608***	−50,987***	−60,312***	−93,116***	−156,222***
Transit	+53,567***	+70,340***	+57,284***	+68,853***	+80,970**
Golf course	+11,521	+25,979***	+22,469***	+32,767***	+22,679
Garage	+17,329***	+42,221***	+58,542***	+76,791***	+71,954**

Notes: Entries are $\pi_k(\tau) = P(\tau)\,[\exp(\hat{\beta}_k(\tau)) - 1]$ (equation (13)), where $P(\tau)$ is the unconditional sample quantile of the asking price (in parentheses in the header, rounded to the nearest thousand). These are capitalisation differentials in asking prices, not welfare measures (Section 4.10). Stars carry over the significance of the underlying coefficient in Table 6. All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.

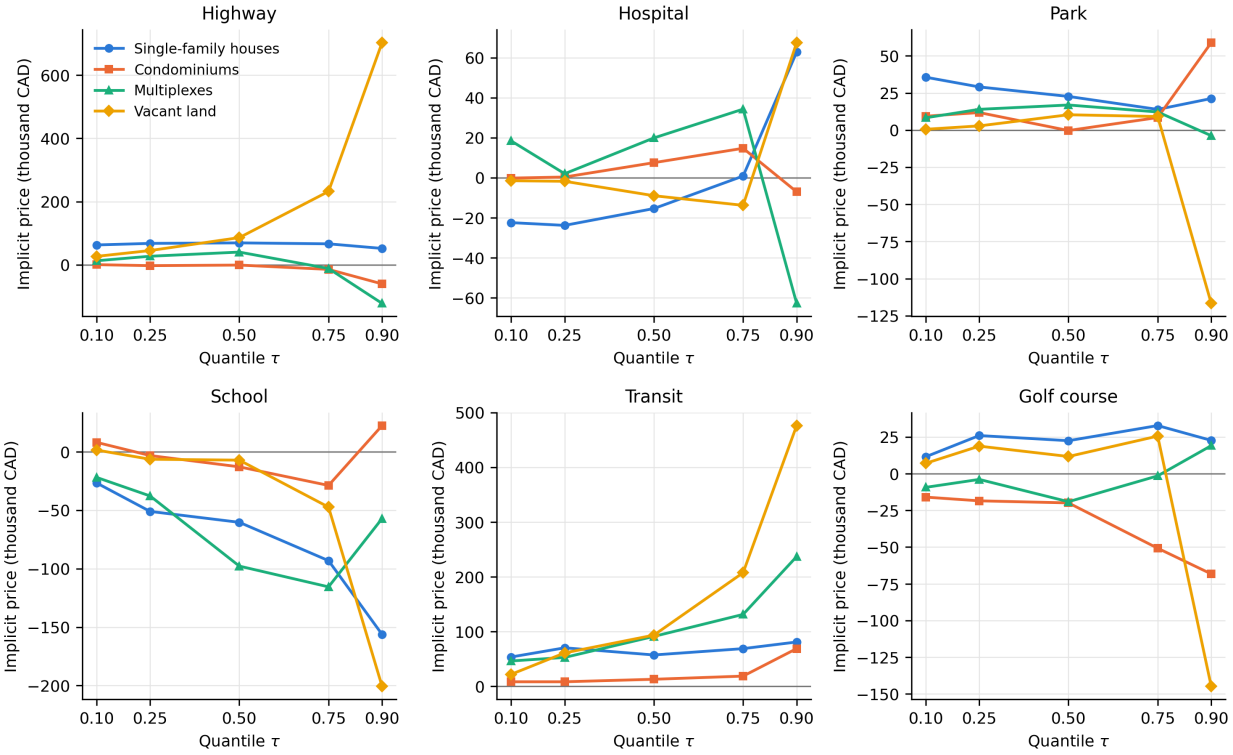


Figure 4. Dollar-equivalent implicit prices by quantile and segment

Notes: $\pi_k(\tau)$ of equation (13), in thousands of Canadian dollars, using the unconditional asking-price quantile of each segment. Rebuilt from the preserved log; the figure replaces the figure of the May 2026 version, whose values could not be traced to the log.

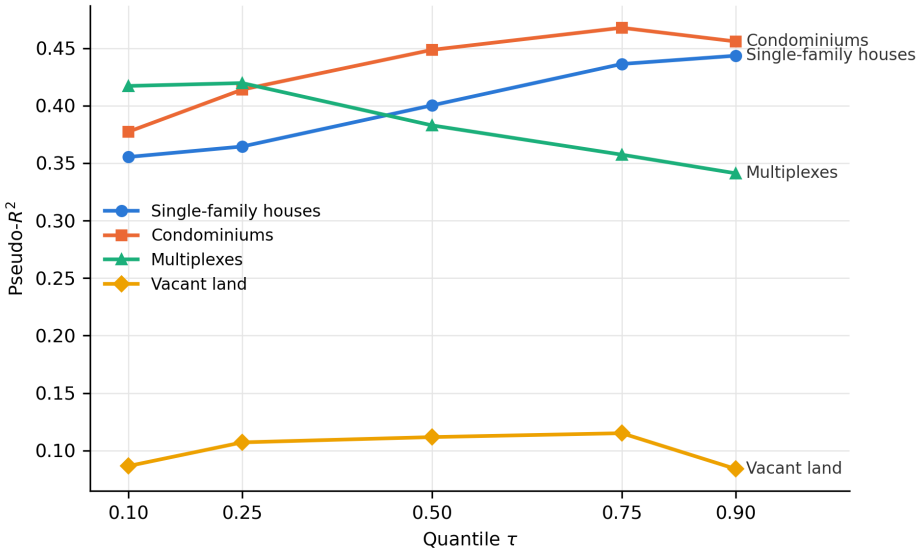


Figure 5. Pseudo-R^2 of the quantile regressions by segment

Notes: $R^1(\tau)$ of equation (4). Rebuilt from the preserved log.

6. Robustness and the limits of the available evidence

A revision of this paper carried out in October 2026 without access to the Centris snapshot (Appendix A) made it necessary to distinguish sharply between robustness evidence that exists in the numerical record and robustness checks that were described in the May 2026 version but for which no output survives. This section reports the former, states the latter as not executed, and ranks the checks that a re-run of the pipeline should perform first.

6.1. Evidence in the record

Three internal checks are available. First, the heterogeneity results for houses do not depend on a single pair of quantiles: the ranking of the amenities by the equality statistic is the same for the $(0.10, 0.90)$, $(0.25, 0.75)$ and $(0.10, 0.50)$ pairs (Table 7), and the adjacent-pair statistics of Figure 2 locate the changes in the lower half of the distribution rather than at a single extreme quantile, which is where estimates would be most sensitive to trimming. Second, the direction of the quantile profiles is consistent with the direction of the OLS-to-2SQR attenuation: the indicators whose coefficients decline most across quantiles (highway, transit) are also those most attenuated by the spatial lag at the median (Table 13), which is what one expects if both phenomena reflect the clustering of lower-priced, amenity-dense listings. Third, the cross-segment comparison is itself a specification check: the transit and school differentials have the same sign in the three built segments, and the segment in which precision is lowest (condominiums) is the one in which no heterogeneity is detected, whereas the two largest samples (houses, land) produce the sharpest rejections. None of these observations substitutes for the formal checks below.

6.2. Checks described in the previous version but not supported by the record

The May 2026 version reported a sensitivity analysis of $\hat{\rho}$ and of the amenity coefficients to the number of neighbours $k \in \{5, 10, 15, 20\}$, a comparison of bootstrap standard errors at $B = 200$ and $B = 500$, re-estimation under 1/99 and 2/98 percent trimming, a split of the house sample into Montreal and non-Montreal listings, variance inflation factors and a condition number for the house design matrix (Belsley et al., 1980; O'Brien, 2007), a partial first-stage F above 100 in every segment, and a statement that the scalar effect decomposition was within 3 percent of the trace-based LeSage–Pace effects on a subsample. None of these outputs is in the preserved log, the figures reported for them could not be reconciled with the log, and the current version of the estimation script computes only the variance inflation factors, the condition number and the overall first-stage F. We therefore withdraw these claims. They are listed in Appendix A as analyses to be executed, and the paper's conclusions are stated so as not to depend on them.

6.3. What a re-run should establish first

The checks below are ordered by their bearing on the conclusions. The first concerns inference. The paired bootstrap ignores spatial dependence that is, by the paper's own diagnostics, strong; a cluster bootstrap by municipality or by spatial block (Cameron et al., 2008; Parente and Santos Silva, 2016; Hagemann, 2017) would widen the confidence bands, and the equality tests of Table 12 should be recomputed with the bootstrap covariance between quantiles, which can only make them less conservative, and with the full stacked covariance matrix to obtain a valid joint test (Koenker and Xiao, 2002). The second concerns the 2SQR itself: the second stage should be bootstrapped jointly with the first to obtain standard errors for $\hat{\rho}(\tau)$ and for the spatially adjusted amenity coefficients, the partial F on $\mathbf{WZ}$ should be reported, the structural residuals should be used for Moran's I, and the quantile-based first stage of Kim and Muller (2004) or the inverse quantile regression of Chernozhukov and Hansen (2006) and Su and Yang (2011) should be compared with the least-squares first stage. The third concerns the spatial weights: k, projected rather than geographic coordinates, a positive distance floor for co-located listings and a distance-band alternative. The fourth concerns the specification: the scaling of the distance variable (Section 5), the replacement of a single province-wide distance gradient by municipality fixed effects (for which the location-scale quantile estimator of Machado and Santos Silva (2019) handles many fixed effects at low cost), and the sensitivity of the tails to the 0.5 percent trimming. The fifth concerns the data: the broker's amenity classification should be validated against geocoded distances to actual facilities for a subsample, and the asking-price results replicated on transaction prices from the Quebec land register (*Registre foncier*). Until at least the first two groups of checks have been run, the heterogeneity results should be regarded as suggestive rather than established, and the spatial-lag estimates as descriptive.

7. Discussion

7.1. Heterogeneous capitalisation as a property of the price schedule

The central finding is that, for single-family houses, the implicit prices of the accessibility-type amenities—highway, transit, park—are large at the lower conditional quantiles and small or nil at the upper ones, while the discount attached to school proximity is flat and the garage and golf premia nearly so. The conditional-mean estimate, which is what a standard hedonic regression would report, lies near the median profile and is therefore a poor description of either tail: it overstates the highway differential at the 90th percentile by a factor of four and understates it at the 10th by about a third. This is the pattern documented for structural attributes by Zietz et al. (2008) and Mak et al. (2010) and for location by Liao and Wang (2012), transposed to broker-reported amenities in a province-wide listing sample. Within the hedonic framework it says that the equilibrium price

schedule is not additively separable in the amenity indicators and the rest of the bundle: the slope of the schedule in the amenity direction depends on where, in the conditional price distribution, the property sits.

Several mechanisms are consistent with the pattern; the design does not separate them. The most natural is sorting on transit dependence and car ownership: households that buy at the lower end of the schedule are more likely to depend on transit and highway access for commuting (Glaeser et al., 2008), so that the bid functions tangent to the schedule in that region are steeper in those attributes, whereas households at the upper end substitute private green space and private mobility for public parks and transit. A second is supply-side: in Quebec, the housing near highways, transit lines and urban parks is older and denser, so that the lower part of the conditional distribution contains most of the amenity-rich stock and the premium reflects, in part, the scarcity of such stock at that price level. A third is measurement: a broker marketing a lower-priced house may be more inclined to list accessibility amenities, and one marketing an expensive house to emphasise other qualities, in which case part of the heterogeneity is in the reporting rather than in the capitalisation. The attenuation of the highway and transit coefficients when the spatial lag is introduced points in the same direction as the second mechanism, since it shows that part of the differential is the price level of the surrounding, similarly situated listings.

7.2. The hospital indicator

The sign change of the hospital coefficient—a discount of 8.5 log points at the 10th percentile and a premium of about 4 log points, marginally significant, at the 90th—is the most distinctive result and the one most in need of corroboration. A plausible reading is that "hospital nearby" describes two different neighbourhood types at the two ends of the schedule: institutional corridors with traffic, parking and emergency-vehicle noise where lower-priced houses adjoin a regional hospital, and the established, well-served residential districts in which the large university hospitals of Montreal and Québec City are embedded. The results of Gu et al. (2024), who find that hospital accessibility capitalises positively in the upper price quantiles and weakly or negatively in the lower ones, are consistent with this reading. The 2SQR estimate at the median (-0.006 against -0.027 without the lag) suggests that the median discount is largely a neighbourhood price-level effect, but without the second-stage estimates at the other quantiles we cannot say whether the sign change survives the spatial adjustment. Distances to actual hospitals, by type, would be required to go further.

7.3. Cross-segment patterns

The four product types differ in the strength of the spatial lag, in the fit of the specification and in the presence of heterogeneity, and the differences are coherent with the nature of the products. Vacant

land, the least differentiated product, has the strongest co-movement with neighbouring asking prices ($\hat{\rho} = 0.32$), the weakest structural model and the largest accessibility differentials: a highway or transit mention is associated with asking prices higher by half or more, at every quantile, which is what one expects when the indicator proxies for serviced, developable land near settlements rather than for an amenity in the residential sense (Colwell and Munneke, 1997). The sign change of the golf indicator for land, from a premium at the lower quantiles to a discount at the top, is plausibly a composition effect—recreational lots near golf courses in the lower part of the schedule, large rural or agricultural parcels at the top—rather than a preference reversal. Multiplexes, priced partly on rental income, show the weakest spatial co-movement ($\hat{\rho} = 0.13$), a transit premium that is stable across the schedule and a garage premium that disappears above the median; both are consistent with an investor's valuation, in which tenant accessibility matters at every price level and a garage adds little to the rent roll of an expensive building. Condominiums show no detectable heterogeneity, but the sample is small, selected on the reporting of interior area, and nearly homogeneous in amenity mentions, so the segment is uninformative on the question rather than evidence of homogeneity.

7.4. Implications

Two implications follow if the pattern is confirmed on transaction data. For mass appraisal and automated valuation, a conditional-mean adjustment for accessibility amenities will systematically over-adjust high-value houses and under-adjust low-value ones; the quantile profiles indicate the direction and rough size of the error, and quantile or segmented models offer a remedy at little cost. For the evaluation of amenity investments, the distribution of capitalisation across the price schedule is itself a policy-relevant object: if transit access capitalises mainly into lower-priced housing, the incidence of a transit investment falls on the owners of that housing and on the renters who may be displaced by the resulting appreciation (Renne et al., 2016), whereas a park investment that capitalises in the lower part of the schedule but not at the top redistributes differently from one with a uniform premium (Wolch et al., 2014). These are statements about the incidence of capitalisation, not about willingness to pay (Kuminoff and Pope, 2014), and they are conditional on the limitations set out below.

7.5. Limitations

The limitations are of four kinds. The data are asking prices from a single snapshot, with binary broker-reported amenity indicators that carry no distance or quality information and that may be reported differently across market segments; the estimates are cross-sectional associations subject to sorting and to the usual omitted-neighbourhood-quality concern, which a single distance gradient and three regional indicators cannot fully address. The 2SQR estimator uses a least-squares first stage

and the fitted lag in the second stage, so that its consistency rests on an independence assumption on the errors, no standard errors exist for the spatial-lag coefficient or for the spatially adjusted amenity coefficients, the second-stage coefficients are available at the median only, and the residual diagnostic was computed on second-stage rather than structural residuals. The inference for the quantile regressions rests on a paired bootstrap that ignores spatial dependence and on equality tests that ignore the covariance between quantiles. And the numerical record consists of a console log produced by an earlier version of the estimation script, which could not be re-run for this revision; the robustness analyses that would ordinarily accompany the results are, for that reason, listed rather than reported. The paper should be read as a documented first estimate whose main qualitative result—that amenity capitalisation in Quebec house prices is not constant across the conditional price distribution—is clear in the record, and whose quantitative details await the re-estimation described in Section 6.3.

8. Conclusion

This paper asked whether the capitalisation of neighbourhood amenities into Quebec residential asking prices is constant across the conditional price distribution and whether it survives an adjustment for spatial dependence. Using 20,275 Centris listings in four product segments, quantile regressions at five quantiles and a two-stage spatial quantile regression with an instrumented spatial lag, it finds that for single-family houses the implicit prices of highway, transit and park proximity decline steadily from the 10th to the 90th conditional percentile and that the hospital indicator changes sign, the equality of the tail coefficients being rejected for all four; that school proximity carries a stable discount of about 11 log points; that condominiums show no detectable heterogeneity while multiplexes and vacant land show it for two amenities each; that residual spatial autocorrelation is strong in every segment; and that the median spatial-lag coefficient ranges from 0.13 for multiplexes to 0.32 for land, its introduction attenuating the median highway, transit and school differentials for houses by a third to a half.

The methodological lesson is that a conditional-mean hedonic coefficient for an accessibility amenity is not a neutral summary: in these data it misstates the differential at both ends of the price schedule and in opposite directions. The substantive lesson is more guarded. The profiles are consistent with sorting on transit dependence and with the concentration of amenity-rich older stock in the lower part of the schedule, but the design cannot separate these from differences in broker reporting, and the estimates are gradients of a listing-price schedule rather than welfare measures.

The agenda that follows is largely one of verification and extension on the same data. Re-running the pipeline would allow a spatially valid bootstrap, standard errors for the spatial-lag coefficients, structural residual diagnostics, the comparison of the least-squares first stage with the quantile-

based and inverse-quantile alternatives, sensitivity to the spatial weights and to trimming, and the restoration of the descriptive statistics and robustness tables that the preserved record does not contain. Beyond that, geocoded distances to actual facilities would replace the broker's binary classification, transaction prices from the land register would replace asking prices, a geographically weighted quantile regression or a spatial Durbin quantile specification would show whether the heterogeneity documented along the price distribution also varies across the territory, and quantile regression forests or gradient boosting (Meinshausen, 2006; Friedman, 2001) would provide flexible benchmarks for the predictive content of the amenity indicators at each quantile. Each of these steps is feasible with the existing code and data, and the audit trail established in this revision is intended to make them straightforward.

References

Amemiya, Takeshi, "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 1982, *50* (3), 689–711.

Anderson, Soren T. and Sarah E. West, "Open Space, Residential Property Values, and Spatial Context," *Regional Science and Urban Economics*, 2006, *36* (6), 773–789.

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*, Dordrecht: Kluwer Academic Publishers, 1988.

_ , "Thirty Years of Spatial Econometrics," *Papers in Regional Science*, 2010, *89* (1), 3–25.

_ **and Nancy Lozano-Gracia**, "Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality," *Empirical Economics*, 2008, *34* (1), 5–34.

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.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

Belsley, David A., Edwin Kuh, and Roy E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: John Wiley & Sons, 1980.

Bentley, Jon Louis, "Multidimensional Binary Search Trees Used for Associative Searching," *Communications of the ACM*, 1975, *18* (9), 509–517.

Bishop, Kelly C., 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.

Bitter, Christopher, Gordon F. Mulligan, and Sandy Dall'erba, "Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method," *Journal of Geographical Systems*, 2007, *9* (1), 7–27.

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.

_ , _ , and _ , "Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods," *Journal of Real Estate Research*, 2010, *32* (2), 139–159.

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

Buchinsky, Moshe, "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 1995, *68* (2), 303–338.

_ , "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources*, 1998, *33* (1), 88–126.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

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

Cassel, Eric and Robert Mendelsohn, "The Choice of Functional Forms for Hedonic Price Equations: Comment," *Journal of Urban Economics*, 1985, *18* (2), 135–142.

Chen, Lin-An and Stephen Portnoy, "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," *Communications in Statistics – Theory and Methods*, 1996, *25* (5), 1005–1032.

Chen, Vivian Yi-Ju, Wen-Shuenn Deng, Tse-Chuan Yang, and Stephen A. Matthews, "Geographically Weighted Quantile Regression (GWQR): An Application to U.S. Mortality Data," *Geographical Analysis*, 2012, *44* (2), 134–150.

Chen, Yulin, Yu Ye, Xiaoping Liu, and Chun Yin, "Examining the Nonlinear and Spatial Heterogeneity of Housing Prices in Urban Beijing: An Application of GeoShapley," *Habitat International*, 2025, *162*, 103439.

Chernozhukov, Victor and Christian Hansen, "An IV Model of Quantile Treatment Effects," *Econometrica*, 2005, *73* (1), 245–261.

_ **and** _ , "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 2006, *132* (2), 491–525.

_ **and** _ , "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 2008, *142* (1), 379–398.

Cliff, Andrew D. and J. Keith Ord, *Spatial Processes: Models and Applications*, London: Pion, 1981.

Colwell, Peter F. and Henry J. Munneke, "The Structure of Urban Land Prices," *Journal of Urban Economics*, 1997, *41* (3), 321–336.

Cribari-Neto, Francisco, "Asymptotic Inference under Heteroskedasticity of Unknown Form," *Computational Statistics & Data Analysis*, 2004, *45* (2), 215–233.

Crompton, John L., "The Impact of Parks on Property Values: A Review of the Empirical Evidence," *Journal of Leisure Research*, 2001, *33* (1), 1–31.

Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell, "On the Choice of Functional Form for Hedonic Price Functions," *Review of Economics and Statistics*, 1988, *70* (4), 668–675.

Debrezion, Ghebreegziabiher, Eric Pels, and Piet Rietveld, "The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-Analysis," *Journal of Real Estate Finance and Economics*, 2007, *35* (2), 161–180.

Do, A. Quang and Gary Grudnitski, "Golf Courses and Residential House Prices: An Empirical Examination," *Journal of Real Estate Finance and Economics*, 1995, *10* (3), 261–270.

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

_ **and** _ , *Spatial Econometrics Using Microdata*, London: ISTE / John Wiley & Sons, 2014.

_ **,** _ **, Marius Thériault, and François Des Rosiers**, "A Spatial Difference-in-Differences Estimator to Evaluate the Effect of Change in Public Mass Transit Systems on House Prices," *Transportation Research Part B: Methodological*, 2014, *64*, 24–40.

_ **, François Des Rosiers, and Nicolas Devaux**, "Yes or Not in My Backyard (YIMBY vs. NIMBY)? The Impact of New Social Housing Construction on Single-Family House Prices in Quebec City (Canada)," *Housing Policy Debate*, 2023, *33* (4), 865–890.

_ **, _ , Marius Thériault, and Patricia Dib**, "Economic Impact of a Supply Change in Mass Transit in Urban Areas: A Canadian Example," *Transportation Research Part A: Policy and Practice*, 2011, *45* (1), 46–62.

_ **, Julie Le Gallo, Franck Chapel, and Mohamed Hilal**, "Is There an Additional Price Premium for Single-Family Houses Exposed to Urban Parks? Insights from Causal Spatio-Temporal Matching in Québec City," *Journal of Environmental Management*, 2025, *394*, 127417.

_ **, 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.

Efron, Bradley and Robert J. Tibshirani, *An Introduction to the Bootstrap*, New York: Chapman & Hall, 1993.

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

Elhorst, J. Paul, "Applied Spatial Econometrics: Raising the Bar," *Spatial Economic Analysis*, 2010, *5* (1), 9–28.

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

Farmer, Michael C. and Clifford A. Lipscomb, "Using Quantile Regression in Hedonic Analysis to Reveal Submarket Competition," *Journal of Real Estate Research*, 2010, *32* (4), 435–460.

fei Lee, Lung, "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," *Econometric Reviews*, 2003, *22* (4), 307–335.

Freedman, David A., "Bootstrapping Regression Models," *Annals of Statistics*, 1981, *9* (6), 1218–1228.

Friedman, Jerome H., "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 2001, *29* (5), 1189–1232.

Getis, Arthur, "Spatial Weights Matrices," *Geographical Analysis*, 2009, *41* (4), 404–410.

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

_ **and Stephen Machin**, "Valuing School Quality, Better Transport, and Lower Crime: Evidence from House Prices," *Oxford Review of Economic Policy*, 2008, *24* (1), 99–119.

_ **, Henry G. Overman, and Eleonora Patacchini**, "Spatial Methods," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Amsterdam: Elsevier, 2015, pp. 115–168.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport, "Why Do the Poor Live in Cities? The Role of Public Transportation," *Journal of Urban Economics*, 2008, *63* (1), 1–24.

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

Gu, Hengyu, Jiaxin Tang, Jinliao Luo, and Ya Feng, "Examination of the Impacts of Hospital Accessibility on Housing Prices: Heterogeneity across Attributes, Space and Price Quantiles," *Journal of Housing and the Built Environment*, 2024, *39* (1), 179–200.

Gyger, Tim, Simon Hauri, Stefan Bühlmann, Matthias Lehner, Jonas Schlesinger, and Fabio Sigrist, "Explainable Spatial Machine Learning for Hedonic Real Estate Modeling," Working Paper 5191260, SSRN 2025.

Hagemann, Andreas, "Cluster-Robust Bootstrap Inference in Quantile Regression Models," *Journal of the American Statistical Association*, 2017, *112* (517), 446–456.

Hahn, Jinyong, "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 1995, *11* (1), 105–121.

Halvorsen, Robert and Henry O. Pollakowski, "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics*, 1981, *10* (1), 37–49.

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

Hao, Lingxin and Daniel Q. Naiman, *Quantile Regression*, Thousand Oaks: Sage, 2007.

Haurin, Donald R., Jessica L. Haurin, Taylor Nadauld, and Anthony Sanders, "List Prices, Sale Prices and Marketing Time: An Application to U.S. Housing Markets," *Real Estate Economics*, 2010, *38* (4), 659–685.

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.

_ , "Bootstrap Methods for Median Regression Models," *Econometrica*, 1998, *66* (6), 1327–1351.

Jin, Fei and Lung fei Lee, "On the Bootstrap for Moran's I Test for Spatial Dependence," *Journal of Econometrics*, 2015, *184* (2), 295–314.

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** _ , "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, 2010, *157* (1), 53–67.

Kestens, Yan, Marius Thériault, and François Des Rosiers, "Heterogeneity in Hedonic Modelling of House Prices: Looking at Buyers' Household Profiles," *Journal of Geographical Systems*, 2006, *8* (1), 61–96.

Kim, Chong Won, Tim T. Phipps, and Luc Anselin, "Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach," *Journal of Environmental Economics and Management*, 2003, *45* (1), 24–39.

Kim, Tae-Hwan and Christophe Muller, "Two-Stage Quantile Regression When the First Stage Is Based on Quantile Regression," *Econometrics Journal*, 2004, *7* (1), 218–231.

Knight, John R., "Listing Price, Time on Market, and Ultimate Selling Price: Causes and Effects of Listing Price Changes," *Real Estate Economics*, 2002, *30* (2), 213–237.

Koenker, Roger, *Quantile Regression*, Cambridge: Cambridge University Press, 2005.

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

_ **and** _ , "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 1982, *50* (1), 43–61.

_ **and José A. F. Machado**, "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 1999, *94* (448), 1296–1310.

_ **and Kevin F. Hallock**, "Quantile Regression," *Journal of Economic Perspectives*, 2001, *15* (4), 143–156.

_ **and Zhijie Xiao**, "Inference on the Quantile Regression Process," *Econometrica*, 2002, *70* (4), 1583–1612.

Kostov, Philip, "A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices," *Spatial Economic Analysis*, 2009, *4* (1), 53–72.

Kuminoff, Nicolai V. and Jaren C. Pope, "Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay?," *International Economic Review*, 2014, *55* (4), 1227–1250.

_ **, 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.

Lahiri, Soumendra N., *Resampling Methods for Dependent Data*, New York: Springer, 2003.

LeSage, James P. and R. Kelley Pace, *Introduction to Spatial Econometrics*, Boca Raton: CRC Press, 2009.

_ **and** _ , "The Biggest Myth in Spatial Econometrics," *Econometrics*, 2014, *2* (4), 217–249.

Liao, Wen-Chi and Xizhu Wang, "Hedonic House Prices and Spatial Quantile Regression," *Journal of Housing Economics*, 2012, *21* (1), 16–27.

Long, J. Scott and Laurie H. Ervin, "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model," *American Statistician*, 2000, *54* (3), 217–224.

Machado, José A. F. and Jo ao M. C. Santos Silva, "Quantiles via Moments," *Journal of Econometrics*, 2019, *213* (1), 145–173.

MacKinnon, James G. and Halbert White, "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 1985, *29* (3), 305–325.

Mak, Stephen, Lennon Choy, and Winky Ho, "Quantile Regression Estimates of Hong Kong Real Estate Prices," *Urban Studies*, 2010, *47* (11), 2461–2472.

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

Marquardt, Donald W., "Comment: You Should Standardize the Predictor Variables in Your Regression Models," *Journal of the American Statistical Association*, 1980, *75* (369), 87–91.

McMillen, Daniel P., "Changes in the Distribution of House Prices over Time: Structural Characteristics, Neighborhood, or Coefficients?," *Journal of Urban Economics*, 2008, *64* (3), 573–589.

_ , *Quantile Regression for Spatial Data* SpringerBriefs in Regional Science, Berlin: Springer, 2013.

Meinshausen, Nicolai, "Quantile Regression Forests," *Journal of Machine Learning Research*, 2006, *7*, 983–999.

Mohammad, Sara I., Daniel J. Graham, Patricia C. Melo, and Richard J. Anderson, "A Meta-Analysis of the Impact of Rail Projects on Land and Property Values," *Transportation Research Part A: Policy and Practice*, 2013, *50*, 158–170.

Nelson, Jon P., "Meta-Analysis of Airport Noise and Hedonic Property Values: Problems and Prospects," *Journal of Transport Economics and Policy*, 2004, *38* (1), 1–27.

O'Brien, Robert M., "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality & Quantity*, 2007, *41* (5), 673–690.

Palm, Risa, "Spatial Segmentation of the Urban Housing Market," *Economic Geography*, 1978, *54* (3), 210–221.

Parente, Paulo M. D. C. and Jo ao M. C. Santos Silva, "Quantile Regression with Clustered Data," *Journal of Econometric Methods*, 2016, *5* (1), 1–15.

Parmeter, Christopher F. and Jaren C. Pope, "Quasi-Experiments and Hedonic Property Value Methods," in John A. List and Michael K. Price, eds., *Handbook on Experimental Economics and the Environment*, Cheltenham: Edward Elgar, 2013, pp. 3–66.

Powell, James L., "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 1983, *51* (5), 1569–1575.

Pryce, Gwilym, "Housing Submarkets and the Lattice of Substitution," *Urban Studies*, 2013, *50* (13), 2682–2699.

Renne, John L., Tara Tolford, Shima Hamidi, and Reid Ewing, "The Cost and Affordability Paradox of Transit-Oriented Development: A Comparison of Housing and Transportation Costs Across Transit-Oriented Development, Hybrid and Transit-Adjacent Development Station Typologies," *Housing Policy Debate*, 2016, *26* (4–5), 819–834.

Rey-Blanco, David, Pelayo Arbués, Fernando A. López, and Antonio Páez, "Using Machine Learning to Identify Spatial Market Segments. A Reproducible Study of Major Spanish Markets," *Environment and Planning B: Urban Analytics and City Science*, 2024, *51* (1), 89–108.

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

Rosiers, François Des, Marius Thériault, and Paul-Y. Villeneuve, "Sorting Out Access and Neighbourhood Factors in Hedonic Price Modelling," *Journal of Property Investment & Finance*, 2000, *18* (3), 291–315.

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

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.

Small, Kenneth A. and Seiji S. C. Steimetz, "Spatial Hedonics and the Willingness to Pay for Residential Amenities," *Journal of Regional Science*, 2012, *52* (4), 635–647.

Spagnolo, Francesco Schirripa, Riccardo Borgoni, Antonella Carcagnì, and Alessandra Michelangeli, "A Spatial Semiparametric M-Quantile Regression for Hedonic Price Modelling," *AStA Advances in Statistical Analysis*, 2024, *108* (1), 159–183.

Squires, Graham, Thao Ngo, Michael McCord, and Daniel Lo, "The Spatial Effect of Airport Proximity on House Prices: A Quantile Regression Analysis for the New Zealand Market," *Applied Economics*, 2026, *58* (43), 9455–9471.

Stakhovych, Stanislav and Tammo H. A. Bijmolt, "Specification of Spatial Models: A Simulation Study on Weights Matrices," *Papers in Regional Science*, 2009, *88* (2), 389–408.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, 2005, pp. 80–108.

Su, Liangjun and Zhenlin Yang, "Instrumental Variable Quantile Estimation of Spatial Autoregressive Models," Working Paper, School of Economics, Singapore Management University, Singapore 2011. https://ink.library.smu.edu.sg/soe_research/1074.

Theebe, Marcel A. J., "Planes, Trains, and Automobiles: The Impact of Traffic Noise on House Prices," *Journal of Real Estate Finance and Economics*, 2004, *28* (2–3), 209–234.

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.

Troy, Austin and J. Morgan Grove, "Property Values, Parks, and Crime: A Hedonic Analysis in Baltimore, MD," *Landscape and Urban Planning*, 2008, *87* (3), 233–245.

Waddell, Paul, Brian J. L. Berry, and Irving Hoch, "Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location," *Journal of Real Estate Finance and Economics*, 1993, *7* (2), 117–141.

Watkins, Craig A., "The Definition and Identification of Housing Submarkets," *Environment and Planning A*, 2001, *33* (12), 2235–2253.

Wen, Haizhen, Yue Xiao, and Ling Zhang, "Spatial Effect of River Landscape on Housing Price: An Empirical Study on the Grand Canal in Hangzhou, China," *Habitat International*, 2017, *63*, 34–44.

Wolch, Jennifer R., Jason Byrne, and Joshua P. Newell, "Urban Green Space, Public Health, and Environmental Justice: The Challenge of Making Cities "Just Green Enough"," *Landscape and Urban Planning*, 2014, *125*, 234–244.

Zhang, Lu, Yiguo Sun, and Thanasis Stengos, "Spatial Dependence in the Residential Canadian Housing Market," *Journal of Real Estate Finance and Economics*, 2019, *58* (2), 223–263.

Zietz, Joachim, Emily Norman 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.

A. Numerical record and audit trail

This appendix documents the provenance of every number in the paper, a matter that became material during the October 2026 revision.

Pipeline. The analysis is produced by `scripts/05_sqr_multi_segment.py`, which loads the four Centris JSON files, builds the estimation samples of Table 2, estimates equations (1), (2) and (8), bootstraps the quantile regressions, computes the Wald statistics, Moran's I, the dollar-equivalent implicit prices and the scalar effect decomposition, and writes seven figures. Scripts 01 to 04 are earlier conditional-mean and single-segment versions kept for reference. The raw snapshot (about 42,000 listings) is not redistributable under the Centris terms of use and is not on the machine used for the revision, so the pipeline could not be re-executed.

The preserved log. The only surviving output of the pipeline is the console log `results/sqr_multi_segment_ou` For each segment it records the sample sizes, the price range and median, the regional counts and medians, the amenity shares, the OLS adjusted R^2 and the Moran statistic on OLS residuals, the pseudo-R^2 at each quantile, the full coefficient table (coefficient, bootstrap standard error, z, p) at each quantile, the Wald statistics for three quantile pairs, Moran's I on the QR and second-stage 2SQR residuals at each quantile, the first-stage R^2, the median spatial-lag coefficient, the dollar-equivalent implicit prices and the median effect decomposition. It does not record the moments of the structural variables, the variance inflation factors, the first-stage F-statistic, an omnibus test, the second-stage coefficients at quantiles other than the median, standard errors for the second stage, or any sensitivity analysis. The log was evidently produced by a version of the script earlier than the one in the repository, which prints several of these quantities; that version was not preserved.

Discrepancies in the May 2026 version. The tables of the first version of this paper could not be reconciled with the log. For houses they reported an OLS adjusted R^2 of 0.685 (log: 0.619), Moran's I of 0.441 (0.506), pseudo-R^2 between 0.440 and 0.483 (0.355–0.444), a median spatial lag of 0.227 (0.291), a highway coefficient of 0.040 at $\tau = 0.10$ (0.205), Moran's I of 0.36–0.40 on 2SQR residuals (0.52–0.55), and a sensitivity table whose $k = 10$ column reproduced the log while the other columns had no source; the descriptive statistics for condominiums, multiplexes and land, including amenity shares and regional composition, differed from the log throughout; the Wald statistics in the text differed from those in the tables, and both differed from the log; an omnibus statistic with 7 degrees of freedom appeared in the abstract and one with 17 degrees of freedom in the results. The original Figure 2 reproduced the tabled rather than the logged values. The most likely explanation is that the tables and that figure came from a second run of a later script version, of

which no log was kept, while the narrative text was written partly from the log; the variations in the distance variable mentioned in the first version ("corrected distance measures") support this reading. Because the log is the only auditable record, the revised paper reports its values throughout, and the earlier figures are documented in the repository's revision changelog rather than reproduced here.

Generated tables and figures. Every table in the paper except the two definitional ones (Tables 1 and 4) and every figure is written by `scripts/06_tables_figures_from_log.py`, which parses the log, checks that each segment has five complete coefficient blocks with the recorded number of regressors, and emits the LaTeX fragments in `paper/tables/` and the PNG files in `figures/`. Three quantities are derived rather than read: the significance stars, recomputed from the recorded coefficients and bootstrap standard errors with the 10/5/1 percent thresholds; the overall first-stage F of Table 13, computed from the recorded R^2, N and the number of first-stage regressors; and the unconditional price quantiles $P(\tau)$ in the header of Table 16, recovered as $\pi_k(\tau)/[\exp(\hat{\gamma}_k(\tau)) - 1]$ from the highway row (the median so recovered, \$579,041, agrees with the recorded median of \$579,000 to within the four-decimal rounding of the coefficient). The parsed log is saved as `results/log_parsed.json`.

Anomalies to resolve on re-execution. Three features of the recorded estimates call for verification. The coefficient on the centred log distance to Montreal is positive and large in every segment (1.6 to 2.5 for houses, up to 8.4 for multiplexes and 5.7 for land, with bootstrap standard errors of 0.4 to 1.7), which is not interpretable as a distance gradient and suggests either a scaling problem in the distance variable of the run that produced the log or near-collinearity with the regional indicators; the amenity coefficients are conditional on this control. Moran's I on the 2SQR residuals exceeds that on the QR residuals, which Section 5.5 attributes to the use of the fitted lag in forming the residuals; the structural residuals should be examined. And the regional coefficient for "other regions" relative to Gatineau changes sign across quantiles for houses (-0.126 at $\tau = 0.10$, $+0.163$ at $\tau = 0.90$), which may be a consequence of the first anomaly.

Outputs to restore. A re-run with the current script and the snapshot would restore, in order of priority: bootstrap standard errors for the 2SQR second stage and the lag coefficients at all five quantiles; the partial first-stage F; Moran's I on structural residuals; a cluster or block bootstrap; the sensitivity of the results to k, to projected coordinates, to the trimming thresholds and to municipality fixed effects; the descriptive statistics of the structural variables; and the variance inflation factors. The paper should not be circulated as a final version before at least the first three items are available.

B. Additional table

Table 17. Structural, locational and regional coefficients by quantile — single-family houses

	Quantile regression, τ				
Variable	0.10	0.25	0.50	0.75	0.90
Constant	12.652***	12.896***	13.160***	13.365***	13.579***
	(0.041)	(0.029)	(0.022)	(0.023)	(0.037)
Bedrooms	−0.066***	−0.051***	−0.036***	−0.037***	−0.051***
	(0.010)	(0.008)	(0.008)	(0.007)	(0.012)
Bathrooms	0.235***	0.240***	0.243***	0.249***	0.244***
	(0.015)	(0.012)	(0.011)	(0.011)	(0.018)
Stories	−0.044**	−0.043***	−0.043***	−0.026**	0.000
	(0.022)	(0.014)	(0.015)	(0.012)	(0.022)
Parking spaces	0.002	0.003*	0.006***	0.011***	0.015***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
ln(floor area)	0.511***	0.509***	0.555***	0.608***	0.688***
	(0.035)	(0.021)	(0.021)	(0.024)	(0.032)
ln(land area)	0.048***	0.042***	0.031***	0.023***	0.023***
	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)
ln(distance to Montreal)	2.453***	1.554***	2.409***	1.965***	1.955**
	(0.552)	(0.402)	(0.480)	(0.459)	(0.830)
Montreal	0.279***	0.293***	0.294***	0.362***	0.441***
	(0.030)	(0.018)	(0.021)	(0.019)	(0.039)
Québec City	0.113*	0.160***	0.179***	0.203***	0.253***
	(0.061)	(0.033)	(0.025)	(0.038)	(0.050)
Other regions	−0.126***	−0.044*	0.025	0.095***	0.163***
	(0.035)	(0.023)	(0.021)	(0.019)	(0.033)

Notes: Same regressions as Table 6 ($N = 6,250$). Continuous variables are centred at their sample means, so the constant is the conditional quantile of log price for a mean-characteristic house in the Gatineau reference region without any amenity indicator. Regional dummies are relative to Gatineau. Bootstrap standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels (two-sided normal test using bootstrap standard errors, $B = 200$ paired resamples). All entries are parsed from `results/sqr_multi_segment_output.txt`, the preserved log of the original run of `scripts/05_sqr_multi_segment.py`, by `scripts/06_tables_figures_from_log.py`.