applied sciences

MDPI

Article

Predicting Commercial Performance from Land Use Configuration: A Spatially and Temporally Validated Machine Learning Framework for Seoul's Commercial District Census

Geiguen Shin [1,2,3] **and Jimin Chae** [4,*]

[1] Department of Public Administration, Kookmin University, Seoul 02707, Republic of Korea; gshin@kookmin.ac.kr
[2] School of Public Administration, University of Central Florida, Orlando, FL 32816, USA
[3] The Convergence Institute of Healthcare and Medical Science, College of Medicine, Catholic Kwandong University, Incheon 22711, Republic of Korea
[4] Graduate School of Convergence Industry, Sungshin Women's University, Seoul 02844, Republic of Korea
* Correspondence: jmchae@sungshin.ac.kr

Abstract

Predictive models built from administrative urban data are increasingly used to guide planning decisions, but their reported accuracy is trustworthy only if it is evaluated on places and years the model has not seen. We develop such a framework for a concrete question: how commercial performance varies with the intensity and extent of urban commercial land. Using two prediction waves from Seoul's commercial district census (2023→2024 and 2024→2025, 3044 district waves from 1530 districts), an XGBoost model interpreted with SHAP and accumulated local effects attains R^2 = 0.325 under nested leave-one-gu-out spatial cross-validation, 0.532 in temporal transfer, and 0.299 in a fully nested combined spatio-temporal holdout, while random cross-validation inflates the spatial figure to 0.517. An untransformed linear model collapses to 0.007, but spline regression on transformed predictors reaches 0.354; the gain comes from nonlinearity, not boosting. Sales per store rise with store density up to an exhaustion threshold near 44 stores/ha and are flat beyond it; the district area profile begins to flatten around 14 ha. The centred ALE curves cross zero near 25 stores/ha and 6 ha, which are sample-relative reference points, not economic thresholds. Alley districts sit mostly below the density reference point, traditional markets are distinguished by small land extent rather than intensity, and developed districts are locally unpredictable. Within-district first differences do not reproduce the cross-sectional density gradient, so these quantities are configurational benchmarks, not causal effects or intervention targets.

Keywords: machine learning; explainable AI; spatial cross-validation; SHAP; accumulated local effects; urban prediction; land use intensity; commercial land; Seoul

Check for updates

Academic Editor: Salvador García-Ayllón Veintimilla

Received: 6 August 2026
Revised: 23 October 2026
Accepted: 26 October 2026
Published: 29 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Predictive models trained on administrative urban data are increasingly used to guide planning and investment decisions, yet their reported accuracy routinely overstates how well they generalize to places and time periods the model has not seen—a gap with direct consequences once the guidance is acted upon. This paper develops and validates such a model for a concrete, high-stakes application. Urban land is scarce—globally, cities have long consumed land faster than they have added people [1]—and how a city uses

the share of it devoted to commerce is a first-order land question. The global land use efficiency agenda—crystallized in Sustainable Development Goal 11.3 and its indicator of the ratio of land consumption rate to population growth rate—has made the productivity of urbanized land an explicit object of measurement and policy [2,3]. Yet at the intra-urban scale on which planning instruments actually operate, the relationship between how intensively commercial land is occupied and how much economic value it yields remains thinly measured. A rapidly growing analytics literature relates the configuration of urban land to "vitality", but vitality is usually proxied by the presence of activity—footfall, mobile phone traces, points of interest—rather than the realized economic output of the establishments occupying the land [4]. For the small businesses whose livelihoods depend on it, and for planners diagnosing commercial districts, realized sales offer information about commercial activity beyond whether people simply pass through; they do not, however, measure merchant profitability.

Commercial districts are the elementary territorial units through which cities organize this allocation. Korea has institutionalized them to an unusual degree: national legislation designates and supports traditional marketplaces [5], and Seoul, the capital of South Korea, operates a Commercial District Analysis Service that delineates the complete population of the city's districts—1090 alley districts, 249 developed districts, and 305 traditional markets—and publishes quarterly, spatially referenced, card-transaction-based estimates of sales for each, alongside demographic and land attributes [6]. Each district in this infrastructure is, in effect, a bounded parcel of urban commercial land with a measured geometry (area), a measured intensity of occupation (stores per hectare), a measured land price, and a measured economic yield. The stakes of understanding this system are high: Korea sustains one of the largest self-employed populations among advanced economies, so the performance of commercial land maps directly onto household livelihood security and a sizeable apparatus of public support [7].

Two gaps limit what the existing evidence can tell a land use planner. The first concerns the outcome. Machine learning studies of urban vitality and commercial space overwhelmingly model activity proxies—social media intensity, street-view populations, review counts—rather than transaction-based sales [4,8,9], while the Korean empirical literature that does use district sales has relied mainly on linear specifications or typological description [10–13]. The second and less visible gap concerns validation. Because urban observations are spatially autocorrelated [14], random cross-validation lets models exploit spatial proximity between training and test cases, systematically overstating how well relationships transfer to unsampled locations [15–17]. The problem is not hypothetical in this literature: a recent machine learning study of commercial land, published in *Land*, reports that the predictive R^2 for land value falls from 0.622 under random splits to 0.050 under spatial cross-validation, yet it retains random splits for its main attribution analysis [18]. For land policy, this matters directly: a predictive regularity that does not transfer across space is not a reliable diagnostic reference for other districts.

This study takes the opposite path and makes transferable prediction—across space and across time—the primary evidentiary criterion. We link each district's complete prior-year land use and demand configuration to its next-year sales per store on a sample drawn from Seoul's complete commercial district census in two temporally separated prediction waves: 2023 conditions predicting 2024 outcomes and 2024 conditions predicting 2025 outcomes (3044 district waves from 1530 districts). An XGBoost model is evaluated on a ladder of increasingly demanding designs: conventional plain random cross-validation; random folds blocked by district, which remove the leakage created when the two waves of one district straddle a split; nested leave-one-gu-out (LOGO) spatial cross-validation across Seoul's 25 autonomous districts, with hyperparameters tuned strictly inside of each

training set; and strict temporal transfer, in which a model tuned and fit entirely on the first wave predicts the second. The ladder does more than rank designs—it decomposes conventional validation optimism into its repeat-unit and spatial components on the same data. Interpretation combines SHAP attributions with accumulated local effects (ALE), which—unlike partial dependence—remain valid when predictors are dependent on construction, as store density and district area are here (density ≡ stores/area) [19–21]. We benchmark the structural model against a persistence model (prior-year sales) and against linear and spline regressions evaluated by the identical nested design, estimate type-specific models, triangulate the machine learning thresholds with a parametric interaction specification whose inference uses the wild cluster bootstrap appropriate to 25 clusters [22], quantify uncertainty in the effect curves with a cluster bootstrap, and diagnose residual spatial structure with Moran's I [23].

Seoul is a particularly informative setting for these questions for four reasons. First, and most important for observing the full fitted intensity profile, Seoul occupies the upper extreme of commercial land use intensity distribution. Some 9.3 million residents live on 605 km^2 in a compact, transit-oriented form, and the intensity of its commercial districts ranges from a handful of stores per hectare to well beyond one hundred. Estimating where the intensity–performance profile flattens requires observations on both sides of that point; in cities whose commercial land rarely approaches such intensities [24,25], only the ascending branch of the curve is realized, and the saturation threshold is not identified by any amount of data. Seoul is one of the few metropolitan systems dense enough to reveal the full geometry of Equation (3). Second, the measurement infrastructure makes commercial land an observable object rather than an inferred one: the boundary geometry, occupation intensity, official land price, and card-transaction-based sales of every district in the city are published on a common quarterly standard [6], so the underlying data infrastructure is census-based rather than sample-based and rests on realized transactions rather than activity proxies. Third, the institutional typology partitions that census into three regimes—organically formed alley clusters, large consolidated developed districts with metropolitan catchments, and statutorily bounded marketplaces [5]—providing an administratively defined basis for examining the type heterogeneity that H4 predicts. Fourth, the stakes are unusually sharp: Korea sustains one of the largest self-employed populations among advanced economies [7], and commercial land in Seoul has been the arena of decade-long politics of rising rents, tenant displacement, and commercial gentrification [11,26,27], which makes the land price channel of Equation (4) a live policy question rather than an academic one.

This study contributes three ways—empirical, methodological, and practical—and around three findings. First, the study offers a stricter evidence standard with a quantified validation premium. The nonlinear model attains out-of-sample $R^2 = 0.325$ across unseen gu and 0.532 across waves, conventional random validation inflates the spatial figure to 0.517 (59%), district-blocked random folds show that nearly all of that optimism is repeat-unit leakage rather than spatial leakage, and a linear model in untransformed levels collapses to $R^2 = 0.007$ while regression with log-transformed predictors and spline terms reaches 0.354–0.361 under the identical nested design—what transfers is a nonlinear structure, which flexible regression and tree ensembles recover alike. Second is configurational benchmarks for commercial land, each labeled by what it identifies: per-store sales rise with log store density up to an exhaustion threshold near 44 stores/ha—located by a centering-invariant criterion and corroborated by slope break estimates in the mid-40s to mid-50s—and are flat beyond it; returns to district area flatten at about 14 ha; and the density and area where the sample-centered ALE curves cross zero (about 25 stores/ha and 6 ha) are reported as relative reference points that shift with the reference population—each estimated with

cluster-bootstrap uncertainty bands and replicated across two independent prediction waves. Finally, this study offers type-differentiated diagnostics: alley districts lie mostly below the density reference point (median 18.4 stores/ha) on the rising segment of the profile; traditional markets are distinguished by land extent rather than intensity (median area 2.8 ha, on the steep part of the area profile); and developed districts are essentially unpredictable from local land configuration, which is itself a policy-relevant fact.

The remainder of the paper proceeds as follows. Section 2 situates commercial districts as micro-units of urban commercial land, develops the agglomeration–saturation–land cost mechanisms, formalizes them in a stylized model of commercial land yield, reviews the spatial validation problem in interpretable machine learning, and states four hypotheses. Section 3 describes the data, variables, models, and validation design. Section 4 reports the results; Section 5 discusses their interpretation and land policy implications, sets out what the Seoul case does and does not license (Section 5.6), and states the limitations (Section 5.7); and Section 6 concludes.

2. Literature Review and Hypotheses

2.1. Commercial Districts as Micro-Units of Urban Commercial Land

The intellectual case for treating the shopping street as a unit of analysis begins with Jacobs, for whom the fine grain of mixed, continuously occupied street frontage was the elementary condition of urban vitality [28]; her propositions have been verified empirically at the scale of Seoul's own streets [29]. Urban economics arrives at a similar unit from the demand side: in the consumer city framework, locally supplied commercial amenities are not a by-product of urban success but one of its drivers, and the district is where those amenities are produced on land [30]. What distinguishes the Korean setting is that this analytical unit is also an administrative one. Statute defines and supports traditional marketplaces [5]; policy distinguishes organically formed alley districts embedded in residential fabric from large, consolidated developed districts with metropolitan drawing power; and Seoul's public data infrastructure fixes the boundary geometry of every district and attaches quarterly measurements to it [6]. The result is rare from an international perspective: a census-based inventory, rather than a sample, of a metropolis's commercial land units, each with area, occupation intensity, land price, and transaction-based sales observed on a common standard.

The deeper root of the district-as-unit idea is central place theory. Christaller's account of retail and service provision organizes commerce into a nested hierarchy of centers, each viable only where the demand of its market area exceeds a threshold and each reaching only as far as the range consumers will travel [31]; Berry's empirical program carried the framework into intra-urban space, mapping the ribbons, specialized areas, and nucleated centers of the city and treating the retail center as a measurable spatial object [32]. Two of the theory's primitives translate directly into the land variables this study estimates. The threshold prefigures a minimum efficient scale for the land base of a commercial center—a theoretical scale requirement below which a full assortment may be difficult to assemble—and the hierarchy of centers anticipates the very typology through which Korean policy now reads its commercial land, from neighborhood alley clusters to metropolitan developed districts. What the classical theory could establish only through cartographic inference, Seoul's census infrastructure renders directly observable: every "center" has a recorded territory, an occupation intensity, a land price, and an economic yield.

A substantial Korean literature has used this and related infrastructure to study what makes districts succeed. Sales levels have been related to locational and compositional characteristics [10,33], survival of establishments to district attributes [34], district trajectories to typologies of growth and decline before and after COVID-19 [11,27], vitality to

diversity indices [13], and the structure of neighborhood commerce to the competitive shocks transmitted by large-format retail [35], with recent work adding social media signals to locational fundamentals [12]. Internationally, the condition of such districts has become a policy concern in its own right, from the adaptive resilience of UK high streets under successive shocks [36] to multidimensional benefit assessment of commercial parcels in Chinese megacities [18]. Yet across these literatures, the two variables that define a district as land—the extent of its territory and the intensity with which stores occupy it—are usually treated as controls rather than the objects of interest. This study inverts that emphasis: area and density are the focal predictors, and their nonlinear returns are the quantities we seek to estimate.

2.2. Agglomeration on Commercial Land: Why Intensity Should Raise the Yield

How can having more stores on a hectare of commercial land increase the sales of each store? Urban economics supplies layered answers. The Marshallian triad—input sharing, labor market pooling, and knowledge spillovers—is the source of the higher productivity of co-located producers [37], and modern micro-foundations formalize the sharing, matching, and learning processes that drive the link between density and productivity [38]. Retail provides an additional externality from the demand side: as consumers economize on search, retailers tend to cluster to attract comparison shoppers, in line with Hotelling's principle of minimum differentiation [39], and retail geography has long regarded this as the mechanism that nucleates shopping streets [32]. Jacobs also introduced a diversity variant: it is not the similarity of adjacent activities but their diversity that creates cross-visits and new combinations [40], and face-to-face "buzz" among proprietors is a distinctly local way of discovering what sells [41].

These micro-mechanisms add up to cumulative causation. Increasing returns geography suggests that market access draws activity and that concentrated activity in turn strengthens market access, so commercial concentration is self-reinforcing up to the point of congestion [42,43]; empirical evidence of agglomeration economies spans industries, cities, and spatial scales [44,45]. At the small scales most relevant to this study, the evidence is direct: block density and building typology explain urban vitality in Shenzhen [46]; street block land use intensity is strongly related to vitality, non-monotonically, across five Chinese megacities [4]; and interpretable machine learning studies have consistently recovered strong, nonlinear built environment effects on street vitality [9]. Within a district, higher intensity of occupation should therefore increase per-store yield over a wide range.

2.3. Saturation, Land Costs, and District Geometry: Why the Returns Must Bend

The curve is bent by three forces. The first is demand pool division, which occurs when a district's customer base is limited to its catchment: beyond a certain point, any further increase in intensity divides a market that is not growing with the district. Organizational ecology formulates this most sharply as density dependence, in which rising organizational density first increases the legitimacy of the population and then diminishes the viability of additional entrants [47]; the retail resilience literature adds that crowded, homogenized centers lose exactly the diversity and independent character that attract consumers [28,36]. The second is congestion as commonly understood: frontage, access, and amenities are competing claims on a given land parcel. Set against the agglomeration mechanisms of Section 2.2, these forces predict a rising then flattening—possibly inverted-U—relationship between land use intensity and per-store yield, a shape already visible in block-level vitality studies [4] and in the threshold effects reported for commercial land benefits [18].

The third force acts within the land market. In the Alonso–Muth tradition, the cost of access is internalized in the price of commercial land [48,49], so the value of commercial

land is an indicator of the occupancy cost that establishments must pay. The prediction is a negative conditional effect on per-store performance—one that is distinct from, and works against, the accessibility value otherwise signaled by high prices. Finally, geometry. According to the threshold logic of central place theory, a district's land area must reach a base level, because a land base that is too small cannot support the range of assortment required to meet the threshold, while further land beyond that level adds assortment at a decreasing rate [31,32]. The modifiable areal unit problem also affects all of these intensity and territory estimates because both are measured on administrative delineations [50], but it is less problematic here than is often the case, as these are the same administrative delineations on which designation and support policy actually operate (see Section 5).

2.4. A Stylized Model of Commercial Land Yield

The mechanisms in Sections 2.2 and 2.3 can be compressed into a stylized model that makes the hypotheses precise. Let a district of type t occupy commercial land of area A (ha) with N establishments so that land use intensity is $D \equiv N/A$ (stores/ha). The district draws expenditure from its catchment in proportion to a demand shifter $\Phi(X)$—summarizing the households, workers, and passing population available to it—and to an attraction function $\alpha_t(D, A)$ that embodies the agglomeration externalities of Section 2.2: comparison shopping and variety gains make the district more than the sum of its stores. Realized expenditure is then divided among the N establishments—the demand division force of Section 2.3—so that per-store sales are

$$S = \frac{\Phi(X)\,\alpha_t(D,\ A)}{N},\ N = D\,A \tag{1}$$

Taking logarithms isolates a composite land term $m_t(D, A)$ that collects everything the district's land configuration contributes to per-store yield:

$$ln\,S\ =\ ln\,\Phi(X)\ +\ ln\,\alpha_t(D,\ A)\ -\ ln\,D\ -\ ln\,A\ \equiv\ ln\,\Phi(X)\ +\ m_t(D,\ A) \tag{2}$$

The geometry of m_t is governed by a race of elasticities. Differentiating Equation (2),

$$\frac{\partial\,ln\,S}{\partial\,ln\,D} = \varepsilon_t^D(D) - 1,\ \frac{\partial\,ln\,S}{\partial\,ln\,A} = \varepsilon_t^A(A) - 1 \tag{3}$$

where ${\varepsilon_t}^D \equiv \partial\,ln\,\alpha_t/\partial\,ln\,D$ and ${\varepsilon_t}^A \equiv \partial\,ln\,\alpha_t/\partial\,ln\,A$. Wherever attraction is density-elastic (${\varepsilon_t}^D > 1$)—the regime in which comparison shopping externalities dominate [32,39]—adding stores raises every store's sales; where it is inelastic, demand division dominates. If ${\varepsilon_t}^D$ declines in D as assortment saturates the catchment [47], the density profile of ln S is single-peaked or, where ${\varepsilon_t}^D$ approaches unity without falling below it, rising and then saturating, with the turning or saturation point ${D_t}^*$ solving ${\varepsilon_t}^D({D_t}^*) = 1$; and if variety complementarities make attraction convex at sub-critical mass [28], the profile is preceded by a region of steep penalties at very low intensity. The same logic applied to A—disciplined by central place thresholds and the bounded range of catchments [31]—yields a minimum efficient scale ${A_t}^{min}$, the upper boundary of the ${\varepsilon_t}^A > 1$ region, with diminishing returns beyond it. The land market enters as a cost: under bid rent capitalization [48,49], establishments facing higher land prices economize on premises, and per-store sales fall with the resulting compression of store scale, so that conditional on Φ the pass-through of $ln\,P$ into $ln\,S$ is negative, $-\eta_t$ with $\eta_t \geq 0$. Collecting terms gives the theory-consistent reduced form the empirical model must recover:

$$ln\,S_i\ =\ ln\,\Phi(X_i)\ +\ m_{t(i)}(D_i,\ A_i)\ -\ \eta_{t(i)}\,ln\,P_i\ +\ u_i \tag{4}$$

Equation (4) isolates the occupancy cost channel; collateral and expectations channels may offset it and cannot be separately identified with the available data (Section 5.3). Equation (4) is otherwise deliberately semi-parametric. Theory restricts the sign pattern and curvature of m_t—sub-critical penalty, elastic ascent, interior exhaustion in D; threshold and satiation in A—but not its functional form. This is precisely why Section 3 estimates m_t with a flexible learner and reads its geometry from ALE curves and why Section 3.5 then checks that geometry against a quadratic parametrization whose turning points estimate D_t^* directly.

2.5. Machine Learning and Explainable AI for Urban Prediction

The empirical strategy belongs to what Breiman called the algorithmic modeling culture: let a flexible learner find the regularities in the data, then interrogate the learner [51]. Tree ensembles—random forests and gradient boosting—are the standard instruments of that culture [21,52,53], with XGBoost the dominant implementation [54], and they have been widely adopted for urban land questions: multidimensional drivers of street vitality [9], social-media-based evaluation of commercial space vitality [8], and multi-indicator benefit evaluation of commercial land parcels [18]. Interpretation rests on two instruments. SHAP values allocate each prediction exactly among the features, with consistent local attributions that aggregate into global importance [19]. Effect curves recover the shape of each relationship, and here the choice of curve matters more than usual. Partial dependence [21] averages predictions over the joint distribution of the remaining features and is therefore biased when features are dependent, whereas accumulated local effects integrate local prediction differences within the observed data envelope and remain valid under dependence [20,55]. Our two focal predictors are dependent on construction—store density is stores divided by area—so ALE is the principled primary instrument in this study, with partial dependence retained only as a robustness overlay.

2.6. The Spatial Validation Problem

Whether any estimated relationship transfers is a separate question, and it is where spatial data discipline machine learning most severely. Spatial observations violate the independence that random cross-validation assumes: near things are more related than distant things [14], spatial autocorrelation contaminates naive error estimates [56], and random splits therefore allow a model to "predict" test cases partly by memorizing their neighbors. The established remedy is structured resampling that holds out contiguous spatial blocks [15], now supported by standard tooling [57]. Applying it has repeatedly exposed severe optimism in large-scale predictive mapping [16] and motivated the concept of a model's area of applicability—the region of predictor space, and hence of geography, within which its error estimates can be trusted [17].

The debate is live rather than settled: for assessing the accuracy of a finished map, Wadoux et al. [58] argue for probability sample validation with design-based inference rather than spatial cross-validation. But the estimand of this study is not map accuracy; it is transferability—whether relationships learned in one part of a city hold in parts the model has never seen, which is exactly the question a planner must answer before applying a threshold city-wide. For that estimand, spatially blocked evaluation is matched to the question by design. Urban applications have nonetheless been slow to adopt it. The Chengdu commercial land study measured the gap directly—its land value R^2 falls from 0.622 under random splits to 0.050 under spatial blocking—yet proceeded with random validation for its main analysis [18]. We invert that choice: blocked evaluation is the primary standard for every number reported in this paper, and the random validation premium is quantified on our own data rather than left implicit.

2.7. Hypotheses

The model and review yield four hypotheses. H1 (transferable nonlinear structure). Because the land term m_t(D, A) in Equation (2) is nonlinear and saturating, any linear-in-levels approximation is globally misspecified; consistent with the algorithmic modeling premise [51] and the spatial leakage critique [15], a specification that allows the land term to bend—a tree ensemble or a spline regression—should retain predictive power for next-year sales per store, across both prediction waves, in gu never seen in training, a specification linear in untransformed levels should not, and random cross-validation should overstate transferability. H2 (intensity thresholds). Reflecting the race of elasticities in Equation (3)—comparison shopping agglomeration [37,39] against density-dependent competition [47]—the density profile of per-store sales is nonlinear and saturating: rising steeply at low intensity, where ${\varepsilon_t}^D$(D) > 1, and exhausted beyond a saturation threshold at which ${\varepsilon_t}^D$ falls to unity. H3 (minimum efficient scale). Echoing the threshold concept of central place theory [31], district land area exhibits a minimum efficient scale ${A_t}^{min}$ below which per-store performance is penalized (${\varepsilon_t}^A > 1$), with diminishing returns beyond it (${\varepsilon_t}^A < 1$). H2 and H3 are statements about changes in slope; the level at which a centered effect curve crosses zero is a relative reference point and is not the estimand of either hypothesis (Section 4.8). H4 (type heterogeneity). Because the attraction technology and cost pass-through are institutional, the type-indexed objects α_t and η_t of Equations (2)–(4) differ across organic street clusters, statutorily bounded marketplaces, and large consolidated developed districts with metropolitan catchments [5]: density-led in alley districts, area-constrained in traditional markets, and weak in developed districts whose demand shifter Φ is dominated by metropolitan-scale conditions the local predictor set does not observe.

Because the three district types vary simultaneously in scale, intensity, property rights structure, policy treatment, and customer base, H4 requires mechanism-level specification rather than a general appeal to institutional difference. We therefore state three distinguishing predictions. First, agglomeration technology: alley districts are organically formed, spatially continuous street frontage in which comparison shopping externalities operate at walking scale, so ${\varepsilon_t}^D$ should be the largest and still rising over the observed intensity range; traditional markets are statutorily bounded precincts whose assortment is already dense and internally substitutable, so additional entrants mainly divide a fixed catchment and ${\varepsilon_t}^D$ should lie near unity throughout. Second, cost transmission: developed districts occupy the highest-priced land and face the most competitive bidding for frontage, so the pass-through parameter η_t should be largest there, whereas statutory protections and potentially different tenure arrangements may weaken short-run cost transmission in traditional markets. Third, demand boundaries: alley and traditional catchments are neighborhood-scale and bounded by walking distance, whereas developed districts draw metropolitan demand through the transit network, so the district-level predictor set observes a much smaller share of their relevant demand, which yields the distinctive prediction that the developed district model should transfer worst. Because district type bundles these mechanisms with morphology and policy, we do not read type differences as the effect of land allocation alone; Section 4.7 reports how much of the raw type gap survives conditioning on measured configuration and treats the remainder as unmeasured institutional and locational difference rather than as a land configuration effect.

3. Materials and Methods

3.1. Study Area and Data

The study covers Seoul, Republic of Korea (25 autonomous districts, gu; 9.3 million residents). Data come from the Seoul Commercial District Analysis Service, which delineates the complete census of the city's 1644 commercial districts—1090 alley districts,

249 developed districts, and 305 traditional markets—and publishes quarterly attributes for each [6]. We assemble all twelve quarters from 2023Q1 to 2025Q4 (19,728 district quarters). Quarterly sales per store are constructed as total district sales, which the service estimates from card transaction data, divided by the number of stores in the district, and enter the analysis in natural logarithms; a minority of cells are suppressed (952 sales cells, 137 working population cells, 221 resident population cells, and 15 floating population cells across the panel). The panel supports two temporally separated prediction waves: wave 1 links 2023 predictor means to 2024 outcome means, and wave 2 links 2024 predictors to 2025 outcomes. Within each wave, the estimation sample retains districts with complete predictor information in all four predictor-year quarters and complete sales in all eight quarters of the two years involved, yielding 1522 districts per wave—3044 district waves from 1530 unique districts, 1514 of which appear in both waves (wave 1: 1022 alley, 243 developed, 257 traditional; wave 2: 1015, 244, and 263). Each gu contributes between 72 and 188 district waves, giving 25 well-populated spatial blocks. District centroids are georeferenced in the Korea Central Belt projection (EPSG:5181); Figure 1 maps the sample.

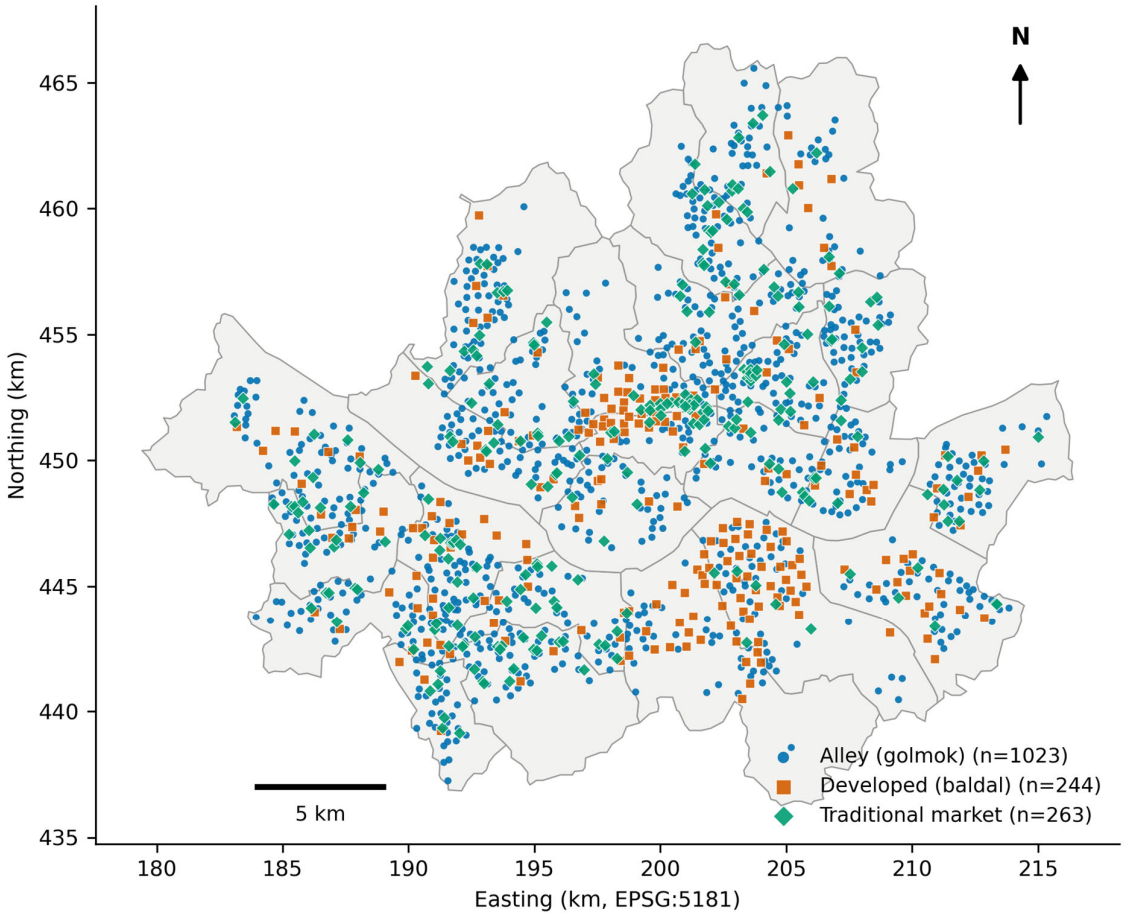


Figure 1. Study area: 1530 commercial districts by type across Seoul's 25 gu.

3.2. Variables

The dependent variable is the district's mean over the four outcome-year quarters of ln quarterly sales per store, a per-establishment measure of the economic yield of commercial land, purged of the mechanical effect of district size on total sales. Predictors are predictor-year four-quarter means of seven variables that together describe a district as a piece of commercial land and its demand environment: store density (stores/ha), the land use intensity measure; district land area (ha), the territorial extent; working population, resident population, and daily floating population, the three demand fundamentals of the

surrounding urban fabric; official land price (KRW/m^2), the land market signal; and mean years in operation of incumbent establishments, a measure of business stock maturity. Two indicator variables distinguish developed districts and traditional markets from the alley baseline, and a wave indicator distinguishes the second prediction wave for ten features in total. The predictor-year mean of ln sales per store is used only in persistence benchmarks, never in the structural model. Table 1 reports definitions and pooled descriptive statistics. The predictor set is deliberately parsimonious and administratively derived; richer built environment inventories—street network centrality, transit access, land use mix, establishment composition [18,59]—are not available on the district census, and their absence is discussed as a limitation in Section 5.7.

Table 1. Variable definitions and descriptive statistics (pooled two-wave sample; n = 3044 district waves).

Variable	Unit	Mean	SD	Median	Min	Max
ln quarterly sales per store, outcome-year mean (dependent variable)	ln KRW	16.78	1.12	16.89	9.45	20.80
Store density, predictor-year mean	stores/ha	36	72	24	1	2108
District land area	ha	10.08	9.83	7.71	0.21	171.36
Working population, predictor-year mean	persons	2478	8299	486	2	134,918
Resident population, predictor-year mean	persons	2418	2286	1756	1	20,525
Daily floating population, predictor-year mean	persons	864,662	863,893	611,554	7467	7,854,667
Official land price, predictor-year mean	KRW/m^2	4,732,803	2,707,357	3,756,000	301,600	15,737,956
Mean years in operation, predictor-year mean	years	8.99	1.87	8.75	3.27	21.25
ln quarterly sales per store, predictor-year mean (benchmarks only)	ln KRW	16.74	1.08	16.85	9.45	20.73
Developed district (baldal) indicator	0/1	0.160				
Traditional market indicator	0/1	0.171				
Wave 2 indicator (2024→2025)	0/1	0.500				

Note: Each observation is one district wave: wave 1 links 2023 predictors to 2024 outcomes and wave 2 links 2024 predictors to 2025 outcomes (1522 districts per wave, 1530 unique districts, 1514 present in both waves). Quarterly sales per store are total district sales divided by store count, entered in natural logarithms. Pooled composition: alley (golmok) n = 2037; developed (baldal) n = 487; traditional market n = 520.

3.3. Model, Attribution, and Effect Curves

We estimate gradient-boosted regression trees (XGBoost [54]). The prediction is an additive ensemble of K regression trees fit by forward stagewise minimization of a regularized objective,

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ \mathcal{L} = \sum_i (y_i - \hat{y}_i)^2 + \sum_k \left[\gamma T_k + \frac{1}{2}\lambda \|w_k\|^2\right] \quad (5)$$

with T_k the number of leaves and w_k the leaf weights of tree k. We use K = 400 trees, learning rate 0.05, row and column subsampling of 0.8, and λ = 1, tuning maximum depth over {3, 4} and minimum child weight over {5, 10} by the nested design of Section 3.4; the modal selected configuration is depth 3 with minimum child weight 10, a strongly regularized setting relative to the 3044 district wave observations. Global attribution uses

exact TreeSHAP values [19], which decompose every prediction as $\hat{f}(x_i) = \varphi_0 + \Sigma_j \varphi_{ij}$ with the Shapley weighting

$$\varphi_{ij} = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left[f_{S\cup\{j\}}(x_i) - f_S(x_i) \right] \quad (6)$$

over feature subsets S of the p-feature set F, and global importance summarized as $I_j = n^{-1}\Sigma_i \mid \varphi_{ij} \mid$. Effect curves are estimated with accumulated local effects [20] over 40 quantile bins per feature: with bin edges $z_{j,0} < \ldots < z_{j,K_j}$ and $k_j(x)$ the bin containing x, the estimator accumulates average local prediction differences within bins,

$$ALE_j(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_k} \sum_{i \in N_k} \left[\hat{f}\left(z_{j,k},\ x_{i,-j}\right) - \hat{f}\left(z_{j,k-1},\ x_{i,-j}\right) \right] - c_j \quad (7)$$

where N_k collects the n_k districts whose x_{ij} falls in bin k and the constant c_j centers the curve to mean zero over the sample. Because of this centering, the level at which an ALE curve crosses zero is defined relative to the sample: it is the value at which the mean-centered ALE contribution of that feature equals zero, not necessarily the value at which the overall predicted sales level equals the sample mean. We therefore treat zero crossings (labeled "average-yield crossings" for continuity) as relative reference points and locate shape-based thresholds separately (Section 3.6). ALE is the primary instrument because store density and district area are dependent on construction (density ≡ stores/area), the condition under which partial dependence [21] can average over poorly supported predictor combinations; centered partial dependence is overlaid as a robustness check. Uncertainty in the ALE curves for the three land variables (density, area, land price) is quantified by a cluster bootstrap that resamples the 25 gu with replacement (120 replications), refits the model, and re-evaluates the curve on the common grid; we report pointwise 90% bands.

3.4. Spatial Cross-Validation Design and Benchmarks

The primary evaluation is nested leave-one-group-out (LOGO) cross-validation with the 25 gu as spatial groups: each gu—both of its waves together—is held out entirely, hyperparameters are selected by an inner three-fold group cross-validation on the remaining 24 gu, the tuned model is refit on all training gu, and predictions are made for the held-out gu, so no information from a test gu, including through tuning, ever reaches the model that predicts it. Formally, with $\hat{f}_{-g}$, the tuned ensemble fit without group g and $g(i)$ the gu of district i, pooled out-of-fold performance is

$$R^2_{LOGO} = 1 - \frac{\sum_i \left(y_i - \hat{f}_{-g(i)}(x_i) \right)^2}{\sum_i (y_i - \overline{y})^2} \quad (8)$$

with RMSE and MAE defined analogously on the same out-of-fold predictions and per-fold R^2 characterizing the geography of transferability. Conventional validation optimism is then measured and decomposed with two random designs run through the identical nested procedure: plain random 10-fold splits of district waves, replicating standard practice, in which the two waves of one district can fall on opposite sides of a split; and district-blocked random 10-fold splits, which keep both waves of a district together and therefore differ from spatial LOGO only in the absence of geographic blocking. The gap between the plain and district-blocked designs isolates repeat-unit leakage; the residual gap to spatial LOGO isolates spatial leakage. A fourth design tests temporal transfer: a model tuned (by inner group cross-validation) and fit entirely on wave 1 predicts wave 2. Four benchmarks complete the design: (i) an OLS model on the identical ten features, in untransformed levels, under the same spatial LOGO evaluation; (ii) a persistence model—OLS on the

predictor-year mean of ln sales per store alone—which bounds how far the outcome is predictable from its own past; (iii) an XGBoost model combining the structural features with the predictor-year sales level, testing whether structure adds forecasting value beyond persistence; and (iv) regression benchmarks that allow nonlinearity without boosting, evaluated by exactly the same nested LOGO design: OLS on the ten features with store density, district area, the three population measures, and land price in natural logarithms; the same with squared terms; an additive spline regression in which each continuous transformed predictor enters through a cubic B-spline basis with quantile knots, estimated by ridge regression with the number of knots (4, 6, 8, 10) and the penalty (0.01, 0.1, 1, 10) tuned by the same inner three-fold group cross-validation; and the same spline regression with separate density and area curves for developed districts and traditional markets. Benchmark (iv) separates the contribution of nonlinearity from that of tree boosting, and its effect curves are read with the same ALE estimator.

3.5. Heterogeneity, Parametric Triangulation, and Diagnostics

Type heterogeneity (H4) is examined two ways. First, separate XGBoost models are estimated within each district type on the seven numeric predictors, evaluated by the same LOGO design, with type-specific SHAP summaries and density ALE curves (bootstrap bands from 80 cluster replications). Second, the machine learning geometry is triangulated against the quadratic parametrization of the land term m_t suggested by Equation (2):

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 \frac{D_i^2}{100} + \sum_t \left[\gamma_t T_{ti} + \beta_{1t} D_i T_{ti} + \beta_{2t} \frac{D_i^2}{100} T_{ti} \right] + \delta' \tilde{x}_i + u_i \quad (9)$$

where T_{ti}, $t \in$ {bal, trad}, indicate developed districts and traditional markets and $\tilde{x}_i$ collects the controls (district area, ln land price, ln floating population, ln working population, ln resident population, years in operation, and the wave indicator). Standard errors are clustered by gu; because 25 clusters is below the range where cluster-robust asymptotics are reliable, inference for the six density terms additionally uses the wild cluster bootstrap with Rademacher weights and 999 replications [22]. The implied type-specific turning points, $D_t^* = -50\,(\beta_1 + \beta_{1t})/(\beta_2 + \beta_{2t})$, estimate the saturation point D_t^* of Equation (3). Spatial diagnostics compute global Moran's I [23] with row-standardized 8-nearest-neighbor weights and 999 permutations,

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}, \; z_i = v_i - \bar{v} \quad (10)$$

applied to the wave-2 cross-section of the outcome and of the out-of-fold residuals of the spatial model (each district enters the weights matrix once), with z_i the mean deviation of the diagnosed variable and w_{ij} the spatial weights. Robustness re-estimates the spatial LOGO evaluation after trimming district waves above the 99th percentile of store density (170 stores/ha), where a small number of micro-area districts generate extreme intensity values. All computations use Python 3.12 with xgboost 3.4, shap 0.52, scikit-learn 1.8, statsmodels 0.15, numpy 2.4, pandas 3.0, and scipy 1.17. All figures were produced with Matplotlib 3.10; the maps additionally use pyproj to transform administrative boundaries from EPSG:4326 to EPSG:5181.

3.6. Additional Validation, Reparameterization, and Robustness Battery

Six additional exercises explore the design. First, a combined spatio-temporal holdout tests the two generalization dimensions simultaneously. For each gu, the model is trained with data from the other 24 gu at wave-1 observations and tested with data at wave-2

observations from the held-out gu, meaning there are no observations in the test location or test period in the training set. Hyperparameters are selected inside of each outer fold using only the wave-1 observations of the 24 training gu by an inner three-fold group cross-validation over the grid of Section 3.3 so that no observation from the test gu or the test year informs the model that predicts it, directly or through tuning. The spatial-only and temporal-only rungs, evaluated on the same wave-2 test observations, are tuned in the same nested way. Uncertainty is quantified by a cluster bootstrap that resamples the 25 held-out gu (2000 replications) and by repeating the entire nested procedure ten times with different random seeds and inner partitions; for comparison, the three rungs are also re-estimated with the modal hyperparameters of the principal analysis held fixed. Second, as the outcome is highly persistent, we re-estimate the model on changes instead of levels, change in ln sales per store between waves, and change in outcome from a regression of the outcome on its own lag, to determine if land configuration has information about improvement or deterioration beyond the persistence of historically strong districts. Third, because store density is constructed as stores divided by area, we re-estimate the model on the primitive pair (store count, district area) and recover the implied density response by setting the store count of every district to d $\times$ $area_i$ over a grid of d, which tests whether the threshold is a property of the data or of the chosen parameterization.

Fourth, the threshold quantities themselves are treated as estimands, and we distinguish two kinds. From each of 200 gu-cluster bootstrap replications we re-extract two quantities from the refitted density ALE curve—the average-yield crossing, defined as the density at which the curve crosses zero, i.e., at which that feature's mean-centered ALE contribution equals zero, which makes it a sample-relative reference point rather than a statement about absolute predicted sales, and the exhaustion threshold, defined as the density at which the curve first attains 95% of the total rise of the curve, measured from its minimum to its maximum. We use the total rise rather than a fraction of the maximum because an ALE curve is centered to mean zero, so a rule expressed as a fraction of the maximum would depend on that arbitrary centering, whereas a rule expressed as a fraction of the rise is invariant to it, and we report percentile intervals; the same extraction is applied to district area and separately within each wave. To show what each kind of quantity depends on, we recompute the crossings for the same fitted curve under alternative reference distributions (each wave, each district type, and a uniform weighting over ln x) and after refitting the model on alternative sample compositions, and we estimate slope break locations in two further ways that do not depend on the level of the curve: a continuous piecewise-linear fit to the ALE curve in ln x with one or two breaks and a segmented regression fitted directly to the data, in which ln sales per store depends on hinge terms in ln density (or ln area), a cubic spline in the other land variable, and the remaining predictors, with breakpoints selected by least squares over a grid of 61 candidate values and the number of breaks compared using the Bayesian information criterion (BIC); both are bootstrapped over gu (200 replications). We also refit two alternative nonlinear learners, a random forest and a histogram-based gradient boosting machine, with the same features and the same leave-one-gu-out evaluation, to establish whether the recovered shapes are features of the data rather than of one algorithm. Fifth, because sales are a revenue-side measure, we re-estimate the model on three alternative outcomes: total district sales, sales per hectare of district land, and a land cost coverage ratio defined as annual sales per store divided by an imputed annual land cost per store, r $\times$ P_i $\times$ A_i/N_i, where r is a capitalization rate. Because r enters the log outcome additively, it shifts the intercept only and leaves the shape of every effect curve unchanged, so the shape of every effect curve and the location of every threshold are invariant to r; only the level of the ratio depends on r, and we set r = 0.04 where levels are reported. Finally, we exploit the panel structure

directly by estimating first differences across the two waves for the 1514 districts observed in both, which removes the time-invariant components of location, accessibility, business format composition, and other persistent district characteristics from the comparison.

4. Results

4.1. The Descriptive Geography of Commercial Land Performance

Table 1 summarizes the pooled sample and Figures 1 and 2 map it. The three district types occupy sharply different positions in land configuration space, and their positions are stable across the two waves. Alley districts are extensive but sparsely occupied: median area 8.0 ha at a median intensity of 18.4 stores/ha. Developed districts are both large and intensively used (medians 15.0 ha and 47.3 stores/ha) and command the highest land prices (mean KRW 6.93 million/m^2 against roughly 4.3 million in the other types). Traditional markets are the mirror image of alley districts: intensively occupied (median 52.8 stores/ha) on very little land (median 2.8 ha). Mean outcome-year ln sales per store runs 16.5 (alley), 17.2 (traditional), and 17.6 (developed), with a pooled standard deviation of 1.119 log points—a wide dispersion in the economic yield of commercial land that Figure 2 shows to be interspersed at fine grain rather than organized into large contiguous zones.



Figure 2. Spatial distribution of the dependent variable: 2025 mean ln quarterly sales per store (wave 2).

4.2. Predictive Performance and the Spatial Validation Gap

Table 2 reports out-of-fold performance and Figure 3 visualizes the validation ladder. The structural XGBoost model attains pooled R^2 = 0.325 (RMSE 0.919, MAE 0.637) across gu never seen in training—prediction error is 82% of the outcome's standard deviation using only land use and demand conditions measured the year before—and R^2 = 0.532 in strict temporal transfer, where a model tuned and fit entirely on the 2023→2024 wave

predicts the 2024→2025 wave without refitting. Transferability across space is real but geographically uneven: per-gu R^2 ranges from −0.03 to 0.53. The ladder's upper rungs quantify what conventional practice would have claimed. Plain random 10-fold validation reports $R^2 = 0.517$—an inflation of 0.192, or 59% of the honest spatial figure—while district-blocked random folds report 0.332. The decomposition is stark: of the total optimism, 0.185 points arise from repeat-unit leakage (the two waves of one district straddling a split, letting the model recognize districts rather than learning structure) and only 0.007 from spatial leakage proper, consistent with the weak spatial autocorrelation of the outcome (Section 4.6). The OLS model on the same features in untransformed levels collapses under spatial evaluation to $R^2 = 0.007$: a linear combination of the raw features predicts unseen gu essentially no better than the training mean. The regression benchmarks show that the cause is the functional form, not the absence of boosting. Entering the skewed predictors in logarithms raises out-of-fold R^2 to 0.265, adding squared terms to 0.329, and an additive spline regression on the transformed predictors reaches 0.354 (0.361 with type-specific density and area curves) against 0.325 for XGBoost; the paired 95% gu-cluster bootstrap interval for the difference from XGBoost is [−0.004, 0.064] for the additive spline and [0.008, 0.066] for the type-specific spline. Allowing the relationship to bend is therefore what makes the model transferable; a flexible regression achieves this at least as well as boosting does, and its density and area curves reproduce the boosted geometry (Section 4.8). The persistence benchmark bounds the ceiling: prior-year sales alone achieve $R^2 = 0.932$, and adding the structural variables yields no improvement (0.924). Structure adds essentially nothing to forecasting—Section 5.4 argues that this is precisely why the structural model, not the persistence model, is the policy-relevant object. Together, these results support H1 in all of its parts.

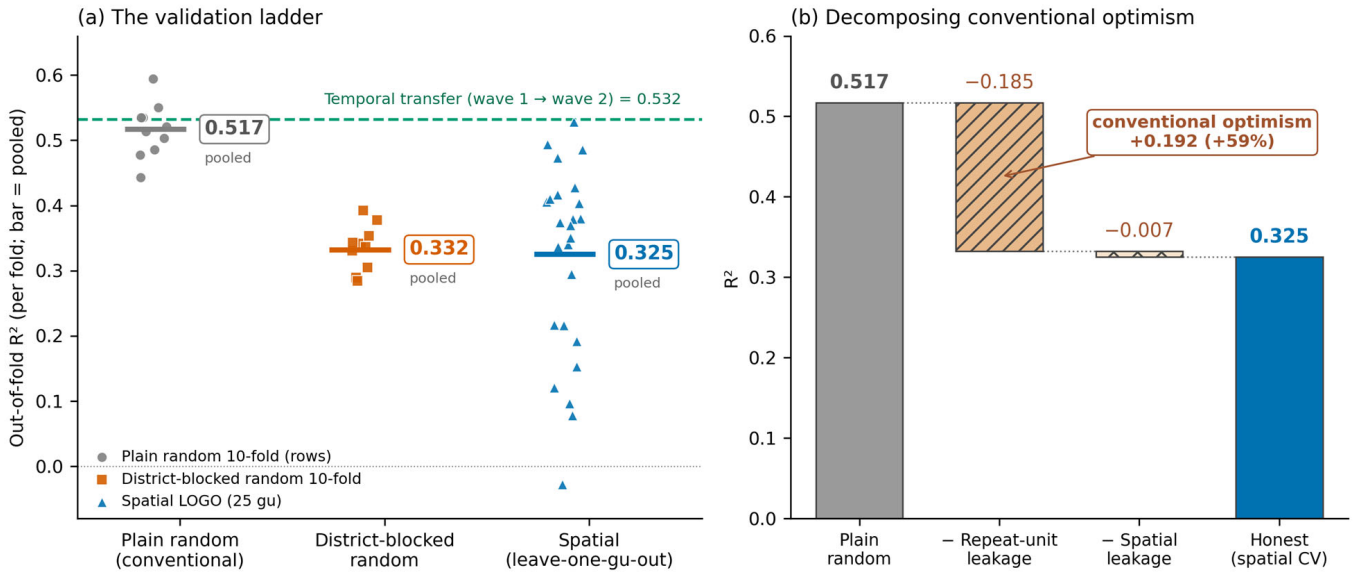


Figure 3. (**a**) Out-of-fold R^2 under plain random, district-blocked random, and spatial cross-validation; (**b**) decomposition of conventional validation optimism into repeat-unit and spatial leakage.

Table 2. Out-of-fold predictive performance under alternative validation designs.

Model	Predictors	Validation	R^2	RMSE	MAE
XGBoost (structural)	10 land use and demand features	Spatial LOGO (25 gu)	0.325	0.919	0.637
XGBoost (structural)	10 land use and demand features	District-blocked random 10-fold	0.332	0.914	0.639
XGBoost (structural)	10 land use and demand features	Plain random 10-fold (rows)	0.517	0.778	0.547
XGBoost (structural)	10 land use and demand features	Temporal transfer (train wave 1, test wave 2)	0.532	0.779	0.557
OLS (structural, untransformed)	10 land use and demand features	Spatial LOGO (25 gu)	0.007	1.115	0.745
OLS, log-transformed predictors	10 features; land and population variables in logs	Spatial LOGO (25 gu)	0.265	0.959	0.664
OLS, log + squared terms	As above plus squares of transformed predictors	Spatial LOGO (25 gu)	0.329	0.917	0.632
Spline regression (additive)	Cubic B-splines of transformed predictors, ridge	Spatial LOGO (25 gu)	0.354	0.900	0.628
Spline regression (type-specific)	As above plus type-specific density and area curves	Spatial LOGO (25 gu)	0.361	0.894	0.626
OLS (persistence)	Predictor-year ln sales per store only	Spatial LOGO (25 gu)	0.932	0.293	0.149
XGBoost (structure + persistence)	10 features + predictor-year ln sales	Spatial LOGO (25 gu)	0.924	0.309	0.164

Note: LOGO = leave-one-group-out by autonomous district (gu); both waves of a gu are held out together. District-blocked random folds keep both waves of a district in the same fold, so their only difference from spatial LOGO is the removal of geographic blocking; plain random folds additionally allow the two waves of one district to fall on opposite sides of the split, replicating conventional practice. All metrics are computed on pooled out-of-fold predictions with hyperparameters tuned inside of each training set (nested design). Regression benchmarks use the same outer folds; logarithms are taken of density, area, the three population measures, and land price, and for the spline models the number of knots and the ridge penalty are tuned by the same inner group cross-validation. Dependent-variable SD = 1.119.

4.3. Global Importance: Land Configuration Dominates, Footfall Ranks Last

Table 3 and Figure 4 summarize global SHAP attributions. The two variables that define the district as land dominate: store density (mean |SHAP| = 0.363) and district area (0.309) jointly account for 50.6% of all attributed importance. The demand fundamentals follow at a distance—resident population (0.163) and working population (0.092)—alongside business stock maturity (0.093) and official land price (0.095), with the type indicators small (0.076 and 0.063) once land configuration is controlled. Two rankings at the bottom are informative. Daily floating population, the canonical footfall measure, is the weakest structural predictor (0.052). For predicting the realized economic yield of commercial land across unseen parts of the city, how the land is configured matters roughly an order of magnitude more than how many people flow past it. And, the wave indicator ranks last overall (0.021): conditional on structure, the 2024→2025 wave differs little from 2023→2024, which is consistent with a limited average shift between waves; the stability of the threshold geometry itself is assessed directly in Section 4.7.

Table 3. Global feature importance: mean absolute SHAP values (full-sample model).

Rank	Feature	Mean |SHAP|	Share of Total
1	Store density (stores/ha)	0.363	27.3%
2	District area (ha)	0.309	23.3%
3	Resident population	0.163	12.3%
4	Official land price	0.095	7.2%
5	Mean years in operation	0.093	7.0%
6	Working population	0.092	6.9%
7	Traditional market (vs. alley)	0.076	5.8%
8	Developed (vs. alley)	0.063	4.7%
9	Floating population	0.052	4.0%
10	Wave 2 (2024→2025)	0.021	1.6%

Note: SHAP values are in units of the dependent variable (ln sales per store). The two land configuration variables (store density and district area) jointly account for 50.6% of total attributed importance; floating population is the weakest structural predictor.

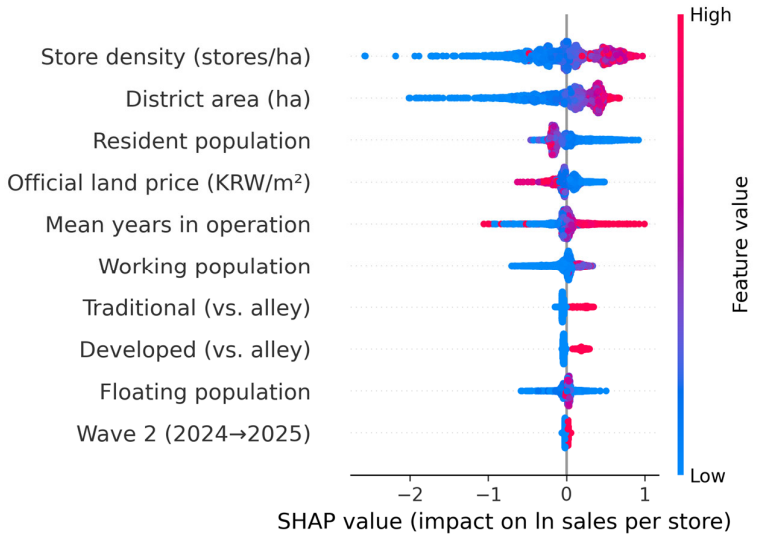


Figure 4. SHAP summary (beeswarm) for the full-sample model.

4.4. Nonlinear Performance Gradients in Land Use Intensity and District Size

Figure 5 presents ALE curves for the six continuous predictors with partial dependence overlaid; the two are nearly indistinguishable throughout, indicating that the construction-induced dependence between density and area does not distort the effect curves in practice, while ALE remains the principled choice. Figure 6 focuses on land use intensity. The density ALE rises steeply over the lower part of the observed range—districts at 5 stores/ha sit 1.16 log points below the centered mean and districts at 10 stores/ha 0.67 below—crosses its sample-centered zero near 25 stores/ha and continues to rise: 0.12 at 30, 0.65 at 45, and 0.69 at 60 stores/ha, where the 90% cluster-bootstrap band spans [0.60, 0.82], and 95% of the total rise is already achieved by 44.4 stores/ha (Section 4.7). Beyond roughly 44, the curve is flat through the entire observed range: 0.69 at 60, 0.66 at 80, a nominal maximum of 0.73 near 100, and 0.71 at 150, with bands that overlap the level at 60 throughout. The curve therefore rises throughout the lower range and stops rising at about 44 stores/ha—the exhaustion threshold, at which the curve attains 95% of its total rise from minimum to maximum, a criterion defined by the shape of the curve and invariant to its centering—with no evidence of further gains at higher intensity. The zero crossing near 25 stores/ha marks where the centered density contribution changes signs: it is a sample-relative reference point whose location depends on the reference population (Section 4.8). The spline regression benchmark of Section 4.2, overlaid in Figure 6, traces the same profile of −1.16 at 5 stores/ha, −0.66 at

10, a zero crossing at 23.1 stores/ha, and a plateau of 0.72, but, being smoother, completes 95% of its rise later, at 54.9 stores/ha. The parametric quadratic (Section 4.5) places the alley district turning point at 62.0 stores/ha, which lies above the nonparametric 95% of total rise point of 44.4 stores/ha. The two figures are also estimated on different footings—the quadratic turning point is alley-specific, while the ALE is pooled across types—so they are indicative of the same region rather than directly comparable point estimates. The two estimands are not the same quantity: forcing a symmetric parabola onto an asymmetric saturating curve pushes the implied peak to the right, whereas the ALE criterion identifies where the rise effectively stops. We therefore take 44.4 stores/ha, which is estimated directly from the nonparametric curve and carries a bootstrap interval, as the operative exhaustion estimate and read the quadratic turning point of 62.0 as an upper bound implied by a restrictive functional form. Both lie well above the average-yield crossing, so the saturation prediction of Equation (3) is supported and H2 holds; only the location of the upper bound is revised.

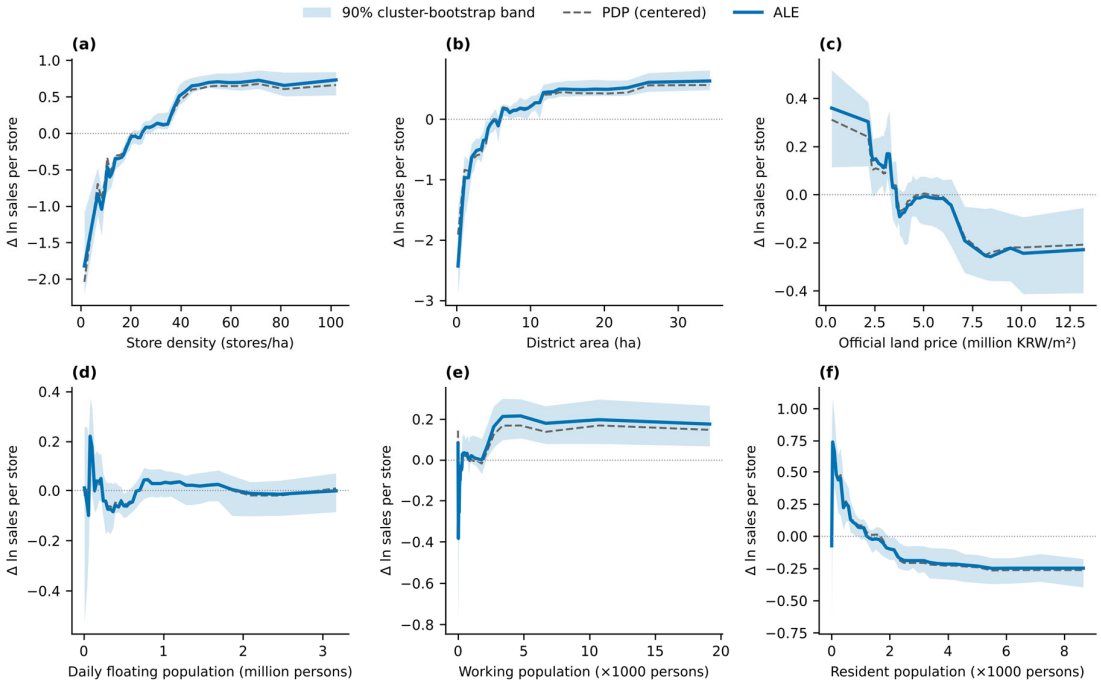


Figure 5. Accumulated local effects for the six continuous predictors (**a**–**f**) (axes truncated at the 99th percentile).

District area traces a generally rising, locally irregular fitted profile. The ALE is severely negative for micro-districts, −1.10 log points at 1 ha and −0.51 at 3 ha, crosses its sample-centered zero near 6 ha (5.9 ha), and rises to 0.21 at 10 ha and 0.49 at 20 ha before reaching its maximum of 0.63 near 34 ha, flat thereafter. Applying the first-attainment 95% of total rise rule to the fitted curve returns 13.9 ha. Because this is an operational criterion applied to a locally irregular curve, it should not be read as a unique economic scale. This is consistent with H3 at the associational level, with the diagnostic flattening region read from the curve shape rather than from its zero crossing: returns to territory are steepest among micro-districts and diminish until the profile flattens at about 14 ha (Section 4.8), whereas the 6 ha crossing marks only where the rising profile crosses the zero of its sample-centered ALE. Official land price, conditional on demand and configuration, traces a broadly declining profile: from 0.34 at 1 million KRW/m^2 to −0.24 at 8 million and essentially flat to 12 million, a decline of 0.58 log points between 1 and 8 million KRW/m^2, consistent with bid-rent cost pressure on per store margins; the parametric elasticity is −0.40. Among

the demand fundamentals, working population saturates quickly, resident population is negative across most of its range—per-store yield is lower, conditional on the rest, in heavily residential settings—and floating population, consistent with its last-place structural SHAP rank, traces a small amplitude, unstable curve.

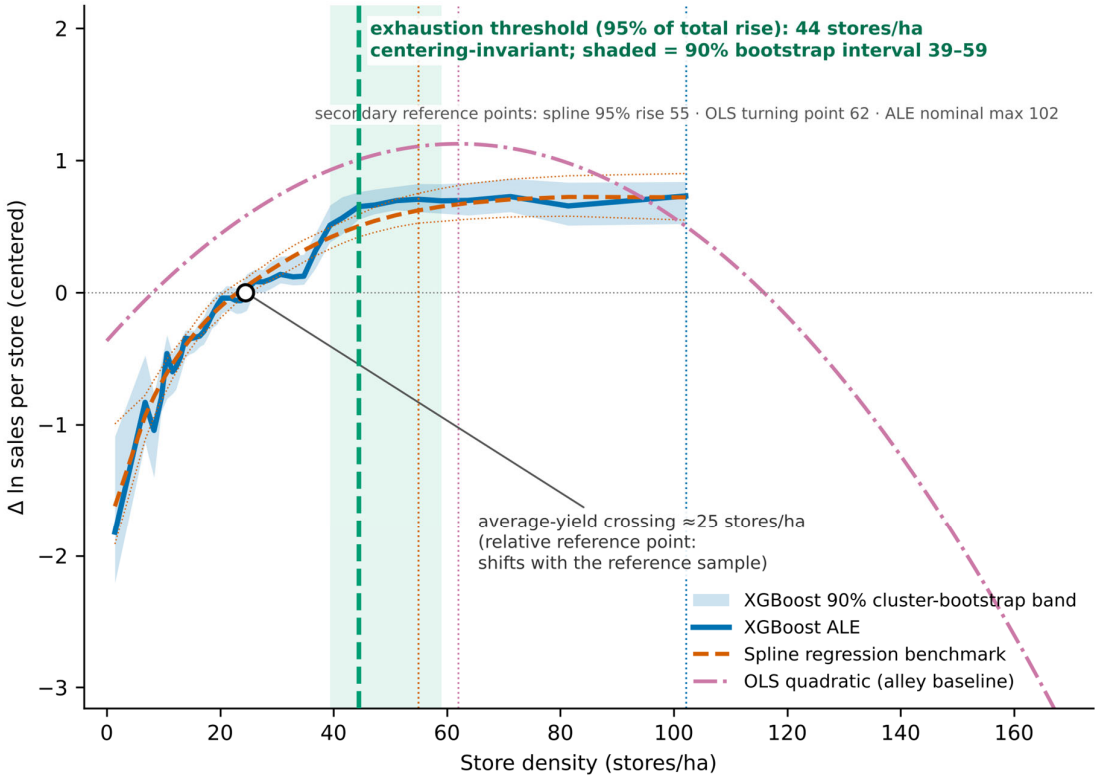


Figure 6. Performance gradient in land use intensity: XGBoost ALE, spline regression benchmark, parametric quadratic, average yield crossing, and exhaustion threshold.

Because density and area describe two dimensions of the same land configuration, their joint fitted pattern is informative for spatial diagnosis, not a prediction of what an intervention would cause. Figure 7 maps the model's predicted yield surface across the observed configuration space, masked to the region where the data provide support. Three features stand out. First, the zero-centered fitted performance contour is convex: among compared configurations, the contour lies near 35 ha at 10 stores/ha, near 6 ha at 20 stores/ha, and below 2 ha at 50 stores/ha. This is not a financial break-even line or a causal substitution rate. Second, the high-yield region is broad in the density direction while still improving mildly with territory—predicted yield is 0.36 at (30 stores/ha, 10 ha), 0.83 at (45, 15), and 0.95 at (50, 30)—which is why the profile is better described by a range than a single value. Third, the three type medians occupy visibly different parts of the space. The alley median (18.4 stores/ha, 8.0 ha) sits essentially on the zero-centered fitted performance contour (−0.03); configurations with the same territory at 30 and 45 stores/ha have predicted yields of 0.32 and 0.70. The traditional market median (52.8 stores/ha, 2.8 ha) reaches only 0.29 despite high intensity and lies on the steep part of the area profile: configurations at the same intensity on 5 and 8 ha have predicted yields 0.32 and 0.44 higher, whereas a configuration at 80 stores/ha on the same 2.8 ha differs by −0.03. The developed median (47.3 stores/ha, 15.0 ha) sits just beyond the exhaustion threshold on the intensity axis and near the point at which the area profile flattens and reaches 0.83. These are differences between configurations observed across the city, not the expected consequence of changing any district's configuration (Sections 4.7 and 5.7); as a description of where

each regime sits relative to the estimated reference points and thresholds, Figure 7 states in one panel what Section 4.5 estimates type by type.

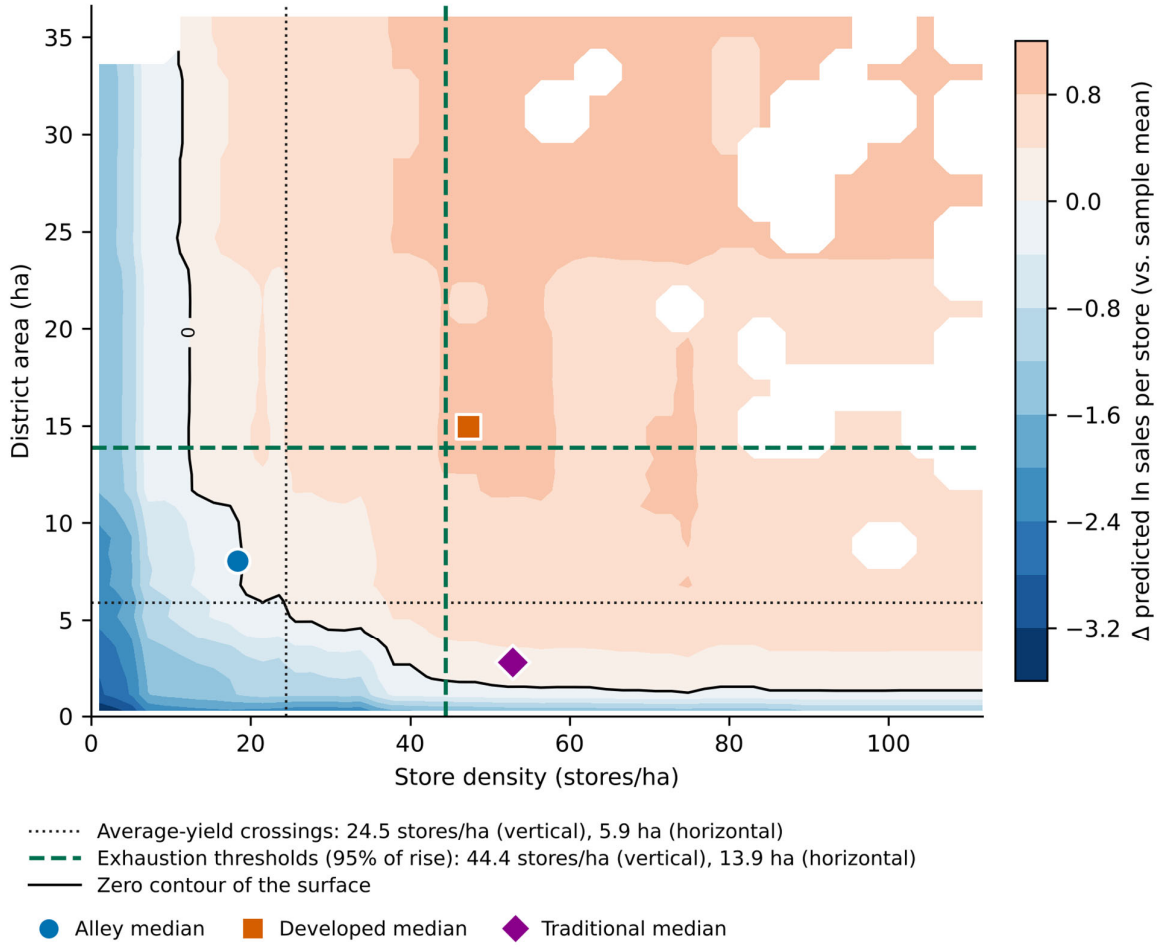


Figure 7. Predicted yield surface over the density–area configuration space, with the average yield crossings, the exhaustion thresholds for density and area, and district type medians (masked to the region of data support).

4.5. Heterogeneity Across District Types

Type-specific models (Table 4, Figure 8) reveal three distinct land regimes, replicated across both waves, supporting H4. Alley districts (n = 2037 district-waves) are the land configuration regime: spatial-CV R^2 = 0.168, with district area (0.318) and store density (0.297) the two dominant features, and a density ALE still rising at the type's upper intensity range—consistent with the fact that the median alley district, at 18.4 stores/ha, sits far below the 44 stores/ha exhaustion point. Most of Seoul's alley districts therefore lie on the rising segment of the cross-sectional density profile, below the average-yield crossing. Traditional markets (n = 520) are the area-dominated regime: the type-specific model transfers best of all (R^2 = 0.218), and district area dominates its attributions (0.521, far ahead of density at 0.307); their density ALE is comparatively flat. Given a median market area of 2.8 ha—below both the area average yield crossing (about 6 ha) and the point at which the area profile flattens (about 14 ha)—territorial extent rather than intensity is the configuration dimension most strongly associated with their per-store performance. Developed districts (n = 487) are the unpredictable regime: spatial-CV R^2 = −0.043. The local land configuration that predicts performance elsewhere carries no transferable information for the large consolidated districts whose customer sheds are metropolitan; their fortunes are set at scales the district-

level predictor set does not observe. This null is informative: it delimits where district-level indicators appear more or less informative.

Table 4. Type-specific models: spatial-CV performance and leading SHAP features.

District Type	n	Spatial LOGO R2	Top Three Features (Mean |SHAP|)
Alley (golmok)	2037	0.168	District area (0.318); store density (0.297); resident population (0.161)
Developed (baldal)	487	−0.043	Store density (0.163); mean years in operation (0.118); resident population (0.107)
Traditional market	520	0.218	District area (0.521); store density (0.307); working population (0.193)

Note: Each model is estimated on the seven numeric predictors within one district type and evaluated by leave-one-gu-out cross-validation with the same hyperparameters as the pooled model.

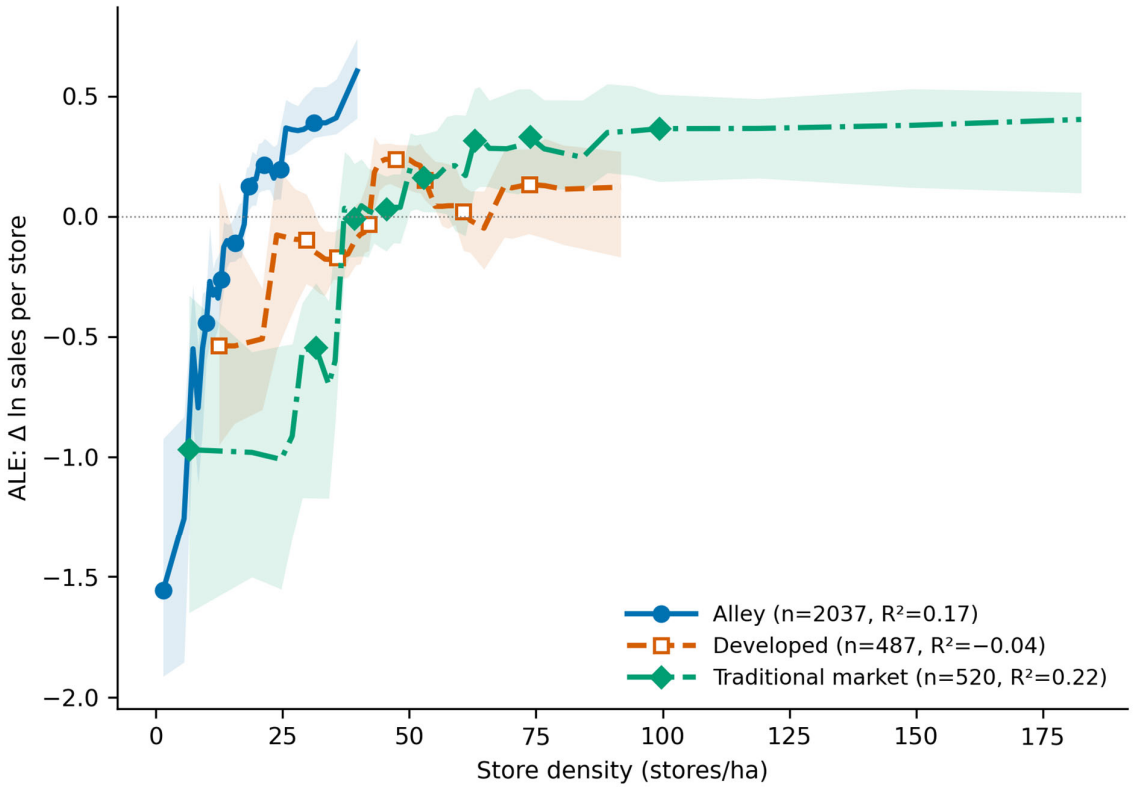


Figure 8. Density ALE by district type (axes truncated at the 97.5th percentile).

The parametric triangulation (Table 5; Equation (9)) confirms the machine learning geometry with conventional inference on the pooled sample (n = 3044; wave fixed effect included). For the alley baseline, density enters positively (0.0481) and its quadratic negatively (−0.0388 per 100), implying a turning point of 62.0 stores/ha; the developed district interactions shift the turning point to 41.4. For traditional markets, the interactions almost exactly offset the baseline terms (−0.0487 against 0.0481, 0.0387 against −0.0388), implying an essentially flat density profile—the parametric mirror of the flat type-specific ALE. All six density terms are significant at the 1% level under the wild cluster bootstrap ($p \leq 0.007$), so the interaction structure survives the small-cluster correction. Two further coefficients merit note: the land price elasticity of −0.40 ($p < 0.001$), and the insignificant linear area term (0.0047, $p = 0.21$)—the size effect is concave—steep among small districts and flat beyond about 14 ha—so a linear term in levels misrepresents it, which is one concrete mechanism behind the collapse of the untransformed OLS model in Table 2 and one that logarithmic and spline terms largely repair (Section 4.2).

Table 5. Parametric triangulation: OLS with density type interactions (dependent variable: outcome-year ln sales per store, pooled waves).

Term	Coefficient	Cluster SE	p (Cluster)	p (Wild Bootstrap)
Density	0.048 ***	(0.007)	<0.001	0.001
Density^2/100	−0.039 ***	(0.010)	<0.001	0.007
Density × developed	−0.042 ***	(0.007)	<0.001	0.001
Density^2/100 × developed	0.031 ***	(0.008)	<0.001	0.001
Density × traditional	−0.049 ***	(0.007)	<0.001	0.001
Density^2/100 × traditional	0.039 ***	(0.010)	<0.001	0.006
Wave 2 (2024→2025)	0.073 ***	(0.013)	<0.001	-
Developed (vs. alley)	1.369 ***	(0.219)	<0.001	-
Traditional (vs. alley)	1.725 ***	(0.160)	<0.001	-
District area (ha)	0.005	(0.004)	0.208	-
ln official land price	−0.396 ***	(0.078)	<0.001	-
ln floating population	0.147 ***	(0.033)	<0.001	-
ln working population	0.204 ***	(0.028)	<0.001	-
ln resident population	−0.052	(0.051)	0.297	-
Mean years in operation	0.034 **	(0.014)	0.019	-
Constant	18.540 ***	(1.226)	<0.001	-

Note: Standard errors clustered by gu (25 clusters); wild cluster bootstrap p-values (Rademacher weights, 999 replications) reported for the six density terms given the small number of clusters. n = 3044; R2 = 0.310. Implied turning points: alley 62.0 stores/ha; developed 41.4 stores/ha; for traditional markets, the linear and quadratic density terms are offset by the interactions, implying an essentially flat profile. *** $p < 0.01$; ** $p < 0.05$.

4.6. Robustness and Spatial Diagnostics

Trimming the 31 district waves above the 99th percentile of store density (170 stores/ha), where micro-area geometry generates extreme intensity values, leaves conclusions unchanged and slightly improves fit (spatial-CV R^2 = 0.327 on n = 3013). Global Moran's I, computed on the wave-2 cross-section so that each district enters the spatial weights once, confirms weak but significant autocorrelation in the outcome (I = 0.057, p = 0.001, 8-nearest-neighbor weights), and the corresponding out-of-fold residuals of the spatial model carry essentially the same structure (I = 0.051, p = 0.001; Figure 9b). The model neither removes nor inflates the outcome's spatial dependence; what neighbors share is precisely the unmeasured common conditions that spatial blocking is designed to keep out of the evaluation. The weakness of that autocorrelation also explains why the spatial component of validation optimism is small here (0.007) while the repeat-unit component is large (0.185), a decomposition available only because the design observes each district twice.

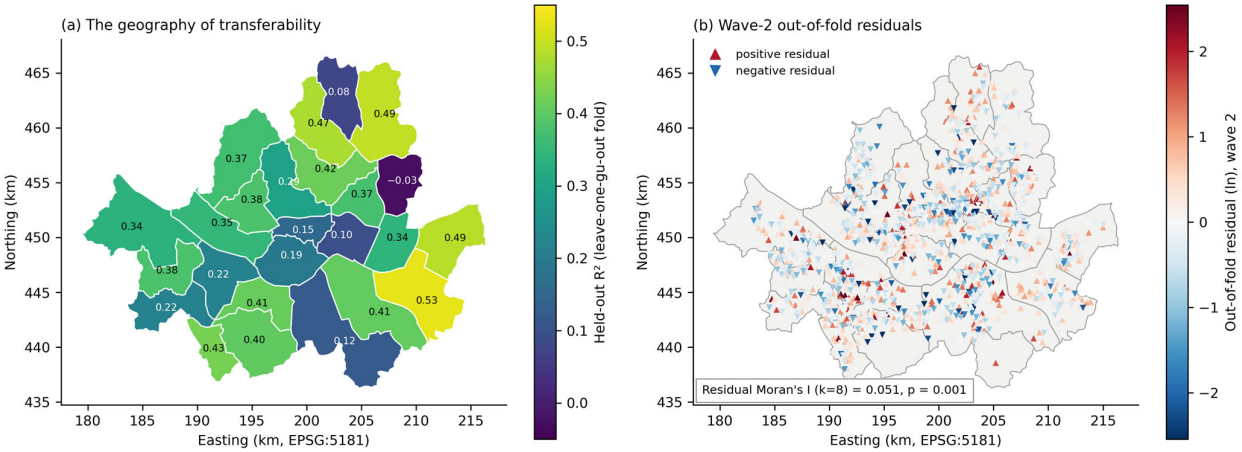


Figure 9. (**a**) Held-out R^2 by gu under leave-one-gu-out validation; (**b**) wave-2 out-of-fold residuals.

4.7. Additional Validation, Reparameterization, and the Scope of Inference

Table 6 collects the battery. The combined spatio-temporal holdout—the strictest design in the paper, in which the model has seen neither the test gu nor the test year and in which hyperparameters are tuned only on wave-1 observations of the 24 training gu—attains R^2 = 0.299 on the 1522 wave-2 district observations (95% gu-cluster bootstrap interval [0.245, 0.348]; RMSE 0.953, MAE 0.666) against 0.326 when only the location is held out and 0.532 when only the period is, both tuned in the same nested way. Across ten repetitions of the entire nested procedure with different random seeds and inner partitions, the combined figure ranges from 0.301 to 0.313 (mean 0.306), so the reported estimate lies at the conservative end of that range. The tuning selects depth 3 in every fold, with minimum child weight 10 in 17 folds and 5 in 8. Holding the modal configuration of the principal analysis fixed instead gives 0.303; that configuration was tuned on both waves of all gu, so the information it borrowed from outside of the combined training samples raised the estimate by less than 0.005 (paired bootstrap interval of the difference [0.000, 0.010]), and 0.299 is the figure we regard as the fully isolated deployment estimate. Twenty-four of the twenty-five folds are positive (range −0.01 to 0.59; median 0.34). Withholding geography and time together costs 0.027 beyond withholding geography alone (interval [0.008, 0.049]), which is small but distinguishable from zero, so the estimated relationships transfer across both dimensions at once, which is the deployment condition the paper claims to evaluate.

Table 6. Additional validation, reparameterization, and robustness battery.

Family	Specification	Result
Validation design	Spatial only, nested (leave-one-gu-out), wave-2 test set	R^2 = 0.326
	Temporal only, nested (train wave 1, test wave 2)	R^2 = 0.532
	Combined spatio-temporal holdout, fully nested (strictest)	R^2 = 0.299 [0.245, 0.348]
	Combined holdout, 10 repetitions over seeds and inner partitions	R^2 = 0.301 to 0.313 (mean 0.306)
	Same three designs, modal hyperparameters held fixed	R^2 = 0.325/0.512/0.303
Change outcomes	Δ ln sales per store, spatial CV	R^2 = −0.058
	Residual after removing lagged sales, spatial CV	R^2 = −0.053
Parameterization	Primitive (store count, area), spatial CV	R^2 = 0.328; crossing 22.4 stores/ha
Alternative learners	Random forest, spatial CV	R^2 = 0.357; crossing 25.1 stores/ha, 5.88 ha
	Histogram gradient boosting, spatial CV	R^2 = 0.324; crossing 25.2 stores/ha, 5.85 ha
	Spline regression (additive), spatial CV	R^2 = 0.354; crossing 23.1 stores/ha, 6.2 ha; exhaustion 54.9
Threshold intervals	Average-yield crossing (density), 200 gu-cluster bootstraps	24.7 [19.8, 25.9] stores/ha
	Exhaustion threshold, 200 gu-cluster bootstraps	44.4 [39.4, 59.0] stores/ha
	Average-yield crossing (area), 200 gu-cluster bootstraps	5.69 [3.94, 6.08] ha
	Area exhaustion threshold, 200 gu-cluster bootstraps	13.9 [11.8, 26.0] ha
	Wave 1/wave 2 estimated separately	crossing 25.2/24.7; exhaustion 47.1/44.5 stores/ha
Alternative outcomes	Total district sales, spatial CV	R^2 = 0.747; crossing 25.1; exhaustion 71.2
	Sales per hectare, spatial CV	R^2 = 0.657; crossing 24.8; exhaustion 71.2
	Land cost coverage ratio, spatial CV	R^2 = 0.653; crossing 24.6; exhaustion 102.1

Table 6. *Cont.*

Family	Specification	Result
Within-district design	First differences, density (n = 1514)	−0.0003 (SE 0.0013, $p = 0.81$)
	First differences, ln land price	−0.286 (SE 0.278, $p = 0.30$)
	Between-district ln land price, for comparison	−0.356 (SE 0.085, $p < 0.001$)

Note: All R^2 values are pooled out-of-fold. "Crossing" (average-yield crossing) is the density (stores/ha) or area (ha) at which the corresponding mean-centered ALE contribution crosses zero—a sample-relative reference point, not necessarily the location where the full model predicts average per-store sales (Section 4.8); "exhaustion" is the density at which it first attains 95% of the total rise of the curve from its minimum to its maximum, a criterion invariant to the mean-zero centering of the ALE. The bracketed interval for the combined holdout is a 95% interval from 2000 gu-cluster bootstrap replications of the held-out predictions; other bracketed values are 90% percentile intervals from 200 gu-cluster bootstrap replications, with the bootstrap median as the point value.

The change models return a clear negative result that we report as such. Predicting the change in ln sales per store between waves from the same land use features yields a spatially cross-validated R^2 of −0.058, and residualizing the level on its own lag before prediction gives −0.053; a linear specification on the same change outcome reaches −0.002. Land configuration thus predicts where a district stands, not how it is about to move. Two features of the data explain this and bound its interpretation. Year-over-year changes are small and noisy relative to levels—their standard deviation is 0.292 log points against 1.119 for the levels—and land configuration itself barely moves over a one-year horizon, as the first difference analysis below shows. The practical implication is that the model is a standing-level instrument suitable for diagnosis and comparison, not a short-horizon early warning system for individual districts.

Reparametrizing the model on the primitive pair (store count, district area) leaves predictive performance unchanged ($R^2 = 0.328$ against 0.325) and reproduces the same geometry: the implied density response crosses zero at 22.4 stores/ha and attains 95% of its total rise at 49.0 stores/ha. This makes a threshold generated solely by the density denominator less plausible while remaining a model-based descriptive regularity; the corresponding joint density–area response surface is shown in Figure 7. The bootstrap intervals locate the average-yield crossing for density at 24.7 stores/ha with a 90% interval of [19.8, 25.9] and that for area at 5.69 ha with an interval of [3.94, 6.08]; the exhaustion threshold is 44.4 stores/ha with a wider but still informative interval of [39.4, 59.0] whose upper bound lies below the value of 60 stores/ha reported in the submitted version. Estimated separately, the two waves agree closely (crossing 25.2 and 24.7 stores/ha; exhaustion 47.1 and 44.5). The two alternative learners recover the same structure—the random forest gives $R^2 = 0.357$ with the crossing at 25.1 and the histogram gradient booster $R^2 = 0.324$ with the crossing at 25.2, and both place the area crossing near 5.9 ha. Figure 10 assembles these nine specifications together with the spline benchmark: the average-yield crossing spans 22.4 to 25.2 stores/ha throughout, a stability that reflects a common reference sample rather than an economic threshold (Section 4.8). For the per-store outcome, the exhaustion threshold is similarly concentrated at 41.5 to 49.0 across the main model for both waves, both alternative learners, and the primitive parameterization, and at 54.9 for the smoother spline benchmark. The much higher values of 71.2 and 102.1 belong to the alternative outcomes discussed next and reflect a change of dependent variable rather than imprecision in estimating a single quantity.

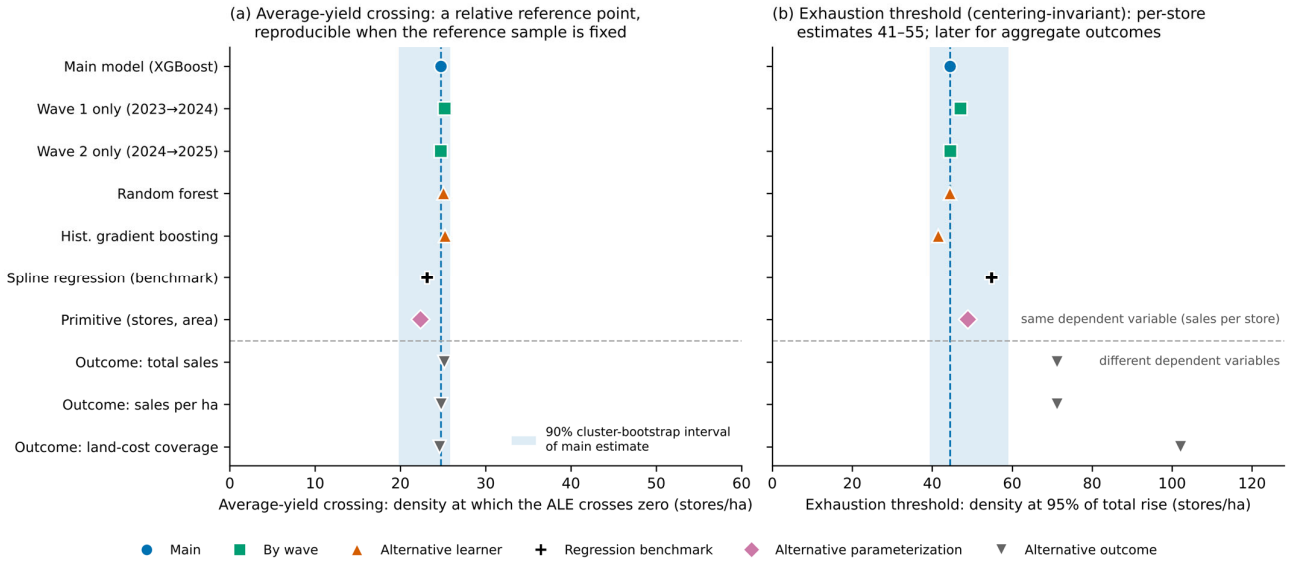


Figure 10. Density reference points and thresholds across ten specifications: (**a**) the average-yield crossing (zero crossing of the ALE under the pooled-sample centering), a relative reference point; (**b**) the exhaustion threshold (95% of total rise), which is invariant to centering. Rows above the dashed line share the per-store dependent variable; rows below it use alternative dependent variables. Shaded bands are 90% gu-cluster bootstrap intervals for the main model.

The alternative outcomes matter for interpretation. Total district sales and sales per hectare are both more predictable than sales per store (R^2 = 0.747 and 0.657) and place the average-yield crossing at the same point (25.1 and 24.8 stores/ha), but their exhaustion thresholds are much higher at 71.2 stores/ha. Saturation in revenue per establishment is therefore not saturation in aggregate commercial activity: past roughly 44 stores/ha, total district sales continue to rise, whereas per-establishment revenue is essentially flat in the fitted cross-sectional profile. A parametric decomposition makes the same point—the elasticity of total district sales with respect to store count is 1.452 (SE 0.156, $p < 0.001$), while the area elasticity is indistinguishable from zero (−0.067, SE 0.095, $p = 0.48$). Scaling revenue by imputed land cost gives R^2 = 0.653 and an average-yield crossing of 24.6 stores/ha, so the configuration result survives an explicit land cost adjustment; at an assumed capitalization rate of r = 0.04, the median coverage ratio is 1.25, indicating that imputed land cost is large relative to revenue; this ratio should not be interpreted as a profit margin or as a direct measure of merchant financial health.

The first-difference estimates define the scope of inference more sharply than any other result in the paper, and they qualify it. Across the 1514 districts observed in both waves, the within-district change in density has a standard deviation of 4.77 stores/ha, 6.6% of the cross-sectional standard deviation of 72.1, and district area does not change at all because the official delineation is fixed. In this within-district design, the density coefficient is −0.0003 (SE 0.0013, $p = 0.81$). The 95% confidence interval on this coefficient is narrow relative to the cross-sectional gradient: it spans roughly ±0.0025 log points per store/ha, about one-thirteenth of the 0.033 log points per store/ha implied by the cross-sectional curve over its steep segment, and therefore excludes an effect of that magnitude over a one-year horizon. For land price the within estimate is −0.286 (SE 0.278, $p = 0.30$) against a between estimate of −0.356 (SE 0.085, $p < 0.001$), the same sign and a similar magnitude, but too imprecise to confirm or refute. We draw the conservative conclusion throughout the discussion. The density and area gradients this paper estimates are cross-sectional regularities that describe which land configurations are associated with stronger per-store performance, and they transfer to unseen districts and unseen years; they are not estimates

of what would happen to a given district within a year of adding stores, and the paper makes no such claim.

Finally, conditioning on measured configuration absorbs part—but only part—of the raw performance gap between district types. The unconditional developed district premium of 1.174 log points falls to 0.611 once density, area, demand, land price, and business stock maturity are included, so roughly 48% of it is accounted for by measured land configuration and demand. The traditional market premium does not shrink at all; at 0.727 unconditionally and 0.875 conditionally, it is if anything larger once their small land base is controlled for. Consistent with the derivation of H4, we therefore treat the residual type gaps as unmeasured institutional, morphological, and locational difference rather than an effect of land allocation.

4.8. Relative Reference Points and Centering-Invariant Shape Benchmarks

The quantities used to summarize the density and area profiles are sample-centered reference crossings and centering-invariant shape benchmarks, and Table 7 and Figure 11 separate them. The first is the point at which an ALE curve crosses zero. Because the ALE is centered to mean zero over the estimation sample (Equation (7)), this identifies where the contribution of the focal feature to the fitted model crosses its sample-centered zero, not necessarily where the full model predicts average per-store sales. We retain "average-yield crossing" as an operational label for this zero crossing and treat it as a sample-relative reference point, not as an economic entry condition or an efficient scale. Its location depends on the reference distribution. Holding the fitted curve fixed and changing only the distribution over which it is centered moves the density crossing from 24.5 stores/ha (pooled sample) to 17.8 when the curve is centered on alley districts to 38.9–39.0 when it is centered on developed districts or traditional markets and to 13.2 when every value of ln density is weighted equally; the area crossing moves from 5.9 ha to between 2.0 and 11.6 ha. Refitting the model on different sample compositions moves the density crossing between 17.5 stores/ha (alley districts only) and 25.2 (wave 1 only), in each case within three stores/ha of the median density of the sample used (18.4–23.8 stores/ha), and the area crossing between 4.5 and 6.2 ha. The crossings are reproducible when the reference population is fixed, which is why they are stable across the specifications of Figure 10, all of which are centered on the same pooled sample, but they are properties of the city's distribution of districts as much as of the density–performance relationship.

Table 7. Relative reference points and centering-invariant shape benchmarks for store density and district area.

Family	Quantity	Store Density (Stores/ha)	District Area (ha)
Relative reference points (depend on the reference sample)	Average-yield crossing, pooled-sample centering	24.5	5.9
	Same curve centered on wave 1/wave 2	24.5/24.5	5.9/5.9
	Same curve centered on alley/developed/traditional	17.8/38.9/39.0	6.1/11.6/2.0
	Same curve, uniform weight over ln x	13.2	2.1
	Model refitted on six alternative samples (range)	17.5–25.2	4.5–6.2
	Median of those six samples (range)	18.4–23.8	6.6–9.2
	Spline regression benchmark	23.1 [20.7, 25.4]	6.2 [5.0, 6.9]

Table 7. *Cont.*

Family	Quantity	Store Density (Stores/ha)	District Area (ha)
Centering-invariant shape benchmarks (invariant to centering)	Exhaustion threshold (95% of total rise), XGBoost	44.4 [39.4, 59.0]	13.9 [11.8, 26.0]
	Exhaustion threshold, six alternative samples (range)	43.5–51.2	13.1–25.1
	Slope break, piecewise-linear fit to the XGBoost curve	55.2 [48.0, 59.9]	12.8 [2.3, 15.1]
	Segmented regression on the data (BIC-selected breaks)	68.8 [50.9, 113.7]	1.1 and 15.3 [1.7, 18.9]
	Spline regression benchmark, 95% of total rise	54.9 [44.4, 71.2]	18.6 [12.5, 26.0]

Note: Point values are full-sample estimates; bracketed values are 90% percentile intervals from 200 gu-cluster bootstrap replications (Table 6 reports the bootstrap medians, 24.7 stores/ha and 5.69 ha, for the pooled sample crossings). Crossings are zero crossings of the centered ALE and depend on the reference distribution by construction; the six alternative samples are wave 1 only, wave 2 only, without traditional markets, without developed districts, alley districts only, and density trimmed at the 99th percentile. Segmented regressions use hinge terms in ln density (ln area), a cubic spline in the other land variable, and the remaining predictors; BIC selects one break for density and two for area (the interval refers to the upper area break).

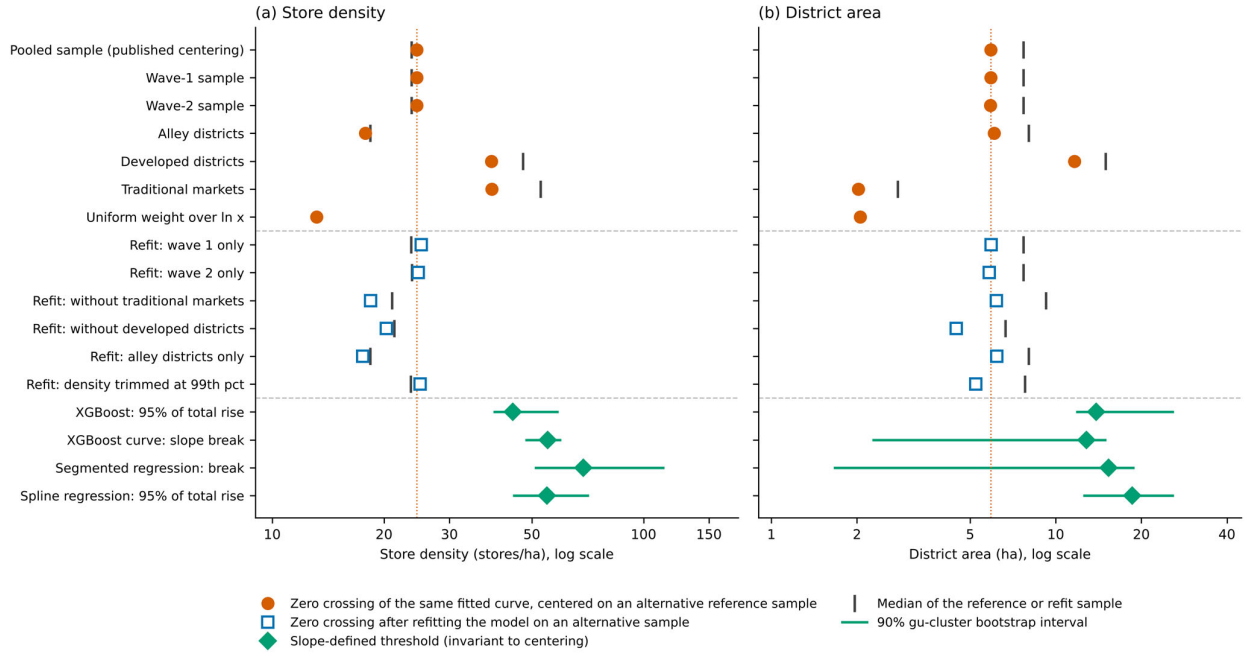


Figure 11. Relative reference points versus curve shape benchmarks for (**a**) store density and (**b**) district area. Filled circles: zero crossings of the same fitted XGBoost curve centered on alternative reference samples; open squares: zero crossings after refitting the model on alternative samples; vertical ticks: median of the corresponding sample; diamonds with bars: centering-invariant curve shape benchmarks or slope break estimates with 90% gu-cluster bootstrap intervals. The dotted line marks the published (pooled-sample) crossing.

The second kind is defined by the shape of the curve and is therefore unaffected by additive re-centering. The 95% of total rise rule is a curve height criterion, not itself a slope break estimator; the piecewise and segmented regressions provide separate tests for slope changes. The exhaustion threshold—the density at which the curve completes 95% of its total rise—is identical under every re-centering (44.4 stores/ha) and lies between 43.5 and

51.2 under refitting on the alternative samples. Two slope break estimators independently locate a change in the rising profile without reference to the level of the curve. A continuous piecewise-linear fit to the boosted curve in ln density places a single break at 55.2 stores/ha (90% interval 48.0–59.9), where the slope falls from 0.81 to 0.02 log points per log unit of density; allowing two breaks isolates a steepest segment between 35.9 and 43.0. The segmented regression fitted directly to the data, which does not use the ALE at all, prefers a single break by BIC at 68.8 stores/ha (interval 50.9–113.7), with a slope of 0.76 before and −0.71 after it; with two breaks, its first break is at 53.5, beyond which the slope is −0.03. The spline benchmark completes 95% of its rise at 54.9 stores/ha (interval 44.4–71.2). The end of the rise is thus a curve shape feature supported by both rise-based and slope break estimators: the 95% criterion on the boosted curve locates it most sharply at about 44 stores/ha (the two-break curve fit agrees at 43.0), the one-break curve fit, the spline benchmark, and the two-break segmented regression at 53–55 stores/ha, and the one-break segmented regression somewhat higher, with intervals that overlap. The lower end of the 25–44 range is different in kind: no estimator finds a change of slope near 25 stores/ha. Below the exhaustion region, per-store sales rise steadily with log density by roughly 0.76 log points per log unit, and 25 stores/ha marks only where that rising profile passes the citywide average.

For district area, the same distinction applies. The average-yield crossing (5.9 ha; bootstrap median 5.7 ha) is a relative reference point. The curve shape and slope break quantities place the flattening of the area profile at a larger scale: the boosted curve completes 95% of its rise at 13.9 ha (interval 11.8–26.0), and the curve-based fit places a break at 12.8 ha (interval 2.3–15.1), where the slope falls from 0.63 to 0.17. The segmented regression on the data, which prefers two breaks by BIC, finds the steepest returns among micro-districts below 1.1 ha, a sustained slope of 0.58 up to 15.3 ha (interval 1.7–18.9), and an essentially flat profile (0.08) beyond; the spline benchmark completes 95% of its rise at 18.6 ha. Per-store sales therefore rise with district area at a diminishing rate until the profile flattens at roughly 13–19 ha—an empirically defined flattening region related to the scale concept in H3—and about 6 ha marks only where the centered area ALE contribution crosses zero. Under refitting on the alternative samples, the area exhaustion point ranges from 13.1 to 25.1 ha, so its location is less precisely identified than its existence. Traditional markets, with a median area of 2.8 ha, lie below both quantities on the steep part of the area profile.

These results change the vocabulary of the paper, not its geometry. Throughout, about 25 stores/ha and about 6 ha are referred to as average-yield crossings (the zero points of feature-level centered ALE curves); the exhaustion threshold near 44 stores/ha (53–55 stores/ha for the smoother estimators) and the flattening of the area profile near 14 ha are the curve shape and slope break quantities used to assess H2 and H3; and the range from 25 to 44 stores/ha is retained only as a description of the segment over which the rising profile passes its sample-centered zero en route to its plateau, not as a target.

5. Discussion

5.1. Intensity Thresholds as Diagnostic Benchmarks

The central empirical object of this paper is a configuration–performance curve for commercial land use intensity—an estimate of the land term $m_t(D, A)$ of Equation (2)—estimated with uncertainty and required to transfer across unseen parts of the city. Its shape—per-store sales rising with log density through the lower range, with the centered density ALE crossing zero near 25 stores/ha and exhausted beyond roughly 44 stores/ha—is the agglomeration–saturation prediction of Sections 2.2 and 2.3 made quantitative. The lower branch is as informative as the upper: sparsely occupied commercial land is asso-

ciated with much lower per-store sales, plausibly because sub-critical store mass cannot generate the comparison shopping externalities and continuous frontage that draw customers [28,37,38]. The upper branch aligns with block-level evidence that land use intensity and vitality are non-monotonically related [4] and with threshold effects reported for commercial land benefits [18] but sharpens both by expressing the outcome as realized sales per establishment—a revenue-side measure relevant to small business activity, though not a measure of profitability—and by certifying the curve out of region. Most of Seoul's alley commercial land lies on the rising segment of the profile, below the average-yield crossing (median 18.4 stores/ha). That is a description of where the alley stock sits relative to citywide regularities; whether additional concentration in any particular alley district is feasible or desirable is a separate, place-specific question that the cross-sectional evidence cannot answer.

5.2. District Size as a Configurational Benchmark: The Traditional Market Case

The area profile—steepest among micro-districts, rising at a diminishing rate until it flattens at about 14 ha—is an empirical counterpart of the threshold concept in central place theory [31], and it makes district geometry a policy-relevant diagnostic dimension rather than a cartographic given. Its sharpest application is to traditional markets, whose median territory (2.8 ha) lies on the steep part of the profile, below both the area average-yield crossing and the point at which the profile flattens and whose type-specific model identifies area as the dominant predictor by a wide margin. For Seoul's statutorily supported marketplaces [5], the variable that carries information is not intensity—their intensity profile is flat—but land extent; compared with the citywide profile, many have a land base smaller than that associated with stronger per-store sales. This reframes the diagnostic question for market support policy in land terms as a matter of description rather than prescription. Land extent is the dimension on which traditional markets differ most from better-performing configurations, and it is therefore the dimension worth examining in place-specific assessment. Whether enlarging the effective land base of a given market through precinct extension, integration with adjacent street commerce, or coordinated designation of contiguous commercial land would raise its per-store sales is not something our design can establish. Two cautions are relevant. Thresholds estimated on administrative delineations partly reflect those delineations—the modifiable areal unit problem [50]—though this limitation is softened here because the same official boundaries are the objects on which designation and support policy actually operate. And, the estimates are associational, as enlarging a district's boundary does not mechanically confer the demand base of organically larger districts.

5.3. Land Price as a Cost Signal

Conditional on demand fundamentals and configuration, official land price is negatively associated with per-store sales across its observed range—an elasticity of -0.40 in the parametric model and an ALE decline of 0.58 log points between 1 and 8 million KRW/m^2. The bid rent reading [48] is that once a district's demand is held constant, higher land cost is consistent with cost pressure on the margins of the establishments occupying the land rather than with a marker of their success. The result carries a monitoring implication: rising official land prices in a district whose demand fundamentals are static may warrant monitoring by commercial land policy as a potential indicator of cost pressure on incumbents, the mechanism repeatedly implicated in Korean debates over commercial gentrification and tenant displacement [11,27]. Interpretation should respect the measure's limits: official appraisal prices track market rents imperfectly, and a compositional channel—

expensive land selecting small-format, lower-turnover establishments—cannot be excluded at the district level.

This occupancy cost reading is, however, not a complete account of how land values reach merchant performance. Real estate can also serve as collateral for borrowing: where firms own pledgeable property, appreciation may ease financing constraints and support investment [59,60]. The net effect of land price changes on merchant performance may therefore depend on the district's mix of owners and tenants, which our data do not observe. Assessed values also serve as a forward-looking asset price: a high assessed value can reflect expected future appreciation of the land even when the current cost burden is high. Our first-difference estimate addresses this question directly but inconclusively—the within-district land price coefficient shows the same negative association of a similar magnitude but is not statistically distinguishable from zero (Section 4.7)—so the negative baseline association cannot be read as a cost effect net of the collateral and expectations channels. These channels can be distinguished only with tenure, credit, and rent microdata, which the district census does not include; obtaining them is the most substantive extension of this analysis.

5.4. Why a Structural Model with $R^2 \approx 0.30$ Is the Policy-Relevant Object

A skeptic will note that prior-year sales predict next-year sales at $R^2 = 0.932$ while the structural model manages 0.325 and ask why the weaker model deserves the paper. The answer is that the two models answer different questions. Persistence is a forecast without levers: it says the hierarchy of commercial land performance is highly stable, but it contains nothing a land use planner can act on and nothing that explains why land performs. The structural model captures the component of performance associated with observable structural conditions that policy can, in part, influence—intensity of occupation, territorial extent, land cost—and certifies that this part transfers to unseen areas ($R^2 = 0.325$) and to a subsequent year (0.532). Four design results discipline the interpretation.

The collapse of the untransformed linear model ($R^2 = 0.007$), set against the recovery of regression with transformed predictors and spline terms (0.265 to 0.361), shows that the transferable structure is nonlinear and that it is recoverable by any specification flexible enough to bend, not only by tree boosting. What boosting adds is that it discovers the shape, including type-specific differences, without pre-specified transformations or knots, and the agreement between the boosted and spline curves is evidence that the geometry belongs to the data. The measured optimism of conventional validation (0.192, or 59%) is our local instance of the general warning from spatial machine learning [15,16], and the two-wave design decomposes it: nearly all is repeat unit leakage (0.185), the mechanism by which a model "recognizes" districts it has already seen, with spatial leakage contributing little (0.007) because outcome autocorrelation is weak ($I = 0.057$). Where autocorrelation is strong, the spatial component alone becomes the chasm documented for commercial land value elsewhere (0.622 to 0.050 [18]). And, the per-fold spread (-0.03 to 0.53; Figure 9a) is a reminder that transferability itself has a geography—a model certified on average still has an area of applicability [17], and district-level guidance should be applied with corresponding humility at the extremes.

5.5. Implications for Commercial Land Use Planning

Four implications follow. They describe how the estimated profiles can be used to read where a district sits relative to citywide regularities; none of them is a recommendation to move a district toward a particular density or area because the evidence is cross-sectional and the within-district analysis does not reproduce the density gradient (Section 4.7). First, use the density profile as a descriptive reference. Alley districts below the density

ALE reference crossing—the bulk of the stock—have a negative density contribution relative to the pooled reference distribution. This does not establish below-average overall predicted sales, nor does it establish a benefit from intensification; rather, it motivates further place-specific examination, not by itself a reason to intensify them, and nothing in the evidence associates densities beyond the exhaustion threshold of roughly 44 stores/ha with further per-store gains, although aggregate revenue generated on the land continues to rise past that point. Second, treat district area as a descriptive dimension. Designation and support programs can read a district's land base against the area profile, and traditional markets, most of which lie on its steep part, may warrant examination of how their effective commercial catchment relates to adjacent commercial streets; our design cannot establish whether enlarging that base would raise per-store sales. Third, monitor land price as a diagnostic screening indicator within commercial district policy, flagging districts where land price growth outruns demand fundamentals. Fourth, distinguish by regime: indicators at the district level provide information in alley districts and traditional markets, while indicators at the metropolitan scale (transit, anchor institutions, destination programming) would better serve the purposes of assessing developed districts, a type of district that is locally unpredictable in our models. All four translate the estimated curves into the language already used in the vocabulary of land use planning and use them as benchmarks for diagnosis, not as intervention targets or predictions of the impact of intervention.

5.6. Why Seoul Matters and What Travels

Because the estimates are certified only for Seoul, it is worth stating precisely what is case-specific and what should generalize. The levels are Seoul's. The average-yield crossings near 25 stores/ha and 6 ha—relative, by construction, to Seoul's own distribution of districts—the exhaustion of intensity returns near 44 stores/ha, and the flattening of the area profile near 14 ha are calibrated to a metropolis with 9.3 million residents on 605 km^2, a dense transit network, an unusually high share of self-employment [7], and an appraisal-based land price series [6]; they should not be transplanted as planning standards to cities with different retail formats, motorization, or land institutions.

The proposed mechanisms are not specific to Korea, but their empirical shapes must be verified in each new setting. The elasticity race of Equation (3)—comparison shopping and variety externalities [28,37,39] running against demand division and density dependence [47]—operates wherever retailers share a bounded catchment, and the threshold logic that generates a minimum efficient scale is the oldest proposition in central place theory [31,32]. What the Seoul census supplies is not a peculiar mechanism but the observational range needed to see the whole of it. In this sense, the city functions as a limiting case: whether other cities occupy the ascending or the saturated portion of the curve is an empirical question, and the value of a dense case is that it locates a ceiling against which they can be assessed. With urban populations projected to grow by more than two billion by mid-century [61] and land consumption historically outpacing population growth [1,24], whether continued urban growth moves more commercial districts toward the saturation region is an empirical question for future comparative research.

Three transferable products follow. The first is the procedure: enumerate commercial land units, separate predictors from outcome in time, evaluate on a ladder that runs from plain random validation through unit-blocked and spatially blocked designs to strict temporal transfer, and read effect curves with ALE. Any city with a business register and a delineation of its commercial areas can execute it, and the resulting centering-invariant shape summaries—the exhaustion criterion for intensity, the point at which the area profile flattens, and land price sensitivity—can be compared across cities subject to comparable measurement and support, whereas average-yield crossings are comparable only relative to

each city's own distribution of districts. The second is the diagnosis of validation optimism: on our data, conventional random validation overstates transferable performance by 59%, and the two-wave design shows that nearly all of the overstatement is repeat unit leakage rather than spatial leakage—a decomposition that panel-style urban data make available anywhere and a warning that applies with special force to the many vitality studies that pool repeated observations of the same places. Where spatial autocorrelation is strong, the spatial component alone produces the collapse documented for commercial land value elsewhere [16,18]. The third is a null worth porting: the finding that large, consolidated districts are locally unpredictable ($R^2 = -0.043$) delimits the scale at which district-level indicators appear informative at all, a boundary condition other cities can test cheaply before designing micro-scale interventions. The same administrative census has also been used to examine spatial stratification and convergence across Seoul's commercial districts [62].

5.7. Limitations and Future Research

Five limitations bound the claims. First, the predictor set is parsimonious: eight administratively derived variables omit built environment attributes such as land use diversity and pedestrian-oriented street design [63], alongside street network centrality, transit accessibility, establishment composition, building form, and market rents. These may carry additional predictive signals; the reported R^2 therefore provides a baseline against which richer commercial land inventories can be evaluated, and assembling one on the district census is the clearest next step. Second, evidence comes from one city and a three-year window spanning two prediction waves; the replication of the threshold geometry across waves is reassuring, but the thresholds are Seoul's, and their portability to systems with different retail formats, e-commerce penetration, and land institutions is an empirical question that the census plus spatial validation template developed here makes testable elsewhere. Third, the analysis is ecological at the district level and associational in nature; the temporal separation of predictors and outcome removes simultaneity but not selection on unobservables, so the curves are decision-relevant descriptions, not causal guarantees. Fourth, results are conditional on official district delineations [50] and appraisal-based land prices, both institutional artifacts—albeit the very artifacts on which Korean policy operates. Fifth, sales per store are constructed from card-transaction-based district sales and store counts; establishment size composition within districts is unobserved. Future work should add multi-city replication, richer land inventories, and quasi-experimental leverage from boundary or designation changes.

Three further restrictions limit the interpretation further, and we list them here because they follow from our own robustness analysis. First, and most importantly, density and area are not exogenous. The configuration variables partly depend on district quality—districts with favorable locational endowments attract entrants—rather than only explaining it, so the unobserved determinants of entry and of local economic level are absorbed into the estimated gradients. The first difference analysis in Section 4.7 confirms the practical force of this concern: the cross-sectional density gradient does not reproduce within districts over a one-year horizon at a precision that would have detected a short-run coefficient as large as the steep cross-sectional gradient.

The thresholds should therefore be read as configurational benchmarks—descriptions of which land configurations are associated with stronger per-store performance and which transfer to unseen districts and years—rather than causal return to intensifying a given district. Credible causal identification would require exogenous variation in configuration, such as zoning reclassification, redevelopment completions, or designation changes, together with a horizon long enough for agglomeration adjustment. Second,

the outcome is revenue-side. Card transaction sales per store capture neither merchant profitability nor cost structure nor return on assets, and if operating costs rise with density and land price in step with revenue, high per-store sales need not indicate either efficient land use or financially healthy merchants. Our land cost coverage measure addresses only the land component of cost and does so under an imputation rather than with observed rents; labor, inventory, and financing costs remain unobserved. Linking the district census to tax or credit bureau microdata to construct per-store profit, rent yield per unit area, and valuation premia is the natural next step, and the comparison between revenue-side and profit-side gradients is an open empirical question. Third, the model predicts levels and not changes; as Section 4.7 reports, land configuration carries no out-of-sample information about year-over-year movement in per-store sales.

6. Conclusions

This paper examined how commercial performance varies with land use intensity, district extent, and land price and required every answer to survive prediction into parts of the city the model had never seen and into a year it had never observed. On a sample drawn from Seoul's complete commercial district census, the answers are concrete. The economic yield of commercial land is associated with land use intensity along a steep, saturating curve: per-store sales rise with log density through the lower range, cross the centered-ALE zero near 25 stores/ha—a relative reference point—and are exhausted beyond roughly 44 stores/ha. Returns to district territory are steepest among micro-districts and flatten at about 14 ha, about 6 ha marks, the sample-relative zero crossing of the centered area ALE. Higher land prices are negatively associated with per-store sales conditional on demand, consistent with cost pressure but not uniquely identifying a cost channel. These relationships are nonlinear—a linear model in untransformed levels has essentially no out-of-region validity, whereas flexible regression and tree ensembles recover them alike—replicated across two independent prediction waves and play out differently across the three land regimes that make up the system: alley districts sitting mostly below the density reference point, traditional markets distinguished by territory rather than intensity, and developed districts whose fortunes local land configuration cannot predict.

Methodologically, the paper demonstrates that honest, spatially blocked evaluation and fully interpretable machine learning are compatible with—and, indeed, sharpen—policy-relevant description of urban land. The template is portable to any city that can enumerate its commercial land units, and the quantities it yields—an exhaustion threshold for intensity, the scale at which the area profile flattens, relative reference points, and land price monitoring signals—are expressed in the vocabulary of land use planning. The paper's contribution is accordingly threefold: it estimates the configuration–performance geometry of commercial land use intensity and district scale on a complete metropolitan census with transaction-based outcomes; it establishes spatially blocked validation as the evidence standard for such estimates and measures what conventional validation would have added; and it converts both into configurational benchmarks—an exhaustion threshold for intensity, relative reference points for density and area, and a land price monitoring signal—for reading where commercial land sits relative to citywide regularities. Two boundaries on these conclusions deserve emphasis. The regularities we report are cross-sectional associations that generalize across unseen districts and unseen years, not causal returns to intensification: within districts, over a one-year horizon, the density gradient is not reproduced. And, the outcome is revenue per establishment, not profit. The benchmarks are accordingly offered for identifying where commercial land sits relative to citywide regularities and as inputs to further investigation, rather than as targets or as guarantees of what a change in configuration would deliver in any particular district.

Author Contributions: Conceptualization, G.S.; Methodology, G.S.; Software, G.S.; Validation, J.C.; Formal analysis, G.S.; Investigation, J.C.; Resources, J.C.; Data curation, J.C.; Writing – original draft, G.S.; Writing – review & editing, J.C.; Visualization, J.C.; Supervision, G.S.; Project administration, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the ANCHOR through the Seoul ANCHOR Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government. (2026-ANCHOR-01-005-11).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seto, K.C.; Fragkias, M.; Güneralp, B.; Reilly, M.K. A meta-analysis of global urban land expansion. *PLoS ONE* **2011**, *6*, e23777. [CrossRef]
2. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*; A/RES/70/1; United Nations: New York, NY, USA, 2015.
3. Koroso, N.H.; Zevenbergen, J.A.; Lengoiboni, M. Urban land use efficiency in Ethiopia: An assessment of urban land use sustainability in Addis Ababa. *Land Use Policy* **2020**, *99*, 105081. [CrossRef]
4. Xia, C.; Yeh, A.G.-O.; Zhang, A. Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. *Landsc. Urban Plan.* **2020**, *193*, 103669. [CrossRef]
5. Korea Ministry of Government Legislation. *Special Act on the Development of Traditional Markets and Shopping Districts*; Act No. 21204; Korea Ministry of Government Legislation: Sejong, Republic of Korea, 2025; effective 17 June 2026. (In Korean)
6. Seoul Metropolitan Government. Seoul Commercial District Analysis Service. Available online: https://golmok.seoul.go.kr (accessed on 8 July 2026).
7. Ministry of SMEs and Startups; Small Enterprise and Market Service. *2023 Survey on Small Business Owners*; Ministry of SMEs and Startups: Sejong, Republic of Korea, 2025. (In Korean)
8. Zhang, Y.; Wang, M.; Yang, X.; Zhang, R. Urban commercial space vitality evaluation method based on social media data: The case of Shanghai. *Land* **2025**, *14*, 697. [CrossRef]
9. Xie, Y.; Zhang, J.; Li, Y.; Zhu, Z.; Deng, J.; Li, Z. Integrating multi-source urban data with interpretable machine learning for uncovering the multidimensional drivers of urban vitality. *Land* **2024**, *13*, 2028. [CrossRef]
10. Kim, H.; Lee, S. A study on the factors affecting the revenue in Seoul's side street trade areas. *Seoul Stud.* **2019**, *20*, 117–134. (In Korean)
11. Park, J.; Jeon, L.; Lee, S. Analysis of growth-decline type and factors influencing growth commercial area using sales data in alley commercial area: Before and after COVID-19. *J. Korean Reg. Sci. Assoc.* **2023**, *39*, 53–66. (In Korean)
12. Lim, S.; Lee, K. An empirical analysis on the impact factors of sales in the alley market of Seoul: Focused on locational characteristics and SNS references. *J. Korea Plan. Assoc.* **2025**, *60*, 110–125. (In Korean) [CrossRef]
13. Kim, K.; Jung, M.; Kim, H.W. Examining the characteristics and dynamic changes in Seoul's alley commercial district using commercial vitality and business diversity indices. *Korea Spat. Plan. Rev.* **2025**, *125*, 19–39. (In Korean)
14. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [CrossRef]
15. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
16. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [CrossRef]
17. Meyer, H.; Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **2021**, *12*, 1620–1633. [CrossRef]
18. Jiang, Y.; Tang, P.; Sun, S.; Liu, Y.; Zhang, X.; Xu, B.; Zheng, G. Benefit evaluation of urban commercial land and XGBoost-SHAP based influence analysis: A case study of Chengdu, China. *Land* **2026**, *15*, 1165. [CrossRef]
19. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
20. Apley, D.W.; Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1059–1086. [CrossRef]

21. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
22. Cameron, A.C.; Gelbach, J.B.; Miller, D.L. Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* **2008**, *90*, 414–427. [CrossRef]
23. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. [CrossRef]
24. Angel, S.; Parent, J.; Civco, D.L.; Blei, A.; Potere, D. The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050. *Prog. Plan.* **2011**, *75*, 53–107. [CrossRef]
25. Bertaud, A. *Order Without Design: How Markets Shape Cities*; MIT Press: Cambridge, MA, USA, 2018.
26. Shin, H.B.; Kim, S.-H. The developmental state, speculative urbanisation and the politics of displacement in gentrifying Seoul. *Urban Stud.* **2016**, *53*, 540–559. [CrossRef]
27. Kim, M.; Lee, S. Categorizing alley commercial districts and analyzing the influencing factors of business openings and closings in Seoul, Korea: Using floating population big data and DTW time series clustering analysis. *J. Korea Plan. Assoc.* **2024**, *59*, 99–116. (In Korean) [CrossRef]
28. Jacobs, J. *The Death and Life of Great American Cities*; Random House: New York, NY, USA, 1961.
29. Sung, H.; Lee, S.; Cheon, S. Operationalizing Jane Jacobs's urban design theory: Empirical verification from the great city of Seoul, Korea. *J. Plan. Educ. Res.* **2015**, *35*, 117–130.
30. Glaeser, E.L.; Kolko, J.; Saiz, A. Consumer city. *J. Econ. Geogr.* **2001**, *1*, 27–50. [CrossRef]
31. Christaller, W. *Central Places in Southern Germany*; Baskin, C.W., Translator; Prentice-Hall: Englewood Cliffs, NJ, USA, 1966.
32. Berry, B.J.L. *Geography of Market Centers and Retail Distribution*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1967.
33. Jung, D.; Kim, H. An analysis of the distribution and factors influencing sales by business types. *GRI Rev.* **2014**, *16*, 101–122. (In Korean)
34. Kim, J.; Seo, K.; Lee, H.-S.; Kim, K.-M. The effect of Seoul retail area characteristics on its survival. *J. Korea Plan. Assoc.* **2022**, *57*, 75–90. (In Korean) [CrossRef]
35. Lee, Y.; Kim, C.; Lee, C. The study of effects of super-store and SSM on the regional commercial district: Based on CBD of the Gangneung·Wonju·Geoje. *J. Korean Reg. Dev. Assoc.* **2015**, *27*, 147–166. (In Korean)
36. Wrigley, N.; Dolega, L. Resilience, fragility, and adaptation: New evidence on the performance of UK high streets during global economic crisis and its policy implications. *Environ. Plan. A* **2011**, *43*, 2337–2363. [CrossRef]
37. Marshall, A. *Principles of Economics*; Macmillan: London, UK, 1890.
38. Duranton, G.; Puga, D. Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*; Henderson, J.V., Thisse, J.-F., Eds.; Elsevier: Amsterdam, The Netherlands, 2004; Volume 4, pp. 2063–2117.
39. Hotelling, H. Stability in competition. *Econ. J.* **1929**, *39*, 41–57. [CrossRef]
40. Jacobs, J. *The Economy of Cities*; Random House: New York, NY, USA, 1969.
41. Storper, M.; Venables, A.J. Buzz: Face-to-face contact and the urban economy. *J. Econ. Geogr.* **2004**, *4*, 351–370. [CrossRef]
42. Krugman, P. Increasing returns and economic geography. *J. Polit. Econ.* **1991**, *99*, 483–499. [CrossRef]
43. Fujita, M.; Krugman, P.; Venables, A.J. *The Spatial Economy: Cities, Regions, and International Trade*; MIT Press: Cambridge, MA, USA, 1999.
44. Rosenthal, S.S.; Strange, W.C. Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*; Henderson, J.V., Thisse, J.-F., Eds.; Elsevier: Amsterdam, The Netherlands, 2004; Volume 4, pp. 2119–2171.
45. Glaeser, E.L.; Kallal, H.D.; Scheinkman, J.A.; Shleifer, A. Growth in cities. *J. Polit. Econ.* **1992**, *100*, 1126–1152. [CrossRef]
46. Ye, Y.; Li, D.; Liu, X. How block density and typology affect urban vitality: An exploratory analysis in Shenzhen, China. *Urban Geogr.* **2018**, *39*, 631–652. [CrossRef]
47. Hannan, M.T.; Freeman, J. The population ecology of organizations. *Am. J. Sociol.* **1977**, *82*, 929–964. [CrossRef]
48. Alonso, W. *Location and Land Use: Toward a General Theory of Land Rent*; Harvard University Press: Cambridge, MA, USA, 1964.
49. Muth, R.F. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*; University of Chicago Press: Chicago, IL, USA, 1969.
50. Openshaw, S. *The Modifiable Areal Unit Problem*; Concepts and Techniques in Modern Geography 38; Geo Books: Norwich, UK, 1984.
51. Breiman, L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Stat. Sci.* **2001**, *16*, 199–231. [CrossRef]
52. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
53. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
54. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
55. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; 2022. Available online: https://christophm.github.io/interpretable-ml-book (accessed on 28 July 2026).

56. Legendre, P. Spatial autocorrelation: Trouble or new paradigm? *Ecology* **1993**, *74*, 1659–1673. [CrossRef]
57. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [CrossRef]
58. Wadoux, A.M.J.-C.; Heuvelink, G.B.M.; de Bruin, S.; Brus, D.J. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol. Model.* **2021**, *457*, 109692. [CrossRef]
59. Kiyotaki, N.; Moore, J. Credit cycles. *J. Polit. Econ.* **1997**, *105*, 211–248. [CrossRef]
60. Chaney, T.; Sraer, D.; Thesmar, D. The collateral channel: How real estate shocks affect corporate investment. *Am. Econ. Rev.* **2012**, *102*, 2381–2409. [CrossRef]
61. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*; United Nations: New York, NY, USA, 2019.
62. Shin, G.; Chae, J. Spatial Stratification of Urban Commercial Land: A Convergence and Diagnostic Framework for Seoul's Commercial Districts. *Land* **2026**, *15*, 1783. [CrossRef]
63. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 199–219. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.