square yards

Approaches to Constructing Residential House Price Index

A comparative methodology
with a focus on the Indian context



Table of Contents

01	Executive Summary	0
02	Introduction	2
03	Methodology Deep Dives	0
04	Comparative Evaluation Framework	3 12
05	Comparison Matrix	0 13
06	The India-Specific Context	7 15
07	Recommendation	20
08	Square Yards and IIM Bangalore: A New Path Forward	22
09	References	23

01 EXECUTIVE SUMMARY

This whitepaper compares six methodologies for constructing a residential house price index: four classical approaches with a long, established track record — Stratification, Hedonic Regression, Repeat-Sales, and Appraisal-Based Methods (commonly implemented via the Sale Price Apprais- al Ratio, or SPAR) — drawn from the Handbook on Residential Property Price Indices (Eurostat, ILO, IMF, OECD, UNECE, and the World Bank); and two newer, still-emerging approaches — AVM/ML Hybrid Approaches and Spatiotemporal Modelling.

Each methodology is evaluated on its own merits, then assessed against India's data landscape — registration practices, valuation benchmarks, and property-identification infrastructure — to determine which are genuinely implementable today, partially constrained, or structurally blocked.



The core finding: no single approach is uniformly superior. India's specific data constraints weigh on the trade-off between data burden, rigor, interpretability, and complexity more heavily than a generic comparison would suggest: repeat-sales is constrained by incomplete property-identification infrastructure and a market skewed toward primary sales; SPAR needs a comprehensive, property-specific appraisal dataset that India lacks; and pure stratification carries a real quality-adjustment risk.

On the evidence in Sections 3–6, spatiotemporal modelling emerges as the strongest-performing approach among the six — provided an organization can manage its two unfavorable characteristics, data burden and implementation complexity. Section 8 discusses an emerging effort positioned to do exactly that.



02 INTRODUCTION

A House Price Index (HPI) is a statistical measure that tracks how residential property prices change over time, typically expressed relative to a fixed base period. Unlike a simple average or median of transaction prices in a given period, a well-constructed HPI is designed to isolate genuine price movement from period-to-period changes in which specific properties happen to transact.

Residential property price indices serve a range of stakeholders — as a macroeconomic and financial-stability indicator, an input into national accounts and household-wealth estimates, a reference for monetary policy, and a practical input into individual buy-or-sell decisions and inflation measurement.

Constructing a reliable HPI is harder than it first appears, because the mix of properties that transact in any given period is never quite the same as the mix that transacted in the last. A period in which more large, premium apartments happen to sell will show rising average prices even if no individual home actually became more expensive — and a period dominated by smaller, peripheral units will show the opposite. Every methodology described in this whitepaper exists, at its core, to solve the same underlying problem: separating genuine price movement from this constantly shifting composition of what happens to be selling.

Understanding where this problem has already been addressed — and where opportunity lies — means looking at both India's own history and the rest of the world's approach.

2.1 India's Own Journey

Figure 2.1 traces the evolution of India's residential and commercial property indices since 2001, showing a market that has experimented widely without any one method becoming a clear standard.

Evolution of India's House & Property Price Indices

Residential, commercial, and academic milestones, with underlying methodology

— verified dates

Year	Type	Entry	Underlying Methodology
2001	RESIDENTIAL	NHB — Delhi pilot study	Hedonic tested alongside Laspeyres; Laspeyres adopted going forward
2004-2007	ACADEMIC	Mumbai residential hedonic study	Time Dummy + Characteristics Price Index; found hedonic index lower than median weighted avg
2007	RESIDENTIAL	NHB RESIDEX launched (5-city pilot)	Market-data based, evolving toward Stratification / Modified Laspeyres
2008-2009	RESIDENTIAL	RBI — one-off Mumbai HPI (base 2002-03=100)	A single Q4 Macroeconomic Review exercise, not an ongoing series

Year	Type	Entry	Underlying Methodology
c. 2010	COMMERCIAL	JLL REIS active	Weighted avg. of quoted developer prices, effective active stock, base 2003
2010-2011	RESIDENTIAL	RBI HPI ongoing series launched, 10 cities	Stratification (Ward/Zone × FSA-category), Laspeyres; base 2010-11 Q1=100
2011	RESIDENTIAL	Magicbricks PropIndex launched	Bottom-up aggregation of listed properties, weighted by locality supply, 11 cities
2013-2015	RESIDENTIAL	NHB RESIDEX — data source shift	Stratification unchanged; source moves to CERSAI mortgage-lien registry
2016-2019	ACADEMIIC	Delhi-NCR high-end housing study	Hedonic + asset/rental-return pricing, stepwise multiple regression
By 2017	CAOCMADMEEM	Knight Frank Prime Office Rental Index	Survey/broker quoted rents, prime/CBD only, 20-market coverage
May 2021	ARECSAIDEMNT	Housing.com / ISB HPI launched	Survey Laspeyres (City × Locality × BHK) + companion Quantity Index, 8 cities
—	GAACPADEMIC	Repeat-sales method	No known Indian implementation — property-identity infrastructure not yet in place
2023	ARECSAIDEMNT	NHB RESIDEX — methodology revision	Stratification refined (City × Category), two-track; base FY2017-18
Nov 2024	CAOCMADMEEM	IIMB-CRE Matrix CPRI-Office published	Spatio-temporal (Gaussian Process + Kriging); already in its 7th edition.
Oct 2025	ARECSAIDEMNT	RBI HPI — rebased, expanded	Formula unchanged; base 2010-11→2022-23; city coverage 10→18
2026	RESIDENTIAL	NHB RESIDEX — rebased	Base revised from FY2017–18 to FY2024–25; linking factors provided for continuity.

Figure 2.1— Evolution of India's house and property price indices, 2001–present.

Among residential indices, RBI's HPI and NHB's RESIDEX are the two long-standing official series. RBI's HPI is built on registered transaction data; NHB's RESIDEX splits into two tracks — Assessment Prices from bank and HFC loan valuations, and Market Prices from a developer survey. Private players take a different route: Magicbricks' PropIndex and Housing.com/ISB's HPI are built from listing and survey data rather than registered transactions, favoring locality-level granularity and speed over completeness. Housing.com also publishes a companion Quantity Index tracking transaction volume.

Hedonic regression has been tested in India in at least three documented instances but not sustained as an ongoing index. No repeat-sales implementation has been identified. CRE Matrix's CPRI-Office, built with IIM Bangalore, is a documented Indian implementation of spatiotemporal modelling — applied to commercial office rents, not residential prices.

2.2 The Global Landscape

Internationally, residential price measurement rests on a more settled foundation. Figure 2.2 surveys major indices from ten countries and regions.

Global Major Indices

Residential price indices, by country/region, with underlying methodology.

Country / Region	Type	Index	Underlying Methodology
	RESIDENTIAL	S&P CoreLogic Case-Shiller	Repeat-Sales — IVWARS, IV regression on matched pairs
	RESIDENTIAL	FHFA House Price Index	Repeat-Sales — weighted repeats, Fannie/Freddie data
	RESIDENTIAL	UK House Price Index (HPI)	HedonicRegression — cross-sectional imputation
	RESIDENTIAL	Japan RPPI	HedonicRegression — time-dummy, Fisher index
	RESIDENTIAL	Teranet-National Bank HPI	Repeat-Sales unit-level IDs — land registry's
	RESIDENTIAL	NBS 70-City Price Index	Stratification — 3 size tiers, same-complex comparison
	RESIDENTIAL	Eurostat Harmonised HPI	Multiplenationalmethods under harmonised definitions
	RESIDENTIAL	National SPAR-based indices	Appraisal-Based (SPAR) — price vs. govt. valuation
	RESIDENTIAL	Zillow ZHVI / Zestimate	AVM/MLHybrid — neural network valuation, tiling
	RESIDENTIAL	CoreLogic / Cotality Hedonic Index	Hedonic+Spatial Smoothing — GAM, thin-plate spline

Figure 2.2 — Major global residential price indices.

Eight of these trace back to one of four classical families: repeat-sales (the US's Case-Shiller and FHFA indices, Canada's Teranet–National Bank HPI), hedonic regression (the UK's HPI, Japan's RPPI), and Appraisal-Based Methods (the Netherlands, New Zealand, Sweden, Denmark). China's NBS 70-City Index uses stratification; the EU's Eurostat Harmonised HPI combines national indices compiled with different methods — hedonic, matched-model, and stratified — under common definitions. Together these form the taxonomy in the Handbook on Residential Property Price Indices.

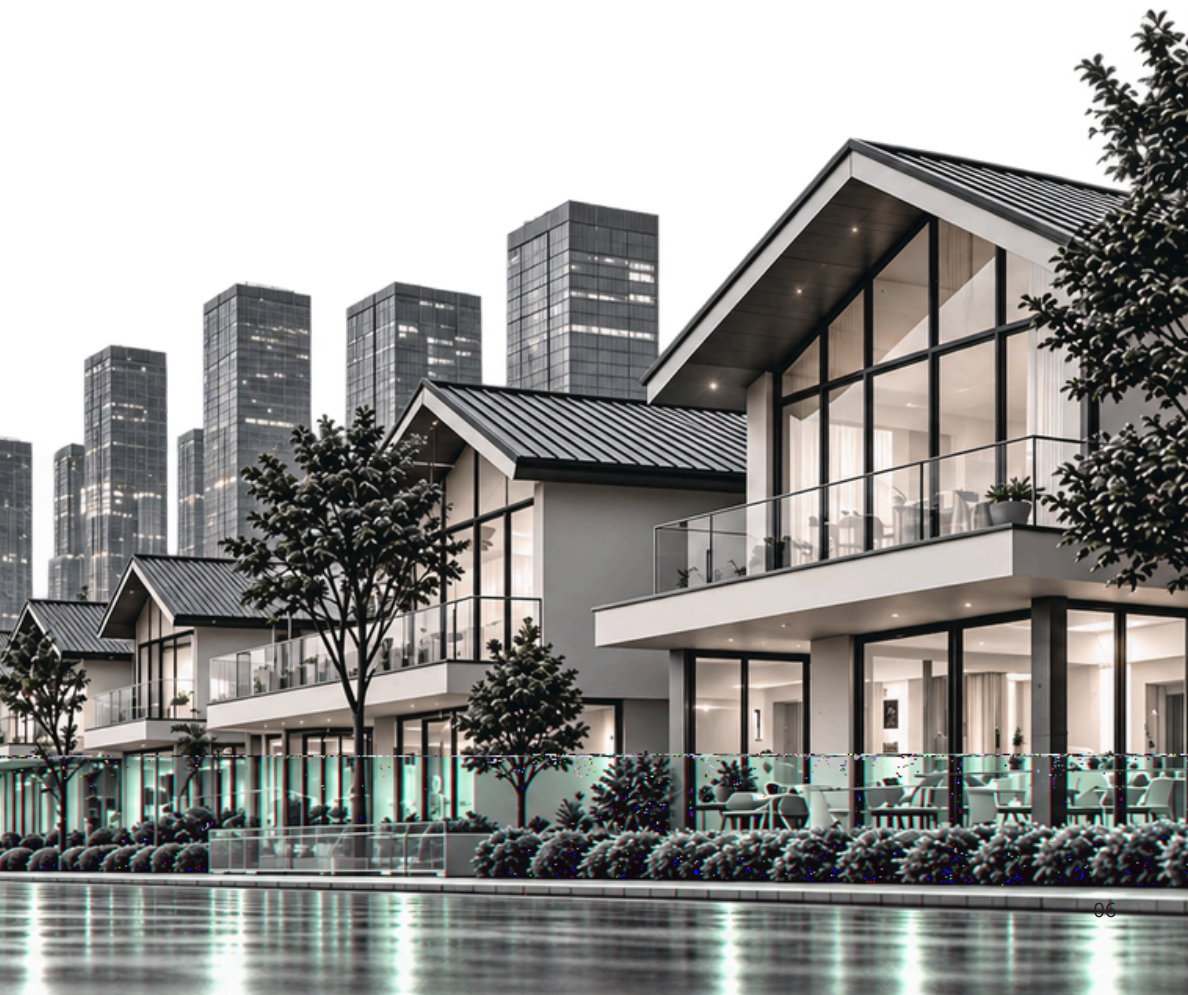
Two entries point toward where the frontier has moved. Zillow's AVM/ML-based approach — the Zestimate valuation, aggregated into the ZHVI index — prices an entire housing

stock continuously, not just transacted homes. CoreLogic/Cotality's Australian index, still hedonic at its core, has adopted spatial-smoothing techniques closer to genuine spatiotemporal modelling — a fuller step that has so far remained largely academic; this review did not identify a country that has yet adopted it as a live residential index.

2.3 Scope of This Whitepaper

This whitepaperis scopedtoresidential real estate only. Commercial property follows a sufficiently different data and market structure — lease-based rather than sale-based transactions, a smaller and more institutionally concentrated buyer base — to warrant separate treatment.

Within that scope, this whitepaper compares six methodologies: four classical families with a long, and Appraisal-Based Methods — and two newer approaches which represent a current emerging frontier: AVM/ML Hybrid Approaches and Spatiotemporal Modelling.



03 METHODOLOGY DEEP DIVES

Thissection describessix approaches toconstructing a price index, each covering its mechanics, data requirements, and canonical strengths and limitations. All six are described generically, as method families rather than as any single named implementation, so the comparison in Sections 4–5 evaluates the underlying statistical approach rather than any one organization's specific execution of it.

Six Approaches to Constructing a Price Index

Approach & Mechanics	Best for:	Breaks down when:
StratiSfitcraattiiofinc a/tMioixn- /A dMjuixs-tAmdejunsttment 1 • Group properties into discrete buckets (location, size, type) • Representative price per bucket (mean/median) • Fixed base-period weights combine buckets	low data availability, need for fast institutional adoption	strata too coarse, or too thin
StratiHficeadtoinonic /RMegixr-eAsdsjiuosntment 2 • Model prices using property characteristics. • Predict a price for every property, average predictions • Cross-sectional or time-dummy variants	markets with rich, structured characteristics data	characteristics data is thin or unreliable
3 StratRiefipceaatti-oSnal e/sMix-Adjustment • Compare the same property's two sale prices • Combine matched sale pairs to estimate price changes over time. • No characteristics data required	deep, long transaction history, persistent property IDs	few repeat sales exist, or units aren't identifiable over time
StratiAficpaptriaoins a/lM-Bixa-sAeddj uMsetmtheondts (e.g. SPAR) 4 • Ratio of sale price to an independent appraisal • Aggregate ratios across the sample • No characteristics data required	markets with reliable, independent, per-property valuations	no genuinely independent appraisal exists at scale
StratiAfiVcaMti o/n M /aMcihxi-nAed Ljuesatrmnienngt based Hybrid Approaches 5 • ML model (tree-based or neural) predicts individual value • Trained on large characteristics + location datasets • Compare modeled values for the same properties over time.	large, continuously-updated proprietary datasets	interpretability or auditability required
StratiSfipcatiiotne /mMpioxr-aAld Mjuosdtmelleingt 6 • Joint covariate + spatial + temporal decomposition • Model how prices are linked across nearby locations and time periods. • Capable of estimating values for unobserved data using information from related locations and time periods.	uneven/sparse markets needing principled uncertainty	data requirements or implementation complexity are high.

Figure 3.1— Overviewofthesixapproachescomparedinthiswhitepaper.

3.1 Stratification/ Mix-Adjustment

Properties are grouped into discrete strata — typically by location, size, and property type — with a representative price (mean or median) computed within each stratum, then combined into an overall index using base-period weights (a Laspeyres-type formula). India's RBI HPI and NHB RESIDEX are both examples of this family, stratifying by ward/floor-area and city/carpet-area categories respectively.

Data requirements

Low: a transaction price, a broad property-type classification, and a location tag. No detailed characteristics or repeat-observation history required.

Strengths

- Lowest data burden of any approach, and can be implemented with data that is already routinely collected.
- Simple, transparent, and easy to audit — a stratum's price and weight can be independently verified.

Canonical limitations

- Residual within-stratum composition bias: a mix-adjusted index only controls for compositional change across the chosen strata, not within them, nor along any dimension the strata don't capture. A
- granularity/reliability tradeoff: finer stratification reduces the quality-mix problem but shrinks the sample per cell, raising standard error — and detailed stratification is only possible if the defining characteristics are available for every observation.
- Fixed base-period weights become stale as the market's composition evolves, until next revised.

3.2 Hedonic Regression

Price is modeled as a function of a property's own characteristics via regression — typically ln (price) = constant + Σ(characteristic coefficients) + error — isolating the pure price change over time from the "implicit price" of each attribute. Two variants exist: a fresh cross-sectional regression each period, with predicted prices compared via their geometric mean (as the UK HPI does); or a single pooled regression with a time-dummy variable, re-estimated as new data arrives (as Japan's RPPI does).

Data requirements

High: a full, structured set of characteristics (size, age, type, condition, location, amenities) attached to every individual transaction.

Strengths

- Directly and explicitly controls for composition, rather than relying on stratum boundaries to approximate it.
- Widely regarded as the most theoretically complete of the classical methods, since it is the only one able to adjust for both a changing mix of transacted properties and changes in an individual property's own condition — such as renovation or depreciation — at the same time.

Canonical limitations

- Omitted variable bias: any price-relevant characteristic left out of the model distorts the coefficients on the variables that are included, if the omitted characteristic is correlated with them.
- Functional form assumptions and fixed characteristic-price coefficients within the estimation window are modelling choices that may not match how the market actually behaves — getting either wrong distorts every derived coefficient.

- Fails to account for spatial and temporal dependence between transactions unless explicitly extended to do so, and can struggle to control for micro-location using characteristics alone; it is also data-hungry, with characteristics often correlated with each other (multicollinearity), and the commonly-used time-dummy variant revises previously published values as new data arrives.

3.3 Repeat-Sales

Rather than comparing different properties to each other, this method compares the same property to itself across two or more sale dates. A repeat-sales regression links matched sale pairs across periods to solve for each period's price-relative; the estimator and weighting scheme vary by implementation (ordinary least squares, weighted least squares, or other variants). S&P Cotality Case-Shiller in the US and the Teranet–National Bank HPI in Canada are both real-world implementations of this method family.

Data requirements

A long transaction history with a reliable way to match a given property's multiple sales to each other over time. No characteristics data is needed, since the property is implicitly compared only to itself.

Strengths

- Controls composition bias: Controls for unchanged property characteristics by comparing the same dwelling across sales, reducing the influence of differences in size, type, and location between properties.
- Because properties are matched at the exact address level, the method automatically holds location constant down to the specific property's micro-location — a degree of locational precision hedonic regression's characteristic variables typically cannot fully replicate.

Canonical limitations

- Sample-selection bias: only ever observes properties that have sold at least twice, which is not necessarily representative of the whole housing stock — frequently-traded segments are systematically over-represented relative to long-held properties.
- Excludes new construction and any property sold only once, and needs a genuinely long transaction history to accumulate enough matched pairs for reliable estimation — both shrinking effective coverage, especially in fast-growing markets.
- Assumes the property is physically unchanged between its two sales — renovations or deterioration are invisible to the method and get misattributed to price movement — and is revision-sensitive, since previously published values can shift as new repeat-sale pairs accumulate and the regression re-estimates.

3.4 Appraisal-Based Methods

This family of methods anchors each transacted property's sale price against an independent, previously-established appraisal for that same property. The most common implementation is the Sale Price Appraisal Ratio (SPAR) method, where each sale price is expressed as a ratio to its independent appraisal, and these ratios are aggregated across the sample (commonly value-weighted) into an index. National SPAR-based indices are in use in the Netherlands and New Zealand.

Data requirements

A reliable, independent, per-property appraisal or assessed value, ideally available for the full housing stock, not just transacted properties.

Strengths

- Uses every transaction in the sample, without the sample-selection concerns repeat-sales carries.

- Conceptually simple and computationally light— no regression to estimate, just a ratio and an aggregation.

Canonical limitations

- Entirely dependent on the quality and currency of the independent appraisal — if the appraisal is stale, biased, or infrequently updated, the ratio inherits that error directly.
- Needs a genuinely independent, property-specific valuation to exist for nearly every property in the sample — a real infrastructural prerequisite — and changes to the appraisal methodology itself over time can introduce artificial breaks in the index even when actual prices haven't moved.
- Doesn't inherently correct for quality change — major repairs, renovations, or depreciation — between the appraisal date and the sale date, unless the appraisal itself is kept current.

3.5 AVM/ML Hybrid Approaches

An automated valuation model — typically a tree-based ensemble or neural network — is trained on transactions, characteristics, and location data to predict an individual value for every property in a coverage universe, transacted or not. These valuations are then aggregated into a market index, commonly via a repeat-measures comparison of the same property's modeled value over time. Zillow's Zestimate is the individual-property valuation underlying this approach; the aggregated ZHVI index is the resulting hybrid method's clearest real-world example.

Data requirements

A large, continuously-updated, high-quality training dataset combining transactions, structured characteristics, and location — the most data- and infrastructure-intensive approach of the six.

Strengths

- Can value any property in its coverage universe, transacted or not — the broadest coverage of any method described here.
- Scales natively to very large datasets using standard machine-learning training infrastructure.

Canonical limitations

- Limited interpretability — harder to explain why the model produced a given number, especially for more complex model architectures, compared to a hedonic coefficient table.
- Model risk and overfitting: a model tuned closely to historical data may not generalize well to new conditions, and the approach is data- and infrastructure-intensive to build and maintain on an ongoing basis.
- Errors can compound badly when the valuation is used for a consequential, binding decision rather than reported as an estimate, and training data can itself encode and perpetuate historical pricing bias if not carefully audited.

3.6 Spatiotemporal Modelling

A spatiotemporal model decomposes property prices into covariate effects (property characteristics), a spatial random effect, and a temporal random effect. Spatial and temporal dependence can be represented through Gaussian Processes, allowing correlations between locations or periods to decay smoothly with distance or elapsed time rather than being forced into discrete strata. This enables the model to estimate plausible property values even when recent transaction data is sparse, using information from related locations and periods through a conditional probability (kriging) mechanism. IIMB–CRE Matrix's CPRI-Office is an example of this approach, currently applied to commercial office rents in India, where

indices are being created at the city level.

However, an additional challenge in the Indian residential housing market is the potential "under-reporting" of transaction values in the registration records. The reported/registered consideration amount might often be lower than the actual transaction amount, thereby reducing the reliability of such registered prices as measures of market value. The proposed method treats the reported price of each property as belonging to one of two possible groups. The first group consists of transactions whose reported values are reasonably close to their true market values. The second group consists of under-reported transactions, where the registered value is lower than the actual amount paid. The model uses "guidance values" (the minimum government-notified values used for stamp-duty assessment) as reference, by exploiting possible clustering/ bunching of reported prices around or just above these values as a statistical signal of under-reporting.

Combining reported prices with property characteristics, geographical location, guidance values, as well as spatial and temporal dependence among transactions, the model jointly estimates the probabilities of under-reporting and the true underlying transaction values. The estimated underlying values can then be aggregated to construct more reliable city-level HPIs that better reflect housing market price movements, while accounting for potential reporting distortions.

Data requirements

Reliable geographic coordinates for each property, plus a structured set of characteristics comparable to what hedonic regression requires; does not require a persistent property identifier or an independent appraisal.

Strengths

- Provides reliable measures by adjusting for registered price distortions like under-reporting and estimating the true underlying prices.
- Explicitly designed to handle sparse or uneven transaction density; estimating a plausible value even where recent data is thin, via the model's own probabilistic structure rather than by shrinking the geography or waiting for more data.
- Produces model-based uncertainty estimates (Bayesian posteriors) alongside every prediction in the proposed implementation, rather than a single point estimate presented without a confidence measure.
- By explicitly modelling spatial and temporal dependencies, it addresses a limitation of plain hedonic regression, which treats every transaction as independent even though nearby properties' prices are often correlated for reasons no characteristic captures, such as a neighbourhood turning desirable or a new metro line opening nearby.
- Through city-level models and indices, it accounts for the differences in market conditions across cities.

Canonical limitations

- Computationally expensive relative to simpler methods, and might require approximation techniques to remain scalable to a large datasets.
- Simplifying assumptions, especially additive rather than interactive spatial and temporal effects, and normally distributed residuals, which might not hold in real markets; violations of these assumptions might understate uncertainty.
- Covariate effects are still held fixed within each estimation window (similar to a limitation of hedonic regression).
- Needs reliable coordinates for every property, which might not always be available.

04 COMPARATIVE EVALUATION FRAMEWORK

To compare six methodologically distinct approaches on a consistent basis, this whitepaper evaluates each against eleven criteria, spanning data requirements, statistical performance, and institutional/operational considerations.

CRITERION	WHAT IT CAPTURES
Data Burden	How much and what kind of data the method requires to function at all
Quality Adjustment	Whether the method separates genuine price movement from both a shifting mix of what's transacting (composition bias) and an individual property's own condition changing over time — renovation, depreciation
Thin-Market Performance	How reliably the method performs where transaction volume is low or uneven
Interpretability	Whether a non-technical stakeholder can be walked through why the index moved
Implementation Complexity	The computational cost, scalability, processing infrastructure, staff expertise, and institutional coordination needed to produce the index on an ongoing basis
Sample Utilisation	What share of available transactions the method actually uses, independent of how deep the overall market is
Values Non-Transacted Property	Whether the method can estimate a value for a property that hasn't recently sold
Uncertainty Quantification	Whether the method produces a confidence measure alongside its point estimate
Revision Stability	How much published values change as new data arrives
Governance / Transparency	The maturity of external validation and disclosure practices associated with the method
Real-World Track Record	The depth of accumulated, real-world operating history across implementations globally

05 COMPARISON MATRIX

The matrix below synthesizes Sections 3 and 4 into a single reference view. Ratings reflect a qualitative synthesis, not an empirical benchmark on a common dataset — the reasoning for each rating is set out in Section 3's limitations.

Comparative Evaluation Matrix

Six approaches, scored across eleven evaluation criteria (qualitative synthesis, not an empirical benchmark)

Approach	Stratification	Hedonic Regression	Repeat-Sales	Appraisal-Based Methods (SPAR)	AVM / ML Hybrid	Spatiotemporal
Data Burden	Low	High	Medium	Medium	High	High
Quality Adjustment	Low	High	Medium	Medium	High	High
Thin-Market Performance	Medium	Medium	Weak	High	Medium	High
Interpretability	High	High	Medium	High	Low	High
Implementation Complexity	Low	Medium	Medium	Low	High	High
Sample Utilisation	High	High	Low	High	High	High
Values Non-Transacted Property	No	Yes	No	No	Yes	Yes
Uncertainty Quantification	Low	Medium	Medium	Low	Medium	High
Revision Stability	Stable	Medium	Low	Stable	Medium	Medium
Governance / Transparency	High	High	High	Medium	Low	High
Real-World Track Record	Extensive	Extensive	Extensive	Moderate	Moderate	Limited

Favorable Mixed / Moderate Unfavorable

Interpretability, governance, and uncertainty ratings assume a suitably documented and validated implementation; they are not guaranteed by the method family. Ratings do not constitute an empirical performance ranking.

Figure 5.1

— Six approaches scored across eleven evaluation criteria.

A few patterns are worth drawing out explicitly:



06 THE INDIA-SPECIFIC CONTEXT

This section evaluates each of the six approaches against India's actual data landscape — registration practices, valuation benchmarks, and property-identification infrastructure — rather than a generic "developing-market data constraints" narrative.

6.1 Data readiness model

India Data Capability Matrix

Is each approach's core data prerequisite currently satisfiable in India?

Approach	Core Data Prerequisite	Currently Satisfiable?	Primary Constraint
Stratification	Price + broad category + location	Yes	Registered price itself may be distorted at source
Hedonic Regression	Structured transaction characteristics	Partial	Registration lacks structured data; project-level disclosures now exist
Repeat-Sales	Persistent, unit-level property ID	Limited	Depends on historical unit matching — IDs not linked to past records
Appraisal-Based (SPAR)	Independent, per-property appraisal	Limited	Not established for broad coverage in India
AVM / ML Hybrid	Large, continuously-updated dataset	Partial	Achievable via private data assembly; no public bulk-access channel
Spatiotemporal	Coordinates + some characteristics	Partial	Coordinates are generally unavailable; gaps in property characteristics are similar to those in the hedonic approach.

Reflects India's public data infrastructure as researched for this whitepaper, not any single organization's private data assets.

Figure 6.1 — Whether each approach's core data prerequisite is currently satisfiable, given India's public data infrastructure.

Figure 6.1 shows what India currently lacks. Figure 6.2 shows the reverse view: what actually enabled other countries to implement their chosen methodology, tracing each case back to the underlying infrastructure and, in most cases, the unrelated purpose it was originally built for.

Understanding The Foundation : Country Context Infrastructure

The data infrastructure that made each real-world implementation possible, and what it was originally built for

Methodology	Country	Enabling Mechanism	What Actually Enables It	Original Purpose of That Infrastructure
Repeat-Sales	United States	TAX-DRIVEN	County Assessor's Parcel Number (APN) — persistent, unit-level ID	Built for property tax administration, decades before any index
Repeat-Sales	Canada (Ontario)	LAND REGISTRY	Property ID (PIN) with condo cross-reference index	Provincial land title registry (Teranet), which also runs the index
SPAR	Netherlands	TAX-DRIVEN	WOZ value — annual, property-specific public valuation	Municipal property tax, independently audited (Waarderingskamer)
SPAR	New Zealand	TAX-DRIVEN	Council Rating Valuation, revalued every 3 years	Local council rates, audited by the Office of the Valuer-General
Hedonic Regression	United Kingdom	DUAL SOURCE	Land Registry (price) + VOA Council Tax/EPC (characteristics)	Land, tax, and energy-efficiency systems, now linked by address
Hedonic Regression	Japan	GOVT SURVEY	Land General Information System, price + characteristics together	A dedicated government survey, purpose-built for price measurement
Hedonic + Spatial Smoothing	Australia	PRIVATE DATA	Nationally assembled title and attribute data, aggregated at scale	Built by CoreLogic/ Cotality from digitized public title records
AVM / ML-based	United States	DUAL SOURCE	MLS listings layered on county tax-assessment characteristics	MLS plus property-tax data — two pre-existing ecosystems combined
Spatiotemporal	India (commercial)	PRIVATE DATA	Lease-level dataset with coordinates, assembled from scratch	Built by CRE Matrix; no comparable public dataset to draw on

The pattern worth noting: 6 of 9 cases reuse infrastructure originally built for an unrelated purpose — usually property tax or land-title registration. Successful indices often combine administrative records with privately assembled data: even Australia's privately-built index draws on digitized public title records. India's commercial spatiotemporal index is the one case here built without a comparable public dataset.

Figure 6.2 — What enables each methodology internationally, and what the underlying infrastructure was originally built for.

The pattern is informative for India's own path forward: in six of the nine cases, the enabling infrastructure was originally built for tax administration or land-title registration, often years or decades before anyone repurposed it for a price index. Many successful indices reuse this kind of administrative infrastructure; where it is genuinely absent, as Figure 6.1 shows it often is in India, private data assembly — the route India's own CPRI-Office has taken — is a proven alternative.

6.2 Approach-by-approach assessment

Stratification

Why it could work:

- Lowest data burden of any approach — needs only a transaction price, broad category, and location, already captured in some form by registration records.
- Requires neither a persistent property identifier nor an independent appraisal, sidestepping India's two most significant current infrastructure gaps.

Limitations in context:

- India's administrative boundaries (wards, zones, municipal limits) are drawn for governance, not to group economically similar properties, and transaction volume varies substantially across and within cities — together risking within-stratum composition bias or unreliably thin cells depending on how fine the stratification is.
- Government-notified valuation benchmarks and related stamp-duty incentives may influence reported values; the dataset should distinguish recorded consideration from stamp-duty assessment value and guidance value, since clustering effects on mean and median are a risk to investigate rather than an established result for every state and dataset.

Hedonic Regression

Why it could work:

- Increasingly feasible given newer, government-mandated project-level disclosure regimes that provide real, structured characteristics data.

Limitations in context:

- Government registration/deed documents traditionally capture price, broad area, and address — not granular per-unit characteristics — and assembling a nationally-representative hedonic dataset requires characteristics from a source other than the registration system itself.
- Registration timing practices allow a significant lag between the transaction and its formal registration — particularly for under-construction property — and a meaningful share of the market also transacts partly or wholly outside formal registration channels, leaving any hedonic model built on registration data working from an incomplete, lagged sample.

Repeat-Sales
Why it could work:

- Digitized registration records in several states extend back multiple decades, meaning the raw historical depth needed to eventually assemble repeat-sale pairs exists in principle.

Limitations in context:

- Municipal property identifiers (PID, UID, or Computer Code, depending on the state) do exist at the individual unit level in many states, following a genuine national push under JnNURM — but they are a relatively recent addition, run as a system separate from property registration, and are not reliably linked to historical registration records, meaning they cannot retroactively support the kind of long-dated repeat-sales matching the method actually needs.
- Registration data is also subject to under-reporting of true transaction values, and the extent of that under-reporting can vary over time — meaning even a successfully matched repeat-sales pair may not reflect the property's true price change, since its two recorded values could be under-reported to different degrees.
- Redevelopment can break property continuity — demolition and reconstruction, sometimes under a different developer or with renumbered units — while new-development-heavy segments may produce relatively few repeat-sale pairs, since the market here skews toward primary sales.

Appraisal-Based Methods (SPAR)
Why it's attractive in principle:

- Conceptually simple and computationally light, if a suitable independent, per-property valuation existed at scale.

Limitations in context:

- SPAR requires reliable property-specific appraisals on a consistent reference basis — the way the Netherlands' WOZ value or New Zealand's rating valuations provide. India lacks a clearly established, comprehensive appraisal dataset suitable for a broad-coverage SPAR index.
- The method cannot deal adequately with quality changes — major repairs, renovations, or depreciation— affecting the dwelling units, and renovation is structurally common in India.

AVM / ML Hybrid Approaches
Why it could work:

- Does not require a persistent property identifier or an independent government appraisal, can be built on privately assembled structured data, and is well suited to handling messy, incomplete, non-standardized characteristics data.

Limitations in context:

- Interpretability risk is elevated given the fragility of public trust in India's existing indices, and the sector has passed through multiple significant regulatory shifts in the past decade that could limit how well a model trained on pre-shift data generalizes afterward.
- There is currently no continuously-published, national-scale residential AVM index to build on, and a model trained substantially on registered prices could invisibly learn and reproduce the valuation-floor distortion described above, embedding a systematic bias rather than surfacing it as a disclosed, citable caveat.

Spatiotemporal Modelling

Why it could work:

- Requires neither a persistent unit-level property identifier nor an independent appraisal, and is designed to handle sparse, uneven data — aligning with India's uneven market depth across cities.
- Can reduce dependence on recent transactions at each individual location by borrowing information from related locations and periods, provided sufficient timely market evidence remains available elsewhere in the dataset.

Limitations in context:

- Inherits the same structured-characteristics gap as hedonic regression, and computational cost is a genuine planning consideration at national scale given India's transaction volume and geographic spread.
- As a newly developed class of methods, its empirical evidence through real-world applications is less extensive than that available for stratification, hedonic regression, and repeat-sales, which have a longer history of international usage.



07 RECOMMENDATION

No single approach satisfies every criterion in Sections 4–6, and India's data landscape rules some approaches in and others out more decisively than a generic comparison would suggest. The framework below is a practical starting point, not a final prescription.

A Practical Starting Point — Not a Final Rule

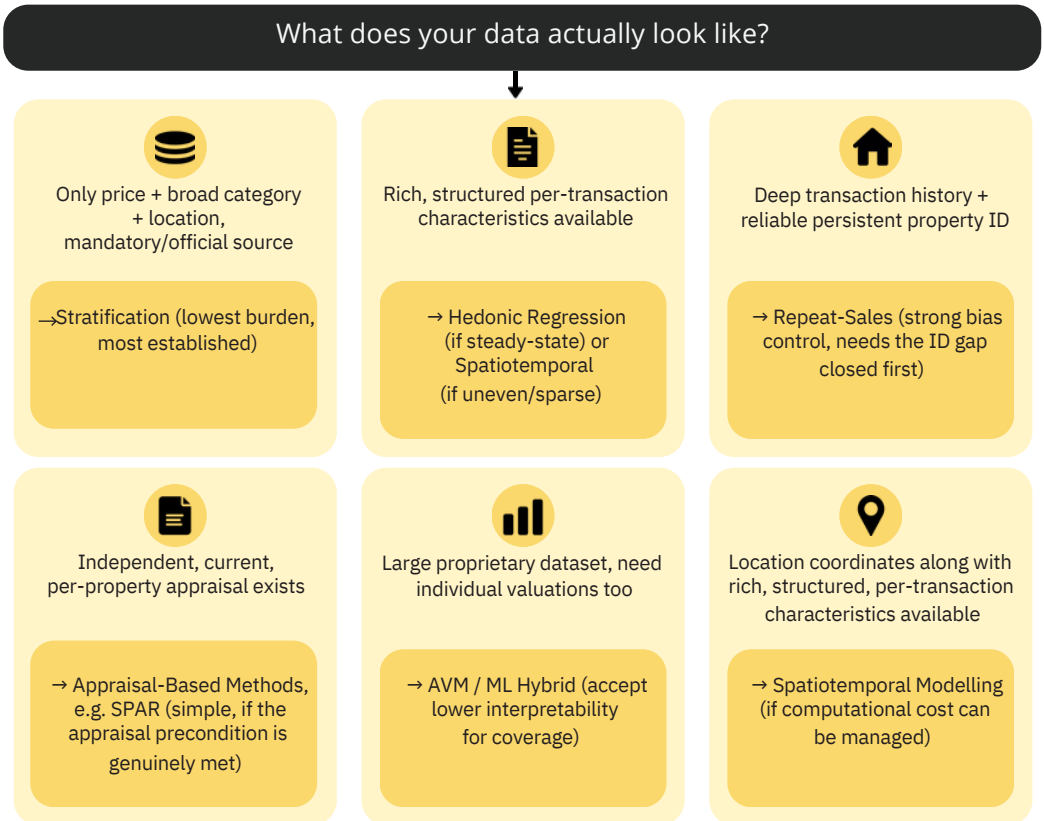


Figure 7.1— A practical starting point for matching data reality to methodology.

Weighing the evidence in Sections 3–6, spatiotemporal modelling is the strongest-performing approach among the six — conditional on managing its two unfavorable characteristics: data burden and implementation complexity. Every alternative carries a more fundamental constraint: Appraisal-Based Methods need an appraisal infrastructure India lacks at scale; repeat-sales is constrained by incomplete property-identification infrastructure and a primary-sales-skewed

market; pure stratification carries a real quality-adjustment risk; and AVM/ML Hybrid Approaches share spatiotemporal modelling's high data burden without its interpretability or governance track record.

Hedonic regression comes closest on paper, sharing spatiotemporal modelling's core requirement of rich, structured, per-transaction characteristics. Relative to the plain hedonic specification considered here, the proposed implementation explicitly models spatial and temporal dependence — addressing a limitation the CPRI-Office literature notes directly, that hedonic models "frequently fail to account for spatial and temporal dependencies, leading to omitted variable bias." Its spatial random effect captures location-driven quality hedonic characteristics alone cannot fully capture, and its Bayesian implementation provides model-based uncertainty estimates alongside every prediction. On the matrix in Section 5, it scores favorably on more criteria simultaneously than any other method compared.

What "managing data burden and implementation complexity" requires is worth being explicit about. Data burden means assembling reliable geographic coordinates alongside structured property characteristics for a large, continuously updated transaction base — an investment closer to what a large private real estate platform maintains than any single Indian government registration system currently provides. Implementation complexity means the specialized statistical capability to build and maintain a Bayesian spatiotemporal model at national scale. Neither cost is trivial or uniquely Indian: this review did not identify a country that has adopted spatiotemporal modelling as an official residential index, plausibly because these two costs have kept it largely academic everywhere. Where an organization can bring extensive transaction-level data together with genuine academic rigor in spatio-temporal statistics — the same combination already working for India's commercial real estate market — spatiotemporal modelling is the strongest available foundation for a new residential index, and Section 8 discusses an effort now underway in that direction.



08 SQUARE YARDS AND IIM BANGALORE: A NEW PATH

On 16 July 2026, IIM Bangalore (IIMB), through its Real Estate Research Initiative (RERI), and Square Yards, one of India's largest integrated proptech platforms, announced a multi-year collaboration to develop and publish a robust, methodologically sound property price index for India. The partnership brings together two complementary strengths mapping onto the two costs identified in Section 7: Square Yards' deep, transaction-level data ecosystem, and IIMB's independent academic expertise in real estate research and spatio-temporal statistics.

Under the collaboration, Square Yards will contribute anonymised and aggregated data — including more than 150 million property registration data points, government guidance values, RERA data, and transaction and listing records — shared periodically for building and maintaining the index. IIMB will lead the index methodology and underlying scientific model, retaining full academic freedom and publication rights. Prof. Venkatesh Panchapagesan, Chairperson of RERI, described the aim as an index that stands up to academic scrutiny and genuinely informs how households, institutions and policymakers understand the direction of property prices in India."

This is not the first attempt by IIMB and RERI to address this problem. The same group developed CPRI-Office, the IIMB–CRE Matrix commercial rental index. Their approach employs a dynamic Bayesian spatio-temporal modelling framework that seamlessly adapts to new quarterly data. The approach addresses computational challenges while preserving accuracy and tackles data sparsity through active lease integration and variance smoothing. Prof. Panchapagesan co-authored the methodology paper with Kapil Gupta, Shagun Srivastava, and Soudeep Deb, establishing a direct link between this initiative and an operational model rather than a merely proposed methodology. Furthermore, Gupta and Deb have published a residential real estate application of spatio-temporal modelling using Greater London transaction data in Annals of Operations Research. This research further demonstrates their expertise in developing advanced spatio-temporal modelling techniques for residential property markets, positioning them well to contribute to the development of robust and scalable residential property price indices in India.

Square Yards' data ecosystem is the kind of large, structured transaction base spatiotemporal modelling's data burden requires, and RERI's track record in CPRI-Office and its residential research abroad reflects the statistical capability its implementation complexity requires. Whether this initiative adopts spatiotemporal modelling remains IIMB's independent academic decision, but the alignment with this whitepaper's strongest-available-approach finding is difficult to overlook.



09 REFERENCES

Core Methodology Frameworks

- Handbookon Residential PropertyPrice Indices (2013), jointly published by Eurostat, International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), United Nations Economic Commission for Europe (UNECE), and the World Bank — https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF
- IMF(2020).Residential PropertyPrice Indices — Practical Compilation Guide.

National & Official Index Sources (India)

- ReserveBank of India, HousePriceIndex — pressrelease on the October 2025 rebasing to 2022-23 and expansion to 18 cities — https://www.rbi.org.in/Scripts/BS_PressReleaseDisplay.aspx?prid=61397
- National Housing Bank, NHB RESIDEX— Methodology White Paper (2023) — https://www.nhb.org.in/wp-content/uploads/2023/12/WHITE-PAPER-NHB-RESIDEX-METHODOLOGY.pdf
- National Housing Bank, About RESIDEX— https://www.nhb.org.in/about-residex/
- Housing.com / Indian School of Business, Housing Price Index— methodology documentation.
- Magicbricks PropIndex— methodologydocumentation.

International Index Methodologies

- S&PDowJones Indices/Cotality,S&PCotality Case-Shiller Home Price Indices — Index Methodology — https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-cotality-cs-home-price-indices.pdf
- Federal Housing Finance Agency (FHFA), House Price Index— https://www.fhfa.gov/data/hpi
- HM Land Registry / Office for National Statistics, UK House Price Index — Quality and Methodology — https://www.gov.uk/government/publications/about-the-uk-house-price-index/quality-and-methodology
- Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan Residential Property Price Index (RPPI) — https://www.mlit.go.jp/common/001360414.pdf
- Zillow Research, Zillow Home Value Index Methodology— https://www.zillow.com/research/methodology-neural-zhvi-32128/
- CoreLogic / Cotality, Residential Property Index Series Methodology (Australia)— https://pages.cotality.com/hubfs/CoreLogic%20AU/Indices/clau-residential-property-index-series.pdf

- Teranet–National Bank, House Price Index — Our Methodology — https://housepriceindex.ca/about/our-methodology/
- National Bureau of Statistics of China. Press Releases (including the 70-City Residential Price Index) https://www.stats.gov.cn/english/PressRelease/index_1.html
- Eurostat, House Price Index — Methodology — https://ec.europa.eu/eurostat/web/housing-price-statistics/methodology

Spatiotemporal Modelling

- Gupta, K., Srivastava, S., Deb, S., Panchapagesan, V. (2025). "CPRI-Office: A new commercial property rental index for Indian cities using spatio-temporal modelling techniques." IIM Bangalore Working Paper No. 727 — https://www.iimb.ac.in/sites/default/files/2025-11/WP%20No.%20727.pdf
- Gupta, K., Deb, S. (2024). "A divide-and-conquer approach for spatio-temporal analysis of large house price data from Greater London." arXiv:2407.15905 — https://arxiv.org/pdf/2407.15905
- Jin, B., Xu, X. (2024, 2025a, 2025b). Gaussian Process Regression approaches to forecasting China's national and city-level housing and rental price indices.
- Muto, S., Sugasawa, S., Suzuki, M. (2023). "Hedonic real estate price estimation with the spatiotemporal geostatistical model."
- Holly, S., Pesaran, M.H., Yamagata, T. (2010). "A spatio-temporal model of house prices in the USA."

India's New Initiative

- IIM Bangalore. "IIM Bangalore and Square Yards partner to build a data-driven Property Price Index for India." Press release, 16 July 2026 — https://www.iimb.ac.in/iimb-square-yards-partner-property-price-index

India Data Infrastructure

- Department of Land Resources, Government of India — Unique Land Parcel Identification Number (ULPIN) — https://dolr.gov.in/en/ulpin/
- Central Registry of Securitisation Asset Reconstruction and Security Interest (CERSAI) — https://www.cersai.org.in/
- State Stamps & Registration Departments — IGR Maharashtra, Kaveri (Karnataka), TNREGINET (Tamil Nadu), IGRS (Telangana/Andhra Pradesh), and other state registration portals.
- Real Estate Regulatory Authority (RERA) — state-level implementation and disclosure records, established under the Real Estate (Regulation and Development) Act, 2016.

Square Yards

Vivek Agarwal
CTO & Co-Founder
vivek.agarwal@squareyards.com

Sunita Mishra
Vice President: Research & Insights
sunita.mishra@squareyards.in

Anirudha Khopade
Associate Principal Partner
anirudha.khopade@squareyards.in

Anand Mohan
Associate Vice President
anand.mohan@squareyards.in

Kanika Narang
Research Analyst
kanika.narang@squareyards.in

Abhivrat Pathak
Research Analyst
abhivrat.pathak@squareyards.in

Neha
Visual Designer
neha.n@squareyards.in

Research Contributors

Kapil Gupta
Assistant Professor
QuantitativeMethodsand Operations Management
Indian Institute of Management Kozhikode
kapil.gupta@iimk.ac.in

Soudeep Deb
Associate Professor
Decision Sciences
Indian Institute of Management Bangalore
soudeep@iimb.ac.in

Shagun Srivastava
Doctoral Student
IndianInstituteofManagement Bangalore
shagun.srivastava24@iimb.ac.in

Venkatesh Panchapagesan
Professor
Finance & Accounting
Indian Institute of Management Bangalore
venky@iimb.ac.in

Scan this QR to read our other reports.



















square
yards