CPRI-Office: A new commercial property rental index for Indian cities using spatio-temporal modeling techniques

Kapil Gupta, Soudeep Deb, Venkatesh Panchapagesan

Indian Institute of Management Bangalore
Bannerghatta Main Road, Bangalore, Karnataka 560076, India.

Abstract

This study presents a novel spatio-temporal rental index designed to analyze rental trends in the commercial office markets of major Indian cities. Moving beyond conventional price-based indices, which often suffer from biases caused by simple averages and outliers, our approach employs a Bayesian spatio-temporal modeling framework to capture nuanced market dynamics. By incorporating both actual transaction data and predicted hypothetical transactions for active leases, the methodology overcomes data sparsity challenges, offering a comprehensive and accurate representation of market trends. A key innovation of this methodology lies in its efficient handling of new data. The model updates the index quarterly using data from the current and previous three quarters while leveraging previously fitted posteriors as priors. This approach significantly reduces computational overhead without compromising accuracy, as validation against full-data models consistently confirms robust results. The resulting index provides actionable insights at various levels of granularity, including city, macromarket, and micro-market scales, making it a valuable tool for investors, developers, tenants, policymakers, and real estate investment trusts (REITs). By reflecting true market patterns and enabling informed, data-driven decisions, this index sets a new standard for analyzing commercial office rental dynamics and advancing market intelligence.

Keywords: Bayesian methodology, Commercial properties, Property index, Real estate, Spatio-temporal models

1. Introduction

The commercial real estate market is a critical driver of economic growth and urban development, influencing investment strategies, corporate operations, and policymaking. Within this sector, office spaces hold a pivotal role as centers of productivity and innovation, making the analysis of rental trends an essential task for stakeholders such as investors, property developers, corporate tenants, and urban planners. Rental indices are indispensable tools for capturing these trends, providing actionable insights into market dynamics, and supporting informed decision-making. The evolution of rental prices has profound implications for macroeconomic stability, investment planning, and infrastructure development. Geltner (1990) emphasized that rental cash flows fundamentally shape the risk-return profile of commercial real estate, further underscoring the importance of robust rental indices in understanding market behavior.

Interestingly, constructing a reliable rental index for the commercial office market is a statistically challenging task. The diverse nature of commercial real estate assets, encompassing variations in property type, location, size, quality, tenant preferences and lease terms, adds complexity to accurately capturing true market dynamics. Deryol et al. (2019) further noted that the paucity of transaction data may exacerbate these challenges, commonly resulting in indices that are unrepresentative or overly sensitive to localized anomalies. Traditional methodologies, such as hedonic pricing models (Rosen, 1974) and repeat-sales methods (Bailey et al., 1963), have been widely used to address these challenges, but each comes with significant limitations. Hedonic models decompose rental prices based on observable property attributes, yet they frequently fail to account for spatial and temporal dependencies, leading to omitted variable bias. Repeat sales methods, on the other hand, rely on properties being leased or sold multiple times. While this approach offers temporal stability, it excludes single transaction properties, reducing representativeness and rendering it less effective in markets with sparse or infrequent transactions. In a related work, Jennen and Brounen (2009) highlighted that locational density significantly influences commercial office rents in Amsterdam, and their approach to account for it considered clusters of properties via a hedonic structure. Kok and Jennen (2012) adopted a fixed-effect model based on precise location-specific determinants, such as pin codes, to analyze office rents. However, while these methods offer valuable insights about the rental markets, they fall short in scenarios with sparse transactions and are inadequate for effectively capturing spatial autocorrelation in the data.

More recent advances, including hybrid approaches and machine learning techniques, have sought to address these limitations to some extent. For example, Oust et al. (2020) proposed a hybrid model, combining hedonic and repeat-sales methods, to leverage the strengths of both, integrating spatial and temporal information for improved accuracy. Calainho et al. (2024) discussed about suitable machine learning methods and their ability to offer flexibility in modeling complex interactions between property attributes. While these methods are becoming popular, their reliance on extensive data and lack of interpretability can limit their applicability in sparse data environments. To that end, spatio-temporal modeling has emerged as a promising alternative, explicitly capturing dependencies across both spatial and temporal dimensions. Previous studies by Francke and Van de Minne (2017); Anundsen et al. (2022) demonstrate the potential of these methods in improving the reliability of rental indices by incorporating spatial effects and addressing compositional biases. Despite these advancements, existing approaches remain limited in their ability to create rental indices which are useful across different settings, including periods with minimal transactions or high volatility. Because of relying heavily on historical transactions, the traditional indices often suffer from distortions during periods of low market activity. Moreover, most models fail to account for active leases that continue to influence market conditions, resulting in indices that reflect only a partial view of market dynamics. The inability to incorporate such granular and time-sensitive information poses a significant challenge for accurately capturing the dynamics of the commercial office market.

We bridge this research gap in the current study and introduce a novel Bayesian spatio-temporal modeling approach to construct a robust rental index. The usefulness of the proposed methodology is illustrated with the construction of indices for commercial office spaces across major Indian cities. Leveraging a comprehensive dataset provided by CRE Matrix[1], a Mumbai-based firm specializing in real estate market data, our proposed methodology overcomes the challenges of data sparsity and heterogeneity. The dataset captures critical attributes such as lease terms, contract dates, building locations, and property grades, forming the foundation for our analysis. Unlike conventional indices, this approach integrates real transaction data with predicted hypothetical transactions for active leases. By including these hypothetical values, the index ensures comprehensive coverage, even in the absence of recent transactions, and minimizes distortions caused by sparse data.

The proposed methodology utilizes a dynamic Bayesian framework designed to seamlessly adapt to new quarterly data or other time-based intervals. Posterior distributions from previously generated indices are leveraged as priors for subsequent updates, allowing the model to recalibrate efficiently using only the most recent year's data. This sequential process significantly reduces computational effort while preserving accuracy, as confirmed through comparisons with full-data models. By isolating location and temporal impacts, the methodology provides a detailed understanding of geographic and time-driven trends affecting rental values. Furthermore, advanced smoothing techniques, such as variance-based weighted averaging, enhance index stability by reducing volatility, ensuring that stakeholders have access to a robust and reliable tool for market analysis.

The flexibility of this approach allows for analyses at multiple levels, including city-wide, macromarket, and granular building-level insights. Additionally, the methodology is extendable to emerging domains such as green building indices, affordable housing markets, and residential rental trends, demonstrating its adaptability and relevance in a variety of contexts. By addressing data gaps, incorporating active leases, and capturing spatial and temporal dependencies, this spatio-temporal model provides a more accurate and comprehensive view of rental market dynamics. This paper aims to fill a critical gap in the analysis of India's commercial office rental market with a scalable and robust methodology for constructing rental indices, thereby advancing the understanding of commercial real estate dynamics. As a consequence, it equips stakeholders with actionable insights for investment planning, policy formulation, and urban development.

The remainder of this paper is structured as follows: Section 2 introduces the Indian office rental data used in the analysis. Section 3 presents the proposed spatio-temporal model and discusses its implementation. Section 4 outlines the methodology for constructing the rental index. Section 5 presents the model outcomes and the resulting indices. Finally, Section 6 provides key conclusions and highlights potential directions for future research.

2. Data

Accurate assessment and official registration of property details are critical in the real estate sector. However, rental agreements are often private contracts and not publicly accessible, making comprehensive analysis challenging. Commercial leases vary widely in duration, market conditions, and terms, often including provisions such as rent-free periods, escalations, and atypical security deposits. These complexities make it difficult to derive meaningful insights directly from rental contract values. Additionally, given the influence of time on market values and the impact of ongoing developments, relying solely on signed contract rental data may not accurately capture market dynamics. To address these challenges, our study uses a comprehensive dataset spanning over a decade, starting from 2010. This dataset provides various information on newly signed leases in the mentioned time period across

[1] https://www.crematrix.com/

top ten markets of India by current office stock: Bengaluru, Mumbai, Navi Mumbai, Pune, Thane, Chennai, New Delhi, Hyderabad, Gurgaon, and Noida. This dataset offers significant depth and scope, accounting for rent-free periods, moratoriums, escalations, and abnormal security deposits, by computing an "effective rent". This measure adjusts for the other factors and serves as the primary variable (scaled by lease area, to reflect effective rent per square foot) in our analysis.

Along with the effective rent per square foot, the dataset encompasses a wide range of features, including property details such as building name, complex name, micromarket, macromarket, city, and geographic coordinates (latitude and longitude). Lease-specific information includes sign date, commencement date, absorption date, lease length, lock-in period, and the size of the leased space. Additional details, such as free rental periods and other specific contract provisions, are also included in some cases. In our main analysis though, few variables such as lock-in period and rental period are excluded due to considerable missingness. Another key feature available in the data is the grade of every building, categorized as Grade A+, Grade A, and Grade B/C. This grading system is based on various parameters, including single ownership versus strata sold, building age, size, location, developer reputation, available amenities, parking facilities, and facade quality. This detailed categorization enables a nuanced analysis of rental trends and facilitates a better understanding of how different building categories perform in the commercial office market. By leveraging this rich dataset, our study addresses limitations associated with incomplete and unstructured data, providing a solid foundation for constructing a spatio-temporal rental index. This index is meant to offer a comprehensive and accurate representation of the rental market, capturing trends across diverse locations and time frames. Note that we develop separate models for each city, acknowledging their inherently differential market behavior. Thus, outliers are handled at the city level. Specifically, we exclude contracts with an effective rent recorded as zero and remove the top and bottom one percentiles of effective rents for each city as well. The summary statistics of the effective rent, along with the total number of transactions, for each city after the cleaning and pre-processing of the data, are presented in Table 1.

Table 1: Summary statistics of effective rents per square feet across different cities (January 2010 to September 2024).

City	Observations (N)	Mean	Standard deviation	Minimum	Maximum
Bengaluru	10151	78.82	35.77	21.60	414.90
Chennai	4255	62.58	23.37	15.00	281.10
Delhi	2004	139.49	68.57	31.00	471.10
Gurgaon	6891	88.71	31.03	16.80	236.20
Hyderabad	4127	58.79	22.71	10.00	216.30
Mumbai	28670	134.12	68.96	11.10	691.10
Navi Mumbai	3203	72.06	30.48	17.30	235.40
Noida	3080	58.66	22.13	14.40	219.00
Pune	8570	73.68	27.24	8.20	278.60
Thane	1854	70.22	26.98	14.80	232.00

From Table 1, we observe that Thane has the fewest transactions, with only 1,854 records, while Mumbai leads with 28,670 transactions. Delhi also records relatively few transactions, with approximately 2,000. Both Delhi and Mumbai exhibit the highest average rents and the greatest variability in rental values, whereas cities like Noida and Hyderabad are on the lower end of the spectrum. Notably, Delhi has the highest minimum rent, suggesting that properties in Delhi have consistently commanded premium rents from the beginning. Lease terms also vary significantly across cities. Mumbai records the shortest average lease terms, while Hyderabad has the longest. These lease terms influence rental prices, with their impact differing across cities. This variation underscores the importance of using separate models for each city to accurately capture these unique dynamics.

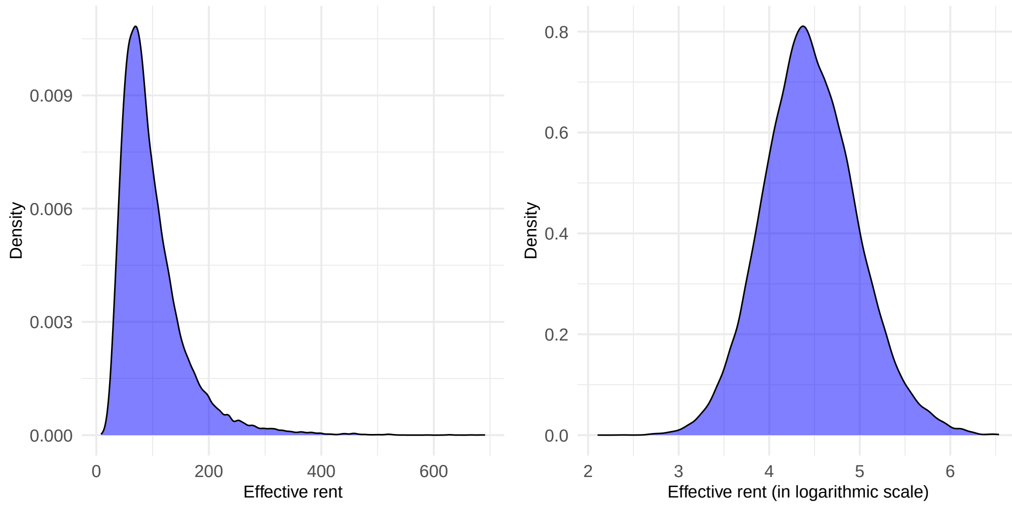

Fig. 1: Density plots comparing the distribution of effective rent (left panel) and log-transformed effective rent (right panel) per square foot across all data.

Figure 1 presents the density plots for rent per square foot and log-transformed rent per square foot, combining data from all cities. The left panel illustrates the highly skewed distribution of raw rent values, exhibiting significant right skewness, a common feature of real estate data. In contrast, the right panel shows the transformed data after applying a logarithmic transformation, resulting in a distribution that more closely approximates normality. This transformation addresses the non-Gaussian nature of the raw data, facilitating more robust modeling. As such, log(rent per square foot) is adopted as the primary response variable for subsequent analyses.

3. Spatio-temporal model for creating the rental index

As mentioned above, each city in our analysis demonstrates unique rental dynamics with varying impacts of different covariates, which motivates us to develop separate models for each city, resulting in a total of 10 distinct models. Within a city, however, we use a single spatio-temporal model for all macro and micromarkets taken together, operating under the assumption that the impacts of covariates remain consistent across different regions, and that the pricing depends on both spatial and temporal proximity to other observations. This approach ensures that our methodology is tailored to the unique characteristics of each city while maintaining a cohesive framework for submarket analysis. A comprehensive list of all the variables which are considered in each model and their descriptions are provided in Table 2 for better understanding.

Table 2: Variables considered in the main analysis and their descriptions.

Category	Variable	Description
Spatial unit	Building	Longitude and latitude of a given building.
Temporal unit	Monthly	Month in which the contract was completed.
Response	Effective rent	Effective rent of the unite per square foot.
	Grade	Based on the quailty of the building, categorized in three levels: A+, A and B/C.
Covariates	Lease Term	Lease duration for which the contract is signed. categorized in 3 levels: cat1 ($\leqslant$ 36 months), cat2 ($>$ 36 months & $\leqslant$60 months), cat3 ($>$ 60 months).
	Area	Square foot of the unit for which the contract is signed.

A critical aspect of constructing the rental index is to ensure computational efficiency and scalability over time. As an ongoing process, the methodology must handle quarterly updates seamlessly, even as the dataset grows. Moreover, the model should effectively capture how building-level impacts evolve with the incorporation of new data. To address these requirements, we begin with a comprehensive model fitted to the first seven years of data (January 2010 to December 2016). For subsequent updates, the model is re-calibrated quarterly using the most recent one-year dataset. The quarterly updation process involves compiling data from the current quarter and the previous three quarters to form a rolling one-year dataset. Posterior distributions from the last fitted model are employed as priors for the new model, creating a streamlined and consistent framework. This sequential approach minimizes computational burden while maintaining accuracy and robustness. By leveraging historical insights and dynamically incorporating new data, the methodology effectively tracks evolving trends and locational impacts, ensuring that the rental index remains adaptive, reliable, and relevant for ongoing market analysis. In the following subsection, we outline the general methodology applicable to a specified set of geographic locations and time points.

Notations: The set of real numbers and the set of positive integers are denoted by $\mathbb{R}$ and $\mathbb{N}$, respectively. We shall use $\mathcal{N}_q(\cdot,\cdot)$ to indicate a q-dimensional Gaussian distribution with appropriate mean and variance parameters. We shall drop the subscript q whenever it is clear from the context. Also, wherever used, $\mathbf{0}$ indicates a vector of 0s of appropriate order, $\boldsymbol{I}_q$ represents the identity matrix of order q, and $\mathbb{I}(A)$ serves as the indicator function for set A.

3.1. Main methodology

In this section and beyond, we refer to "location" as specific buildings acting as geographical units within our database, each characterized by longitude and latitude coordinates, denoted by $s_i \in \mathbb{R}^2$ for $i = 1, \ldots, S$. Here, S indicates the total number of individual office buildings in the dataset. Monthly time-points, represented by $t_j \in \mathbb{N}$ for $j = 1, \ldots, T$, mark instances when data are recorded. Each combination of location and time-point, (s_i, t_j), is associated with k_{ij} observations. The value of k_{ij} is greater than 1 if there are multiple rental contracts signed in a building for a given month, while it remains zero when there is no rental agreement signed for that particular combination, which is the predominant case in our dataset. It is imperative to point out that many spatio-temporal studies in the extant literature require unique observation per location and time-point, thereby forcing a need to aggregate multiple observations. For example, Holly et al. (2010) took a panel data approach and considered the price at state level in the spatial domain. Fotheringham et al. (2015), on the other hand, took random samples from every time-point to ensure non-overlapping observations. In contrast to these approaches, notably, we do not perform any aggregation based on time or location; and instead focus on analyzing each rental contract individually to extract better insights.

We reiterate that the number of unique locations and unique time-points in the data are denoted by S and T, respectively. Let each unique location be linked with l_i transactions, and each time-point with m_j transactions. Then, total number of observations is given by

$$N = \sum_{i=1}^{S}\sum_{j=1}^{T} k_{ij} = \sum_{i=1}^{S} l_i = \sum_{j=1}^{T} m_j.$$

We represent the response variable, the logarithm of effective rent per square foot, as $y(s_i, t_j, r_{ij})$, for an office r_{ij} in building s_i at time point t_j, with $i = 1, \ldots, S;\ j = 1, \ldots, T;\ r_{ij} = 1, \ldots, k_{ij}$. Our proposed model takes the form:

$$y(s_i, t_j, r_{ij}) = \mu(s_i, t_j, r_{ij}) + w(s_i, r_{ij}) + v(t_j, r_{ij}) + \epsilon(s_i, t_j, r_{ij}). \tag{1}$$

Here, $\mu(s_i, t_j, r_{ij})$ signifies the mean value of the response, encompassing all covariate effects and trend patterns. The error terms $\epsilon(s_i, t_j, r_{ij})$ are assumed to be independent and identically distributed Gaussian variables with mean 0 and variance σ_ϵ^2. Next, it is important to note that we opt to analyze temporal and spatial dependencies separately. While an alternative approach is to allow interaction between the two types of dependencies (see, e.g., Muto et al., 2023; Gupta and Deb, 2024), we choose the additive form of spatial and temporal processes for better explainability and computational ease. In this regard, we further assume that all transactions at a given location share the same spatial characteristics. An identical assumption is made in case of transactions at the same time-point as well. Hence, we are going to assume $w(s_i, r_{ij}) = w(s_i)$ and $v(t_j, r_{ij}) = v(t_j)$, where $w(s_i)$ denotes a zero-mean Gaussian process capturing spatial dependence, while $v(t_j)$ represents another zero-mean Gaussian process capturing temporal dependence. More details of these two processes are going to be discussed below. Regarding the mean structure, we consider the additive effect of all covariates, leading to the following model representation:

$$y(s_i, t_j, r_{ij}) = \sum_{p=0}^{m} \beta_p x_p(s_i, t_j, r_{ij}) + w(s_i) + v(t_j) + \epsilon(s_i, t_j, r_{ij}). \tag{2}$$

Here, x_p indicates the values of the p^{th} covariate, and we set $x_0(s_i, t_j, r_{ij}) = 1$ to let β_0 serve as the intercept in the model. Since $w(s_i)$ and $v(t_j)$ are recurrently used for the same combination of location i and time point j respectively, we can represent the above model in matrix form as

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{W} + \boldsymbol{A}V + \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{Y}$ is an $N \times 1$ vector representing all observations of the response variable, arranged first by location and then by time; and $\mathbb{X}$ is the corresponding $N \times (m+1)$ design matrix, with the first column containing all 1s and subsequent columns containing information about the regressors. The parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^\top$ captures the covariate effects. The term $\boldsymbol{B}$ is a sparse $N \times S$ binary matrix where each row contains 1 in the columns corresponding to the appropriate locations. Similarly, $\boldsymbol{A}$ is a sparse $N \times T$ binary matrix with 1 denoting presence of transactions at appropriate time-points. Cleaerly, the count of 1s in each column of $\boldsymbol{B}$ matches l_i for that specific location, and the count of 1s in each column of $\boldsymbol{A}$ matches m_j for that specific time-point. Finally, $\boldsymbol{\epsilon}$ is the N-dimensional white noise vector satisfying $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$.

To model spatial and temporal dependence separately in observed outcomes, we assume $w(s_i)$ and $v(t_j)$ to be zero-mean Gaussian processes. Specifically, we consider $\boldsymbol{W} \sim \mathcal{N}_S(\boldsymbol{0}, \Sigma_s)$, where Σ_s is the covariance matrix with spatial dependence, and $\boldsymbol{V} \sim \mathcal{N}_T(\boldsymbol{0}, \Sigma_t)$, where Σ_t is the covariance matrix with temporal dependence. These two covariance matrices are defined using the following relationships:

$$\begin{aligned} \text{Cov}\left\{w(s_i), w(s_j)\right\} &= \sigma_w^2 \rho_s \left(\|s_i - s_j\|\,; \phi_s\right), \\ \text{Cov}\left\{v(t_u), v(t_v)\right\} &= \sigma_v^2 \rho_t \left(\|t_u - t_v\|\,; \phi_t\right) \end{aligned} \tag{4}$$

In the above, $\|t_u - t_v\|$ is the absolute time difference, and $\|s_i - s_j\|$ is calculated using the Vincenty Ellipsoid method, renowned for its superior accuracy (Hijmans, 2021). For both covariance functions, we adopt the Matérn structure with $\nu = 0.5$, resulting in an exponentially decaying pattern:

$$\begin{aligned} \rho_s\left(||s_i - s_j||; \phi_s\right) &= \exp\left(-\phi_s \|s_i - s_j\|\right), \\ \rho_t\left(||t_u - t_v||; \phi_t\right) &= \exp\left(-\phi_t |t_u - t_v|\right). \end{aligned} \tag{5}$$

3.2. Bayesian Implementation

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_w^2, \sigma_v^2, \sigma_\epsilon^2, \phi_s, \phi_t)^\top$ denote the vector of all unknown parameters in the proposed model. Our goal is to estimate these parameters using a Bayesian approach, which requires careful specification of prior distributions. For the coefficients in $\boldsymbol{\beta}$, we choose independent Gaussian priors with a mean of 0 and a relatively large variance c, typically around 10^4. For the error variances σ_ϵ^2, σ_w^2, and σ_v^2, we adopt independent inverse gamma priors denoted as $IG(a, \lambda)$, where a and λ are the shape and scale parameters, respectively. To maintain non-informativeness in our priors, we set $a = 2$. Furthermore, our analysis indicates that the parameter estimates and model selection criteria are robust to variations in the scale parameter. Therefore, we fix λ to 1 throughout this study. We note that these choices are fairly common in related spatio-temporal studies from other domains, such as Sahu et al. (2006); Rawat and Deb (2023), and many others.

Moving on to $\boldsymbol{\phi} = (\phi_s, \phi_t)^\top$, the spatial and temporal decay parameters in the covariance functions, we utilize the single-variable slice sampling algorithm (Neal, 2003) to directly estimate $\boldsymbol{\phi}$ within the Bayesian framework. Albeit a common alternative is to employ a cross-validation scheme, it is restrictive, whereas the slice sampling algorithm facilitates a thorough exploration of the parameter space, enabling us to consider a broad range of values without overlooking potential solutions. We specify uniform priors within the intervals $(0.25, 6)$ and $(0.1, 1)$ for the spatial and temporal decay parameters, respectively. These priors reflect our belief that the extent of significant spatial dependence may vary from 0.5 km to a considerable distance (approximately 12 km), while the same for temporal dependence may range from 3 months to 30 months.

To utilize Bayesian computations techniques for estimation and prediction of the proposed model, we begin by expressing the proposed model as

$$\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{V} \sim \mathcal{N}\left(\mathbb{X}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{W} + \boldsymbol{A}\boldsymbol{V}, \sigma_\epsilon^2 \mathbf{I}_N\right), \tag{6}$$

with the prior specifications

$$\begin{aligned} &\boldsymbol{\beta} \sim \mathcal{N}_{m+1}\left(0, 10^4 \mathbf{I}_{m+1}\right), \\ &\sigma_\epsilon^2 \sim IG(2,1),\ \sigma_w^2 \sim IG(2,1),\ \sigma_v^2 \sim IG(2,1), \\ &\phi_s \sim U(0.25, 6),\ \phi_t \sim U(0.1, 1), \\ &\boldsymbol{W} \sim \mathcal{N}_S\left(0, \sigma_w^2 \Sigma_s\right),\ \boldsymbol{V} \sim \mathcal{N}_T\left(0, \sigma_v^2 \Sigma_t\right). \end{aligned} \tag{7}$$

Subsequently, the complete posterior distribution can be written as

$$\begin{aligned} f(\boldsymbol{\beta}, \boldsymbol{W}, \boldsymbol{V}, \sigma_\epsilon^2, \sigma_w^2, \sigma_v^2, \phi_s, \phi_t \mid \boldsymbol{Y}) \propto\ & f(\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{W}, \boldsymbol{V}, \sigma_\epsilon^2, \sigma_w^2, \sigma_v^2, \phi_s, \phi_t) \\ & f(\boldsymbol{W} \mid \sigma_w^2, \phi_s) f(\sigma_w^2 \mid \phi_s) \\ & f(\boldsymbol{V} \mid \sigma_v^2, \phi_t) f(\sigma_v^2 \mid \phi_t) \\ & f(\phi_s \mid \phi_t) f(\phi_t) f(\sigma_\epsilon^2) f(\boldsymbol{\beta}). \end{aligned} \tag{8}$$

The Bayesian computation makes use of Gibbs sampling (Gelfand, 2000) and slice sampling steps, both of which require the conditional posterior distribution of each parameter, taking into account the data and information about other parameters. To simplify notation, for each such conditional posterior distribution written below, we consistently represent all available information as $\mathcal{F}$. Then, we can express the conditional posterior distributions for $\boldsymbol{\beta}, \sigma_w^2, \sigma_v^2, \sigma_\epsilon^2$, ϕ_s, ϕ_t, $\boldsymbol{W}$, and $\boldsymbol{V}$ as follows:

$$
\begin{aligned}
(\boldsymbol{\beta} \mid \mathcal{F}) &\sim \mathcal{N}_{m+1}\left(\left(\frac{\mathbb{X}^\top\mathbb{X}}{\sigma_\epsilon^2} + \frac{\mathbf{I}_{m+1}}{10^4}\right)^{-1}\left(\frac{\mathbb{X}^\top\left(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{W} - \boldsymbol{A}\boldsymbol{V}\right)}{\sigma_\epsilon^2}\right), \left(\frac{\mathbb{X}^\top\mathbb{X}}{\sigma_\epsilon^2} + \frac{\mathbf{I}_{m+1}}{10^4}\right)^{-1}\right), \\
(\sigma_\epsilon^2 \mid \mathcal{F}) &\sim IG\left(a + \frac{N}{2}, \frac{\left\|\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{W} - \boldsymbol{A}\boldsymbol{V}\right\|^2}{2} + \lambda\right), \\
(\sigma_w^2 \mid \mathcal{F}) &\sim IG\left(a + \frac{S}{2}, \frac{\boldsymbol{W}^\top\Sigma_s^{-1}\boldsymbol{W}}{2} + \lambda\right), \\
(\sigma_v^2 \mid \mathcal{F}) &\sim IG\left(a + \frac{T}{2}, \frac{\boldsymbol{V}^\top\Sigma_t^{-1}\boldsymbol{V}}{2} + \lambda\right), \\
(\phi_s \mid \mathcal{F}) &\propto |\Sigma_s|^{-1/2}\exp\left(-\frac{\boldsymbol{W}^\top\Sigma_s^{-1}\boldsymbol{W})}{2\sigma_w^2}\right)\mathbb{I}(0.25 < \phi_s < 6), \\
(\phi_t \mid \mathcal{F}) &\propto |\Sigma_t|^{-1/2}\exp\left(-\frac{\boldsymbol{V}^\top\Sigma_t^{-1}\boldsymbol{V}}{2\sigma_v^2}\right)\mathbb{I}(0.1 < \phi_t < 1), \\
(\boldsymbol{W} \mid \mathcal{F}) &\sim \mathcal{N}_S\left(\left(\frac{\Sigma_s^{-1}}{\sigma_w^2} + \frac{\boldsymbol{D_s}}{\sigma_\epsilon^2}\right)^{-1}\left(\frac{\boldsymbol{B}^\top(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta} - \boldsymbol{A}\boldsymbol{V})}{\sigma_\epsilon^2}\right), \left(\frac{\Sigma_s^{-1}}{\sigma_w^2} + \frac{\boldsymbol{D_s}}{\sigma_\epsilon^2}\right)^{-1}\right), \\
(\boldsymbol{V} \mid \mathcal{F}) &\sim \mathcal{N}_T\left(\left(\frac{\Sigma_t^{-1}}{\sigma_v^2} + \frac{\boldsymbol{D_t}}{\sigma_\epsilon^2}\right)^{-1}\left(\frac{\boldsymbol{A}^\top(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{W})}{\sigma_\epsilon^2}\right), \left(\frac{\Sigma_t^{-1}}{\sigma_v^2} + \frac{\boldsymbol{D_t}}{\sigma_\epsilon^2}\right)^{-1}\right).
\end{aligned}
\tag{9}
$$

In the above, $\boldsymbol{D_s}$ is a diagonal matrix with dimensions $S \times S$, where each diagonal entry corresponds to the value l_i associated with location s_i and $\boldsymbol{D_t}$ is a diagonal matrix with dimensions $T \times T$, where each diagonal entry corresponds to the value m_j associated with time point t_j.

To conclude this section, we will clearly outline the essential steps for implementing our main analysis. The model integrates the variables listed in Table 2, along with the squared logarithm of the area, leading to a total of six covariates and an intercept term in the additive mean structure. For parameter estimation, we apply non-informative prior distributions as described in (7). We then employ Gibbs and slice samplers, as previously discussed. The convergence of the process is assessed using the Geweke diagnostic (Geweke et al., 1991). Once convergence is confirmed, we run 2000 iterations of the Markov Chain Monte Carlo method, selecting every 10th sample to ensure independence among observations. Finally, these posterior samples are used for inference.

4. Construction of the Rental Index

After fitting the proposed model to the data, the estimates corresponding to all components are obtained. To create a rental index, we work under the philosophy that it should reflect the changing market behavior based on location and time. At this point, we recognize that the features such as lease term, building category, and area are specific to each building and influence rental prices based on changes in these characteristics. On the other hand, we estimate the vectors $\boldsymbol{W}$ and $\boldsymbol{V}$, which capture the spatial and temporal effects based on the data, after removing the effects of the building-specific features. Therefore, we propose to utilize only these random effects in the creation of rental indices as they accurately reflect location and time-dependent changes. For better understanding, Figure 2 illustrates the structure of our procedure, where the logarithm of effective rent is modeled based on covariate effects, spatial and temporal error processes, along with a white noise error component. Then, the rental index for a particular city, macromarket, or micromarket is derived based on the estimated spatial and temporal error components. The mathematical underpinning of this is explained below.

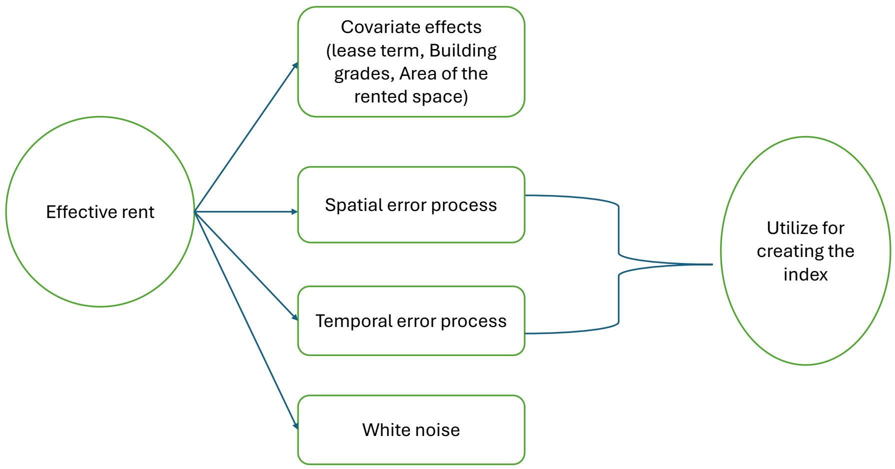

Fig. 2: Structure of our proposed method of rental index creation.

As the first step of the rental index creation, we consider the estimate of the vector $\boldsymbol{W}$, representing the spatial error process, and use it to assess the spatial impact of all buildings with active leases at a given time point. We recall that the index updation process takes advantage of the Bayesian framework and implements the model using data from a rolling one-year window during which some of the active leases may not have been renewed. For those cases, we evaluate the spatial impact using available information and the estimated model. This allows us to capture location information comprehensively, ensuring that the rental index reflects the entire market, not just new transactions in a particular time window. In fact, this is one of the key advantages of our framework, as this approach ensures robust estimation and prediction, providing insights into market trends even in sparse data scenarios.

Mathematically, at a given time point t, consider an office building with location coordinates $s^* \in \mathbb{R}^2$, which did not observe a new transaction but has been active during the period. For obtaining the spatial impact $w(s^*)$, we note that it is dependent on the observed spatial impact $\boldsymbol{W}$, following the correlation structure mentioned in Section 3.1. Using multivariate distribution theory and the concept of partitioning a matrix, the joint distribution can be expressed as

$$\begin{pmatrix} w(s^*) \\ \boldsymbol{W} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_w^2 \begin{bmatrix} 1 & \Sigma^{*\top} \\ \Sigma^* & \Sigma_s \end{bmatrix}\right), \tag{10}$$

where $w(s^*)$ is the spatial impact of the target location, the second sub-vector $\boldsymbol{W}$ is the spatial impact estimated in the fitted model, and Σ^* is a vector of with i^{th} entry as $\rho_s\left(||s_i - s^*||; \phi_s\right)$, which provides the spatial dependence of the target location on other transactions. Let $\hat{\boldsymbol{w}}$ be the posterior estimates of $\boldsymbol{W}$ obtained from the fitted model. Then, from the multivariate normal theory, the conditional distribution of $w(s^*)$ given $\boldsymbol{W}$ is obtained as

$$\left(w(s^*) \mid \boldsymbol{W} = \hat{\boldsymbol{w}}\right) \sim \mathcal{N}\left(\Sigma^* \Sigma_s^{-1} \hat{\boldsymbol{w}}, \sigma_w^2 (1 - \Sigma^* \Sigma_s^{-1} \Sigma^{*\top})\right). \tag{11}$$

For ease of understanding of the proposed approach, turn attention to Figure 3 which illustrates the philosophy using a simple case study focused on Bengaluru, as depicted on the map.

At time point $T = 1$, there are 7 active transactions corresponding to distinct building locations, for which we can estimate the spatial impacts. Then, at $T = 2$ and $T = 3$, we observe 8 and 10 new

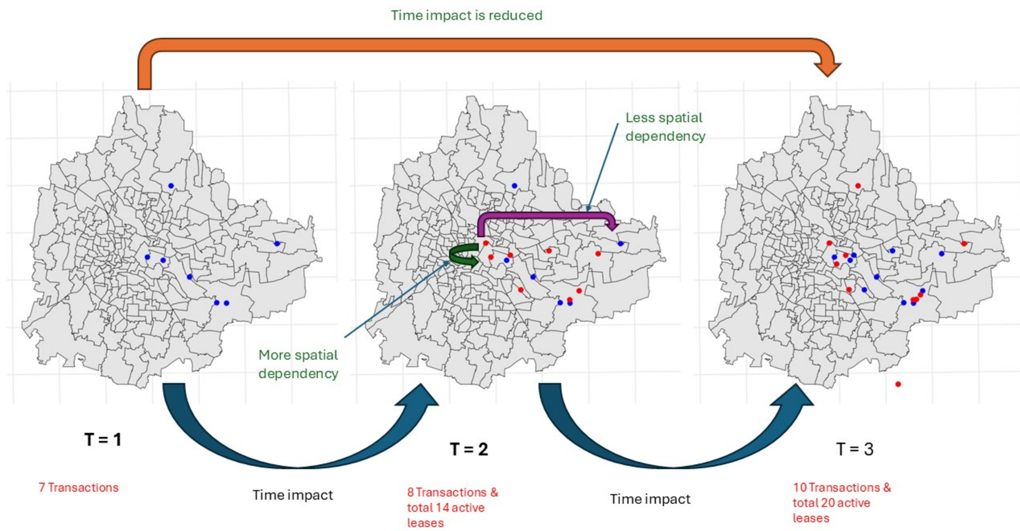

Fig. 3: Illustration of incorporating active leases from unobserved buildings into the analysis.

transactions respectively, and estimate the spatial impacts for those buildings. As we explained above, our proposed methodology extends this analysis to evaluate the impact for unobserved locations with active leases as well. For example, at $T = 2$, in addition to the observed spatial impacts for the 8 transactions, we use (11) to estimate the same for the 6 unobserved locations where leases remain active. Similarly, at $T = 3$, we evaluate the hypothetical impact of location for the 10 buildings where leases remained active but there was no new transaction. This continuous assessment ensures that the location impact reflects the dynamics of all active leases, even when transactions are not directly observed. The visualization in the above figure further demonstrates how spatial dependency is captured, with closer locations exhibiting stronger correlation, while distant locations are assumed to have weaker spatial correlation. Similarly, in the temporal dimension, the dependency diminishes as the time gap increases. The continuous tracking of all transactions is performed iteratively, beginning with an initial model fitted to the historical data of seven years, and then updated with each new quarter's data. The process for constructing and updating the rental index using the robust Bayesian framework is illustrated in Figure 4.

As is shown here, initially, the model is fitted to the historical dataset to estimate posteriors and generate the rental index based on location-specific $w(s)$ and time-specific $v(t)$ impacts. When new quarterly data becomes available, the dataset is updated by incorporating information from the current quarter (say, q^{th} quarter) and the previous quarter (i.e., $q-1$), thereby adding six months of new data. This accounts for leases that may have been recorded late for the most recent quarter, while older quarters remain unchanged. The updated dataset is constructed as

$$\mathcal{D}_{\text{new}} = \{\mathcal{D}_{\text{old}} - \mathcal{D}_{q-1}\} \cup \mathcal{D}^*_{q-1} \cup \mathcal{D}_q, \tag{12}$$

where $\mathcal{D}^*_{q-1}$ represents the revised data for the previous quarter and $\mathcal{D}_q$ is the newly added data. Data from the earlier quarters $\{\mathcal{D}_{q-2}, \mathcal{D}_{q-3}, \ldots\}$ remain unchanged.

With this updated dataset, we employ two distinct models using a rolling one-year window that includes the quarter in question. Specifically, this involves two scenarios: one where the updated data from the previous quarter, along with the three quarters preceding it, is used; and another where the current quarter, with its preceding three quarters (including one updated quarter), is considered. The priors for location impacts are utilized to get the priors for location impact of newly added buildings in

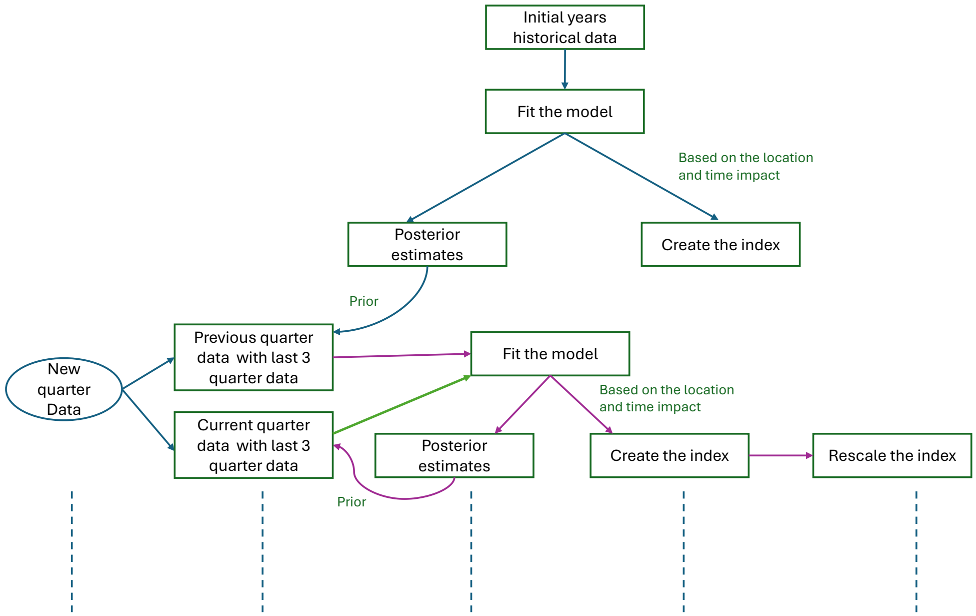

Fig. 4: Robust Bayesian Framework to Construct the Index

the updated data, following the same methodology discussed for continuous tracking of transactions. Similarly, the priors of time impacts for new time points are derived using the prior estimates of temporal effects. Once the model is fitted, the total impact (γ_{total}) is re-estimated for the entire year, covering 12 time points. Mathematically, the estimated impact for a location s and time-point t is

$$\hat{\gamma}_{\text{total};s,t} = \hat{w}(s) + \hat{v}(t), \tag{13}$$

which provides the foundation for the rental index at any granularity. Specifically, to construct the city-level rental index (say, for a city $\mathcal{C}$) for quarter q we summarize the total impacts from relevant locations and time-points as

$$\hat{\gamma}_{\mathcal{C},q} = \underset{s\in\mathcal{C},\, t\in q}{\text{median}} \left(\hat{\gamma}_{\text{total};s,t}\right). \tag{14}$$

A common practice in such an index construction is to standardize the relevant metric. We do that by ensuring that the total impact at the first time point (say, $\hat{\gamma}_{\mathcal{C},1}$) based on the respective model is set to zero; that is, we define a new adjusted metric $\hat{\gamma}_{\text{adj},\mathcal{C},q} = \hat{\gamma}_{\mathcal{C},q} - \hat{\gamma}_{\mathcal{C},1}$. Subsequently, the adjusted impacts are exponentiated and scaled by 100 to define the rental index,

$$I_{\mathcal{C},q} = 100 \exp\left(\hat{\gamma}_{\text{adj},\mathcal{C},q}\right). \tag{15}$$

As the final step of our methodology, to maintain continuity with the previously obtained index for the earlier quarter (note that the estimates for the earlier quarter may change after new data are incorporated), the new values are rescaled to align with the prior values. Specifically, we adjust the index obtained for the last quarter using the updated dataset by applying an appropriate factor, ensuring it matches the index for the same quarter as previously determined. This scaling factor is then applied to the remaining values to maintain consistency across all time points. We apply this approach to the outcomes of both fitted models, updating the index values for both the last quarter and the current quarter accordingly.

While our model integrates all active leases to create the rental index, it is important to acknowledge that these leases often reflect prices agreed upon at the time of signing. Such prices may not fully capture the ongoing market conditions but rather the localized or short-term volatility that prevailed at the time. For instance, a surge in demand or a temporary lack of supply in a specific region could lead to lease agreements at unusually high or low rental rates. Over time, these anomalies can skew the rental index, reducing its ability to represent the broader market accurately. Therefore, a smoothing mechanism becomes essential to mitigate such effects and ensure that the index provides a stable and reliable representation of rental trends.

To address these challenges, we employ a variance-based smoothing approach that utilizes the full set of Bayesian posterior estimates. For each city and time point, the model generates 200 indices from posterior samples, capturing uncertainties in the estimation process. The mean and variance of these indices are calculated at each time point to construct the final index. This involves combining the current mean index value with the mean values from the previous three quarters, weighted by the reciprocal of the variance. Indices with lower variance, reflecting higher confidence, are given greater weight, while those with higher variance are weighted less. This method balances recent trends with historical stability, providing a more nuanced representation of market dynamics. Unlike traditional techniques such as exponential averaging, which do not differentiate based on data certainty, this variance-based approach explicitly incorporates uncertainty. By doing so, it enhances the index's robustness and reliability, offering stakeholders a credible tool for informed decision-making.

5. Results and discussion

5.1. Estimated model and interpretation

As outlined earlier, we initially fit the models using the first seven years of data. Subsequently, the models are updated and re-estimated for two quarters when new quarterly data becomes available, using a rolling one-year dataset. Consequently, the coefficients of each parameter are subject to change with each update. For demonstration purposes, we present the model results based on the initial seven years of data in Table 3, showcasing the model fitting outcomes across all 10 cities. The statistical metrics reported in the Table 3 provide crucial insights into the model's performance. The R^2, or the coefficient of determination, indicates the proportion of variance in the dependent variable that can be predicted from the independent variables, with a higher R^2 suggesting a better model fit. The Mean Absolute Error (MAE) reflects the average magnitude of errors in predictions, disregarding direction, and serves as an average of absolute differences between predictions and actual observations. Meanwhile, the Mean Absolute Percentage Error (MAPE) expresses the forecast accuracy as a percentage, offering a relative measure of error that is particularly useful for understanding prediction accuracy in proportional terms. Lastly, the Root Mean Square Error (RMSE) provides the square root of the average of squared differences between predicted and actual values, offering a sense of the model's absolute fit by illustrating how closely the observed data align with the predicted values. Together, these metrics offer a comprehensive evaluation of the model's accuracy and predictive capabilities.

It is commonly assumed that shorter leases command higher rents due to the transaction costs associated with more frequent renewals. However, our model, which considers effective rent (a comprehensive measure accounting for factors like rent-free periods and escalations), reveals a more nuanced picture. For leases between 3 and 5 years, we observe no significant price difference compared to shorter leases (within 3 years) in cities like Bengaluru, Chennai, Delhi, Gurgaon, Hyderabad, Navi Mumbai, and Thane. In contrast, cities like Mumbai, Noida, and Pune demonstrate higher rents for this lease category, with Noida exhibiting the highest impact. When considering leases longer than 5 years, the results diverge further. While it might be expected that extended lease terms result in lower rents due to market uncertainty or a preference for flexibility, our findings indicate otherwise. In most cities, leases exceeding 5 years attract higher rents, with the exceptions being Delhi and Thane.

Table 3: Posterior means and credible intervals for parameters across 10 major cities of India (January 2010 to December 2016).

Variable	Bengaluru	Chennai	Delhi	Gurgaon	Hyderabad
Intercept	3.401 (3.106, 3.747)	3.733 (3.202, 4.237)	4.11 (2.78, 5.481)	4.188 (3.859, 4.53)	4.347 (3.891, 4.841)
lease (cat 1)	0.008 (−0.014, 0.03)	−0.016 (−0.05, 0.013)	0.057 (−0.037, 0.174)	−0.003 (−0.029, 0.024)	−0.005 (−0.034, 0.032)
lease (cat 2)	0.074 (0.047, 0.107)	0.074 (0.037, 0.1)	0.066 (−0.049, 0.174)	0.092 (0.064, 0.114)	0.101 (0.058, 0.136)
Grade A	0.143 (0.075, 0.197)	0.054 (−0.046, 0.127)	0.122 (0.01, 0.244)	0.034 (−0.023, 0.097)	0.027 (−0.109, 0.166)
Grade A+	0.142 (0.047, 0.234)	0.119 (−0.02, 0.238)	0.331 (0.087, 0.575)	0.068 (−0.048, 0.171)	0.013 (−0.125, 0.156)
log (area)	0.137 (0.072, 0.201)	0.022 (−0.062, 0.1)	0.141 (−0.199, 0.443)	0.023 (−0.058, 0.098)	−0.141 (−0.229, −0.05)
$\log(area)^2$	−0.009 (−0.012, −0.005)	−0.003 (−0.007, 0.002)	−0.009 (−0.027, 0.012)	−0.003 (−0.007, 0.002)	0.006 (0.002, 0.011)
σ_e^2	0.048 (0.045, 0.05)	0.033 (0.031, 0.036)	0.07 (0.06, 0.081)	0.034 (0.032, 0.035)	0.031 (0.028, 0.034)
σ_w^2	0.123 (0.097, 0.156)	0.154 (0.115, 0.209)	0.221 (0.142, 0.355)	0.19 (0.123, 0.314)	0.141 (0.081, 0.254)
σ_v^2	0.045 (0.031, 0.059)	0.044 (0.031, 0.058)	0.107 (0.064, 0.176)	0.049 (0.033, 0.069)	0.056 (0.038, 0.08)
ϕ_s	1.676 (1.023, 2.485)	1.811 (0.891, 3.244)	1.73 (0.625, 3.814)	1.218 (0.571, 2.339)	1.681 (0.436, 4.525)
ϕ_t	0.106 (0.1, 0.125)	0.106 (0.1, 0.126)	0.139 (0.101, 0.322)	0.106 (0.1, 0.123)	0.108 (0.1, 0.134)
N	3342	1559	318	2278	1302
R^2	0.716	0.775	0.864	0.728	0.604
MAE	0.149	0.117	0.139	0.128	0.117
MAPE	3.681	3.051	2.905	2.996	3.156
RMSE	0.203	0.167	0.187	0.167	0.16

Variable	Mumbai	Navi Mumbai	Noida	Pune	Thane
Intercept	4.76 (4.463, 5.013)	3.61 (2.793, 4.448)	2.631 (1.831, 3.507)	4.819 (4.452, 5.196)	4.387 (2.993, 5.819)
lease (cat 1)	0.027 (0.013, 0.042)	−0.037 (−0.091, 0.023)	0.06 (0.012, 0.11)	0.042 (0.014, 0.067)	−0.058 (−0.164, 0.045)
lease (cat 2)	0.149 (0.11, 0.192)	0.139 (0.024, 0.247)	0.081 (0.031, 0.127)	0.085 (0.036, 0.134)	−0.2 (−0.499, 0.2)
Grade A	0.205 (0.154, 0.253)	0.123 (−0.057, 0.315)	0.049 (−0.046, 0.13)	0.143 (0.079, 0.203)	−0.079 (−0.305, 0.135)
Grade A+	0.413 (0.331, 0.502)	−0.058 (−0.32, 0.24)	0.014 (−0.121, 0.166)	0.182 (0.095, 0.251)	0.307 (−0.199, 0.9)
log (area)	−0.059 (−0.109, −0.009)	0.046 (−0.124, 0.238)	0.249 (0.077, 0.433)	−0.203 (−0.278, −0.115)	−0.058 (−0.37, 0.28)
$\log(area)^2$	0.002 (−0.001, 0.005)	−0.002 (−0.013, 0.006)	−0.016 (−0.026, −0.006)	0.008 (0.003, 0.013)	0 (−0.019, 0.018)
σ_e^2	0.077 (0.075, 0.078)	0.075 (0.069, 0.082)	0.053 (0.049, 0.057)	0.049 (0.046, 0.052)	0.111 (0.094, 0.126)
σ_w^2	0.133 (0.111, 0.159)	0.171 (0.104, 0.298)	0.238 (0.145, 0.46)	0.117 (0.079, 0.169)	0.158 (0.083, 0.347)
σ_v^2	0.037 (0.025, 0.049)	0.078 (0.049, 0.117)	0.065 (0.041, 0.092)	0.048 (0.033, 0.07)	0.117 (0.071, 0.206)
ϕ_s	2.841 (2.11, 3.565)	2.953 (0.539, 5.763)	1.246 (0.438, 2.278)	0.975 (0.483, 1.792)	3.106 (0.366, 5.831)
ϕ_t	0.105 (0.1, 0.12)	0.117 (0.1, 0.173)	0.113 (0.1, 0.14)	0.106 (0.1, 0.122)	0.143 (0.101, 0.318)
N	9165	726	881	2125	317
R^2	0.706	0.557	0.759	0.626	0.554
MAE	0.198	0.193	0.154	0.16	0.204
MAPE	4.34	4.894	3.997	4.004	5.24
RMSE	0.266	0.251	0.204	0.207	0.283

Notably, Mumbai displays the highest premium for long-term leases compared to short-term leases, highlighting its unique market dynamics. These findings underscore the variability of lease term impacts across cities, illustrating how our model effectively captures city-specific rental behaviors. This differentiation provides valuable insights into how lease terms influence the log of effective rent, challenging typical assumptions and emphasizing the importance of granular, location-specific analysis.

Expanding on the impact of grades, we observe significant variations across markets. In our analysis, Grades B and C are modeled jointly as the base category. Interestingly, cities like Chennai, Gurgaon, Hyderabad, Navi Mumbai, Noida, and Thane do not exhibit any statistically significant impact of better-grade buildings (A or A+) on effective rents. This finding is particularly surprising, given that many of these cities are emerging commercial markets where higher-grade buildings are expected to command a premium. In contrast, other cities, especially established markets like Mumbai, Bengaluru, Pune, and Delhi, show a clear premium for better-grade buildings. Mumbai, in particular, stands out with the highest difference in rental prices for Grade A and A+ buildings compared to the base category. This reflects the heightened demand for premium office spaces in Mumbai's highly competitive real estate market, where quality and location play a critical role in determining rental values. These insights highlight the varying dynamics of commercial office markets across cities. While new markets may not yet fully capitalize on building grades, established markets demonstrate that higher-grade properties consistently fetch higher rents.

The impact of area and its squared term on effective rents reveals diverse trends across cities, reflecting varying market preferences and demand dynamics. In Chennai, Delhi, Gurgaon, Navi Mumbai, and Thane, neither log(area) nor its squared term has a significant impact on rental prices. This indicates that the size of the rental property does not influence rents in these cities, suggesting that other factors, such as location, building grade, or market-specific attributes, play a more critical role in determining rental values. In contrast, Bengaluru and Noida show a distinct non-linear relationship between area and rents. Here, log(area) has a positive significant impact, while squared log(area) has a negative significant impact. This suggests that rents initially increase with property size, reflecting strong demand for moderately sized spaces. However, beyond a certain threshold, rents decline for very large areas, potentially due to lower demand for expansive spaces or discounts offered to attract tenants for large leases. This pattern indicates that mid-sized spaces are more desirable in these markets. Hyderabad and Pune exhibit the opposite trend. In these cities, log(area) has a negative significant impact, while squared log(area) has a positive significant impact. This suggests that rents decrease initially with increasing property size, likely reflecting weaker demand for mid-sized spaces. However, after reaching a certain threshold, rents increase, indicating a premium for very large office spaces and possibly a niche demand for expansive properties in these markets. In Mumbai, the results highlight a strong preference for compact spaces. Log(area) has a negative significant impact, while squared log(area) is not significant. This suggests that larger office spaces are generally associated with lower rents, possibly due to bulk discounts or weaker demand for larger areas. Conversely, smaller office spaces command higher rents, reflecting strong demand for compact properties in Mumbai's dense and competitive market. Overall, the analysis demonstrates that the impact of area on effective rents varies significantly across cities. While some cities are size-neutral, others show clear preferences for either moderately sized spaces or very large areas. These insights underscore the importance of understanding city-specific dynamics when developing pricing strategies, as market preferences for property size can vary widely and influence rental trends.

The error parameters in our model provide significant insights into the underlying dynamics of spatial and temporal dependencies in the rental market across different cities. A key observation is that, in all cases, the total variance attributed to the spatial and temporal error processes is substantially higher than the variance associated with the white noise error process. This highlights the importance of incorporating spatial and temporal components in the model, as they account for most of the unexplained variability in rental prices, justifying the use of our proposed spatio-temporal framework. Examining the decay parameters further reveals intriguing patterns. The temporal decay parameter remains relatively consistent across all cities, suggesting that the influence of time on rental dynamics behaves similarly across markets. However, the spatial decay parameter varies significantly, reflecting differences in how location impacts rental prices in different cities. For instance, in Bengaluru, Chennai, Delhi, and Hyderabad, the spatial dependency spans a range of 1.65 km to 1.85 km. This indicates that rental prices in these cities are influenced by properties within this radius, pointing to moderate spatial clustering. In Gurgaon, Noida, and Pune, the spatial dependency extends across a broader range, approximately 2.4 to 3 kilometers, indicating a wider radius of influence on rental values. This could be attributed to the expansive and relatively less dense commercial layouts in these cities. On the other hand, in Mumbai, Navi Mumbai, and Thane, the spatial dependency is around 1 km. This shorter range is indicative of highly localized rental dynamics, likely driven by the dense urban structure and high competition for compact office spaces in these markets. Interestingly, the spatial decay patterns show a clear grouping: Gurgaon and Noida exhibit similar spatial behavior, as do Mumbai, Navi Mumbai, and Thane. These findings have potential implications for future modeling strategies. For instance, given the similar spatial behaviors, a combined model could be developed for Mumbai, Navi Mumbai, and Thane, while another combined model could be explored for Gurgaon and Noida. Such an approach could streamline modeling efforts and provide more robust insights for cities with shared characteristics.

Examining the model fit metrics, such as R^2, we observe that the model performs well across all cities. However, in cities like Thane, Navi Mumbai, Noida, and Hyderabad, where the R^2 value is below 0.7, we note that the number of observations is relatively low. This lower data availability could be a contributing factor to the reduced model fit in these cases. All of the discussed results, however, are based on the first seven years of data analyzed in this study and may evolve as more recent data is incorporated. This emphasizes the need for continuous updates to the index to capture changing market dynamics and emerging trends effectively.

5.2. Rental indices for different cities

Let us now move on to the rental index, which, as discussed before, can be provided for each city, categorized by grade, macromarket or any other necessary specifications. For this study, we are going to focus on the Grade A/A+ rental index for each city, as that is of primary interests to various stakeholders. This index is calculated as a weighted average of the Grade A and Grade A+ indices, with weights determined by the number of active leases in a given quarter. Since comprehensive data is available for most cities from 2012 onward, these indices have been constructed starting from that year. The city-wise quarterly rental indices are illustrated in Figure 5.

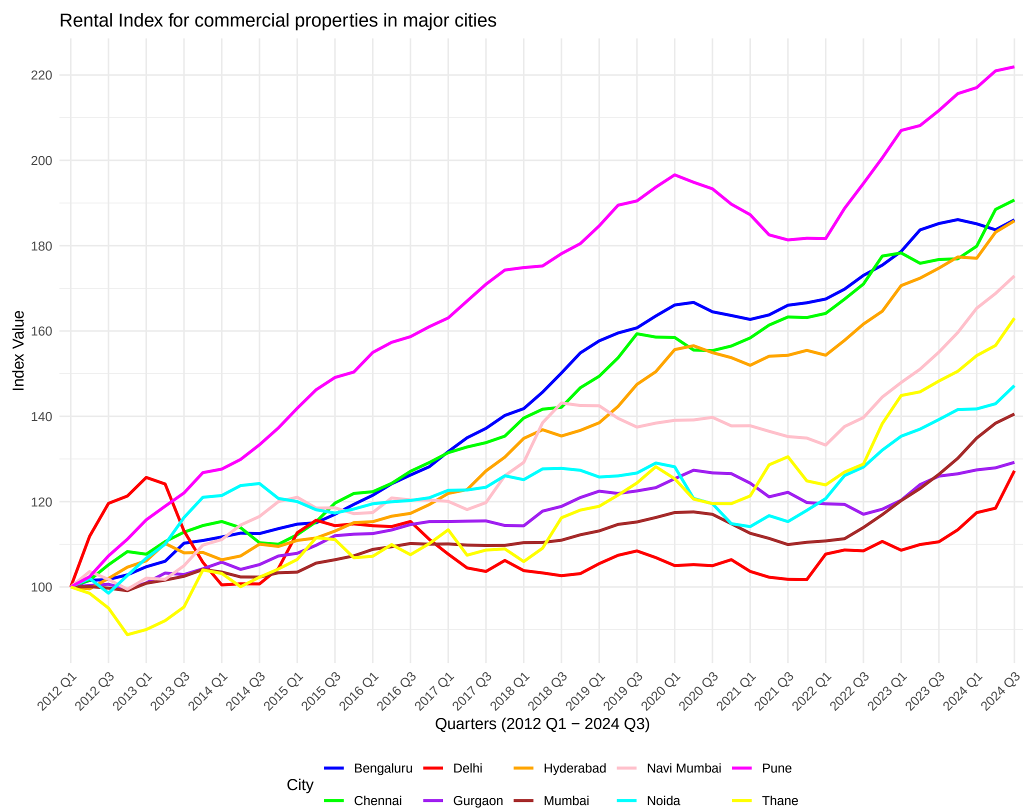

Fig. 5: Quarterly rental indices for the commercial properties in major cities in India

From the indices, we can observe year-by-year growth trends. Notably, Delhi's rental market exhibited significant growth initially but plateaued in subsequent years. This stagnation may be attributed to Delhi's already high rental prices and the rapid development of neighboring regions like Gurgaon and Noida, which attracted many companies and reduced overall demand in Delhi, limiting further price increases. Pune, on the other hand, shows the highest growth in its rental index, reflecting its increasing demand as an emerging commercial hub. The impact of the COVID-19 pandemic is evident across all cities, with a noticeable decline in rental indices during that period.

While some cities took longer to recover, others rebounded relatively quickly. Emerging markets like Navi Mumbai and Thane have also shown substantial improvements in their index values, reflecting growing commercial activity in these areas. Conversely, Mumbai, being a well-established market, has not experienced significant jumps in its rental index, likely due to its already mature status.

To further demonstrate the efficacy of our approach in capturing rental trends at a more granular level, using Bengaluru as a case study, we next compute the rental indices for its macromarkets. In Figure 6, the indices for six prominent macromarkets – CBD-Bengaluru, Off-CBD Bengaluru, South Bengaluru, North Bengaluru, Outer Ring Road, and Whitefield – are presented. These areas represent diverse commercial zones within the city, showcasing varying trends and growth patterns over the years.

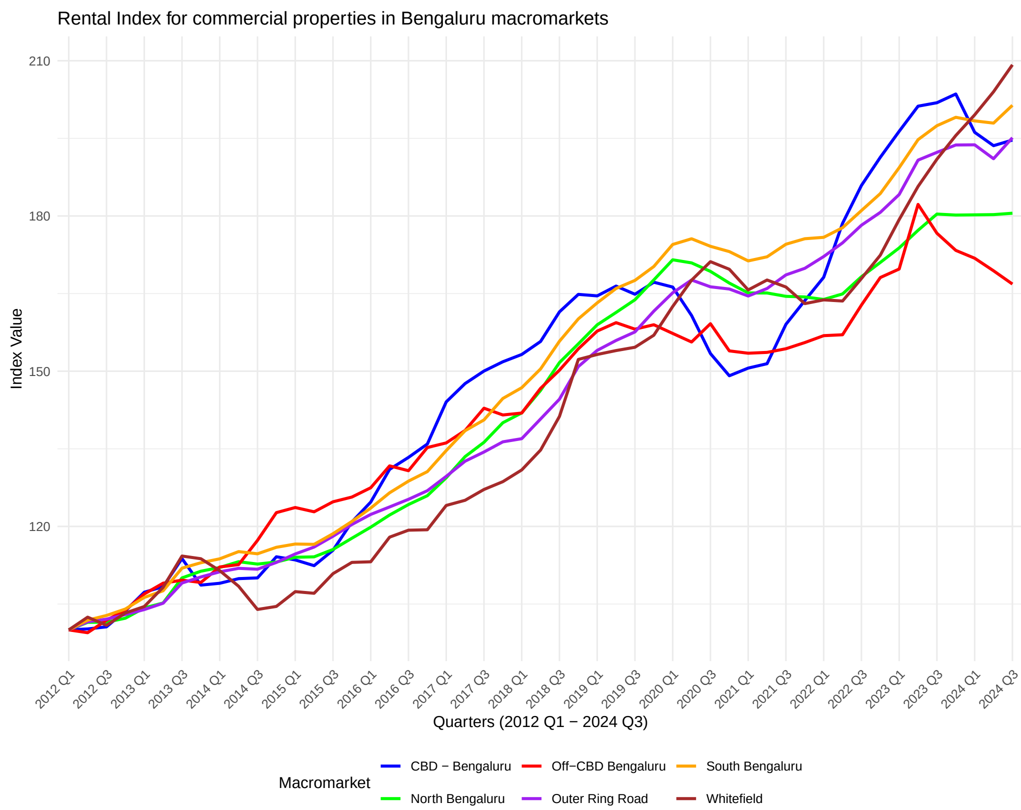

Fig. 6: Quarterly rental indices for the commercial properties in macromarkets in Bengaluru

The impact of COVID-19 is evident across all macromarkets, with a noticeable decline in rental indices during the pandemic period. This pattern reflects the overall slowdown in commercial activity and the challenges faced by the office rental market during that time. Among the macromarkets, Whitefield stands out due to its remarkable development trajectory. Over the years, its rental index has shown steady progress, particularly after the second quarter of 2022, where a significant upward trend is observed. Starting from first quarter of 2023, this growth becomes even more pronounced, with sharp increases in rental values. This surge can be attributed to the launch of the metro line in January 2023, which enhanced connectivity and boosted the market's attractiveness. Interestingly, the impact of the metro project began to manifest in the rental index before its official launch, as the anticipation of improved accessibility likely influenced market dynamics. A similar development trend is observed in the Outer Ring Road macromarket, which has also experienced consistent growth over the years. As a rapidly evolving commercial hub, this area has seen increasing demand for office spaces, driving up rental values. Conversely, already established macromarkets such as Off-CBD Bengaluru have not experienced significant growth in rental indices. These areas are relatively saturated, with limited room for further development or substantial shifts in demand. This highlights the contrast between mature markets, which exhibit stability, and emerging markets like Whitefield and the Outer Ring Road, where growth opportunities are still abundant. It is evident that the macromarket-level rental indices in Bengaluru not only underscores the varying dynamics within different zones of the city but also demonstrates the ability of our model to capture location-specific impacts, such as the influence of infrastructure developments like the metro. This granular approach provides stakeholders with actionable insights into localized market trends, enabling more informed decision-making.

We also present macromarket-level indices for other major cities in the appendix, highlighting several notable trends that stakeholders can relate to. Over the past few years, rental price patterns across various regions have shown significant variability. In Chennai, the Secondary Business District (SBD) has experienced a remarkable rise in rental prices, reflecting increased demand and development, whereas the Central Business District (CBD) has seen much slower growth. In Delhi, South Delhi has shown consistent progress in terms of rental prices, indicating steady demand. Meanwhile, in Gurgaon, the Golf Course Road continues to dominate with consistently high rental values and significant increases over time. Interestingly, Sohna Road has also seen a rise in rental prices, likely due to improved connectivity and its proximity to Golf Course Road. In Hyderabad, both Gachibowli and HITEC City have displayed similar rental growth patterns, emphasizing their shared importance as commercial and residential hubs. In Mumbai, the western suburbs and Bandra Kurla Complex (BKC) have exhibited substantial rental growth in recent years, while the northern macromarket in Navi Mumbai has also shown promising performance. Noida, on the other hand, has seen strong rental performance in Central Noida since its early days, but in recent years, suburban areas have begun to gain traction. Finally, in Pune, both the CBD and northeastern regions have performed exceptionally well, recording significant rental price growth. These observations underscore the dynamic nature of rental markets, driven by factors such as infrastructure development, connectivity, and evolving demand, offering valuable insights for stakeholders navigating these trends.

6. Conclusion

In summary, this paper introduces a spatio-temporal model to analyze the commercial rental market across India, addressing the limitations of traditional methods such as hedonic and repeat-sales approaches. Unlike models relying on fixed-location impacts or aggregated data, our methodology considers each transaction individually, dynamically capturing location and temporal trends. By continuously tracking active leases and estimating impacts for properties without recent transactions, the model provides a comprehensive view of the market. Quarterly updates are achieved using a Bayesian framework, which integrates new data while maintaining computational efficiency. Variance-based smoothing minimizes volatility, ensuring a stable and reliable index. The approach is adaptable to different levels of analysis, as demonstrated through city-level indices for 10 Indian cities and macromarket-level results for Bangalore, revealing distinct patterns across and within cities. A major strength of the methodology is its ability to reflect market-wide behavior, even for areas with limited data, ensuring unbiased insights. While the current implementation uses separate models for each city, future work could explore unified frameworks with spatially varying coefficients, though this would require addressing computational complexities. In summary, the proposed model effectively constructs rental indices that provide valuable insights into market trends and dynamics. Its flexibility extends to applications beyond commercial offices, including green buildings, affordable housing, and residential markets, making it a forward-looking tool for real estate analysis.

References

Anundsen, A. K., Bjørland, C., and Hagen, M. (2022). Location, location, location!*: a quality-adjusted rent index for the oslo office market. *Journal of European Real Estate Research*, 15(2):208–226.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942.

Calainho, F. D., van de Minne, A. M., and Francke, M. K. (2024). A machine learning approach to price indices: Applications in commercial real estate. *The Journal of Real Estate Finance and Economics*, 68(4):624–653.

Deryol, E., Konukçu, D., Szemere, R., and Tissot, B. (2019). Mind the data gap: commercial property prices for policy. Technical report, Bank for International Settlements.

Fotheringham, A. S., Crespo, R., and Yao, J. (2015). Exploring, modelling and predicting spatiotemporal variations in house prices. *The Annals of Regional Science*, 54:417–436.

Francke, M. K. and Van de Minne, A. (2017). The hierarchical repeat sales model for granular commercial real estate and residential price indices. *The journal of real estate finance and economics*, 55:511–532.

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American statistical Association*, 95(452):1300–1304.

Geltner, D. (1990). Return risk and cash flow risk with long-term riskless leases in commercial real estate. *Real Estate Economics*, 18(4):377–402.

Geweke, J. F. et al. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis.

Gupta, K. and Deb, S. (2024). A divide-and-conquer approach for spatio-temporal analysis of large house price data from greater london. *arXiv preprint arXiv:2407.15905*.

Hijmans, R. J. (2021). Introduction to the "geosphere" package (version 1.5-14).

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158(1):160–173.

Jennen, M. G. and Brounen, D. (2009). The effect of clustering on office rents: evidence from the amsterdam market. *Real Estate Economics*, 37(2):185–208.

Kok, N. and Jennen, M. (2012). The impact of energy labels and accessibility on office rents. *Energy policy*, 46:489–497.

Muto, S., Sugasawa, S., and Suzuki, M. (2023). Hedonic real estate price estimation with the spatiotemporal geostatistical model. *Journal of Spatial Econometrics*, 4(1):10.

Neal, R. M. (2003). Slice sampling. *The annals of statistics*, 31(3):705–767.

Oust, A., Hansen, S. N., and Pettrem, T. R. (2020). Combining property price predictions from repeat sales and spatially enhanced hedonic regressions. *The Journal of Real Estate Finance and Economics*, 61(2):183–207.

Rawat, S. and Deb, S. (2023). A spatio-temporal statistical model to analyze covid-19 spread in the usa. *Journal of Applied Statistics*, 50(11-12):2310–2329.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.

Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:61–86.

Appendix A. Macromarket Level Indices

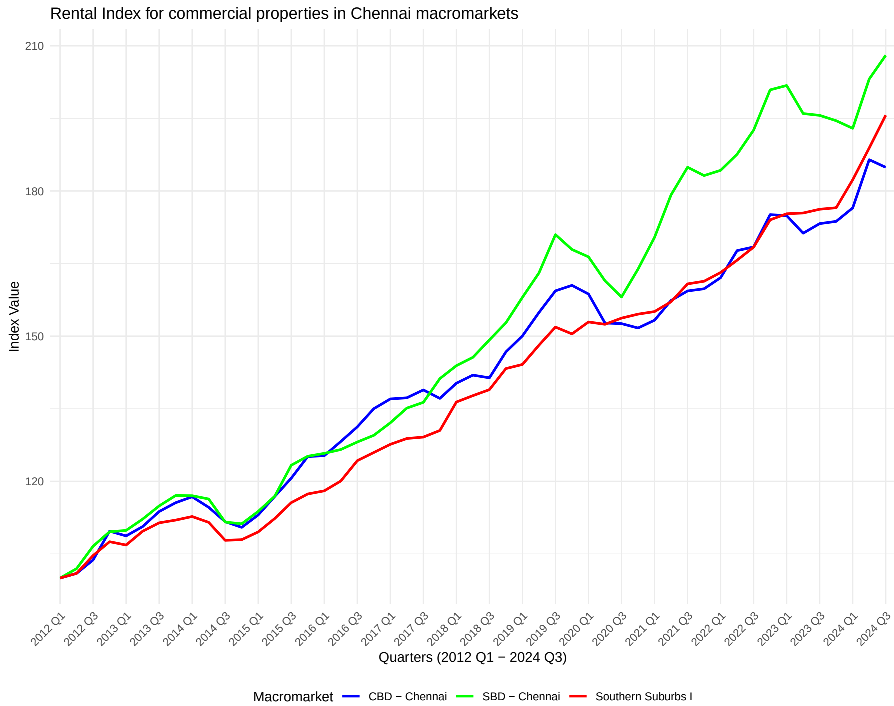

Fig. A.1: Quarterly rental indices for the commercial properties in macromarkets in Chennai

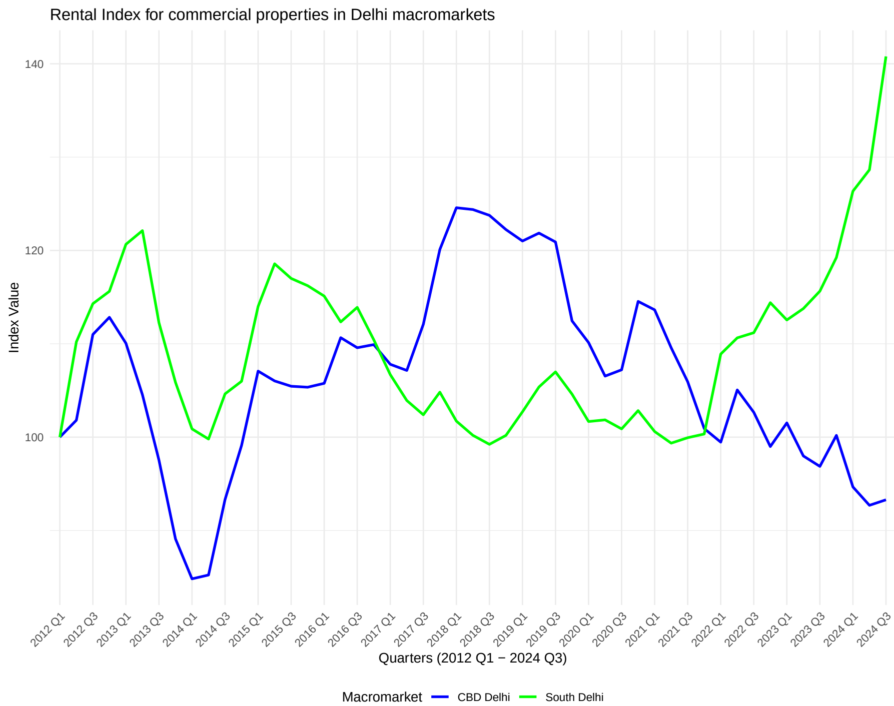

Fig. A.2: Quarterly rental indices for the commercial properties in macromarkets in Delhi

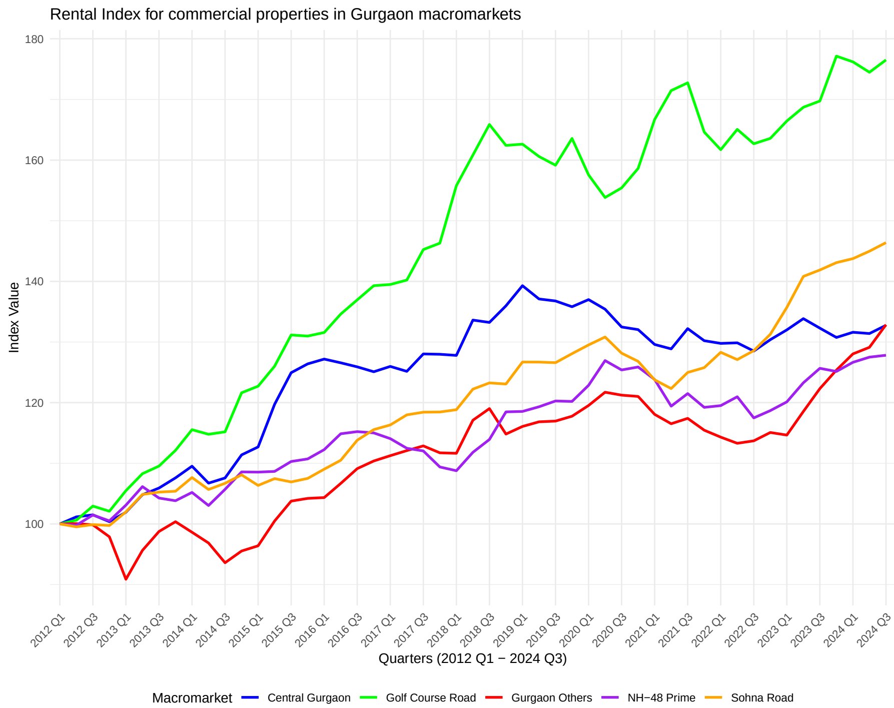

Fig. A.3: Quarterly rental indices for the commercial properties in macromarkets in Gurgaon

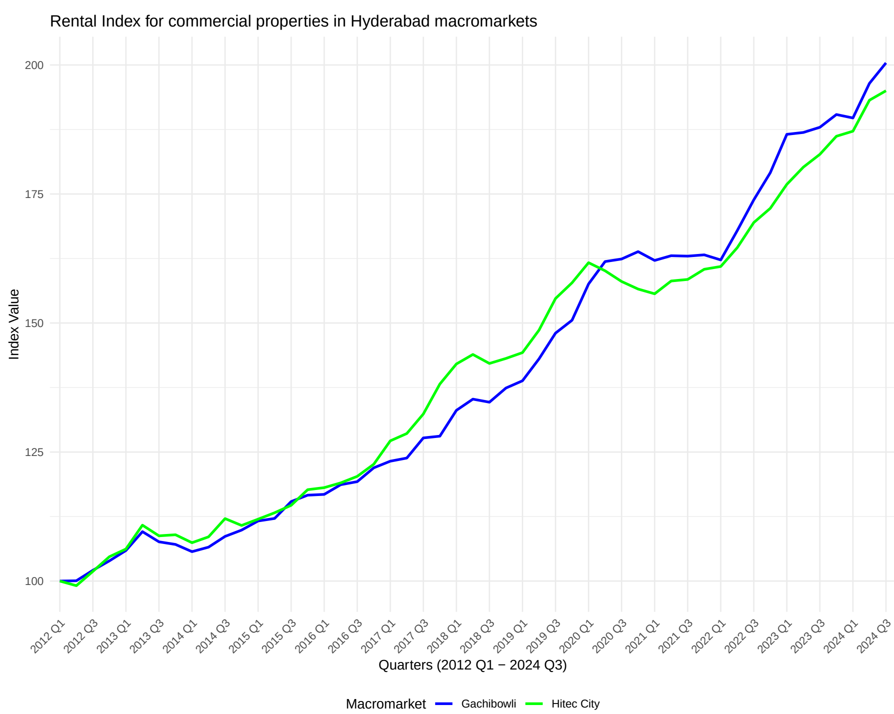

Fig. A.4: Quarterly rental indices for the commercial properties in macromarkets in Hyderabad

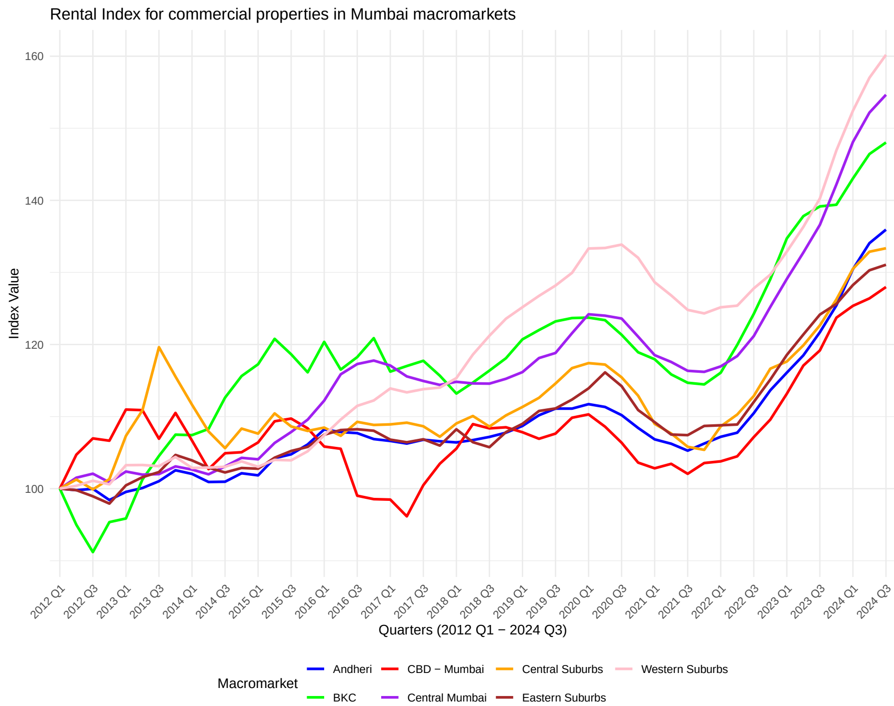

Fig. A.5: Quarterly rental indices for the commercial properties in macromarkets in Mumbai

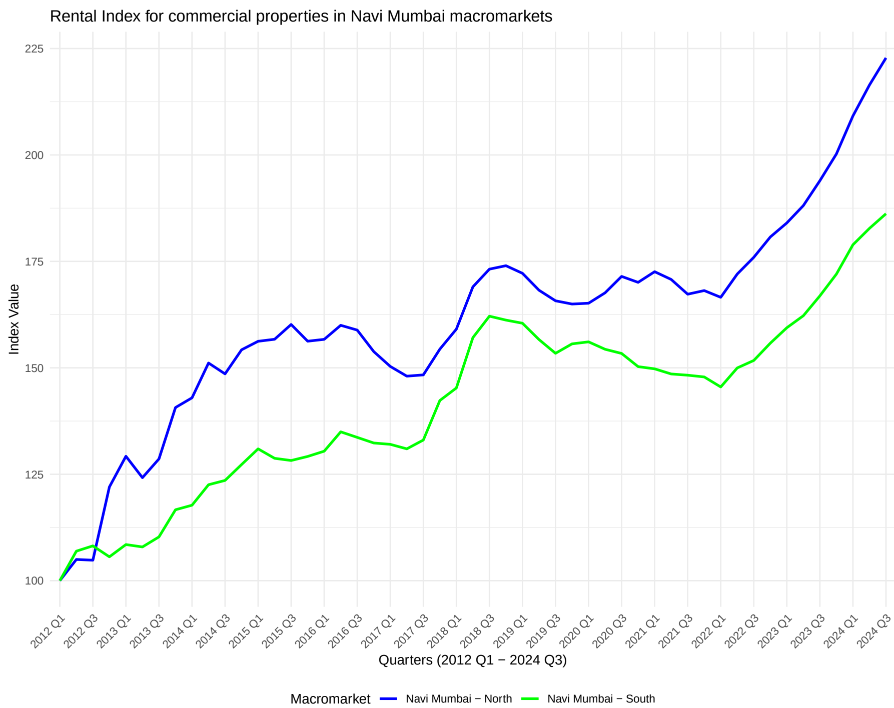

Fig. A.6: Quarterly rental indices for the commercial properties in macromarkets in Navi Mumbai

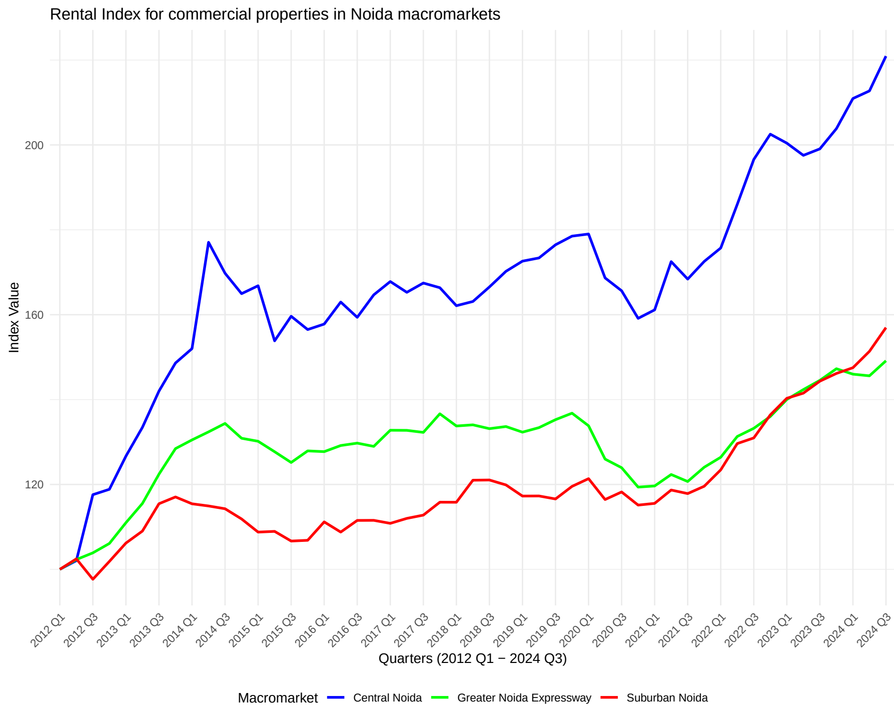

Fig. A.7: Quarterly rental indices for the commercial properties in macromarkets in Noida

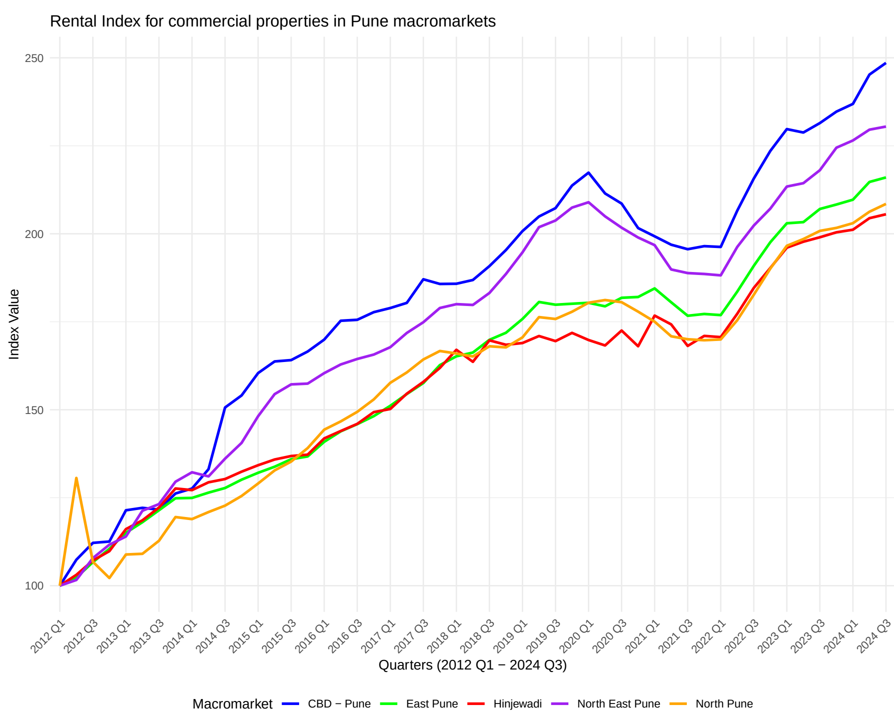

Fig. A.8: Quarterly rental indices for the commercial properties in macromarkets in Pune

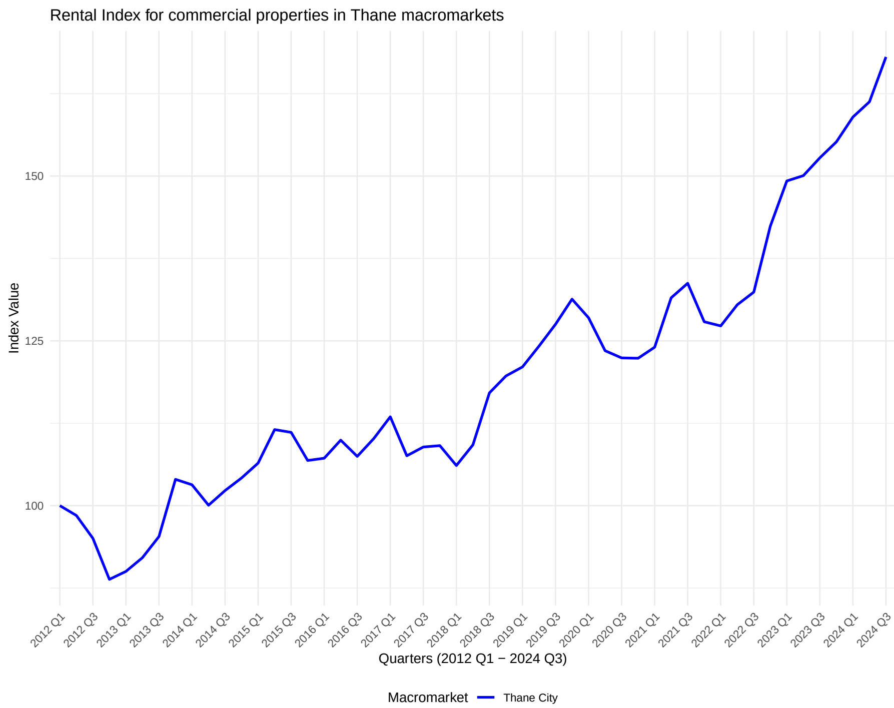

Fig. A.9: Quarterly rental indices for the commercial properties in macromarket in Thane