Université du Québec en Outaouais

Département des sciences administratives

Working Paper No. 5

# Airbnb, Residential Rents, and Housing Market Pressure: A Hedonic and Spatial Econometric Analysis

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**

This paper examines the relationship between short-term rental (STR) activity and residential rents in Quebec, Canada, using a 2026 cross-section of 3,456 Airbnb listings and 8,303 residential rental listings with precise coordinates. For each rental listing we count the Airbnb listings within 250 m, 500 m, 1 km and 2 km (Haversine distances) and estimate hedonic rent equations with building-type and location fixed effects. With city fixed effects, one additional Airbnb listing within 500 m is associated with monthly rents about 0.4% higher (0.3–0.5% across specifications), the per-listing coefficient decays with distance, the association is larger at the upper quantiles of the conditional rent distribution, and it survives spatial lag and spatial error models (total impact 0.40%). Two further results qualify this picture. First, heteroskedasticity-robust standard errors overstate precision: Conley spatial HAC and neighbourhood-clustered standard errors are about three times larger, although the city-fixed-effects estimate remains significant. Second, with fixed effects for forward sortation areas (postal-code neighbourhoods), the coefficient falls to 0.08% per listing and is not distinguishable from zero under spatial inference, with the same pattern for logarithmic and extensive-margin exposure measures. The Airbnb–rent association in these data is therefore predominantly a between-neighbourhood phenomenon, consistent with STR supply concentrating in high-amenity, high-rent neighbourhoods rather than with a within-neighbourhood causal effect. Spatially blocked cross-validation shows that the apparent advantage of tree-based machine-learning models over the linear hedonic model disappears when test neighbourhoods are unseen, and a hedonic model of Airbnb nightly prices finds no premium attached to city-level rents. We discuss the implications for the interpretation of cross-sectional STR–rent correlations and for Quebec's registration-based regulation.

**Keywords:** Airbnb, short-term rentals, housing rents, hedonic price function, neighbourhood fixed effects, spatial econometrics, Quebec

**JEL Classification:** R21, R31, R52, L83, C21

## 1. Introduction

Short-term rental (STR) platforms have changed the way urban housing can be used. Since 2008 Airbnb has grown from a small home-sharing service into a global accommodation platform, an archetypal case of disruptive innovation in tourism (Guttentag, 2015) and of the broader rise of peer-to-peer markets (Einav et al., 2016; Sundararajan, 2016). By allowing owners to let a dwelling, or part of it, to visitors at nightly rates that can exceed what a long-term tenant pays per month, the platform creates an incentive to divert housing units from the residential rental market. The flexible peer supply that generates welfare gains in the accommodation market (Zervas et al., 2017; Farronato and Fradkin, 2022) is, from the housing market's perspective, a competing use for the residential stock.

This diversion mechanism underlies a policy concern that STR activity raises residential rents and reduces affordability, particularly in cities with strong tourist appeal. The empirical literature largely supports the concern. Studies exploiting cross-market variation (Barron et al., 2021; Horn and Merante, 2017), regulatory discontinuities (Koster et al., 2021; Valentin, 2021; Duso et al., 2024) and within-city diffusion (Garcia-López et al., 2020; Franco and Santos, 2021) find that STR activity raises rents and house prices, with effects that are modest per listing but spatially concentrated. The economic logic is simple: when a unit earns more on Airbnb than under a lease, the effective supply of long-term rental housing contracts. On the demand side, STR guests may raise foot traffic and consumption amenities in a neighbourhood (Hidalgo et al., 2024; Almagro and Domínguez-Iino, 2025), so the net effect on rents depends on the local context, the composition of listings and the degree of segmentation between the short-term and long-term segments.

The province of Quebec, and Montreal in particular, is a natural setting for studying these questions. Montreal is Canada's second-largest city and one of its main tourist destinations; national analyses identify it as one of the country's largest and most commercialised STR markets (Combs et al., 2020). The province has also experienced a tight rental market in recent years, and policymakers have responded with a provincial registration regime for STR operators and with municipal zoning and seasonal restrictions in Montreal (Nieuwland and van Melik, 2020). Yet econometric evidence on the housing-market correlates of Airbnb in Quebec is scarce: the Canadian literature is either national and descriptive (Combs et al., 2020) or focused on the political economy of displacement in Toronto (Grisdale, 2021). To our knowledge, no study has estimated the Airbnb–rent relationship at the level of individual listings for Quebec.

Estimating the effect of STR platforms on rents is difficult, and the central difficulty is endogeneity. Airbnb listings are not randomly located: hosts list where tourist demand, amenities and rents are already high, so a positive cross-sectional correlation between Airbnb density and rents may reflect reverse causality or omitted neighbourhood characteristics rather than a causal effect. The

quasi-experimental literature addresses this with regulatory shocks and panel data (Koster et al., 2021; Valentin, 2021; Duso et al., 2024) or with instruments for Airbnb penetration (Barron et al., 2021; Garcia-López et al., 2020). Our data are a single 2026 cross-section of Airbnb and rental listings, and we do not claim to identify a causal effect. Instead, we ask what a carefully specified cross-sectional hedonic design can and cannot establish, and we let the data speak on the identification problem itself: because every rental listing carries a postal code, we can compare the Airbnb–rent association across progressively finer location fixed effects, from the city to the forward sortation area (FSA, a postal-code neighbourhood), and we can compute standard errors that respect the spatial dependence built into a buffer-based exposure measure.

Our empirical strategy proceeds in stages. We first estimate a hedonic rent equation in which the logarithm of the monthly asking rent is regressed on the number of Airbnb listings within a buffer of the listing (our reference radius is 500 m), controlling for bedrooms, bathrooms, building type and city fixed effects; the coefficient is a semi-elasticity, the approximate percentage change in rent associated with one additional nearby listing. We then replace city fixed effects with FSA fixed effects, and compare heteroskedasticity-robust, cluster-robust and Conley spatial HAC standard errors. We estimate spatial autoregressive and spatial error models by generalised moments and report the direct, indirect and total impacts that the spatial lag implies. We vary the buffer radius, decompose exposure into rings and into extensive and intensive margins, estimate quantile regressions across the conditional rent distribution, and benchmark the linear hedonic model against regularised and tree-based learners under both random and spatially blocked cross-validation. A complementary hedonic model of Airbnb nightly prices examines whether STR pricing responds to city-level rents.

The results have two layers. With city fixed effects, each additional Airbnb listing within 500 m is associated with monthly rents about 0.4% higher; the per-listing coefficient falls monotonically with the radius, the association is largest at the upper quantiles, and it is essentially unchanged in spatial lag and spatial error models, with a total impact of 0.40% per listing. These are the results a reader of the existing cross-sectional literature would expect. The second layer qualifies them. Standard errors that account for spatial dependence are about three times larger than the heteroskedasticity-robust errors usually reported, so that the $t$-statistic of the city-fixed-effects estimate falls from 17 to about 5. More importantly, once FSA fixed effects absorb neighbourhood-level unobservables, the coefficient falls to 0.08% per listing and is no longer distinguishable from zero under spatial inference; the logarithmic and extensive-margin exposure measures behave in the same way. About a quarter of the variance in the 500 m count is within FSAs, so the attenuation does not reflect a lack of identifying variation. The Airbnb–rent association in these data is thus predominantly a between-neighbourhood phenomenon. Spatially blocked cross-validation delivers a parallel message for the machine-learning benchmark: the sizeable advantage of random forests and gradient boosting

over the linear model under random splits disappears when whole neighbourhoods are held out, which indicates that the flexible learners were interpolating neighbourhood rent levels rather than uncovering structure in the exposure–rent relationship.

The paper makes three contributions. First, it provides the first listing-level econometric analysis of the Airbnb–rent relationship for Quebec, and to our knowledge for any Canadian market, using coordinates precise enough for distance-based matching between the two sets of listings; the Canadian evidence to date is descriptive (Combs et al., 2020) or qualitative (Grisdale, 2021). Second, it documents, within a single data set, how much of a typical cross-sectional STR–rent correlation survives neighbourhood fixed effects and spatially robust inference. The answer, about one fifth of the point estimate and little of its statistical significance, is informative for the reading of other cross-sectional studies and is consistent with the location-sorting mechanism emphasised by Almagro and Domínguez-Iino (2025). Third, it brings this evidence to bear on the design of Quebec's registration-based regime and Montreal's zoning rules, while being explicit that cross-sectional listing data cannot establish the causal effect that would justify a given level of restriction.

The remainder of the paper is organised as follows. Section 2 reviews the literature on STR platforms and housing markets and the methods we use. Section 3 describes the data, their limitations and the spatial merge. Section 4 sets out the econometric framework, including the fixed-effects, spatial and inference choices. Section 5 reports the results, Section 6 the robustness analyses, Section 7 the interpretation, policy implications and limitations, and Section 8 concludes.

## 2. Literature Review

This section reviews the empirical and methodological literatures on which the paper builds. We first survey the evidence on STR platforms and housing costs, from early correlational studies to quasi-experimental designs (Section 2.1), then the Canadian and Quebec evidence (Section 2.2), the economics of peer-to-peer accommodation (Section 2.3) and the pricing of STR listings (Section 2.4). The remaining subsections cover the methodological foundations: hedonic pricing and its identification problems (Section 2.5), spatial econometrics and spatial inference (Section 2.6), quantile regression (Section 2.7) and machine learning in housing economics (Section 2.8). Section 2.9 reviews the tourism-gentrification and regulation literature, and Section 2.10 states the gap the paper addresses.

### 2.1. Short-Term Rentals and Housing Costs

Housing costs respond to shifts in effective supply, and markets with inelastic supply capitalise demand shocks into prices and rents (Glaeser et al., 2006; Saiz, 2010; Gyourko et al., 2013). STR platforms create such a shift at the margin: by raising the return to allocating a dwelling to visitors

rather than tenants, they can withdraw units from the long-term stock. The empirical literature testing this mechanism has developed in two waves.

**Correlational and panel evidence.** The first large-scale study is Barron et al. (2021), who exploit zipcode-level variation in Airbnb penetration across the United States and instrument Airbnb growth with the interaction of Google search interest for Airbnb and a measure of touristiness. They estimate that a 1% increase in Airbnb listings raises rents by 0.018% and house prices by 0.026%, and attribute the effect to the reallocation of units from long-term to short-term supply. Horn and Merante (2017) use listing-level asking rents in Boston and find that a one-standard-deviation increase in Airbnb listings is associated with asking rents 0.4% higher, with the effect concentrated where listings are entire homes. Sheppard and Udell (2016) document property-value capitalisation in New York City; Lee (2016) documents the removal of units from the Los Angeles rental stock; and Wachsmuth and Weisler (2018) show that the "rent gap" opened by STR revenue potential is systematically related to gentrification pressure. Ayouba et al. (2020) apply spatial panel methods to eight French cities and find significant rent effects of Airbnb density in Lyon, Montpellier and Paris but not elsewhere, early evidence that the association is heterogeneous across markets. Using platform data and a structural model of landlord choice, Li et al. (2022) show that Airbnb entry shifts landlords at the margin between long-term and short-term supply, with the incidence falling on long-term tenants in high-demand markets. Calder-Wang (2021) estimates a structural model of the New York City rental market and finds that the rent burden created by Airbnb falls disproportionately on higher-income renters.

**Quasi-experimental evidence.** A second wave exploits policy discontinuities and regulatory shocks. Garcia-López et al. (2020) use the staggered, spatially uneven diffusion of Airbnb across Barcelona neighbourhoods together with an instrumental-variables strategy and estimate that Airbnb activity raised rents by 1.9%, transaction prices by 4.6% and posted prices by 3.7% in the average neighbourhood, and rents by about 7% in the neighbourhoods with the most Airbnb activity. Koster et al. (2021) exploit Home Sharing Ordinances adopted by some Los Angeles-area cities but not others in a spatial regression-discontinuity design: the ordinances reduced Airbnb listings by about 50% and lowered house prices and rents by about 3% in treated areas, with much larger effects close to tourist attractions. Valentin (2021) finds that New Orleans' licensing reform capitalised into property values where STR use was legalised. Duso et al. (2024) use two Berlin regulatory shocks and estimate that each additional commercial Airbnb listing displaces 0.23–0.37 long-term rental units and raises rents per square metre by 1.3–2.4%. Franco and Santos (2021) document positive effects of Airbnb density on both prices and rents across Portuguese municipalities, strongest in tourist-intensive locations, and Shabrina et al. (2022) link neighbourhood-level commercial Airbnb

use to higher rents in London.

Taken together, this literature finds positive, spatially concentrated associations between STR activity and housing costs, with quasi-experimental designs generally confirming, and sometimes exceeding, the magnitudes suggested by correlational work. Two features of the quasi-experimental studies matter for the present paper. First, they rely on within-neighbourhood variation over time, which removes time-invariant neighbourhood amenities from the comparison. Second, their per-listing effects are expressed for well-defined treatments, such as commercial or high-availability listings. A cross-sectional design cannot replicate the first feature, but it can test how much of its estimate depends on between-neighbourhood variation, which is what we do.

### 2.2. Canadian and Quebec Evidence

Evidence for Canada is comparatively scarce and largely descriptive. Combs et al. (2020) provide the first national analysis of Canadian STR markets, documenting rapid but spatially uneven growth, strong revenue concentration among multi-listing operators, and between 17,000 and 43,000 entire homes rented frequently enough in 2018 to be unlikely to house a permanent resident, with Montreal among the three dominant markets. Grisdale (2021) traces the political economy of STR-driven displacement in Toronto, linking commercial Airbnb operation to rental-supply loss in central neighbourhoods. Beyond market measurement and critical urban geography there is little listing-level econometric evidence for Canadian markets, and none, to our knowledge, for Quebec, although the province combines a large STR market, a distinctive provincial registration regime and a tight rental market.

### 2.3. The Economics of Peer-to-Peer Accommodation

A complementary literature studies STR platforms as peer-to-peer markets. Einav et al. (2016) analyse how platforms reduce transaction costs and enable trade in underutilised assets, and Sundararajan (2016) surveys the rise of "crowd-based capitalism." Guttentag (2015) frames Airbnb as a disruptive innovation in tourism accommodation. On market impacts, Zervas et al. (2017) estimate that Airbnb entry reduced hotel revenues in the most affected Texas segments by 8–10%, and Farronato and Fradkin (2022) show that flexible peer supply expands capacity when demand peaks, generating consumer surplus while eroding hotel profits. This literature establishes the incentive at the heart of the housing channel: hosts respond elastically to STR returns, and the same responsiveness that creates accommodation-market surplus reallocates dwellings away from long-term tenants when STR returns exceed residential rents. It also points to a second channel: STR guests change local consumption amenities, which Hidalgo et al. (2024) document for Madrid and Almagro and Domínguez-Iino (2025) embed in a model of location sorting for Amsterdam. Endogenous amenities are a reason why STR density and rents can be correlated even in the absence

of supply withdrawal, and a reason to control for neighbourhood characteristics as finely as the data permit.

### 2.4. Pricing of Short-Term Rental Listings

A distinct strand estimates hedonic models of Airbnb nightly prices, which informs our Model 2. Wang and Nicolau (2017) study some 180,000 listings in 33 cities and find that entire-home status, capacity and host attributes, notably superhost status, are dominant price determinants. Gibbs et al. (2018) estimate hedonic price equations for five Canadian Airbnb markets, the study closest to our Model 2 in geographic scope, and document significant premia for entire homes and location. Ert et al. (2016) show that host reputation and trust signals are capitalised into prices. These studies consistently report positive superhost premia and explanatory power well above what we obtain on the full sample of scraped prices; we show in Section 5 that the discrepancy is a data-quality issue in the scraped price field rather than a Quebec-specific pricing pattern.

### 2.5. Hedonic Pricing: Theory, Practice and Identification

The hedonic framework descends from Lancaster (1966), who recast consumer demand as demand for characteristics, and Rosen (1974), who formalised the market equilibrium in which the prices of differentiated goods reveal implicit characteristic prices. In housing applications the implicit price of a locational attribute, here exposure to nearby STR activity, is identified from cross-sectional price variation conditional on structural characteristics (Palmquist, 2005). Surveys of empirical practice document both the workhorse status and the fragility of hedonic estimates: Sirmans et al. (2005) catalogue the composition of hundreds of published models, Malpezzi (2003) reviews specification choices, and Parmeter et al. (2007) illustrate the gains from flexible functional forms. The identification literature is more sobering. Kuminoff et al. (2010) show in large-scale simulations that spatial fixed effects are the most effective safeguard against omitted-variable bias in cross-sectional hedonics; Kuminoff and Pope (2014) show that capitalisation estimates recover marginal willingness to pay only under restrictive conditions; and Bishop et al. (2020) summarise current best practice, including the use of fine spatial fixed effects and quasi-experimental variation. Oster (2019) formalises the use of coefficient movements across control sets to gauge the potential importance of selection on unobservables. Our design follows this guidance as far as a cross-section permits: we compare city and neighbourhood fixed effects, report how the coefficient moves as controls are added, and interpret the city-fixed-effects estimate as an upper bound.

### 2.6. Spatial Econometrics and Spatial Inference

Housing prices are spatially dependent: nearby dwellings share unobserved amenities, and valuation spillovers propagate through comparable-based pricing. Ignoring this dependence invalidates

inference and can bias coefficients when spatially correlated unobservables are also correlated with regressors (Dubin, 1988; Can, 1992). The canonical modelling responses are the spatial autoregressive (SAR) and spatial error (SEM) models (Anselin, 1988), with feasible estimation by spatial two-stage least squares and generalised moments developed by Kelejian and Prucha (1998, 1999), the estimators we employ. LeSage and Pace (2009) emphasise that in a SAR model the regression coefficients are not marginal effects; the appropriate summary measures are the average direct, indirect and total impacts implied by the spatial multiplier, which we report. A separate literature addresses inference under spatial dependence without modelling it: Conley (1999) proposes a spatial heteroskedasticity-and-autocorrelation-consistent (HAC) covariance estimator with a distance kernel, Kelejian and Prucha (2007) extend it to spatial regression models, and Cameron and Miller (2015) discuss cluster-robust alternatives and their failure when a single cluster dominates the sample. Our buffer-count exposure variable is itself a spatial moving average of Airbnb locations, so neighbouring rental listings have mechanically correlated regressors; this makes spatially robust inference a requirement rather than a refinement. Because any buffer or neighbourhood definition is a discretisation of a continuous spatial process, the modifiable areal unit problem (Fotheringham and Wong, 1991) also applies, which motivates the use of several radii and ring decompositions.

### 2.7. Distributional Effects and Quantile Regression

Mean regression can mask heterogeneity across the conditional price distribution. Quantile regression (Koenker and Bassett, 1978; Koenker and Hallock, 2001) recovers implicit prices at arbitrary quantiles, and housing applications consistently find that implicit prices differ across the distribution: Zietz et al. (2008) show that buyers of higher-priced homes value certain attributes differently from buyers of lower-priced homes, and McMillen (2008) decomposes changes in the house-price distribution into characteristics and coefficient effects. In the STR context, distributional heterogeneity matters for policy: if STR pressure is concentrated where rents are already high, the affordability consequences differ from those of a uniform shift. Our Model 5 provides, to our knowledge, the first quantile-level evidence on the Airbnb–rent association for a Canadian market.

### 2.8. Machine Learning in Housing Economics

Machine-learning methods enter hedonic practice as prediction benchmarks and as data-driven specification checks (Mullainathan and Spiess, 2017; Athey and Imbens, 2019). We use the LASSO (Tibshirani, 1996) and the elastic net (Zou and Hastie, 2005) for regularised variable selection, random forests (Breiman, 2001) and gradient boosting (Friedman, 2001) as flexible nonlinear benchmarks, and SHAP values (Lundberg and Lee, 2017) for feature attribution; all models are implemented in scikit-learn (Pedregosa et al., 2011). The comparison between OLS and the flexible

learners serves an econometric purpose: if a linear hedonic model attains predictive accuracy close to the flexible frontier, linearity is an adequate approximation for inference; where it falls short, the gap signals nonlinearities that fixed effects and exposure construction must absorb. With geographic coordinates among the features, however, tree-based learners can reproduce neighbourhood rent levels from the locations of nearby training listings, so that random train–test splits overstate their accuracy for new neighbourhoods; spatially blocked cross-validation, in which whole neighbourhoods are held out, addresses this and is the counterpart in prediction of the fixed-effects comparison in estimation.

### 2.9. Tourism Gentrification and Short-Term Rental Regulation

A critical urban literature situates STRs within longer-running processes of tourism-driven neighbourhood change. Gotham (2005) coined "tourism gentrification" to describe the transformation of New Orleans' Vieux Carré; Cocola-Gant (2016) identifies holiday rentals as a distinct gentrification front in which transient visitors replace residents; Cocola-Gant and Gago (2021) document the buy-to-let investment channel and displacement in Lisbon; and Wachsmuth and Weisler (2018) connect this process to platform-enabled rent gaps. The commercialisation of nominally peer-to-peer platforms is central to this critique: Ke (2017) shows that multi-listing hosts account for a disproportionate share of Airbnb supply and revenue. On the policy side, Gurran and Phibbs (2017) ask how planners should respond to Airbnb, and Nieuwland and van Melik (2020) compare regulatory regimes across cities, from prohibition to registration and cap systems. Quebec's regime combines provincial registration of all STR establishments, including primary residences, with municipal zoning; Montreal further restricts primary-residence STRs to a summer window and bans them in some boroughs (Section 7).

### 2.10. Positioning and Contribution

Three gaps emerge from this review. First, despite a mature international literature spanning correlational and quasi-experimental designs, there is no listing-level econometric evidence on the STR–rent relationship for Quebec, and Canadian evidence more broadly stops at market measurement (Combs et al., 2020) and critical urban geography (Grisdale, 2021). Second, the cross-sectional studies that dominate the evidence base for most markets rarely report how their estimates respond to neighbourhood fixed effects or to spatially robust inference, although the identification literature identifies these as the first-order checks. Third, the distributional incidence of the association is seldom examined. This paper addresses the three gaps with a single data set: it provides listing-level estimates for Quebec, documents the attenuation of the Airbnb–rent coefficient from city to neighbourhood fixed effects and the inflation of standard errors under spatial dependence, and characterises the association across the conditional rent distribution. We are explicit about what the paper does not do: with a single cross-section we do not identify a causal effect, and we read our

estimates against, rather than as substitutes for, the quasi-experimental magnitudes surveyed above.

## 3. Data

This section describes the two data sets, the cleaning procedures, the spatial merge that links them, the construction of the exposure and neighbourhood variables, and the limitations of the data. Both data sets were collected in early 2026; the Realtor.ca extraction is dated 8 January 2026, and the Airbnb file carries no timestamp.

### 3.1. Airbnb Listings

The Airbnb data were scraped from the public platform for the province of Quebec and comprise 5,000 listings. Each record contains the listing name, city, latitude and longitude, property type, displayed price, star rating, number of reviews, number of bedrooms and bathrooms, guest capacity, amenities count, superhost status, guest-favourite designation and pet policy. The listings span the province, with the largest concentrations in Montreal, Quebec City, Mont-Tremblant, Gatineau and resort communities in the Laurentians, Charlevoix and the Eastern Townships. Property types are *Rental unit* (the most common), *Cabin/Chalet*, *House*, *Other*, *Condo* and *Apartment*.

**Cleaning.** We standardise city names by merging accented and unaccented variants and consolidating aliases (Appendix A). The price field is missing for 1,544 listings (30.9%), which we drop, leaving 3,456 listings; prices are then winsorised at the 1st and 99th percentiles. Missing ratings (16%) are imputed with the sample median, on the assumption that unrated listings are recent and of average quality, and missing review counts are set to zero. An indicator `is_entire_home` identifies the property types House, Cabin/Chalet and Condo, which are typically let in their entirety; the platform's room-type field was not captured by the scraper, so this is a proxy. Superhosts account for 53% of the cleaned listings and entire homes for 47% (Table 2).

**A caveat on the price field.** The distribution of the scraped price is bimodal (Figure 4). The median is $220 per night and 71% of listings are priced below $600, but a second mode lies between roughly $1,400 and $3,300 and 26.5% of listings carry a "nightly" price above $1,000, a category that the scraper itself labelled "Ultra-Luxury." Prices of this order for one night are implausible for most of the listings concerned, which include ordinary rental units in Montreal, and the pattern most plausibly reflects total-stay or multi-night prices captured for part of the sample. We cannot recover the intended nightly price from the available fields. The price variable therefore enters the analysis in two ways. For the Airbnb pricing model (Model 2) we report estimates on the full sample and on the subsample priced at or below $1,000 per night (1,699 listings with a matched city), and we treat the latter as the more credible. For the rent regressions the price field plays no role in the main

exposure measure, which is a count; it affects only the mean nearby Airbnb price used as a secondary exposure variable and as a machine-learning feature, which we interpret with corresponding caution.

### 3.2. Residential Rental Listings

The rental data consist of 8,356 listings scraped from Realtor.ca, Canada's main real-estate listing platform, for the province of Quebec. Each record contains the address, latitude and longitude, property type, lease rent, building type, number of bedrooms and bathrooms, interior size, number of storeys and postal code. After restricting the sample to single-family residential listings with a monthly rental period and a positive parsed rent, 8,303 listings remain, all with valid coordinates. Building types are *Apartment* ($n$ = 7,513), *House* ($n$ = 742) and *Row/Townhouse* ($n$ = 48). The geographic distribution is concentrated in the Montreal metropolitan area (Montreal proper accounts for 59% of listings), with clusters in Laval, Longueuil, Gatineau, Trois-Rivières, Quebec City and Sherbrooke; 253 municipalities are represented.

**Cleaning.** Monthly rents are extracted from the lease-rent string and winsorised at the 1st and 99th percentiles; the log of monthly rent is the dependent variable of the hedonic rent model. Bedroom and bathroom counts are converted to numeric values; bedrooms are missing for 378 listings, which are excluded from regressions that use them, so that the baseline estimation sample comprises 7,925 listings. Interior size is reported for 3,699 listings and is used only in robustness checks; it contains implausible extreme values at both ends (Table 3), which we do not correct. The postal code is complete and defines the forward sortation area (FSA, its first three characters), a Canada Post delivery zone that we use as a neighbourhood identifier: the sample spans 333 FSAs, 180 of which contain at least ten listings and together account for 93% of the sample; Montreal alone has 103.

### 3.3. Spatial Merge Strategy

The two data sets share no common identifier, so we link them through their coordinates. For each rental listing $i$ we compute the Haversine (great-circle) distance $d(i, j)$ to every Airbnb listing $j$ (Appendix B.1) and count the Airbnb listings within four concentric radii, $r \in \{250\,\text{m}, 500\,\text{m}, 1\,\text{km}, 2\,\text{km}\}$:

$$A_{i,r} = \sum_{j=1}^{N_A} \mathbf{1}\left[d(i, j) \le r\right], \tag{1}$$

where $N_A$ = 3,456 and $\mathbf{1}[\cdot]$ is the indicator function. Within each radius we also construct the Airbnb density $A_{i,r}/(\pi r^2)$ (listings per km$^2$), the mean nightly price, the share of entire-home listings, the mean rating and the superhost share; the composition variables are undefined when $A_{i,r} = 0$.

Two properties of this construction matter for the econometrics. First, the buffers are nested: $A_{i,2\text{km}}$ contains $A_{i,1\text{km}}$, which contains $A_{i,500\text{m}}$, and so on, so coefficients at different radii are not independent pieces of evidence and a ring decomposition (Section 6) is needed to separate near from far exposure. Second, the buffers of neighbouring rental listings overlap, so that the same Airbnb listing enters the exposure of many rental listings. This is not double counting in the regression, each rental listing's exposure being correctly measured, but it makes the regressor a spatial moving average of Airbnb locations and induces strong spatial correlation in both the regressor and, plausibly, the error. Section 4.9 addresses the consequences for inference.

Table 1 summarises the exposure counts. The distribution is highly skewed and zero-inflated: at 500 m the mean count is 7.9 and the standard deviation 14.2, but the median is 1 and 44% of rental listings have no Airbnb listing within 500 m (61% at 250 m, 27% at 1 km, 13% at 2 km). The extensive margin, having any Airbnb nearby, is therefore a substantial part of the variation in exposure, which motivates the extensive/intensive decomposition and the logarithmic specification in Sections 5 and 6. Within FSAs, the standard deviation of the 500 m count is 7.2, and about 26% of its total variance is within-FSA variation, so neighbourhood fixed effects leave meaningful identifying variation.

**Table 1.** Airbnb Exposure of Rental Listings by Buffer Radius

| Radius | N | Mean | SD | Median | P90 | Max | Share zero |
|---|---|---|---|---|---|---|---|
| 250 m | 8,303 | 2.41 | 4.94 | 0 | 9 | 33 | 0.612 |
| 500 m | 8,303 | 7.92 | 14.20 | 1 | 30 | 84 | 0.441 |
| 1 km | 8,303 | 26.73 | 41.96 | 4 | 103 | 206 | 0.274 |
| 2 km | 8,303 | 81.45 | 118.90 | 16 | 281 | 438 | 0.130 |

Number of Airbnb listings within each radius of a rental listing (Haversine distance), over all rental listings with valid coordinates. "Share zero" is the fraction of rental listings with no Airbnb listing within the radius.

### 3.4. City-Level Aggregation

As a complementary linkage, we aggregate the Airbnb data to the city level and merge the total Airbnb count, the mean nightly price and the entire-home share onto the rental data by standardised city name. Of the 253 municipalities in the rental data, 153 have at least one Airbnb listing and enter the city-level regressions of Section 5.7; for the Airbnb pricing model, 2,437 Airbnb listings are located in a city with rental listings and thus have a city mean rent. This linkage discards the within-city spatial variation that drives the buffer-based analysis and is used only as a descriptive benchmark.

### 3.5. Summary Statistics

Tables 2 and 3 report summary statistics for the two cleaned data sets. The median monthly rent is $1,950 (mean $2,142), with an interquartile range of $1,615 to $2,495 after winsorisation; the typical listing has two bedrooms and one bathroom. The Airbnb listings are heterogeneous, from budget rooms to large chalets sleeping up to sixteen guests; the mean nightly price of $1,233 reflects the bimodality discussed above and should not be read as a typical price. Appendix Table A3 reports the pairwise correlations between log rent, the 500 m exposure measures and the dwelling controls. The raw correlation between log rent and the 500 m count is 0.14; the count is negatively correlated with bedrooms (−0.19), reflecting the smaller units found in dense, Airbnb-rich neighbourhoods, a pattern that explains why the Airbnb coefficient rises when dwelling controls are added (Section 5.1).

**Table 2.** Summary Statistics — Airbnb Listings

| Variable | N | Mean | SD | Min | P25 | Median | P75 | Max |
|---|---|---|---|---|---|---|---|---|
| Nightly price (CAD) | 3456 | 1233.35 | 2624.71 | 30.96 | 125.00 | 219.82 | 1211.76 | 18205.04 |
| Rating | 3456 | 4.80 | 0.22 | 2.25 | 4.75 | 4.86 | 4.94 | 5.00 |
| Bedrooms | 3456 | 2.12 | 1.56 | 0.00 | 1.00 | 2.00 | 3.00 | 25.00 |
| Bathrooms | 3456 | 1.30 | 1.09 | 0.00 | 1.00 | 1.00 | 2.00 | 24.00 |
| Max guests | 3456 | 5.40 | 3.41 | 1.00 | 2.00 | 4.00 | 7.00 | 16.00 |
| Amenities count | 3456 | 45.78 | 15.74 | 7.00 | 35.00 | 46.00 | 57.00 | 109.00 |
| Number of images | 3456 | 28.20 | 19.17 | 0.00 | 15.00 | 24.00 | 36.00 | 199.00 |
| Number of reviews | 3456 | 53.40 | 74.80 | 0.00 | 6.00 | 26.00 | 71.00 | 986.00 |
| Quality score | 3456 | 96.82 | 38.79 | 0.00 | 95.60 | 108.40 | 121.10 | 139.80 |
| Superhost (share) | 3456 | 0.53 | 0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| Entire home (share) | 3456 | 0.47 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |

Cleaned sample of 3,456 listings with a non-missing price (winsorised at the 1st and 99th percentiles). Ratings are median-imputed for 16% of listings. The nightly-price distribution is bimodal (Figure 4); see Section 3.1.

**Table 3.** Summary Statistics — Residential Rental Listings

| Variable | N | Mean | SD | Min | P25 | Median | P75 | Max |
|---|---|---|---|---|---|---|---|---|
| Monthly rent (CAD) | 8303 | 2141.81 | 764.87 | 995.08 | 1615.00 | 1950.00 | 2495.00 | 5000.00 |
| Bedrooms | 7925 | 2.05 | 0.94 | 1.00 | 1.00 | 2.00 | 3.00 | 10.00 |
| Bathrooms | 8303 | 1.22 | 0.47 | 0.00 | 1.00 | 1.00 | 1.00 | 8.00 |
| Interior size (sq ft) | 3699 | 955.81 | 706.43 | 1.05 | 643.00 | 880.00 | 1099.00 | 22029.00 |

Cleaned sample of 8,303 monthly rental listings (rent winsorised at the 1st and 99th percentiles). Bedrooms are missing for 378 listings; interior size is reported for 3,699 listings and contains implausible extreme values that are not corrected.

### 3.6. Geographic Distribution

Figures 1 and 2 display the geographic distribution of the two samples. Their overlap is concentrated in the Montreal metropolitan area and, to a lesser extent, in Quebec City and Gatineau; many Airbnb

listings are in resort areas with few or no rental listings, and the rental sample in turn covers suburban municipalities with little STR activity. Figure 3 shows the distributions of Airbnb nightly prices and monthly rents in levels, and Figure 4 the nightly-price distribution on a logarithmic scale, where the two modes are apparent.

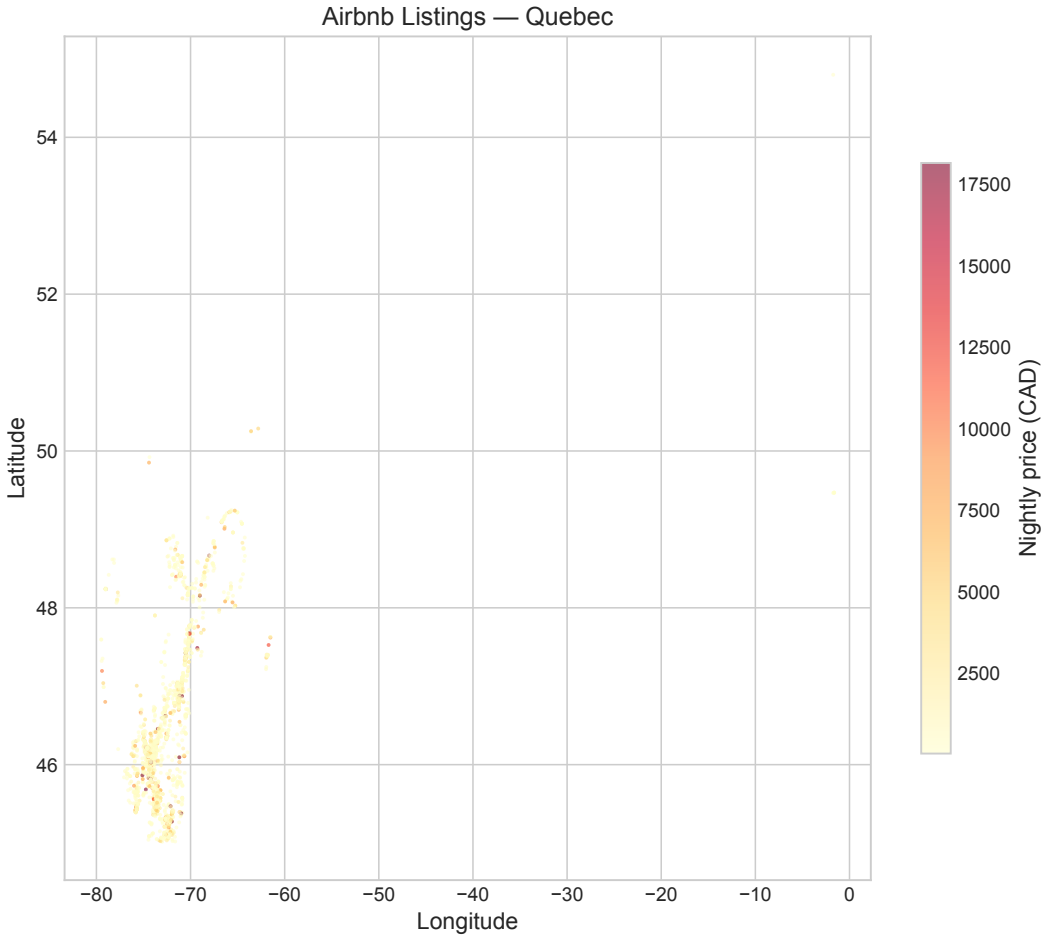



**Figure 1.** Geographic Distribution of Airbnb Listings in Quebec

### 3.7. Data Limitations

Several limitations should be stated at the outset. First, the data are a single cross-section: we observe neither the entry and exit of Airbnb listings nor the evolution of rents, and we cannot control for time-varying confounders. Second, we do not observe host identifiers, which precludes multi-listing indicators distinguishing commercial from casual hosts, a distinction that matters in the Canadian debate (Combs et al., 2020). Third, the Airbnb data are scraped from the public platform and omit listings that were inactive, calendar-blocked or filtered out at the time of collection; the price field is affected by the measurement problem described above; and the room type is not observed. The collection took place in winter, when the composition of active listings (resort chalets

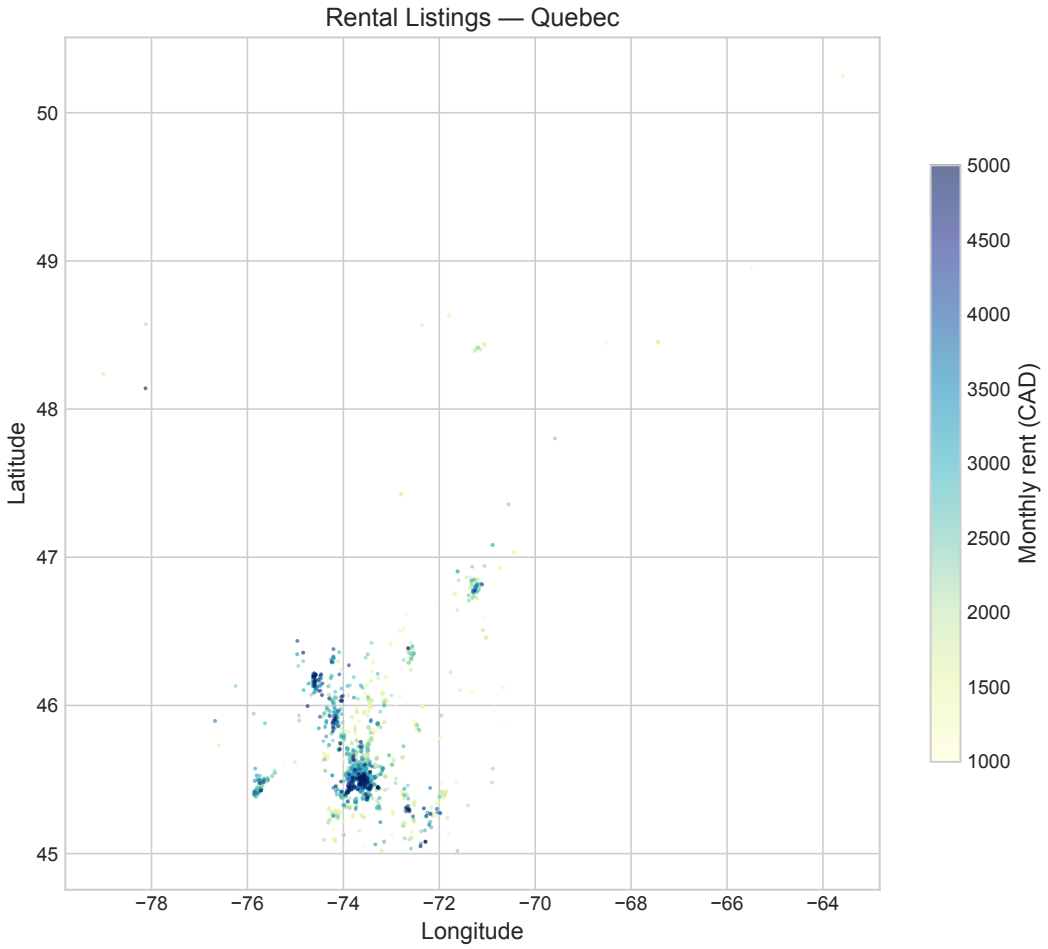
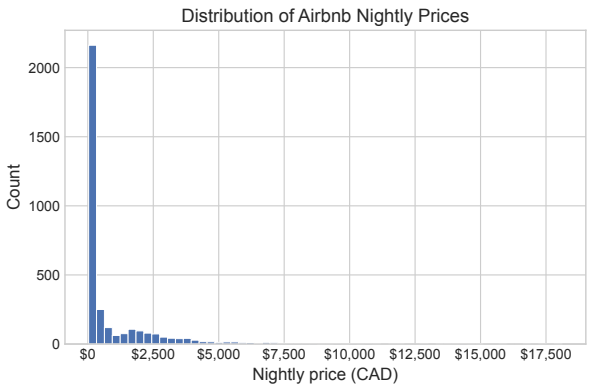
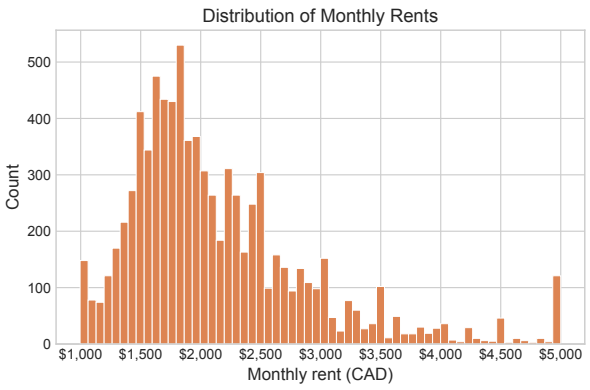



**Figure 2.** Geographic Distribution of Residential Rental Listings in Quebec




**(a)** Airbnb nightly price (CAD, trimmed at the 99th percentile)




**(b)** Monthly rent (CAD, trimmed at the 99th percentile)

**Figure 3.** Distributions of Airbnb Nightly Prices and Monthly Rents

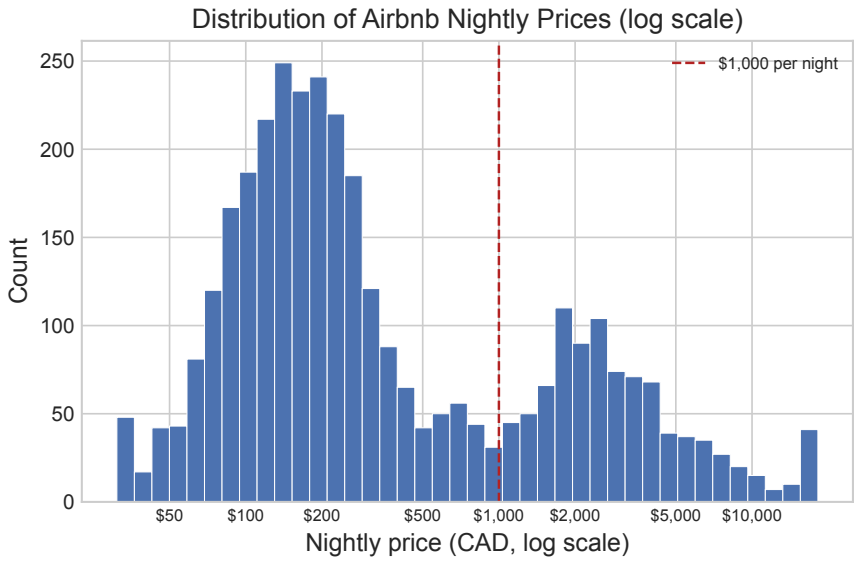



**Figure 4.** Distribution of Scraped Airbnb Nightly Prices on a Logarithmic Scale

The dashed line marks $1,000 per night. The second mode above this threshold (26.5% of listings) most plausibly reflects total-stay prices captured by the scraper; see Section 3.1.

versus urban apartments) differs from the summer peak and when Montreal's seasonal rules for primary-residence STRs were not in their permitted window, so the Montreal listings observed are those operating in commercial zones or outside the rules. Fourth, the rental data are *asking* rents for units advertised on Realtor.ca rather than contract rents for the occupied stock, and Realtor.ca covers the agent-listed segment of the market, which over-represents newer and higher-priced units relative to the stock as a whole; the coverage of both platforms is also denser in urban areas, which limits generalisation to peripheral markets. Fifth, because the composition measures are undefined for rental listings with no Airbnb within the radius, specifications that use them (Model 1e, the machine-learning benchmark) are estimated on the 4,370 listings with positive 500 m exposure, a selected subsample. We return to these limitations in Section 7.

## 4. Methodology

This section presents the econometric models, the fixed-effects strategy, the treatment of spatial dependence in estimation and in inference, and the machine-learning benchmark. Throughout, $i$ indexes rental listings, $j$ Airbnb listings, $c$ cities and $g$ forward sortation areas (FSAs).

### 4.1. Model 1: Hedonic Rent Function with Airbnb Exposure

The baseline specification is a hedonic rent function (Rosen, 1974; Malpezzi, 2003) in which the log of the monthly asking rent is regressed on the Airbnb count within radius $r$ and a vector of dwelling

characteristics:

$$\ln(\text{rent}_i) = \alpha + \beta\, A_{i,r} + \mathbf{X}_i'\boldsymbol{\gamma} + \delta_{c(i)} + \varepsilon_i, \tag{2}$$

where $A_{i,r}$ is defined in Equation (1) (the reference radius is $r = 500$ m); $\mathbf{X}_i$ contains the number of bedrooms, the number of bathrooms and building-type indicators (House and Row/Townhouse, with Apartment as the reference category), with interior size added in a robustness check; $\delta_{c(i)}$ is a fixed effect for the city of listing $i$; and $\varepsilon_i$ is an error term. The coefficient $\beta$ is a semi-elasticity: $100\beta$ is approximately the percentage difference in rent associated with one additional Airbnb listing within the radius, conditional on dwelling characteristics and city, and $100(e^{\beta} - 1)$ is the exact percentage difference. Columns (1d) and (1e) of Table 4 replace $A_{i,r}$ with the Airbnb density, which at a fixed radius is the count divided by $\pi r^2$ and therefore a rescaling, and with the entire-home share within the buffer, which is defined only when $A_{i,r} > 0$.

City fixed effects absorb all city-level determinants of rent, including average neighbourhood quality, labour-market conditions, housing-market tightness, tourism infrastructure and municipal regulation. The identifying variation in Equation (2) is therefore the within-city, between-neighbourhood and between-block variation in Airbnb density.

### 4.2. Neighbourhood Fixed Effects

Within a city, Airbnb listings concentrate in neighbourhoods with tourist attractions, restaurants, transit and architectural character, which also command higher rents. City fixed effects do not absorb these neighbourhood-level unobservables. Following the specification guidance of Kuminoff et al. (2010) and Bishop et al. (2020), we therefore also estimate

$$\ln(\text{rent}_i) = \alpha + \beta^F A_{i,r} + \mathbf{X}_i'\boldsymbol{\gamma}^F + \theta_{g(i)} + \varepsilon_i, \tag{3}$$

where $\theta_{g(i)}$ is a fixed effect for the FSA of listing $i$. FSAs are nested within cities in practice (the first character identifies the region, the second the urban or rural character and the third a delivery zone), so Equation (3) compares rental listings within the same postal-code neighbourhood that differ in the number of Airbnb listings within 500 m. The fixed effects are absorbed by the within transformation, and FSAs with a single listing are dropped. Comparing $\hat{\beta}$ and $\hat{\beta}^F$ is a direct test of how much of the city-level association is driven by between-neighbourhood sorting. In the spirit of Oster (2019), we also report how the coefficient moves from the bivariate regression to the specifications with dwelling controls, city fixed effects and FSA fixed effects; a coefficient that is stable across control sets is more credible than one that moves substantially, and the direction of movement indicates the direction of the omitted-variable bias that the added controls remove.

Within Equation (3) we also decompose exposure into an extensive and an intensive margin, replacing $\beta^F A_{i,r}$ with $\beta_0^F \mathbf{1}[A_{i,r} > 0] + \beta_1^F A_{i,r}$, and we use the concave transformation $\ln(1 + A_{i,r})$,

under which the coefficient is approximately the percentage change in rent associated with a one-percent increase in $1 + A_{i,r}$. These forms address the zero inflation documented in Table 1.

### 4.3. Model 2: Hedonic Airbnb Pricing Model

To examine whether STR pricing responds to local long-term rents, we estimate a hedonic model of Airbnb nightly prices:

$$\ln(\text{price}_j) = \alpha' + \theta\,\overline{\ln(\text{rent})}_{c(j)} + \mathbf{Z}_j'\boldsymbol{\lambda} + u_j, \tag{4}$$

where $\overline{\ln(\text{rent})}_{c(j)}$ is the mean log monthly rent in the city of listing $j$, computed from the rental data, and $\mathbf{Z}_j$ contains bedrooms, bathrooms, guest capacity, the amenities count, superhost status and the entire-home indicator. Because the two price variables are in logarithms, $\theta$ is an elasticity. A positive $\theta$ would be consistent with an opportunity-cost channel in which hosts facing higher forgone rents set higher nightly prices. When city fixed effects are added, $\overline{\ln(\text{rent})}_{c(j)}$ is constant within city and perfectly collinear with them, so $\theta$ is not identified; the fixed-effects columns therefore omit it and serve to estimate the within-city implicit prices of listing characteristics. Given the measurement problem in the price field (Section 3.1), we estimate Equation (4) on all 2,437 listings with a matched city and on the 1,699 listings priced at or below $1,000 per night.

### 4.4. Model 3: City-Level Regressions

To describe the cross-city co-movement of Airbnb activity and rents, we estimate a forward and a reverse aggregate regression,

$$\overline{\ln(\text{rent})}_c = a_1 + b_1\,A_c + \mathbf{W}_c'\boldsymbol{\phi}_1 + e_{1c}, \tag{5}$$

$$A_c = a_2 + b_2\,\overline{\ln(\text{rent})}_c + \mathbf{W}_c'\boldsymbol{\phi}_2 + e_{2c}, \tag{6}$$

where $A_c$ is the total number of Airbnb listings in city $c$ and $\mathbf{W}_c$ contains the city means of bedrooms and bathrooms. The 153 cities with at least one Airbnb listing enter these regressions; most contribute few underlying listings, and the regressions are descriptive.

### 4.5. Model 4: Spatial Lag and Spatial Error Models

Rents are spatially dependent beyond what observed characteristics and fixed effects capture. We estimate a spatial autoregressive (SAR) model and a spatial error model (SEM) alongside OLS. Let $\mathbf{W}$ be an $n \times n$ row-standardised $k$-nearest-neighbour matrix with $k = 5$: $w_{ik} = 1/5$ if rental listing $k$ is among the five listings closest to $i$ in projected (kilometre) coordinates, excluding $i$ itself, and

$w_{ik} = 0$ otherwise (Appendix B.2). The SAR model is

$$\ln(\text{rent}_i) = \alpha'' + \rho \sum_k w_{ik} \ln(\text{rent}_k) + \beta' A_{i,r} + \mathbf{X}_i' \boldsymbol{\gamma}' + \delta''_{c(i)} + \eta_i, \tag{7}$$

so that $\sum_k w_{ik} \ln(\text{rent}_k)$ is the mean log rent of the five nearest rental listings and $\rho$ measures the strength of spatial co-movement. The SEM places the spatial process in the disturbance, $\eta_i = \lambda \sum_k w_{ik} \eta_k + \nu_i$, capturing spatially correlated unobservables. Because the spatial lag is correlated with $\eta_i$, OLS estimation of Equation (7) is inconsistent; we use the spatial two-stage least squares estimator of Kelejian and Prucha (1998), which instruments the lag with spatially lagged exogenous regressors, with heteroskedasticity-robust standard errors, and the generalised-moments estimator of Kelejian and Prucha (1999) for the SEM (`GM_Lag` and `GM_Error` in `spreg`). To keep the instrument matrix well conditioned, the spatial models include fixed effects for the fifteen largest cities with the remaining cities grouped, and the OLS column of the spatial table uses the same fixed effects.

In a SAR model the coefficient $\beta'$ is not the marginal effect of exposure. Writing Equation (7) in matrix form and solving, $\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta})$, so the matrix of partial derivatives of $\mathbf{y}$ with respect to a regressor $x$ is $(\mathbf{I} - \rho \mathbf{W})^{-1} \beta_x$. Following LeSage and Pace (2009) we report the average direct impact, $n^{-1} \operatorname{tr}[(\mathbf{I} - \rho \mathbf{W})^{-1}]\, \beta_x$, the average total impact, which for a row-standardised $\mathbf{W}$ equals $\beta_x / (1 - \rho)$, and the average indirect (spillover) impact, their difference:

$$\text{Direct}_x = \frac{\beta_x}{n} \operatorname{tr}\left[(\mathbf{I} - \rho \mathbf{W})^{-1}\right], \qquad \text{Total}_x = \frac{\beta_x}{1 - \rho}, \qquad \text{Indirect}_x = \text{Total}_x - \text{Direct}_x. \tag{8}$$

Standard errors for the impacts are obtained by simulation from the estimated asymptotic distribution of $(\hat{\beta}_x, \hat{\rho})$ (Appendix B.2). The spatial models serve two purposes: they absorb residual spatial dependence that would otherwise distort inference, and the comparison of the SEM coefficient with OLS indicates whether the Airbnb coefficient proxies for spatially correlated unobservables at the scale of the five nearest listings. They do not, however, substitute for neighbourhood fixed effects, because a five-neighbour lag operates at the block scale whereas the sorting of STR supply operates at the neighbourhood scale.

### 4.6. Model 5: Quantile Regression

To examine heterogeneity across the conditional rent distribution we estimate quantile regressions (Koenker and Bassett, 1978; Koenker and Hallock, 2001),

$$Q_\tau\left[\ln(\text{rent}_i) \mid A_{i,r}, \mathbf{X}_i, c(i)\right] = \alpha_\tau + \beta_\tau A_{i,r} + \mathbf{X}_i' \boldsymbol{\gamma}_\tau + \delta_{c(i),\tau}, \tag{9}$$

for $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and on a finer grid for the figure. To keep the linear-programming problem well conditioned, the city fixed effects are restricted to the five largest cities with the remainder grouped; the companion OLS column uses the same fixed effects. The coefficient $\beta_\tau$ is the marginal association between exposure and the $\tau$th conditional quantile of log rent. Standard errors are the kernel-based asymptotic estimates implemented in `statsmodels` (Appendix B).

### 4.7. Model 6: Machine-Learning Benchmark

We complement the parametric analysis with the LASSO (Tibshirani, 1996), the elastic net (Zou and Hastie, 2005), random forests (Breiman, 2001) and gradient boosting (Friedman, 2001), alongside OLS, all implemented in scikit-learn (Pedregosa et al., 2011). The models predict $\ln(\text{rent}_i)$ from the 500 m exposure measures (count, density, mean nearby price, entire-home share), bedrooms, bathrooms and the geographic coordinates; because two of the exposure measures are undefined for listings without nearby Airbnb, the sample is the 4,370 listings with positive 500 m exposure. Performance is evaluated in two ways. First, on a random 80/20 hold-out split, as in the original design. Second, by five-fold cross-validation under two fold designs: random folds, and spatially blocked folds in which whole FSAs are assigned to folds so that no test listing shares a neighbourhood with a training listing. The difference between the two designs measures how much of a learner's apparent accuracy comes from interpolating the rent levels of nearby training listings rather than from generalisable structure (Mullainathan and Spiess, 2017). SHAP values (Lundberg and Lee, 2017) computed on the hold-out set describe the contribution of each feature to the gradient-boosting predictions.

### 4.8. Identification

We are explicit about what the cross-sectional design can identify. Three channels make Airbnb exposure endogenous in Equation (2). Reverse causality: high rents may attract hosts, because the opportunity cost of a vacant unit is high and STR income can offset carrying costs. Omitted neighbourhood amenities: proximity to attractions, restaurants and transit raises both tourist demand and residential rents. Spatial sorting: hosts may select locations on unobserved characteristics, such as architectural character, that also raise rents. All three channels bias $\hat{\beta}$ upward. The quasi-experimental literature addresses them with regulatory shocks or instruments (Koster et al., 2021; Barron et al., 2021; Garcia-López et al., 2020; Duso et al., 2024); neither is available in a single cross-section of listings.

Our strategy is instead to make the identification problem measurable. City fixed effects remove city-level confounders; FSA fixed effects additionally remove neighbourhood-level confounders that are constant within a postal-code zone, leaving within-neighbourhood variation that is closer to,

though still not, the variation exploited by quasi-experimental designs. The remaining threat under FSA fixed effects is block-level sorting within neighbourhoods, which the spatial error model partly addresses and which we cannot rule out. We therefore interpret the city-fixed-effects estimate as an upper bound on any causal effect, the FSA-fixed-effects estimate as the closest the data come to a within-neighbourhood comparison, and both as conditional associations rather than causal effects.

### 4.9. Standard Errors

The exposure variable is a spatial moving average and the error term is spatially correlated (Table 7 reports $\hat{\lambda} \approx 0.54$), so the independence assumption underlying heteroskedasticity-robust standard errors fails. We report four alternatives for the key coefficient. HC1 standard errors are the baseline, for comparability with the original analysis and with most of the cross-sectional literature. Cluster-robust standard errors by city (Cameron and Miller, 2015) allow arbitrary dependence within cities, but Montreal contains 59% of the sample, and cluster-robust inference is unreliable when one cluster dominates; we therefore also cluster by FSA, which yields several hundred more balanced clusters. Finally, the Conley (1999) spatial HAC estimator weights the cross-products of scores for pairs of listings by a Bartlett kernel $K(d_{ik}) = \max\{0,\ 1 - d_{ik}/\bar{d}\}$ of their great-circle distance, with cutoffs $\bar{d}$ of 1, 2, 5 and 10 km (Kelejian and Prucha, 2007); it imposes no partition and lets dependence decline smoothly with distance. For the quantile regressions we report the asymptotic kernel-based standard errors, and for the machine-learning models held-out performance with regularisation parameters chosen by five-fold cross-validation.

## 5. Results

This section presents the estimates in the order of the identification argument: the baseline hedonic rent model with city fixed effects, the neighbourhood-fixed-effects estimates, the comparison of standard errors, the spatial models and their impacts, the quantile regressions, the Airbnb pricing and city-level models, and the machine-learning benchmark. Throughout, we describe estimates as conditional associations; the terms "effect" and "impact" are used only in their technical spatial-econometric sense.

### 5.1. Baseline Hedonic Rent Model

Table 4 reports OLS estimates of Equation (2). Column (1a) is the bivariate regression of log rent on the 500 m Airbnb count; column (1b) adds bedrooms, bathrooms and building type; column (1c), the reference specification, adds city fixed effects; columns (1d) and (1e) replace the count with the Airbnb density and with the entire-home share within the buffer.

The coefficient on the 500 m count is positive and statistically significant in columns (1a)–(1d). In the reference specification each additional Airbnb listing within 500 m is associated with a monthly

**Table 4.** Hedonic Rent Model — Baseline OLS Results

| Dep. var: | **(1a)** | **(1b)** | **(1c)** *log rent* | **(1d)** | **(1e)** |
|---|---|---|---|---|---|
| Constant | 7.5990*** | 7.0390*** | 7.2380*** | 7.2380*** | 7.2766*** |
| | (0.0040) | (0.0139) | (0.0677) | (0.0677) | (0.0332) |
| Airbnb count (500 m) | 0.0020*** | 0.0048*** | 0.0039*** | | |
| | (0.0003) | (0.0002) | (0.0002) | | |
| Airbnb density (500 m, per km$^2$) | | | | 0.0030*** | |
| | | | | (0.0002) | |
| Entire-home share (500 m) | | | | | -0.0041 |
| | | | | | (0.0107) |
| Bedrooms | | 0.1118*** | 0.1276*** | 0.1276*** | 0.1215*** |
| | | (0.0046) | (0.0046) | (0.0046) | (0.0057) |
| Bathrooms | | 0.2604*** | 0.2287*** | 0.2287*** | 0.3033*** |
| | | (0.0128) | (0.0126) | (0.0126) | (0.0104) |
| House | | 0.0880*** | 0.0825*** | 0.0825*** | 0.0305 |
| | | (0.0128) | (0.0137) | (0.0137) | (0.0260) |
| Row/townhouse | | 0.1493*** | 0.0981*** | 0.0981*** | 0.1187* |
| | | (0.0375) | (0.0366) | (0.0366) | (0.0695) |
| Observations | 8,303 | 7,925 | 7,925 | 7,925 | 4,370 |
| $R^2$ | 0.0075 | 0.4705 | 0.5697 | 0.5697 | 0.4961 |
| Adj. $R^2$ | 0.0074 | 0.4702 | 0.5554 | 0.5554 | 0.4857 |

Models (1c)-(1e) include city fixed effects (not shown).

Robust (HC1) standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

rent approximately 0.39% higher ($\hat{\beta}$ = 0.0039, HC1 standard error 0.0002), holding dwelling characteristics and city constant; at the median rent of $1,950 this is about $8 per month per listing, and one standard deviation of the count (14.2 listings) corresponds to about 5.5%. The coefficient moves substantially across control sets: it is 0.0020 without controls, 0.0048 with dwelling controls and 0.0039 with city fixed effects. The rise from (1a) to (1b) reflects the negative correlation between Airbnb density and unit size documented in Appendix Table A3: dense, Airbnb-rich neighbourhoods have smaller units, so the bivariate coefficient understates the conditional association. The fall from (1b) to (1c) reflects the opposite force at the city level, since cities with many Airbnb listings also have higher rents for reasons unrelated to STR activity. Both movements are large relative to the coefficient, which already signals that the estimate is sensitive to the set of location controls, a point Section 5.2 develops.

Among the controls, dwelling size dominates: each additional bedroom is associated with 12–14% higher rent and each additional bathroom with 26–30% higher rent in columns (1b)–(1d) (exact percentages, $100(e^{\hat{\gamma}} - 1)$). Houses rent for about 9% more and row houses for about 10% more than apartments in column (1c). City fixed effects raise the $R^2$ from 0.47 to 0.57, confirming substantial cross-city variation in rent levels; the adjusted $R^2$ of 0.56 leaves a considerable residual, consistent with unobserved unit quality and micro-location. The density coefficient in column (1d) is the count coefficient rescaled by the buffer area and carries the same information. The entire-home share in column (1e) is small, negative and insignificant; this column is estimated on the 4,370 listings with at least one Airbnb within 500 m, for which the share is defined, and compares the composition of nearby STR supply conditional on its presence.

### 5.2. Neighbourhood Fixed Effects

Table 5 replaces city fixed effects with FSA fixed effects (Equation 3). Column (F1) re-estimates the reference specification with the fixed effects absorbed and singleton cities dropped, and reproduces the coefficient of 0.0039 on a sample of 7,860 listings in 187 cities. Column (F2) uses fixed effects for 289 FSAs. The $R^2$ rises from 0.56 to 0.63, and the coefficient on the 500 m count falls to 0.0008 (HC1 standard error 0.0004), about one fifth of the city-fixed-effects estimate. At the median rent this is $1.6 per month per listing; one standard deviation of the count corresponds to 1.2%, and one within-FSA standard deviation (7.2 listings) to 0.6%.

The remaining columns show that this attenuation is not specific to the linear 500 m count. The 250 m count, the most local exposure measure and the one with the largest city-fixed-effects coefficient (Section 6.1), has a within-FSA coefficient of 0.0008 with a standard error of 0.0009 (column F3). Decomposing exposure into an extensive and an intensive margin (column F4) yields a negative and significant coefficient on the indicator for any Airbnb within 500 m (−0.028) and a count coefficient of 0.0010: within a neighbourhood, rental listings with at least one Airbnb nearby

**Table 5.** Neighbourhood (FSA) Fixed Effects — Hedonic Rent Model

| | **(F1)** | **(F2)** | **(F3)** | **(F4)** | **(F5)** | **(F6)** |
|---|---|---|---|---|---|---|
| Dep. var: | | | *log rent* | | | |
| Any Airbnb within 500 m (0/1) | | | | -0.0279*** | | |
| | | | | (0.0088) | | |
| Airbnb count (500 m) | 0.0039*** | 0.0008** | | 0.0010*** | | 0.0003 |
| | (0.0002) | (0.0004) | | (0.0004) | | (0.0006) |
| Airbnb count (250 m) | | | 0.0008 | | | |
| | | | (0.0009) | | | |
| log(1 + Airbnb count, 500 m) | | | | | 0.0022 | |
| | | | | | (0.0053) | |
| Bedrooms | 0.1276*** | 0.1501*** | 0.1502*** | 0.1500*** | 0.1502*** | 0.1479*** |
| | (0.0046) | (0.0047) | (0.0047) | (0.0047) | (0.0047) | (0.0083) |
| Bathrooms | 0.2287*** | 0.1980*** | 0.1978*** | 0.1979*** | 0.1978*** | 0.1892*** |
| | (0.0125) | (0.0120) | (0.0120) | (0.0120) | (0.0120) | (0.0200) |
| Interior size (100 sq ft) | | | | | | 0.0040** |
| | | | | | | (0.0016) |
| Fixed effects | City | FSA | FSA | FSA | FSA | FSA |
| FE groups | 187 | 289 | 289 | 289 | 289 | 222 |
| Observations | 7,860 | 7,883 | 7,883 | 7,883 | 7,883 | 3,485 |
| $R^2$ | 0.5636 | 0.6306 | 0.6303 | 0.6313 | 0.6303 | 0.6505 |
| Within $R^2$ | 0.4703 | 0.5189 | 0.5185 | 0.5197 | 0.5184 | 0.5326 |

Heteroskedasticity-robust (HC1) standard errors in parentheses. All columns control for bedrooms, bathrooms and building type; fixed effects are absorbed (within transformation) and singleton groups are dropped. FSA = forward sortation area (first three characters of the postal code). Column (F3) uses the 250 m count; column (F6) adds interior size on the subsample reporting it. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

rent for about 2.8% less than otherwise comparable listings with none, and the intensive-margin association is small. The logarithmic exposure measure (column F5) has a coefficient of 0.002 with a standard error of 0.005, against 0.048 with city fixed effects (Table 17). Adding interior size on the subsample that reports it (column F6) reduces the count coefficient to 0.0003 with a standard error of 0.0006.

Three readings of these results are possible, and the data do not distinguish them. The first is that the city-fixed-effects association is driven by neighbourhood-level sorting: Airbnb supply concentrates in neighbourhoods that command high rents for reasons unrelated to STR activity, and comparing listings within neighbourhoods removes this confound along with most of the association. The second is that STR-induced rent pressure operates at the neighbourhood rather than the block scale, so that fixed effects at the FSA level absorb the effect itself; this reading requires that supply withdrawal within a neighbourhood raise rents uniformly across its blocks, which is possible but leaves the between-neighbourhood estimate exposed to the first reading. The third is attenuation from measurement error, which is more severe within neighbourhoods where the signal-to-noise ratio of the exposure count is lower; given that a quarter of the variance of the count is within FSAs, this can explain part but not all of a fivefold reduction. Under any reading, the city-fixed-effects estimate is an upper bound on the within-neighbourhood association, and it is the within-neighbourhood association that corresponds most closely to the variation exploited by quasi-experimental studies. The negative extensive-margin coefficient in column (F4) is consistent with the first reading: within neighbourhoods, Airbnb listings cluster on blocks with commercial frontage or less residential character, which rent for less.

### 5.3. Inference under Spatial Dependence

Table 6 reports, for the city- and FSA-fixed-effects estimates of the 500 m coefficient, the HC1 standard error, cluster-robust standard errors by city and by FSA, and Conley spatial HAC standard errors at cutoffs of 1, 2, 5 and 10 km.

For the city-fixed-effects estimate the HC1 standard error is 0.0002 and the $t$-statistic 17. Clustering by city reduces the standard error to 0.0001 and raises the $t$-statistic to 36, an implausible tightening that illustrates the unreliability of cluster-robust inference when one cluster (Montreal) contains 59% of the observations (Cameron and Miller, 2015); we do not rely on it. Clustering by FSA and the Conley estimator at every cutoff agree closely: the standard error is 0.0007, three to three and a half times the HC1 value, and the $t$-statistic falls to between 5 and 6. The city-fixed-effects association is therefore statistically robust to spatial dependence, but conventional heteroskedasticity-robust errors overstate its precision by a factor of about three, and the same is likely true of cross-sectional STR–rent estimates elsewhere that report robust standard errors on buffer-based exposure measures. For the FSA-fixed-effects estimate the HC1 $t$-statistic of 2.3 falls

**Table 6.** Alternative Standard Errors for the Airbnb Coefficient

| | **City FE (Model 1c)** $\hat{\beta} = 0.0039$ | | **FSA FE (F2)** $\hat{\beta} = 0.0008$ | |
|---|---|---|---|---|
| Standard errors | SE | $t$ | SE | $t$ |
| Heteroskedasticity-robust (HC1) | (0.0002) | 17.3 | (0.0004) | 2.3 |
| Clustered by city | (0.0001) | 35.7 | (0.0004) | 2.3 |
| Clustered by FSA | (0.0007) | 5.9 | (0.0007) | 1.1 |
| Conley spatial HAC, 1 km cutoff | (0.0007) | 5.6 | (0.0007) | 1.1 |
| Conley spatial HAC, 2 km cutoff | (0.0007) | 5.3 | (0.0007) | 1.2 |
| Conley spatial HAC, 5 km cutoff | (0.0007) | 5.9 | (0.0005) | 1.7 |
| Conley spatial HAC, 10 km cutoff | (0.0007) | 5.4 | (0.0004) | 2.0 |
| City clusters | 187 | | 239 | |
| FSA clusters | 318 | | 289 | |

Coefficient on the Airbnb count within 500 m from the specification in the column (controls: bedrooms, bathrooms, building type; fixed effects absorbed, singleton groups dropped). Each row reports the standard error (in parentheses) and the implied $t$-statistic under a different covariance estimator: heteroskedasticity-robust (HC1); clustered by city and by FSA (Cameron and Miller, 2015); and Conley spatial HAC with a Bartlett kernel declining linearly to zero at the stated great-circle cutoff (Conley, 1999; Kelejian and Prucha, 2007). Listings in the same building (identical coordinates) receive full weight in the Conley estimator.

to 1.1–1.2 with FSA-clustered or short-range Conley standard errors and to 1.7–2.0 at the 5 and 10 km cutoffs. The within-neighbourhood association is thus marginal at best, and we describe it below as not statistically distinguishable from zero under spatial inference.

### 5.4. Spatial Lag and Spatial Error Models

Table 7 reports OLS, SAR and SEM estimates of Equation (7), all with fixed effects for the fifteen largest cities, and Table 8 the direct, indirect and total impacts implied by the SAR estimates (Equation 8).

Spatial dependence is substantial. The spatial autoregressive parameter is $\hat{\rho} = 0.134$ (standard error 0.022) and the spatial error parameter $\hat{\lambda} = 0.544$: the rents of the five nearest listings carry systematic information about a listing's rent after controlling for its characteristics and city, and the disturbances are strongly spatially correlated, which is the dependence that the Conley standard errors above account for. The coefficient on the 500 m count is 0.0034 in the SAR model and 0.0039 in the SEM, against 0.0038 by OLS on the same sample. The SAR coefficient is not comparable to the OLS coefficient without the spatial multiplier: the average direct impact is 0.0034, the indirect impact 0.0005 and the total impact 0.0040 (standard error 0.0002), so that once feedback through neighbouring rents is included the SAR model implies the same marginal association as OLS. With a five-nearest-neighbour weights matrix the multiplier is modest ($n^{-1}\mathrm{tr}[(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}] = 1.003$, $1/(1 - \hat{\rho}) = 1.155$), so direct impacts nearly coincide with coefficients and indirect impacts are about 13% of the total. The SEM estimate is essentially the OLS estimate with a larger standard error (0.0004), which is the expected consequence of modelling the error correlation.

**Table 7.** Spatial Regression Models

| | **OLS** | **SAR (GM_Lag)** | **SEM (GM_Error)** |
|---|---|---|---|
| Dep. var: | | *log rent* | |
| Constant | 7.1114*** | 6.0958*** | 7.0918*** |
| | (0.0211) | (0.1653) | (0.0331) |
| Airbnb count (500 m) | 0.0038*** | 0.0034*** | 0.0039*** |
| | (0.0002) | (0.0002) | (0.0004) |
| Bedrooms | 0.1194*** | 0.1222*** | 0.1549*** |
| | (0.0045) | (0.0044) | (0.0030) |
| Bathrooms | 0.2432*** | 0.2286*** | 0.1797*** |
| | (0.0123) | (0.0127) | (0.0057) |
| House | 0.1342*** | 0.1182*** | 0.1261*** |
| | (0.0131) | (0.0124) | (0.0097) |
| Row/townhouse | 0.1782*** | 0.1615*** | 0.1533*** |
| | (0.0359) | (0.0333) | (0.0295) |
| $W$ ln(rent) | | 0.1340*** | |
| | | (0.0222) | |
| $\lambda$ (spatial error) | | | 0.5437 |
| Observations | 7925 | 7925 | 7925 |
| $R^2$ / pseudo-$R^2$ | 0.5081 | 0.5508 | 0.4973 |

Standard errors in parentheses: HC1 (OLS), White-robust (SAR, spatial two-stage least squares, `GM_Lag`), and the classical GM standard errors of `GM_Error` (SEM; no standard error is reported for $\lambda$). $W$ is a row-standardised $k = 5$ nearest-neighbour matrix on projected coordinates. All columns include building-type dummies and fixed effects for the 15 largest cities (remaining cities grouped), so the OLS column differs slightly from column (1c) of Table 4. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

**Table 8.** Average Direct, Indirect and Total Impacts in the SAR Model

| | **Direct** | **Indirect** | **Total** |
|---|---|---|---|
| Airbnb count (500 m) | 0.0034*** | 0.0005*** | 0.0040*** |
| | (0.0002) | (0.0001) | (0.0002) |
| Bedrooms | 0.1225*** | 0.0186*** | 0.1411*** |
| | (0.0043) | (0.0037) | (0.0065) |
| Bathrooms | 0.2291*** | 0.0348*** | 0.2639*** |
| | (0.0125) | (0.0060) | (0.0127) |
| $\hat{\rho}$ | | 0.1340 | |
| $n^{-1}\,\mathrm{tr}[(I - \hat{\rho}W)^{-1}]$ | | 1.0026 | |
| Observations | | 7,925 | |

Average direct, indirect and total impacts of the SAR model in Table 7 (LeSage and Pace, 2009): total = $\hat{\beta}/(1-\hat{\rho})$, direct = $\hat{\beta}\, n^{-1}\mathrm{tr}[(I - \hat{\rho}W)^{-1}]$, indirect = total − direct. Standard errors in parentheses from 2,000 draws of $(\hat{\beta}, \hat{\rho})$ from their estimated asymptotic normal distribution. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

These results qualify, rather than contradict, the neighbourhood-fixed-effects findings. Spatial lag and error processes defined on the five nearest listings operate at the scale of a block, and conditioning on them leaves the Airbnb coefficient intact; fixed effects defined at the scale of a postal-code neighbourhood remove most of it. The unobservables correlated with Airbnb exposure therefore vary across neighbourhoods more than across adjacent blocks, which is what one expects of tourist amenities and neighbourhood reputation.

### 5.5. Quantile Regression Results

Table 9 and Figure 5 report the quantile regression estimates of $\beta_\tau$ (Equation 9) with fixed effects for the five largest cities.

**Table 9.** Quantile Regression Results — Coefficient on the Airbnb Count (500 m)

| Dep. var: | $\tau = 0.10$ *log rent* | $\tau = 0.25$ *log rent* | $\tau = 0.50$ *log rent* | $\tau = 0.75$ *log rent* | $\tau = 0.90$ *log rent* | **OLS** *log rent* |
|---|---|---|---|---|---|---|
| Constant | 6.8249*** | 6.9035*** | 7.0339*** | 7.1237*** | 7.2461*** | 7.0508*** |
| | (0.0258) | (0.0211) | (0.0209) | (0.0221) | (0.0328) | (0.0197) |
| Airbnb count (500 m) | 0.0037*** | 0.0034*** | 0.0035*** | 0.0041*** | 0.0047*** | 0.0038*** |
| | (0.0003) | (0.0003) | (0.0002) | (0.0002) | (0.0003) | (0.0002) |
| Bedrooms | 0.1236*** | 0.1034*** | 0.1071*** | 0.1360*** | 0.1418*** | 0.1152*** |
| | (0.0058) | (0.0043) | (0.0039) | (0.0038) | (0.0055) | (0.0046) |
| Bathrooms | 0.2411*** | 0.2796*** | 0.2748*** | 0.2669*** | 0.2807*** | 0.2571*** |
| | (0.0103) | (0.0076) | (0.0072) | (0.0076) | (0.0115) | (0.0125) |
| House | 0.0608*** | 0.1079*** | 0.1382*** | 0.1079*** | 0.1094*** | 0.1201*** |
| | (0.0139) | (0.0119) | (0.0120) | (0.0131) | (0.0205) | (0.0131) |
| Row/townhouse | 0.1310*** | 0.0858** | 0.1602*** | 0.2501*** | 0.2274*** | 0.1753*** |
| | (0.0468) | (0.0383) | (0.0383) | (0.0417) | (0.0625) | (0.0378) |
| City FE (top 5) | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 7925 | 7925 | 7925 | 7925 | 7925 | 7925 |
| (Pseudo-)$R^2$ | 0.2251 | 0.2529 | 0.2976 | 0.3294 | 0.3402 | 0.4849 |

Standard errors in parentheses. OLS uses HC1 robust SE. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. City FE limited to the 5 largest cities (others grouped).

The coefficient is positive at every quantile. It is roughly flat over the lower half of the distribution (0.0037 at $\tau = 0.10$, 0.0034–0.0035 at $\tau = 0.25$ and $\tau = 0.50$) and rises in the upper tail to 0.0041 at $\tau = 0.75$ and 0.0047 at $\tau = 0.90$, about 25% above the OLS estimate of 0.0038 on the same specification. The standard errors (0.0002–0.0003) are of the kernel-based asymptotic type and, like HC1, do not account for spatial dependence; the differences between quantiles are modest relative to the sampling uncertainty that Table 6 suggests, and we have not tested them formally. Taken at face value, the pattern indicates that the association is strongest in the upper segment of the rental

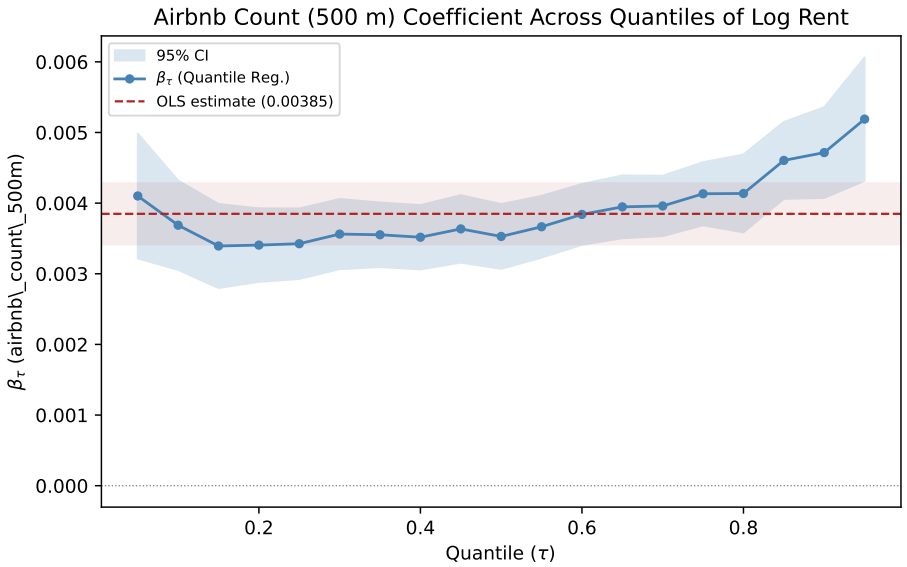



**Figure 5.** Quantile Regression Coefficients on the Airbnb Count (500 m) with 95% Confidence Intervals

market, where listings are located in the desirable neighbourhoods in which STR activity is also most intense (Zietz et al., 2008; McMillen, 2008). Because these regressions condition only on coarse city groups, the quantile pattern inherits the between-neighbourhood character of the mean estimate; quantile regressions with FSA fixed effects would require penalised fixed-effects methods given the number of groups and are left for future work.

### 5.6. Hedonic Airbnb Pricing Model

Table 10 reports the Airbnb pricing model (Equation 4). Columns (2a)–(2c) use the 2,437 listings with a matched city; columns (2d)–(2e) restrict the sample to the 1,699 listings priced at or below $1,000 per night (Section 3.1). Columns (2c) and (2e) include city fixed effects, which absorb the city mean rent.

The city mean log rent does not enter significantly in any specification: the elasticity $\hat{\theta}$ is −0.03 and −0.05 in columns (2a) and (2b), with standard errors above 0.11, and 0.003 (standard error 0.05) in the restricted sample. Airbnb nightly prices are not systematically higher in higher-rent cities once listing characteristics are taken into account, so the data offer no support for a simple opportunity-cost pricing channel at the city level. Part of the explanation is composition: the cross-city variation in Airbnb prices is dominated by resort municipalities with high nightly prices and few or inexpensive long-term rentals.

The listing-level coefficients depend on the sample in a way that illustrates the price-field problem. On the full sample, explanatory power is very low ($R^2$ of 0.03 without and 0.14 with city fixed effects), guest capacity enters negatively, the amenities count is insignificant, and superhost status

**Table 10.** Hedonic Airbnb Pricing Model — OLS Results

| Dep. var: | **(2a)** | **(2b)** | **(2c)** | **(2d)** | **(2e)** |
|---|---|---|---|---|---|
| | *log nightly price* | | | | |
| Constant | 6.1958*** | 6.3698*** | 5.0498*** | 4.4077*** | 4.3390*** |
| | (0.8696) | (0.8889) | (0.2040) | (0.3910) | (0.1535) |
| City mean log rent | -0.0296 | -0.0531 | | 0.0029 | |
| | (0.1129) | (0.1180) | | (0.0520) | |
| Bedrooms | | 0.0923** | 0.0852* | 0.1404*** | 0.1512*** |
| | | (0.0448) | (0.0457) | (0.0271) | (0.0283) |
| Bathrooms | | 0.1555*** | 0.1226*** | 0.1654*** | 0.1494*** |
| | | (0.0428) | (0.0458) | (0.0325) | (0.0359) |
| Guest capacity | | -0.0409** | -0.0187 | -0.0005 | -0.0014 |
| | | (0.0182) | (0.0190) | (0.0107) | (0.0115) |
| Amenities count | | 0.0008 | 0.0019 | 0.0032*** | 0.0035*** |
| | | (0.0022) | (0.0024) | (0.0012) | (0.0013) |
| Superhost (0/1) | | -0.3467*** | -0.2766*** | 0.0141 | 0.0420 |
| | | (0.0621) | (0.0651) | (0.0336) | (0.0358) |
| Entire home (0/1) | | -0.0722 | 0.1670** | 0.0396 | 0.1368*** |
| | | (0.0661) | (0.0749) | (0.0374) | (0.0464) |
| Observations | 2,437 | 2,437 | 2,437 | 1,699 | 1,699 |
| $R^2$ | 0.0000 | 0.0300 | 0.1404 | 0.2055 | 0.2981 |
| Adj. $R^2$ | -0.0004 | 0.0272 | 0.0804 | 0.2022 | 0.2340 |

Columns (2c) and (2e) include city fixed effects (not shown), which absorb the city mean log rent. Columns (2a)–(2c) use all listings with a matched city; columns (2d)–(2e) restrict the sample to listings with a nightly price of at most $1,000 (see Section 3).

Robust (HC1) standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

carries a large negative coefficient (−0.35 and −0.28), a discount of about 25–30% that contradicts the positive reputation premia found by Wang and Nicolau (2017), Ert et al. (2016) and Gibbs et al. (2018). On the restricted sample the picture is conventional: the $R^2$ rises to 0.21 and 0.30, bedrooms and bathrooms carry premia of about 15–16% each, each additional amenity adds about 0.3%, the entire-home indicator carries a 14.7% premium within cities ($e^{0.137} - 1$), and the superhost coefficient is small, positive and insignificant. The superhost discount in the full sample is therefore an artefact of the bimodal price field, in which urban rental units operated by superhosts are more often recorded with plausible nightly prices while large chalets appear with multi-night totals, and not evidence of a Quebec-specific pricing pattern. We regard the restricted-sample estimates as the credible ones and note that even their explanatory power is below that of hedonic Airbnb models estimated on platform data with reliable prices and room types.

### 5.7. City-Level Results

Table 11 presents the city-level regressions (Equations 5 and 6) on the 153 cities with Airbnb listings.

**Table 11.** City-Level Regressions

| Dep. var: | **(3-fwd)** *mean log rent* | **(3-rev)** *Airbnb listings in city* |
|---|---|---|
| Constant | 6.9341*** | -40.9296 |
| | (0.0707) | (58.2750) |
| Airbnb listings in city | 0.0001** | |
| | (0.0000) | |
| Mean bedrooms | 0.0882*** | -14.4332 |
| | (0.0241) | (9.4579) |
| Mean bathrooms | 0.3741*** | 11.6631* |
| | (0.0528) | (6.4498) |
| Mean log rent | | 10.2254 |
| | | (10.9040) |
| Observations | 153 | 153 |
| $R^2$ | 0.4916 | 0.0136 |

Robust (HC1) standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

The forward regression shows a positive and statistically significant but small cross-city association: one hundred additional Airbnb listings in a city are associated with a mean log rent about 1% higher, conditional on average dwelling size. The reverse regression has no explanatory power ($R^2 = 0.01$) and the mean rent does not predict the city's Airbnb count. Most of the 153 cities contribute few listings to either data set, so these regressions are descriptive and serve only to motivate the listing-level analysis.

### 5.8. Machine-Learning Benchmark

Table 12 reports hold-out performance on a random 80/20 split for the five models, estimated on the 4,370 listings with positive 500 m exposure, and Table 13 the five-fold cross-validation results under random and spatially blocked folds.

**Table 12.** Machine-Learning Model Performance (Random Hold-Out Test Set)

| Model | $R^2$ (Train) | $R^2$ (Test) | RMSE (Test) | MAE (Test) |
|---|---|---|---|---|
| OLS | 0.4741 | 0.5094 | 0.2242 | 0.1727 |
| LASSO | 0.4740 | 0.5099 | 0.2241 | 0.1725 |
| Elastic Net | 0.4740 | 0.5099 | 0.2241 | 0.1725 |
| Random Forest | 0.9092 | 0.7079 | 0.1730 | 0.1284 |
| Gradient Boosting | 0.8762 | 0.6935 | 0.1772 | 0.1319 |

*Notes:* OLS, LASSO, and Elastic Net use standardized features. Tree-based models use raw features. Train/test split is 80/20 with random_state=42.

**Table 13.** Random versus Spatially Blocked Cross-Validation

| Model | $R^2$ (mean) | $R^2$ (s.d.) | RMSE | MAE |
|---|---|---|---|---|
| *Random 5-fold* | | | | |
| OLS | 0.4767 | 0.0405 | 0.2256 | 0.1737 |
| LASSO | 0.4767 | 0.0406 | 0.2256 | 0.1737 |
| Elastic Net | 0.4767 | 0.0405 | 0.2256 | 0.1737 |
| Random Forest | 0.6473 | 0.0445 | 0.1849 | 0.1343 |
| Gradient Boosting | 0.6538 | 0.0334 | 0.1834 | 0.1340 |
| *Spatial 5-fold (FSA blocks)* | | | | |
| OLS | 0.4667 | 0.0539 | 0.2261 | 0.1752 |
| LASSO | 0.4668 | 0.0538 | 0.2261 | 0.1752 |
| Elastic Net | 0.4668 | 0.0538 | 0.2261 | 0.1752 |
| Random Forest | 0.4307 | 0.1594 | 0.2306 | 0.1803 |
| Gradient Boosting | 0.4140 | 0.1076 | 0.2355 | 0.1842 |

Five-fold cross-validation of the models in Table 12 on the same features and sample. Random folds shuffle listings; spatial folds assign whole forward sortation areas (181 FSA groups) to folds, so no test listing shares a neighbourhood with a training listing. Metrics are averaged over the five test folds; $R^2$ (s.d.) is the standard deviation across folds.

On the random hold-out split the random forest attains the highest test $R^2$ (0.71), followed by gradient boosting (0.69); the regularised linear models perform on par with OLS (0.51), and the LASSO retains the Airbnb count with a positive coefficient. Taken alone, the gap of twenty percentage points would suggest important nonlinearities. The spatially blocked cross-validation shows that the gap is an artefact of the fold design. Under random folds the tree-based models again outperform the linear models (cross-validated $R^2$ of 0.65 against 0.48). When whole FSAs are held out, so that the learners cannot draw on training listings in the test neighbourhoods, the linear

models are essentially unchanged (0.47) while the random forest and gradient boosting fall to 0.43 and 0.41, below OLS and with much larger variability across folds. With coordinates among the features, the tree-based learners were reconstructing neighbourhood rent levels from the locations of nearby training listings; they do not generalise to unseen neighbourhoods better than a linear hedonic model. This is the prediction counterpart of the fixed-effects result: the structure that flexible learners exploit, and that the city-fixed-effects regression attributes to Airbnb exposure, is neighbourhood-level variation in rent.

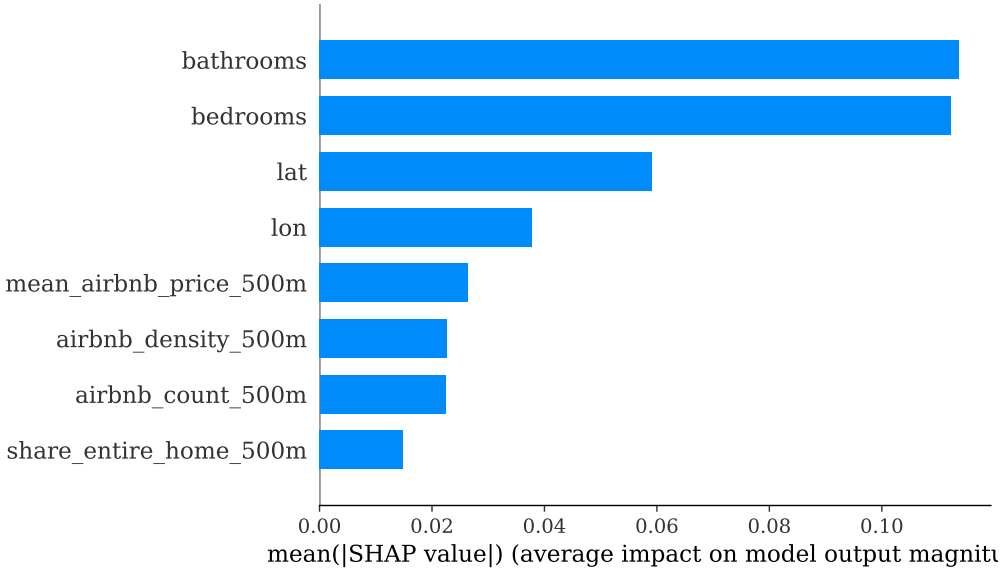



**Figure 6.** Mean Absolute SHAP Values, Gradient Boosting Model (Random Hold-Out Set)

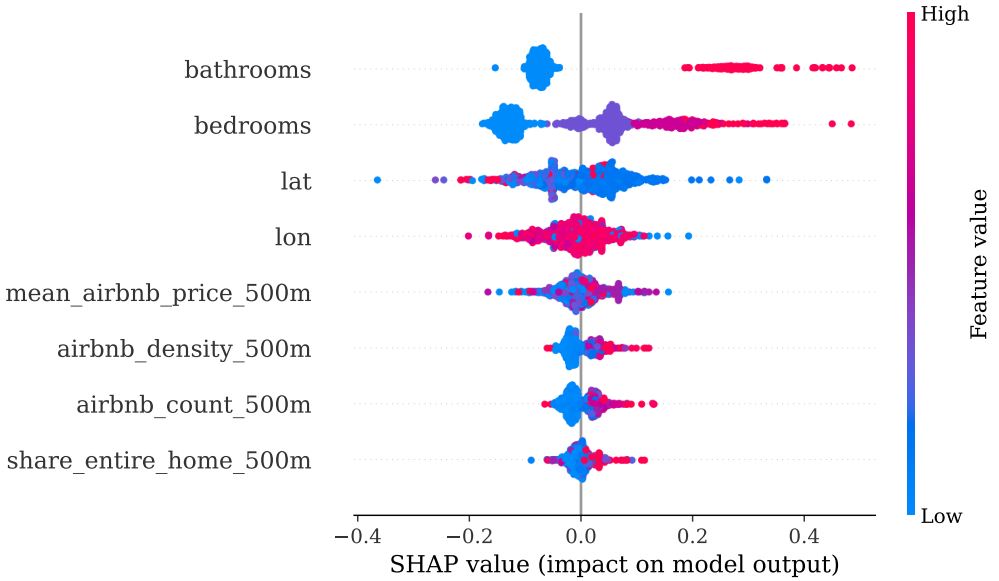



**Figure 7.** SHAP Summary Plot, Gradient Boosting Model (Random Hold-Out Set)

Figures 6 and 7 describe the gradient-boosting model fitted on the random split. Bathrooms, the geographic coordinates and bedrooms are the most important features; the exposure variables rank below them, and higher Airbnb counts contribute positively to predicted rents, consistent with the parametric estimates on the same (city-level) variation. Given the cross-validation results, the prominence of the coordinates should be read as the model's way of encoding neighbourhood rent levels, and the SHAP attributions as descriptions of the fitted model rather than as evidence about mechanisms.

## 6. Robustness Checks

This section assesses the sensitivity of the city-fixed-effects estimate to the buffer radius, the exposure measure, the sample, outliers, additional controls, the functional form and the exclusion of individual cities. The checks establish that the between-neighbourhood association of Section 5.1 is a stable feature of the data; they do not, and cannot, address the identification issue raised by the neighbourhood-fixed-effects estimates, which Section 7 takes up.

### 6.1. Buffer Radius

Table 14 and Figure 8 report the coefficient on the Airbnb count for radii of 250 m, 500 m, 1 km and 2 km, with dwelling controls and the fifteen-city fixed effects used in the spatial models.

**Table 14.** Buffer Radius Robustness — Hedonic Rent Model

| | **250m** | **500m** | **1km** | **2km** |
|---|---|---|---|---|
| Dep. var: | *log rent* | *log rent* | *log rent* | *log rent* |
| Airbnb count | 0.009272*** | 0.003826*** | 0.001589*** | 0.000612*** |
| | (0.000636) | (0.000221) | (0.000080) | (0.000028) |
| Controls | Yes | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes | Yes |
| Observations | 7,925 | 7,925 | 7,925 | 7,925 |
| $R^2$ | 0.5026 | 0.5081 | 0.5167 | 0.5205 |

Robust (HC1) standard errors in parentheses. Controls: bedrooms, bathrooms, building-type dummies; fixed effects for the 15 largest cities (remaining cities grouped), as in Table 7. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

The per-listing coefficient is largest at 250 m (0.0093) and declines monotonically with the radius (0.0038 at 500 m, 0.0016 at 1 km, 0.0006 at 2 km), remaining significant at the 1% level throughout. Part of this decline is mechanical: a listing counted within a wide buffer is a weaker indicator of proximity than one counted within a narrow buffer, and the counts themselves grow with the area. Multiplying each coefficient by the mean count at its radius gives the association for a listing with average exposure: 2.2% at 250 m, 3.0% at 500 m, 4.2% at 1 km and 5.0% at 2 km.

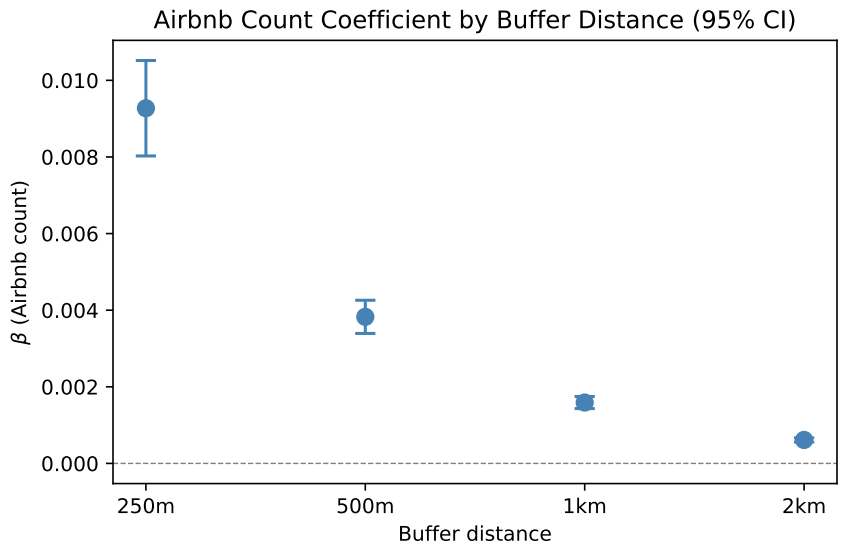



**Figure 8.** Airbnb Count Coefficient Across Buffer Radii (95% HC1 Confidence Intervals)

The total association therefore accumulates with the radius while the per-listing association decays, which is what one expects if the relevant variation is the overall STR intensity of the surrounding neighbourhood rather than the presence of particular listings within a few hundred metres. The $R^2$ also rises with the radius (0.503 to 0.521). Because the buffers are nested, these four estimates are not independent, and the ring decomposition in Section 6.5 separates them.

### 6.2. Alternative Exposure Measures

Columns (1d) and (1e) of Table 4 replace the count by the Airbnb density and by the entire-home share within 500 m. The density is the count divided by the buffer area and yields the same inference. The entire-home share is not significantly associated with rents conditional on dwelling characteristics and city fixed effects (point estimate −0.004, standard error 0.011), on the subsample of 4,370 listings with at least one Airbnb nearby. Conditional on STR presence, it is the intensity of nearby activity rather than its tilt toward entire homes that co-varies with rents. This does not support the compositional prediction of the supply-withdrawal hypothesis, under which entire-home listings, the closest substitutes for long-term units, should matter most (Horn and Merante, 2017; Duso et al., 2024). Two caveats apply: the entire-home indicator is a property-type proxy rather than the platform's room type, and the share is measured with considerable noise where few listings fall within the buffer.

### 6.3. Subsample Analysis

Table 15 estimates the model with bedrooms, bathrooms and city fixed effects on four subsamples: Montreal and the rest of the province, and apartments and houses.

**Table 15.** Subsample Analysis — Hedonic Rent Model

| | (1) Montreal | (2) Outside Mtl | (3) Apartments | (4) Houses |
|---|---|---|---|---|
| Airbnb count (500m) | 0.003865*** | 0.003046* | 0.003860*** | 0.005857** |
| | (0.000226) | (0.001580) | (0.000226) | (0.002348) |
| N | 4,578 | 3,347 | 7,136 | 741 |
| $R^2$ | 0.4587 | 0.6959 | 0.5046 | 0.6441 |
| City FE | Yes | Yes | Yes | Yes |

*Notes:* Robust standard errors (HC1) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include city fixed effects, bedrooms, and bathrooms as controls.

The Montreal subsample (4,578 listings) yields a coefficient of 0.0039, identical to the full-sample estimate; the non-Montreal subsample yields 0.0030 with a much larger standard error (0.0016), reflecting the smaller sample and the heterogeneity of the other markets, which include suburbs, regional cities and resort communities. The apartment subsample mirrors the full sample, as expected given that apartments are 90% of it. The 741 houses yield a larger coefficient (0.0059, standard error 0.0023), significant at the 5% level; houses are a heterogeneous category and the estimate is imprecise.

### 6.4. Outlier Sensitivity

Table 16 re-estimates the model on two trimmed samples: listings with monthly rents between the 5th and 95th percentiles, and listings outside the top and bottom 1% of the 500 m count.

**Table 16.** Outlier Sensitivity — Hedonic Rent Model

| | (1) Full Sample | (2) Rent 5/95 | (3) Airbnb 1/99 |
|---|---|---|---|
| Airbnb count (500m) | 0.003891*** | 0.003057*** | 0.004222*** |
| | (0.000225) | (0.000186) | (0.000242) |
| N | 7,925 | 7,254 | 7,868 |
| $R^2$ | 0.5664 | 0.4966 | 0.5673 |
| City FE | Yes | Yes | Yes |

*Notes:* Robust standard errors (HC1) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column (2) winsorizes monthly rent at the 5th and 95th percentiles. Column (3) trims the top and bottom 1% of Airbnb listing counts.

Trimming the rent distribution reduces the coefficient by about a fifth, from 0.0039 to 0.0031, consistent with the quantile results, which locate the strongest association in the upper tail; trimming the most Airbnb-saturated locations raises it slightly, to 0.0042. The coefficient remains significant at the 1% level in every case, so the baseline result is not driven by a small number of influential observations.

### 6.5. Additional Specification Checks

Table 17 subjects the reference specification to four further checks; column (R1) reproduces the baseline.

**Table 17.** Extended Robustness — Reference Hedonic Specification

| Dep. var: | **(R1)** *log rent* | **(R2)** *log rent* | **(R3)** *log rent* | **(R4)** *log rent* | **(R5)** *log rent* |
|---|---|---|---|---|---|
| Airbnb count (500m) | 0.0039*** | 0.0039*** | 0.0032*** | | 0.0013*** |
| | (0.0002) | (0.0001) | (0.0003) | | (0.0003) |
| log(1 + Airbnb count 500m) | | | | 0.0484*** | |
| | | | | (0.0025) | |
| Airbnb count (500m–1km ring) | | | | | 0.0018*** |
| | | | | | (0.0002) |
| Bedrooms | 0.1276*** | 0.1276*** | 0.1271*** | 0.1296*** | 0.1322*** |
| | (0.0046) | (0.0043) | (0.0076) | (0.0046) | (0.0047) |
| Bathrooms | 0.2287*** | 0.2287*** | 0.2103*** | 0.2274*** | 0.2246*** |
| | (0.0126) | (0.0291) | (0.0202) | (0.0125) | (0.0125) |
| Interior size (100 sq ft) | | | 0.0044*** | | |
| | | | (0.0017) | | |
| Standard errors | HC1 | City cluster | HC1 | HC1 | HC1 |
| Clusters | | 252 | | | |
| Building type & city FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 7,925 | 7,925 | 3,531 | 7,925 | 7,925 |
| $R^2$ | 0.5697 | 0.5697 | 0.6044 | 0.5726 | 0.5786 |

Standard errors in parentheses: HC1 except column (R2), which clusters by city. All columns control for bedrooms, bathrooms, building type, and city fixed effects; column (R3) adds interior size (per 100 sq ft) on the subsample reporting it. Column (R5) includes the 500m count and the count in the 500m–1km annulus jointly. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

**City-clustered standard errors.** Column (R2) clusters standard errors by city (252 clusters). Clustering tightens the standard error on the Airbnb coefficient from 0.0002 to 0.0001 while widening it on bathrooms. As discussed in Section 5.3, this tightening is not credible given Montreal's weight in the sample, and FSA-clustered or Conley standard errors, which are three times the HC1 value, are the appropriate reference.

**Interior size.** Column (R3) adds interior floor area (per 100 square feet) on the 3,531 listings that report it. Size enters positively and significantly, and the Airbnb coefficient falls to 0.0032, about a fifth below the baseline, consistent with size absorbing some quality variation; it remains significant.

**Functional form.** Column (R4) replaces the count with $\ln(1 + A_{i,500\text{m}})$. The coefficient of 0.048 implies that doubling the nearby count (plus one) is associated with about 3.4% higher rent, and the fit improves marginally over the linear form, indicating concavity in the exposure–rent relationship. With FSA fixed effects the same coefficient is 0.002 and insignificant (Table 5).

**Ring exposure.** Column (R5) includes the 500 m count jointly with the count in the 500 m–1 km annulus. Both enter positively and significantly, and the per-listing coefficients are of similar size (0.0013 within 500 m, 0.0018 in the annulus). The association is thus not confined to the immediate 500 m but extends to the kilometre scale, in line with the accumulation pattern of Section 6.1. Because the two counts are strongly correlated, the split should be read as a decomposition of a common neighbourhood signal rather than as separate gradients; a strictly local effect would have produced a much larger inner-ring than outer-ring coefficient.

### 6.6. Leave-One-City-Out Sensitivity

Figure 9 re-estimates the reference specification excluding, in turn, each of the eight cities with the most rental listings.

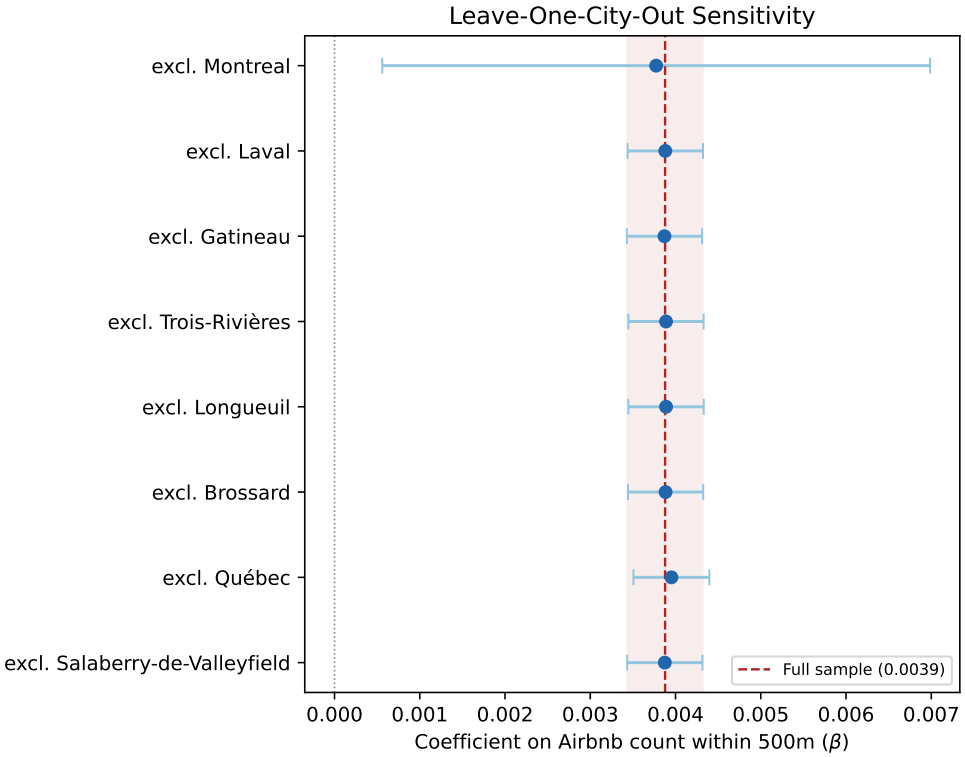



**Figure 9.** Leave-One-City-Out Sensitivity of the Airbnb Coefficient (95% HC1 Confidence Intervals)

The coefficient ranges from 0.0038 to 0.0040 across exclusions. Excluding Montreal, which

accounts for 58% of the estimation sample, leaves the point estimate at 0.0038 with a wider confidence interval; the between-neighbourhood association is not an artefact of a single market.

### 6.7. Summary of Robustness

Figure 10 summarises the Airbnb coefficient across the buffer and subsample specifications.

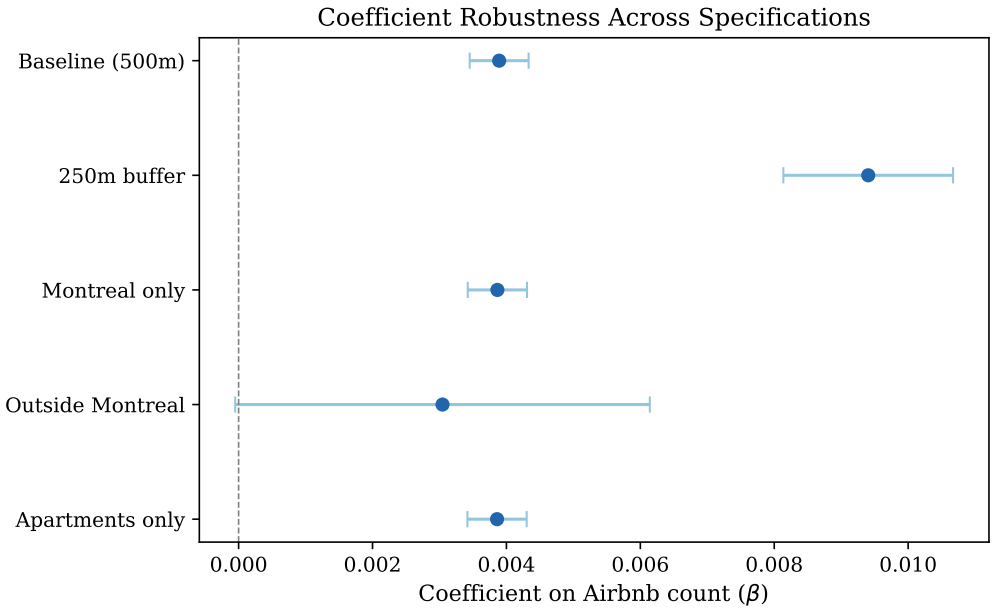



**Figure 10.** Summary of Airbnb Count Coefficients Across Specifications (95% HC1 Confidence Intervals)

Across radii, exposure measures, functional forms, control sets, subsamples, trimming rules and the exclusion of individual cities, the association between Airbnb exposure and rents conditional on city fixed effects is consistently positive and of similar magnitude. The stability of the estimate across these dimensions is informative about its reliability as a description of the data; it is not informative about its interpretation, which depends on the single dimension along which the estimate is not stable, the spatial resolution of the location fixed effects.

## 7. Discussion

This section places the estimates in the context of the international literature, interprets the contrast between the city- and neighbourhood-fixed-effects results, draws the implications for housing policy and STR regulation, and sets out the limitations of the analysis.

### 7.1. Magnitudes in the Context of the Literature

The city-fixed-effects estimate implies that one additional Airbnb listing within 500 m is associated with a rent about 0.4% higher, and one within 250 m with a rent about 0.9% higher. How do

these compare with the quasi-experimental literature? Duso et al. (2024) estimate, from Berlin's regulatory shocks, that each additional commercial Airbnb listing raises rents per square metre by 1.3–2.4%. Per listing, this is larger than our 500 m coefficient, but their treatment is a commercial listing, whereas our count includes every listing, casual or commercial, so the two are not directly comparable. Barron et al. (2021) report an elasticity of rents with respect to Airbnb listings of 0.018; our logarithmic specification with city fixed effects gives 0.048, more than twice as large, while the FSA-fixed-effects version is 0.002 and insignificant. Horn and Merante (2017) find that a one-standard-deviation increase in Airbnb listings in Boston is associated with asking rents 0.4% higher; the corresponding standardised association in our data is 5.5% with city fixed effects and 1.2% with FSA fixed effects. Garcia-López et al. (2020) estimate average rent effects of 1.9% in Barcelona and about 7% in the most exposed neighbourhoods; for a listing with the average 500 m exposure of eight Airbnb listings our estimates imply 3.1% with city fixed effects and 0.6% with FSA fixed effects, and for a listing at the 90th percentile of exposure (30 listings) 11.7% and 2.4%. Koster et al. (2021) find that ordinances which halved Airbnb listings lowered rents by about 3%; halving the average exposure in our sample would correspond to 1.5% and 0.3%.

The pattern is consistent. When standardised, the city-fixed-effects estimates are larger than the quasi-experimental magnitudes, often by a factor of several, while the FSA-fixed-effects estimates are of the same order as, or smaller than, those magnitudes. This is what one expects if the city-fixed-effects estimate combines a causal component with the between-neighbourhood sorting that the quasi-experimental designs remove. The earlier version of this paper read the proximity of its point estimate to the Berlin per-listing figure as mutual corroboration; the comparison in standardised terms and the neighbourhood-fixed-effects results do not support that reading. Our estimates are best described as bracketing the quasi-experimental magnitudes from above (city fixed effects) and from below or around (FSA fixed effects), with the latter too imprecise to pin down.

Two further findings are worth placing against the literature. First, the superhost discount and the implausibly low explanatory power of the Airbnb pricing model on the full sample, which the earlier version interpreted as a Quebec-specific composition effect, are artefacts of the scraped price field; on the plausibly priced subsample, the model recovers the entire-home premium and the size gradients that Wang and Nicolau (2017) and Gibbs et al. (2018) report, and no superhost discount. Second, the compositional prediction of the supply-withdrawal hypothesis, that entire-home listings should matter most (Horn and Merante, 2017; Duso et al., 2024), is not supported: conditional on STR presence, the entire-home share within the buffer is unrelated to rents. Given the coarse property-type proxy for entire-home status, we regard this as inconclusive rather than as evidence against the hypothesis.

### 7.2. Between-Neighbourhood and Within-Neighbourhood Variation

The central empirical fact of the paper is the contrast between a stable, precisely estimated association with city fixed effects and a small, imprecise one with FSA fixed effects. We offered three readings in Section 5.2; here we weigh them.

The sorting reading has the most support. It is the mechanism that the identification literature anticipates (Kuminoff et al., 2010; Bishop et al., 2020), it is consistent with the direct evidence on endogenous amenities and location sorting in tourist cities (Hidalgo et al., 2024; Almagro and Domínguez-Iino, 2025), it is consistent with the finding that block-scale spatial lag and error processes leave the coefficient intact while neighbourhood-scale fixed effects remove it, and it is consistent with the negative within-neighbourhood coefficient on the presence of any Airbnb, which suggests that within a neighbourhood STR listings sit on the less residential blocks. It is also consistent with the behaviour of the machine-learning benchmark, where the predictive advantage of flexible learners vanishes once neighbourhoods are held out.

The scale reading, under which STR-induced rent pressure is a neighbourhood-level phenomenon absorbed by the fixed effects, cannot be excluded. Supply withdrawal does operate on a neighbourhood-wide rental market, and a tenant displaced from one block searches on neighbouring blocks. But the same argument implies that the between-neighbourhood comparison, which is where the estimate lives, is exactly the comparison most exposed to amenity confounding. The quasi-experimental studies resolve this tension with time variation within neighbourhoods; a cross-section cannot.

The measurement-error reading explains part of the attenuation. The scraped Airbnb sample omits inactive and filtered listings and other platforms, so the count measures true STR exposure with error, and the share of signal in the within-FSA variation is smaller than in the total variation. Classical measurement error attenuates both estimates toward zero and the within-FSA estimate more; it cannot, however, account for a fivefold reduction when a quarter of the variance is within neighbourhoods, nor for the sign of the extensive-margin coefficient.

On balance, we conclude that most of the Airbnb–rent association in these data is a between-neighbourhood association, that the city-fixed-effects estimate of 0.4% per listing is an upper bound on any causal within-neighbourhood effect, and that the within-neighbourhood estimate of 0.08% per listing, with a 95% Conley confidence interval that includes zero and extends to about 0.2%, is the more conservative summary of what the data show.

### 7.3. Implications for Housing Affordability

The implications for affordability policy follow from the two bounds. At the upper bound, a dwelling with twenty Airbnb listings within 500 m, roughly the 85th percentile of exposure, would rent for about 8% more than an otherwise identical dwelling in an Airbnb-free location of the same

city; at the FSA-fixed-effects estimate the differential would be about 1.6% and not statistically distinguishable from zero. The first figure is the one a reader of the cross-sectional literature would take away, and it is economically meaningful in a market with low vacancy rates; the second is the one that survives the closest approximation to a within-neighbourhood comparison the data permit. In either case, dwelling characteristics and neighbourhood location explain far more of the variation in rents than STR exposure does, so STR activity is at most one factor among several in the rental affordability equation. The cross-sectional evidence from Quebec, read carefully, therefore does not by itself establish that STR activity raises rents within neighbourhoods; it establishes that STR activity is concentrated in high-rent neighbourhoods, which is relevant to where the costs of any effect fall but is a different proposition.

### 7.4. Short-Term Rental Regulation

Quebec regulates STRs at two levels.[1] Since 2022 the provincial Tourist Accommodation Act requires every STR establishment, including a primary residence let occasionally, to hold a registration certificate, to display the registration number in every advertisement and to transmit it to the platform; a 2023 amendment made platforms liable for listings without a valid number, with substantial fines. Montreal adds zoning rules that confine commercial STRs to designated zones and, since 2025, restricts primary-residence STRs to a summer window (10 June to 10 September) and prohibits them in some boroughs. Our data were collected in January 2026, outside that window, so the Montreal listings we observe were either in commercial zones or non-compliant, which is itself a reminder that enforcement, not only rules, shapes STR supply.

The results bear on this regime in three ways. First, the spatial pattern of the between-neighbourhood association, decaying per listing and accumulating with the radius, indicates that STR activity is a neighbourhood-scale phenomenon, which favours geographically targeted instruments such as zoning and borough-level rules over province-wide ones, as Gurran and Phibbs (2017) and Nieuwland and van Melik (2020) argue more generally. Second, the quantile results locate the association in the upper segment of the rental market, so that any rent pressure attributable to STRs would fall first on neighbourhoods that are already expensive; whether this is an argument for or against restriction depends on distributional objectives (Calder-Wang, 2021). Third, and most importantly, the within-neighbourhood estimates imply that the cross-sectional correlation that often motivates regulation substantially overstates what can be attributed to STR activity itself. A regulatory rationale that rests on such correlations is weaker than it appears; one that rests on the quasi-experimental literature (Koster et al., 2021; Duso et al., 2024) remains intact but has not been established for Quebec. Regulation also involves trade-offs: STRs generate income for hosts, tax

[1]The regulatory description reflects the rules in force in 2025–2026 as published by the Government of Quebec (`https://www.quebec.ca/tourisme-et-loisirs/hebergement-touristique`) and the City of Montreal; it is not based on the academic literature and should be verified against the current texts before being relied upon.

revenue for municipalities and consumer surplus for travellers (Farronato and Fradkin, 2022), and restrictive rules may push activity underground. A framework that targets commercial multi-listing operation while preserving occasional home-sharing, which is the direction Quebec's registration regime has taken, is consistent with the evidence, but its effects on rents are an open empirical question.

### 7.5. Professional Hosts and Commercialisation

The literature documents the increasing commercialisation of Airbnb, with a growing share of listings operated by multi-listing hosts (Ke, 2017; Wachsmuth and Weisler, 2018; Combs et al., 2020). Our data contain no host identifiers, so we cannot measure multi-listing activity directly, and our compositional proxy, the entire-home share within the buffer, is unrelated to rents conditional on STR presence. Either the property-type proxy is too coarse to isolate commercial operation, or within Quebec's market casual and commercial listings are sufficiently co-located that composition adds little information beyond the count. Host-resolved data would allow the association to be decomposed by host type and would sharpen the evidence base for a regime that explicitly distinguishes commercial from occasional hosts.

### 7.6. Limitations

The limitations of the analysis are of three kinds: design, data and specification.

**Design.** The data are a single cross-section. We cannot separate the effect of Airbnb on rents from reverse causality or from unobserved neighbourhood characteristics, and the fixed-effects comparison shows that the latter matter. A credible causal estimate requires panel data, ideally combined with a policy shock that shifted STR supply in some locations but not others, as in Koster et al. (2021), Valentin (2021) and Duso et al. (2024), or an instrument for Airbnb penetration in the spirit of Barron et al. (2021) and Garcia-López et al. (2020). The analysis is also static and cannot capture supply responses to rising rents (Glaeser et al., 2006; Saiz, 2010) or host exit in response to enforcement.

**Data.** The scraped Airbnb price field is bimodal and partly records total-stay rather than nightly prices, which invalidates the full-sample Airbnb pricing model and adds noise to the mean-nearby-price exposure variable; the count-based main results are unaffected. Room type and host identity are not observed, so entire-home status is a property-type proxy and commercial operation cannot be measured. The Airbnb sample omits inactive, blocked and filtered listings and other platforms, and was collected in winter, when the composition of active listings differs from the summer peak; measurement error in exposure attenuates the estimates, more so within neighbourhoods. The rental

data are asking rents on Realtor.ca, which over-represents agent-listed, newer and higher-priced units relative to the occupied stock; several benchmark studies share the asking-rent feature (Horn and Merante, 2017; Duso et al., 2024), which aids comparability but not representativeness. The composition-based specifications and the machine-learning benchmark are estimated on the selected subsample of listings with positive exposure. Interior size is available for less than half the sample and contains implausible values.

**Specification.** Any distance-based exposure measure is a discretisation of a continuous spatial process and is subject to the modifiable areal unit problem (Fotheringham and Wong, 1991); we report four radii, a ring decomposition and alternative functional forms, but the set of buffers remains a simplification (Can, 1992). FSAs are administrative delivery zones rather than housing submarkets, and their size varies; the attenuation we document would plausibly be larger with finer neighbourhood definitions and smaller with coarser ones. The spatial models use a single weights specification ($k = 5$ nearest neighbours) and coarse city groups; the quantile regressions use even coarser groups and standard errors that ignore spatial dependence. Finally, the sample is dominated by Montreal and a handful of other cities, which limits generalisation to peripheral and rural markets, where STR dynamics differ (Ayouba et al., 2020; Combs et al., 2020).

### 7.7. Future Work

Four extensions follow directly. First, repeated scrapes of the same platforms would yield a listing-level panel for Quebec and allow fixed-effects and event-study designs around the province's registration regime and Montreal's 2025 seasonal rule, the most direct route to within-neighbourhood causal estimates for Canada. Second, host-resolved data would permit the commercial–casual decomposition that the entire-home proxy cannot deliver. Third, administrative contract rents, for instance from lease registries, would resolve the asking-rent and selection concerns and allow welfare analysis in the spirit of Li et al. (2022) and Calder-Wang (2021). Fourth, on the present data, quantile regressions with neighbourhood fixed effects, estimated with penalised fixed-effects methods, and spatial models with neighbourhood fixed effects would complete the within-neighbourhood picture.

## 8. Conclusion

This paper has examined the relationship between Airbnb activity and residential rents in Quebec using a 2026 cross-section of 3,456 Airbnb listings and 8,303 rental listings linked by exact distance-based matching. It is, to our knowledge, the first listing-level econometric analysis of the question for Quebec and for a Canadian market, and it has been designed to make the identification problem of cross-sectional STR–rent studies measurable rather than merely acknowledged.

The findings have two layers. With city fixed effects, each additional Airbnb listing within

500 m is associated with a monthly rent about 0.4% higher, the per-listing association decays with distance while the total association accumulates with the radius, the association is largest at the upper quantiles of the conditional rent distribution, and it is unchanged in spatial lag and spatial error models, with a total impact of 0.40% per listing. These results are stable across exposure measures, functional forms, subsamples, trimming rules and the exclusion of individual cities, and they are of the kind that the cross-sectional literature on STR platforms has generally reported.

The second layer changes their interpretation. Standard errors that account for spatial dependence in a buffer-based exposure measure are about three times larger than the heteroskedasticity-robust errors conventionally reported, although the city-fixed-effects estimate remains significant. More importantly, once fixed effects for postal-code neighbourhoods absorb neighbourhood-level unobservables, the coefficient falls to about 0.08% per listing and is no longer distinguishable from zero under spatial inference; the logarithmic and extensive-margin exposure measures behave in the same way, and within neighbourhoods the presence of any Airbnb listing is associated with lower rather than higher rents. The predictive advantage of tree-based machine-learning models over the linear hedonic model, which a random hold-out sample suggests, disappears when whole neighbourhoods are held out. In these data, the Airbnb–rent association is predominantly a between-neighbourhood phenomenon, consistent with STR supply concentrating in high-amenity, high-rent neighbourhoods. When standardised, the city-fixed-effects estimates exceed the magnitudes found by quasi-experimental studies in Berlin, Los Angeles, Barcelona and Boston, while the neighbourhood-fixed-effects estimates are of the same order or smaller; the two bracket the causal literature from above and below.

Two subsidiary results concern the Airbnb side of the market. A hedonic model of nightly prices finds no association between city-level rents and STR prices, so the data offer no support for a simple opportunity-cost pricing channel. The model also reveals that the scraped price field is bimodal and partly records total-stay prices; on the plausibly priced subsample the conventional entire-home and size premia reappear and the superhost discount reported in an earlier version of this paper disappears.

For policy, the results counsel care in reading cross-sectional correlations between STR density and rents as evidence of STR-induced rent pressure. Such correlations describe where STR activity is located, which matters for the incidence of any effect, but in Quebec most of the correlation does not survive a within-neighbourhood comparison. The case for regulating STRs on affordability grounds therefore rests on the quasi-experimental evidence from other markets, and establishing it for Quebec requires panel data, ideally combined with the province's registration regime and Montreal's seasonal rules as sources of variation, together with host-resolved listings and contract rents. The present paper provides the baseline against which such work can be measured, and a template, neighbourhood fixed effects and spatially robust inference, for what cross-sectional STR

studies should report.

## References

**Almagro, Milena and Tomás Domínguez-Iino**, "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Econometrica*, 2025, *93* (3), 1031–1071.

**Àngel Garcia-López, Miquel, Jordi Jofre-Monseny, Rodrigo Martínez-Mazza, and Mariona Segú**, "Do Short-Term Rental Platforms Affect Housing Markets? Evidence from Airbnb in Barcelona," *Journal of Urban Economics*, 2020, *119*, 103278.

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

**Athey, Susan and Guido W. Imbens**, "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 2019, *11*, 685–725.

**Ayouba, Kassoum, Marie-Laure Breuillé, Camille Grivault, and Julie Le Gallo**, "Does Airbnb Disrupt the Private Rental Market? An Empirical Analysis for French Cities," *International Regional Science Review*, 2020, *43* (1–2), 76–104.

**Barron, Kyle, Edward Kung, and Davide Proserpio**, "The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb," *Marketing Science*, 2021, *40* (1), 23–47.

**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.

**Breiman, Leo**, "Random Forests," *Machine Learning*, 2001, *45* (1), 5–32.

**Calder-Wang, Sophie**, "The Distributional Impact of the Sharing Economy on the Housing Market," 2021. Working paper, The Wharton School, University of Pennsylvania; SSRN 3908062.

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

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

**Cocola-Gant, Agustin**, "Holiday Rentals: The New Gentrification Battlefront," *Sociological Research Online*, 2016, *21* (3), 112–120.

_ **and Ana Gago**, "Airbnb, Buy-to-Let Investment and Tourism-Driven Displacement: A Case Study in Lisbon," *Environment and Planning A: Economy and Space*, 2021, *53* (7), 1671–1688.

**Combs, Jennifer, Danielle Kerrigan, and David Wachsmuth**, "Short-Term Rentals in Canada: Uneven Growth, Uneven Impacts," *Canadian Journal of Urban Research*, 2020, *29* (1), 119–135.

**Conley, Timothy G.**, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

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

**Duso, Tomaso, Claus Michelsen, Maximilian Schäfer, and Kevin Ducbao Tran**, "Airbnb and Rental Markets: Evidence from Berlin," *Regional Science and Urban Economics*, 2024, *106*, 104007.

**Einav, Liran, Chiara Farronato, and Jonathan Levin**, "Peer-to-Peer Markets," *Annual Review of Economics*, 2016, *8*, 615–635.

**Ert, Eyal, Aliza Fleischer, and Nathan Magen**, "Trust and Reputation in the Sharing Economy: The Role of Personal Photos in Airbnb," *Tourism Management*, 2016, *55*, 62–73.

**Farronato, Chiara and Andrey Fradkin**, "The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry," *American Economic Review*, 2022, *112* (6), 1782–1817.

**Fotheringham, A. Stewart and David W. S. Wong**, "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis," *Environment and Planning A: Economy and Space*, 1991, *23* (7), 1025–1044.

**Franco, Sofia F. and Carlos Daniel Santos**, "The Impact of Airbnb on Residential Property Values and Rents: Evidence from Portugal," *Regional Science and Urban Economics*, 2021, *88*, 103667.

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

**Gibbs, Chris, Daniel Guttentag, Ulrike Gretzel, Jym Morton, and Alasdair Goodwill**, "Pricing in the Sharing Economy: A Hedonic Pricing Model Applied to Airbnb Listings," *Journal of Travel & Tourism Marketing*, 2018, *35* (1), 46–56.

**Glaeser, Edward L., Joseph Gyourko, and Raven E. Saks**, "Urban Growth and Housing Supply," *Journal of Economic Geography*, 2006, *6* (1), 71–89.

**Gotham, Kevin Fox**, "Tourism Gentrification: The Case of New Orleans' Vieux Carré (French Quarter)," *Urban Studies*, 2005, *42* (7), 1099–1121.

**Grisdale, Sean**, "Displacement by Disruption: Short-Term Rentals and the Political Economy of "Belonging Anywhere" in Toronto," *Urban Geography*, 2021, *42* (5), 654–680.

**Gurran, Nicole and Peter Phibbs**, "When Tourists Move In: How Should Urban Planners Respond to Airbnb?," *Journal of the American Planning Association*, 2017, *83* (1), 80–92.

**Guttentag, Daniel**, "Airbnb: Disruptive Innovation and the Rise of an Informal Tourism Accommodation Sector," *Current Issues in Tourism*, 2015, *18* (12), 1192–1217.

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

**Hidalgo, Alberto, Massimo Riccaboni, and Francisco J. Velázquez**, "The Effect of Short-Term Rentals on Local Consumption Amenities: Evidence from Madrid," *Journal of Regional Science*, 2024, *64* (3), 621–648.

**Horn, Keren and Mark Merante**, "Is Home Sharing Driving Up Rents? Evidence from Airbnb in Boston," *Journal of Housing Economics*, 2017, *38*, 14–24.

**Ke, Qing**, "Sharing Means Renting? An Entire-Marketplace Analysis of Airbnb," in "Proceedings of the 2017 ACM on Web Science Conference (WebSci '17)" ACM Troy, NY 2017, pp. 131–139.

**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** _ , "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 1999, *40* (2), 509–533.

_ **and** _ , "HAC Estimation in a Spatial Framework," *Journal of Econometrics*, 2007, *140* (1), 131–154.

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

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

**Koster, Hans R. A., Jos van Ommeren, and Nicolas Volkhausen**, "Short-Term Rentals and the Housing Market: Quasi-Experimental Evidence from Airbnb in Los Angeles," *Journal of Urban Economics*, 2021, *124*, 103356.

**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.

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

**Lee, Dayne**, "How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations," *Harvard Law & Policy Review*, 2016, *10* (1), 229–253.

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

**Li, Hui, Yijin Kim, and Kannan Srinivasan**, "Market Shifts in the Sharing Economy: The Impact of Airbnb on Housing Rentals," *Management Science*, 2022, *68* (11), 8015–8044.

**Lundberg, Scott M. and Su-In Lee**, "A Unified Approach to Interpreting Model Predictions," in "Advances in Neural Information Processing Systems 30" 2017, pp. 4765–4774.

**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 Science, 2003, pp. 67–89.

**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.

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

**Nieuwland, Shirley and Rianne van Melik**, "Regulating Airbnb: How Cities Deal with Perceived Negative Externalities of Short-Term Rentals," *Current Issues in Tourism*, 2020, *23* (7), 811–825.

**Oster, Emily**, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

**Palmquist, Raymond B.**, "Property Value Models," in Karl-Göran Mäler and Jeffrey R. Vincent, eds., *Handbook of Environmental Economics*, Vol. 2, Amsterdam: Elsevier, 2005, pp. 763–819.

**Parmeter, Christopher F., Daniel J. Henderson, and Subal C. Kumbhakar**, "Nonparametric Estimation of a Hedonic Price Function," *Journal of Applied Econometrics*, 2007, *22* (3), 695–699.

**Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay**, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011, *12*, 2825–2830.

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

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

**Shabrina, Zahratu, Elsa Arcaute, and Michael Batty**, "Airbnb and Its Potential Impact on the London Housing Market," *Urban Studies*, 2022, *59* (1), 197–221.

**Sheppard, Stephen and Andrew Udell**, "Do Airbnb Properties Affect House Prices?," Working Paper 2016-03, Department of Economics, Williams College, Williamstown, MA 2016.

**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.

**Sundararajan, Arun**, *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*, Cambridge, MA: MIT Press, 2016.

**Tibshirani, Robert**, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, *58* (1), 267–288.

**Valentin, Maxence**, "Regulating Short-Term Rental Housing: Evidence from New Orleans," *Real Estate Economics*, 2021, *49* (1), 152–186.

**Wachsmuth, David and Alexander Weisler**, "Airbnb and the Rent Gap: Gentrification Through the Sharing Economy," *Environment and Planning A: Economy and Space*, 2018, *50* (6), 1147–1170.

**Wang, Dan and Juan L. Nicolau**, "Price Determinants of Sharing Economy Based Accommodation Rental: A Study of Listings from 33 Cities on Airbnb.com," *International Journal of Hospitality Management*, 2017, *62*, 120–131.

**Zervas, Georgios, Davide Proserpio, and John W. Byers**, "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," *Journal of Marketing Research*, 2017, *54* (5), 687–705.

**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.

**Zou, Hui and Trevor Hastie**, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (2), 301–320.

# A. Data Appendix

## A.1. Variable Definitions

Table A1 lists the variables used in the analysis, with their sources, definitions and units.

## A.2. City Name Standardisation

The Airbnb data contain city names with inconsistent formatting, and the Realtor.ca addresses identify Montreal boroughs, Quebec City and Longueuil arrondissements and similar subdivisions, which are collapsed to the municipality. Table A2 lists the explicit mappings applied to the Airbnb data; the rental data are standardised by prefix rules in `scripts/03_clean_rent.py`.

## A.3. Sample Construction

The Airbnb file contains 5,000 listings, of which 1,544 (30.9%) have no price and are dropped, leaving 3,456 listings, all with coordinates; 2,437 of them are located in a municipality with rental listings and enter the Airbnb pricing model, and 1,699 of those are priced at or below $1,000 per night. The Realtor.ca file contains 8,356 records; restricting to single-family residential listings with a monthly rental period and a positive parsed rent leaves 8,303 listings, all with valid coordinates and postal codes. All 8,303 enter the spatial merge; the Airbnb count is zero for listings with no Airbnb within the radius, and the composition variables are undefined in that case (4,370 listings have positive 500 m exposure). The baseline regression sample comprises the 7,925 listings with non-missing bedrooms; the FSA-fixed-effects sample drops singleton FSAs (7,883 listings in 289 FSAs); the interior-size specifications use 3,531 (city fixed effects) and 3,485 (FSA fixed effects) listings.

# B. Methodological Appendix

## B.1. Haversine Distance

The Haversine formula gives the great-circle distance between two points on a sphere from their latitudes and longitudes. For points $(\phi_1, \lambda_1)$ and $(\phi_2, \lambda_2)$ in radians,

$$d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right), \tag{A.1}$$

where $R = 6{,}371$ km is the mean Earth radius, $\Delta\phi = \phi_2 - \phi_1$ and $\Delta\lambda = \lambda_2 - \lambda_1$. For distances below a few kilometres the difference from the geodesic distance on the reference ellipsoid is negligible. In the spatial merge (Section 3.3) the distance matrix between each chunk of 500 rental listings and all 3,456 Airbnb listings is computed in vectorised form, and buffer membership is determined

**Table A1.** Variable Definitions

| Variable | Source | Definition | Unit |
|---|---|---|---|
| *Rental listing variables* | | | |
| `rent` | Realtor.ca | Monthly asking rent (winsorised at p1/p99) | CAD |
| `log_rent` | Constructed | Natural log of monthly rent | — |
| `bedrooms` | Realtor.ca | Number of bedrooms | Count |
| `bathrooms` | Realtor.ca | Number of bathrooms (total) | Count |
| `building_type` | Realtor.ca | Building type (Apartment, House, Row/Townhouse) | Category |
| `size_interior` | Realtor.ca | Interior floor area (reported for 3,699 listings) | sq ft |
| `lat`, `lon` | Realtor.ca | Latitude and longitude of the listing | Degrees |
| `city` | Realtor.ca | Municipality (standardised; boroughs collapsed) | Category |
| `postal_code` | Realtor.ca | Six-character postal code | Category |
| `fsa` | Constructed | Forward sortation area (first three characters of the postal code) | Category |
| *Airbnb listing variables* | | | |
| `price_numeric` | Airbnb | Displayed price (winsorised at p1/p99; see Section 3.1) | CAD |
| `log_price` | Constructed | Natural log of displayed price | — |
| `property_type` | Airbnb | Property type (Rental unit, House, Cabin/Chalet, Condo, Apartment, Other) | Category |
| `rating` | Airbnb | Star rating (median-imputed when missing) | 0–5 |
| `num_reviews` | Airbnb | Number of guest reviews (zero when missing) | Count |
| `bedrooms`, `bathrooms` | Airbnb | Number of bedrooms and bathrooms | Count |
| `guests_count` | Airbnb | Maximum number of guests | Count |
| `amenities_count` | Airbnb | Number of listed amenities | Count |
| `is_superhost` | Airbnb | Superhost status indicator | Binary |
| `is_entire_home` | Constructed | Property type in {House, Cabin/Chalet, Condo} | Binary |
| `lat`, `lon` | Airbnb | Latitude and longitude of the listing | Degrees |
| `city` | Airbnb | City name (standardised) | Category |
| *Airbnb exposure variables (spatial buffer merge, $r \in \{250\,m, 500\,m, 1\,km, 2\,km\}$)* | | | |
| `airbnb_count_`$r$ ($A_{i,r}$) | Constructed | Number of Airbnb listings within radius $r$ | Count |
| `airbnb_density_`$r$ | Constructed | $A_{i,r}/(\pi r^2)$ | Per km$^2$ |
| `mean_airbnb_price_`$r$ | Constructed | Mean displayed price of Airbnb listings within $r$ (undefined if $A_{i,r} = 0$) | CAD |
| `share_entire_home_`$r$ | Constructed | Share of entire-home listings within $r$ (undefined if $A_{i,r} = 0$) | Proportion |
| `mean_rating_`$r$ | Constructed | Mean rating within $r$ | 0–5 |
| `superhost_share_`$r$ | Constructed | Share of superhost listings within $r$ | Proportion |
| `any_airbnb_500m` | Constructed | $\mathbf{1}[A_{i,500\text{m}} > 0]$ | Binary |
| `airbnb_ring_500m_1km` | Constructed | $A_{i,1\text{km}} - A_{i,500\text{m}}$ | Count |
| *City-level variables* | | | |
| `airbnb_count_city` ($A_c$) | Constructed | Total Airbnb listings in the city | Count |
| `mean_rent_city` | Constructed | City mean of `log_rent` | — |
| `mean_airbnb_price_city` | Constructed | City mean displayed Airbnb price | CAD |
| `share_entire_home_city` | Constructed | City share of entire-home listings | Proportion |

*Notes:* All monetary values are in Canadian dollars. "Constructed" indicates variables derived during cleaning and merging.

**Table A2.** City Name Standardisation Mapping (Airbnb Data)

| Original name(s) | Standardised name |
|---|---|
| Montréal | Montreal |
| Québec City, Quebec City, Quebec | Québec |
| Levis | Lévis |
| Saint Come | Saint-Côme |
| Sainte Adele, Sainte-Adele, Ste-Adèle | Sainte-Adèle |
| Saint Sauveur, Saint-Sauveur-des-Monts | Saint-Sauveur |
| Ste-Agathe-des-Monts | Sainte-Agathe-des-Monts |

by comparing each element against the radius. The same routine computes the pairwise distances between rental listings used by the Conley estimator.

### B.2. Spatial Weights, Estimation and Impacts

**Projection.** Nearest neighbours are determined on coordinates projected to kilometres by an equirectangular projection centred on the sample centroid, $x = 111.32 \cos(\bar{\phi})\,(\lambda - \bar{\lambda})$ and $y = 111.32\,(\phi - \bar{\phi})$ with angles in degrees. An earlier version of the analysis searched for neighbours in raw degree coordinates; at the latitude of Quebec one degree of longitude spans about 0.69 of a degree of latitude, so those neighbourhoods were stretched east–west. The correction changes the spatial estimates only marginally ($\hat{\rho}$ from 0.137 to 0.134, $\hat{\lambda}$ from 0.539 to 0.544).

**Weights.** The weights matrix $\mathbf{W}$ is a $k$-nearest-neighbour matrix with $k = 5$: for each rental listing $i$, $\mathcal{N}_i$ is the set of the five other rental listings closest to $i$ in projected coordinates, and

$$w_{ik} = \begin{cases} 1/5 & \text{if } k \in \mathcal{N}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{A.2}$$

so that $\sum_k w_{ik} \ln(\text{rent}_k)$ is the mean log rent of the five nearest listings. A nearest-neighbour definition guarantees that every listing has a neighbourhood, which a fixed-radius definition would not in sparse areas; the resulting graph is not fully connected (140 components), which does not affect the generalised-moments estimators. Listings in the same building share coordinates and are each other's nearest neighbours. The choice $k = 5$ balances responsiveness against noise; in dense urban areas the five nearest listings lie within a few hundred metres.

**Estimation.** The spatial lag $\sum_k w_{ik} \ln(\text{rent}_k)$ is correlated with the disturbance, so OLS estimation of Equation (7) is inconsistent. The SAR model is estimated by the spatial two-stage least squares estimator of Kelejian and Prucha (1998), which instruments the lag with the spatially lagged exogenous regressors $\mathbf{WX}$ (`GM_Lag` in `spreg`, White-robust standard errors); the SEM

by the generalised-moments estimator of Kelejian and Prucha (1999) for $\lambda$ followed by feasible generalised least squares (`GM_Error`), whose standard errors are the classical GM ones. Both include building-type dummies and fixed effects for the fifteen largest cities with the remaining cities grouped.

**Impacts.** With $\mathbf{S} = (\mathbf{I} - \rho\mathbf{W})^{-1}$, the average direct impact of regressor $x$ is $\beta_x \, n^{-1}\mathrm{tr}(\mathbf{S})$, the average total impact is $\beta_x \, n^{-1}\boldsymbol{\iota}'\mathbf{S}\boldsymbol{\iota} = \beta_x/(1-\rho)$ for a row-standardised $\mathbf{W}$, and the average indirect impact is their difference (LeSage and Pace, 2009). At the point estimates the trace is computed exactly from a sparse LU factorisation of $\mathbf{I} - \hat{\rho}\mathbf{W}$. Standard errors are obtained from 2,000 draws of $(\beta_x, \rho)$ from the estimated asymptotic normal distribution of the spatial two-stage least squares estimator; for each draw the trace is evaluated through the series $\mathrm{tr}(\mathbf{S}) = \sum_{k \geq 0} \rho^k \, \mathrm{tr}(\mathbf{W}^k)$, truncated at $k = 60$, with $\mathrm{tr}(\mathbf{W}^0) = n$, $\mathrm{tr}(\mathbf{W}) = 0$ and the higher-order traces estimated by Hutchinson's stochastic estimator with 50 Rademacher vectors. The series and the exact trace agree to five decimals at the point estimate.

### B.3. Fixed-Effects Estimation and Spatial HAC Standard Errors

The city- and FSA-fixed-effects specifications in Tables 5 and 6 are estimated by the within transformation: $\tilde{y}_i = y_i - \bar{y}_{g(i)}$ and $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}_{g(i)}$, where $g(i)$ is the city or FSA of listing $i$, and $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$. Groups with a single listing are dropped. With $\hat{e}_i$ the residuals and $\mathbf{s}_i = \tilde{\mathbf{x}}_i \hat{e}_i$ the scores, the covariance estimators share the sandwich form $\hat{\mathbf{V}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \, \hat{\mathbf{M}} \, (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ and differ in the meat $\hat{\mathbf{M}}$:

$$\hat{\mathbf{M}}_{\mathrm{HC1}} = \frac{n}{n-k-G} \sum_i \mathbf{s}_i \mathbf{s}_i', \tag{A.3}$$

$$\hat{\mathbf{M}}_{\mathrm{cluster}} = \frac{C}{C-1} \, \frac{n-1}{n-k} \sum_{c=1}^{C} \Big( \sum_{i \in c} \mathbf{s}_i \Big) \Big( \sum_{i \in c} \mathbf{s}_i \Big)', \tag{A.4}$$

$$\hat{\mathbf{M}}_{\mathrm{Conley}} = \sum_i \sum_j K(d_{ij}) \, \mathbf{s}_i \mathbf{s}_j', \qquad K(d) = \max\{0, \, 1 - d/\bar{d}\}, \tag{A.5}$$

where $G$ is the number of absorbed groups, $C$ the number of clusters, $d_{ij}$ the great-circle distance between listings $i$ and $j$, and $\bar{d}$ the cutoff (1, 2, 5 or 10 km). The Bartlett kernel guarantees a positive semi-definite estimate (Conley, 1999; Kelejian and Prucha, 2007). Pairs at zero distance (units in the same building) receive weight one, so the Conley estimator also absorbs within-building correlation. The double sum is evaluated exactly in chunks of 500 listings against all others, without a sparsity approximation.

### B.4. Quantile Regression Estimation

The quantile regression at quantile $\tau$ (Equation 9) minimises the asymmetrically weighted sum of absolute residuals,

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau\big(\ln(\text{rent}_i) - \mathbf{X}_i'\boldsymbol{\beta}\big)\,, \qquad \rho_\tau(u) = u\,(\tau - \mathbf{1}[u < 0]), \tag{A.6}$$

(Koenker and Bassett, 1978), solved by iteratively reweighted least squares in `statsmodels`. Standard errors are the asymptotic estimates based on a kernel estimate of the conditional density of the response at the fitted quantile (the `statsmodels` default), which are consistent under independent but not necessarily identically distributed errors and do not account for spatial dependence.

### B.5. Machine-Learning Models

**Regularised linear models.** The LASSO (Tibshirani, 1996) solves $\min_{\boldsymbol{\beta}}\{\frac{1}{2n}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\}$, and the elastic net (Zou and Hastie, 2005) adds an $\ell_2$ penalty with a mixing parameter; both regularisation parameters are chosen by five-fold cross-validation within the training data, and features are standardised.

**Tree ensembles.** The random forest (Breiman, 2001) averages 500 regression trees of maximum depth 15, each grown on a bootstrap sample with random feature subsampling. Gradient boosting (Friedman, 2001) fits 500 trees of maximum depth 5 sequentially to the residuals of the current ensemble with a learning rate of 0.05. SHAP values (Lundberg and Lee, 2017) are computed on the hold-out set with the tree explainer.

**Evaluation.** The hold-out evaluation uses a random 80/20 split with a fixed seed. The cross-validation uses five folds under two designs: random folds (shuffled with the same seed) and spatially blocked folds, in which `GroupKFold` assigns whole FSAs (181 groups in the 4,370-listing sample) to folds, so that no test listing shares a neighbourhood with a training listing. Within each training fold the scaler and the regularisation parameters are re-estimated.

## C. Additional Tables

This appendix reports the correlation matrix of the 500 m exposure measures and dwelling controls referred to in Section 3. Full regression output for every specification, including the fixed-effect coefficients, the quantile-regression control coefficients and the machine-learning cross-validation folds, is produced by the replication scripts in the paper's repository (`scripts/05` to `scripts/12`) and is not reproduced here.

**Table A3.** Correlation Matrix — Log Rent, Airbnb Exposure (500 m) and Dwelling Controls

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| (1) Log monthly rent | 1.00 | | | | | | |
| (2) Airbnb count (500 m) | 0.14 | 1.00 | | | | | |
| (3) Airbnb density (500 m) | 0.14 | 1.00 | 1.00 | | | | |
| (4) Mean Airbnb price (500 m) | 0.07 | 0.07 | 0.07 | 1.00 | | | |
| (5) Entire-home share (500 m) | 0.07 | -0.19 | -0.19 | 0.14 | 1.00 | | |
| (6) Bedrooms | 0.53 | -0.19 | -0.19 | 0.02 | 0.08 | 1.00 | |
| (7) Bathrooms | 0.60 | 0.00 | 0.00 | 0.08 | 0.08 | 0.48 | 1.00 |

Pearson correlations on the 4,370 rental listings with at least one Airbnb listing within 500 m (the composition variables are undefined otherwise). At a fixed radius the density is a rescaling of the count, hence the correlation of 1.00.

The exposure count is positively but weakly correlated with log rent (0.14) and negatively correlated with bedrooms (−0.19), which accounts for the increase in the Airbnb coefficient when dwelling controls are added (Section 5.1). The entire-home share is negatively correlated with the count (−0.19): neighbourhoods with many Airbnb listings have a lower share of houses, chalets and condos among them, consistent with urban rental units dominating dense STR clusters.