

The NIVAS Index

A Continuous Asking-Price Index for Bangkok Condominiums Across Changes in the Underlying Data

NIVAS Research

Version 1.0 (index formula v1c)
October 2026

Abstract

We present the NIVAS Index, a monthly index of condominium asking prices in Bangkok, measured in Thai baht per square metre and built from published district-level asking-price averages. The index runs continuously from June 2019 to September 2026 and carries a 95% band in every month. Its history is joined across a gap of eight months in 2023, where two forms of the data meet; across that gap district averages step up by a factor of 1.20 on average (0.97 to 1.59 by district), while the asking prices of the same 98 building pairs, read on the same archived pages in both eras, are flat (-0.9% , 95% confidence interval -3.2% to $+1.5\%$). We model the index as a Gaussian state-space panel in logarithms: a latent random-walk city level, district fixed effects, one measurement offset per district drawn around a common break, and aggregate repeat-measurement anchors. Penalised weighted least squares, equivalent to the fixed-interval Kalman smoother, estimates the common break at $+22.8\%$ (95% interval $+16.2\%$ to $+29.7\%$) and removes it. The corrected index is essentially flat over seven years. Back-tests recover injected 40% breaks, six additional months of data revise history by at most 2.13%, and no single district moves the break estimate outside 21.2–23.9%.

1 Introduction

Residential price indices are built from repeated, comparable measurements. In emerging property markets the only measurement available over a long horizon is often a published aggregate. When a break appears in such a series, a naive index inherits it as if it were a market move. This paper describes how the NIVAS Index treats such a break: it is estimated with its uncertainty and removed, rather than hidden or cut out.

The index measures the asking sale price per square metre of Bangkok condominiums. Its source for 2019 to today is a published district-level average of asking prices, available as archived copies for 2019–2022 (a sparse history) and as a monthly chart from late 2023; nothing at all covers January to August 2023. Chaining the two eras across that gap produces a $+19.8\%$ jump between late 2022 and late 2023. We show that the jump is not a market move, and we build an index that is continuous across it.

The index is an *asking-price benchmark*: it measures what sellers ask for comparable units, not what buyers eventually pay. It is designed to be read with its band and to be revised in a controlled, versioned way.

Contributions.

1. *Diagnosis.* We document three independent signatures of a measurement break in the published figures (Section 3.2): a step in the published city figure between two periods, a large and heterogeneous step across districts, and flat asking prices for the same buildings read on the same pages on both sides of the gap.
2. *Model.* We specify a state-space panel with district-specific measurement offsets and aggregate anchors (Section 4), and prove exactly which part of the break the district data cannot identify and what closes it (Proposition 1 and Corollary 1).
3. *Estimator.* We solve the model as one penalised weighted least-squares problem, equal to the Kalman and Rauch–Tung–Striebel smoother, with method-of-moments variance components and posterior bands (Section 5). The reference implementation has no dependency beyond a standard library.
4. *Evidence and governance.* We report the estimated break, the corrected index and a battery of backtests (Sections 6 and 7), and a versioning and revision policy for publication (Section 9).

Section 2 reviews the methods the design draws on. Section 3 describes the data and the evidence of the break. Section 8 states the limitations and Section 10 concludes.

2 Background and related work

Elementary indices and housing indices. A matched-item index that chains the geometric mean of price relatives is the Jevons form [24]; it is symmetric under inversion and insensitive to the very different price levels of its items, which is why it is the standard elementary aggregate. Housing practice offers three families for dealing with heterogeneity and composition [18]: hedonic regression with time dummies [14, 19, 21], repeat sales [6, 9–11, 31], and stratification or mix adjustment. Our inputs are already aggregates, so hedonics in the strict sense do not apply. Two constructions do transfer. The *time-product dummy* regression of the multilateral scanner-data literature [15, 23, 26] regresses log prices on time dummies and item fixed effects; our observation equation is that regression with districts as items. Repeat-sales indices regress the log ratio of the same unit’s price at two dates on the difference of two time effects [6]; our anchors are aggregate observations of exactly that form. The splicing methods of the multilateral literature (window, movement and mean splices; [13, 26]) extend an index without revising history; they inform our revision policy rather than the break itself.

Breaks and interventions. Methods for structural change [4, 5] estimate the *dates* of unknown breaks. Here the date is known to lie in the 2023 gap, so the relevant tool is intervention analysis [7]: a level-shift regressor with a known onset, whose size is estimated jointly with the rest of the model. Our regime offsets are interventions of that kind, one per district.

State-space models and smoothing. Structural time-series models [16, 20] represent an unobserved level as a random walk, observed through equations with their own noise; missing observations require no special treatment, and the Kalman filter [25] with the fixed-interval smoother of Rauch et al. [29] delivers the conditional mean of the states given all data. Dynamic factor models [32, 33] share this structure with a common factor and idiosyncratic components. For Gaussian linear models the smoother’s output also solves

Table 1: Sources of the district series. A point is one district-month after the latest-reading rule.

Regime	Page form	Points	Districts	Months
A	Archived text sentence, one figure per page	122	27	2019-06 to 2022-12
B	Archived 13-month chart	567	31	2023-09 to 2025-09
B	Live 13-month chart, weekly reads	546	42	2025-09 to 2026-09
All sources		1235	45	2019-06 to 2026-09

a penalised least-squares problem; the Whittaker–Henderson graduation [34, 35] and the Hodrick–Prescott filter [22, 27] are the best known instances. We exploit this equivalence to estimate the model with a single linear solve.

Lessons from on-chain price oracles. Price oracles face a closely related problem: publishing one number from several imperfect feeds. Four design choices transfer. *Geometric accumulation:* Uniswap v2 accumulates prices so that a time-weighted arithmetic mean can be read [1], while Uniswap v3 accumulates log prices so that the mean is geometric and symmetric under inversion [2]; the NIVAS Index lives in logarithms for the same reasons. *A number with its confidence:* Pyth’s aggregation combines each publisher’s price and confidence interval into an aggregate price and an aggregate confidence [28]; every NIVAS month is published with its band. *Robustness to one feed:* Chainlink’s off-chain reporting exposes the median of the node observations so that a minority of faulty reports cannot move the answer [8, 12]; our equivalent is leave-one-district-out and source drop-out testing, and a median across sources once a second district-level source exists. *Delayed, versioned updates:* the Oracle Security Module holds a new price for a configurable delay so that a faulty update can be caught before it acts [30]; our builds are versioned and screened for revisions before release. Curve’s StableSwap invariant [17] blends two pricing rules through one parameter rather than switching between them, an apt image for joining two measurement regimes through one latent state. What does not transfer is the adversarial setting: manipulation-cost arguments [3] and block-time weighting address traders; our defence is source diversity and independent anchors.

3 Data

3.1 Source and measure

Published condominium project pages show the average asking sale price per square metre of the project’s district and of Bangkok as a whole. The measure of the NIVAS Index is that district figure, in THB/m². We read it from two kinds of pages: archived copies of those pages and live pages read weekly since October 2026. Only these published aggregates are kept; no listing, lister or listing text enters the index. The source revises its most recent months; when a district-month has several readings, the latest one is used.

The page has taken two forms (Table 1). From 2019 to 2022 it printed one sentence with the district’s “current average”, which we date to the month of the archived copy (*regime A*). From late 2024 it carries a chart of the district’s last thirteen months, which archived and live copies together extend back to September 2023 (*regime B*). No page form covers January to August 2023. Of Bangkok’s 50 districts, 45 are observed at least once, 26 of them in both regimes; 5 have no history.

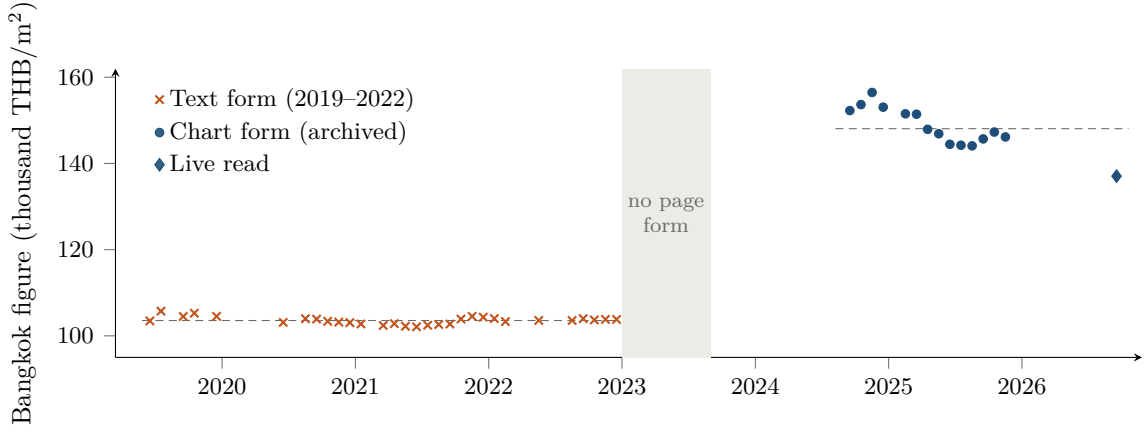


Figure 1: The published Bangkok figure. Two periods separated by a step of 43% (dashed lines: geometric means of each period, 103,555 and 148,066 THB/m²).

Weights. Districts are weighted by their stock of condominium projects in a Bangkok project directory (2,414 projects across the 50 districts), normalised to mean one over the observed districts. An observation whose move into the month exceeds 3% and is reversed by at least half the next month stays in the fit with its weight multiplied by 0.15; the last month of a series is never judged.

3.2 Evidence of a measurement break

The city figure. The Bangkok figure that the same pages print is flat in the text era (103,459 THB/m² in June 2019, 103,808 in December 2022, +0.3%), then reads 152,272 on the first chart page (September 2024), +46.7% over December 2022 (Figure 1). Its geometric mean is 103,555 over the text era and 148,066 over the chart era, a ratio of 1.43. The city figure also changed its relation to the districts: in the text era it agrees with the stock-weighted district level; in the chart era it sits well above it. The stock-weighted geometric mean of district figures moves from 96,715 THB/m² (2021–2022, 20 districts) to 113,809 (September 2024 to June 2025, 31 districts), +18%, while the city figure moves +43% over the same eras. The published city figure is not used in the index.

The district step. For the 26 districts observed on both sides of the gap, the ratio of the first regime-B figure to the last regime-A figure has geometric mean 1.20, median 1.23 and stock-weighted mean 1.23, with a standard deviation of 0.122 in logarithms and a range from 0.97 (Khlong San) to 1.59 (Bang Kho Laem) (Figure 2). Within regime B, the month-to-month log change of a district has a standard deviation of 0.035 (1058 changes) and a mean of 0.00% per month. A step that is large on average and that differs by more than ten points between districts within a few months is consistent with a change of definition rather than a market move. The step is only weakly related to the district’s price level (correlation 0.20 in logarithms).

The same buildings on both sides. The decisive evidence comes from the buildings themselves. Archived project pages of both eras also print the building’s own average asking price per square metre. Two samples of archived pages yield 98 pairs of readings of the same building in both eras, on 57 distinct buildings (the archive holds 75 buildings with a copy in each era; 18 printed no building figure in one of them). The building figure in the text era equals the last point of the page’s own quarterly building chart on 43 of 43 pages

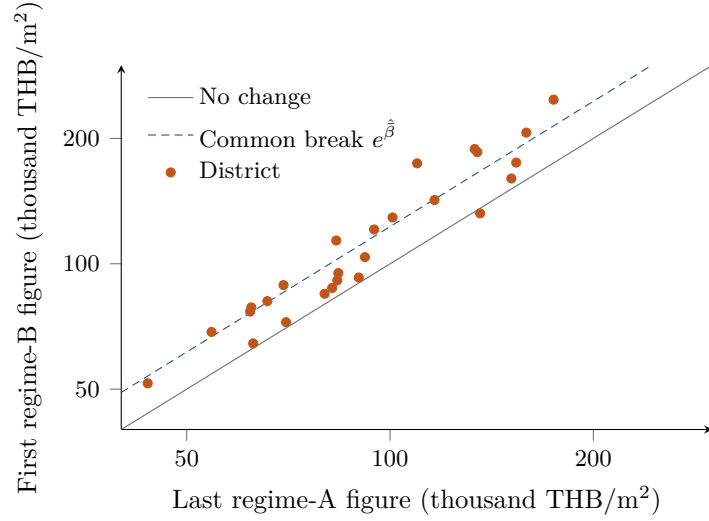


Figure 2: District figures on either side of the 2023 gap (log scales). Points above the solid line stepped up; the dashed line is the common break estimated in Section 6. The dispersion around it is the district-specific part of the break.

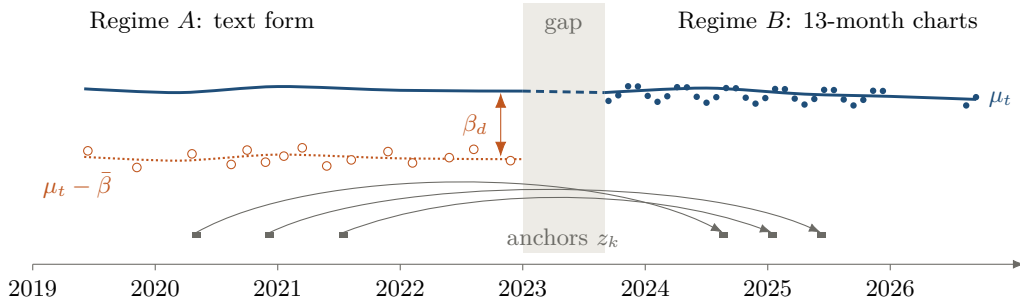


Figure 3: Observation design. Old-regime district observations (open circles) measure the latent level shifted by a district offset; new-regime observations (dots) measure it directly; no page covers the gap. Anchors (arcs) are the same buildings read in both eras: they observe the market’s move across the gap without passing through the district measure.

where both exist, so the building measure did not change with the page form. Across the 98 pairs, the building figure moved by -0.9% (95% confidence interval -3.2% to $+1.5\%$; first sample -1.4% on 54 pairs, second sample -0.2% on 44 pairs), while the district figures printed on the same pages rose by 25.2% (27.6% in the first sample).¹ Same pages, same buildings: these buildings’ asking prices were flat while the district averages stepped up.

3.3 What the district data can identify

District series that stop before a gap and restart after it identify each district’s movement within each era and the *differences* between districts’ steps (Figure 3). They cannot separate a market move during the eight empty months from a rescaling of the measure common to all districts (Proposition 1). The model therefore needs either a prior on the market’s path through the gap or outside data on that path; we use both and report each.

¹The building-level statistics are computed from the 39 aggregate rows of Appendix B. The district change on the same pages was recorded when the samples were collected; per-building lists were not retained, by design.

Table 2: Notation.

$\mathcal{D}, d$	Observed districts ($ \mathcal{D} = 45$); d^* is the reference district (largest stock), with $\alpha_{d^*} = 0$.
$t = 1, \dots, T$	Months on a regular grid, June 2019 to September 2026 ($T = 88$); t_0 is the base month, January 2024.
Ω	Observed district-months; $ \Omega = 1235$.
$r(d, t) \in \{A, B\}$	Measurement regime of an observation; B (current measure) is the reference.
$\mathcal{D}_A$	Districts observed at least once in regime A .
y_{dt}	Logarithm of the published district asking price per square metre.
μ_t	Latent Bangkok log level, on the scale of the current measure.
α_d	District fixed effect (level relative to Bangkok).
$\beta_d, \bar{\beta}$	Measurement offset of district d between regimes, and the common break.
$\tilde{w}_d$	Stock weight of district d , normalised to mean one.
$\mathcal{K}; (s_k, t_k, n_k, z_k)$	Anchors: n_k building pairs read in months $s_k < t_k$, with mean log change z_k .
$\sigma_A, \sigma_B, \sigma_\mu, \tau, \sigma_z$	Observation, level, offset and anchor standard deviations.
I_t	The index, 100 in month t_0 .

4 Model

Table 2 collects the notation. The model has four equations.

Definition 1 (Measurement model). For every observed district-month $(d, t) \in \Omega$,

$$y_{dt} = \mu_t + \alpha_d - \beta_d \mathbf{1}\{r(d, t) = A\} + \varepsilon_{dt}, \quad \varepsilon_{dt} \sim \mathcal{N}\left(0, \sigma_{r(d, t)}^2 / \tilde{w}_d\right). \quad (1)$$

The Bangkok level follows a random walk,

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\mu^2), \quad t = 2, \dots, T, \quad (2)$$

the district offsets are drawn around a common break,

$$\beta_d = \bar{\beta} + \nu_d, \quad \nu_d \sim \mathcal{N}(0, \tau^2), \quad d \in \mathcal{D}_A, \quad (3)$$

and each anchor observes the market move between two months,

$$z_k = \mu_{t_k} - \mu_{s_k} + \xi_k, \quad \xi_k \sim \mathcal{N}\left(0, \sigma_z^2 / n_k\right), \quad k \in \mathcal{K}. \quad (4)$$

All disturbances are mutually independent. The parameters μ_1 , $(\alpha_d)_{d \neq d^*}$ and $\bar{\beta}$ carry flat priors.

Equation (1) is a time-dummy panel regression with district fixed effects and a known-onset intervention for each district observed in the old regime. Equation (2) is the only link between months without data; it states that the city level is expected to move by zero from one month to the next, with standard deviation σ_μ . Equation (3) shrinks each district's offset toward the common break; a district observed in regime A only is tied to $\bar{\beta}$ by this equation alone. Equation (4) is a repeat-measurement observation in the form of Bailey et al. [6], with no offset: it rests on the following assumption.

Assumption 1 (Stable building measure). The building-level average asking price per square metre was computed the same way in both periods.

Assumption 1 is supported by the equality of the text-era sentence and the building chart on 43 of 43 pages and by the flat change of the building figure itself (Section 3.2).

Definition 2 (Index and band). Let $\hat{\mu}$ be the posterior mean of the level and Σ its posterior covariance. The index and its 95% band are

$$I_t = 100 \exp(\hat{\mu}_t - \hat{\mu}_{t_0}), \quad \left[I_t e^{-1.96 s_t}, I_t e^{+1.96 s_t} \right], \quad s_t^2 = \Sigma_{tt} + \Sigma_{t_0 t_0} - 2 \Sigma_{tt_0}. \quad (5)$$

The band is zero at the base month by construction. The ratio of the index between any two months t and u does not depend on the base, and its band follows from the same covariance with $s_{tu}^2 = \Sigma_{tt} + \Sigma_{uu} - 2 \Sigma_{tu}$.

4.1 Identification of the common break

The data in hand have a particular design: every regime- A observation precedes every regime- B observation, with the gap in between. Let t_A be the last month with a regime- A observation and t_B the first month with a regime- B observation (t_A = December 2022, t_B = September 2023).

Proposition 1 (What the district data cannot see). *Let $\theta = (\mu, \alpha, \beta, \bar{\beta})$ and let $J(\theta)$ be minus twice the log posterior of Definition 1, up to a constant. Fix a month t^* with $t_A < t^* \leq t_B$ and define the direction v by*

$$v_{\mu_t} = -\mathbf{1}\{t < t^*\}, \quad v_{\beta_d} = -1 \ (d \in \mathcal{D}_A), \quad v_{\bar{\beta}} = -1, \quad v_{\alpha_d} = 0.$$

Then the observation and offset terms of J are invariant along v , and for every $c \in \mathbb{R}$

$$J(\theta + cv) - J(\theta) = 2c \left(\frac{\Delta_{t^*}}{\sigma_\mu^2} - \sum_{k \in \mathcal{K}^*} \frac{n_k \rho_k}{\sigma_z^2} \right) + c^2 \left(\frac{1}{\sigma_\mu^2} + \sum_{k \in \mathcal{K}^*} \frac{n_k}{\sigma_z^2} \right), \quad (6)$$

where $\Delta_{t^} = \mu_{t^*} - \mu_{t^*-1}$, $\rho_k = z_k - (\mu_{t_k} - \mu_{s_k})$ is the anchor residual, and $\mathcal{K}^* = \{k : s_k < t^* \leq t_k\}$ is the set of anchors that straddle t^* .*

Proof. A regime- A observation has $t \leq t_A < t^*$, so its fitted part $\mu_t + \alpha_d - \beta_d$ changes by $-c + c = 0$. A regime- B observation has $t \geq t_B \geq t^*$, so neither μ_t nor α_d moves. Each offset residual $\beta_d - \bar{\beta}$ is unchanged. In the state term only the increment at t^* moves, from Δ_{t^*} to $\Delta_{t^*} + c$. An anchor residual changes by $-c$ when the anchor straddles t^* and is otherwise unchanged. Expanding the two affected squares gives (6). $\square$

Moving the whole old era down by c and every offset by the same amount leaves every district observation unchanged: the common part of the break is invisible to the district data. Only the random-walk prior and the anchors that straddle the gap price that direction. Setting the derivative of (6) to zero at the optimum gives a sharp description of the estimated path.

Corollary 1 (The path through the gap). *At the posterior mean, for every $t^* \in (t_A, t_B]$,*

$$\hat{\Delta}_{t^*} = \frac{\sigma_\mu^2}{\sigma_z^2} \sum_{k \in \mathcal{K}^*} n_k \hat{\rho}_k. \quad (7)$$

Without anchors the estimated level is exactly constant from t_A to t_B , and the common break is the one that makes the market flat across the gap. When every anchor straddles the gap, the estimated level is linear across it, with a slope set by the precision-weighted anchor residuals.

On the data, all 39 anchors straddle the gap and the slope is -0.1122% per month, a market move of -1.00% over the 9 increments from December 2022 to September 2023; without anchors the level is constant over the same months, to machine precision. The curvature in (6) also shows the relative weight of the two sources of information along v : $1/\sigma_\mu^2 \approx 7,290$ for the prior and $\sum_k n_k/\sigma_z^2 \approx 6,050$ for the anchors. The anchors therefore weigh about as much as the prior and *test* it; that they agree with it (Section 6) is the substantive finding.

5 Estimation

5.1 Penalised weighted least squares

Write the four equations of Definition 1 as rows $x_i^\top \theta = y_i$ with weights ω_i equal to their inverse variances: $\tilde{w}_d/\sigma_{r(d,t)}^2$ for an observation (times the factor of the Weights paragraph where it applies), $1/\sigma_\mu^2$ for a level increment (with $y_i = 0$), $1/\tau^2$ for an offset (with $y_i = 0$) and n_k/σ_z^2 for an anchor. Minus twice the log posterior is, up to a constant,

$$\begin{aligned} J(\theta) = & \sum_{(d,t) \in \Omega} \frac{\tilde{w}_d}{\sigma_{r(d,t)}^2} (y_{dt} - \mu_t - \alpha_d + \beta_d \mathbf{1}\{r(d,t) = A\})^2 + \sum_{t=2}^T \frac{(\mu_t - \mu_{t-1})^2}{\sigma_\mu^2} \\ & + \sum_{d \in \mathcal{D}_A} \frac{(\beta_d - \bar{\beta})^2}{\tau^2} + \sum_{k \in \mathcal{K}} \frac{n_k}{\sigma_z^2} (z_k - \mu_{t_k} + \mu_{s_k})^2, \end{aligned} \quad (8)$$

a penalised weighted least-squares criterion whose penalty is the first-difference (Whittaker) smoother on the level. Its minimiser and the posterior covariance are

$$\hat{\theta} = (X^\top W X)^{-1} X^\top W y, \quad \Sigma_\theta = (X^\top W X)^{-1}, \quad (9)$$

with X the stacked rows, $W = \text{diag}(\omega_i)$ and y the stacked targets.

Proposition 2 (Equivalence with the Kalman smoother). *Under Definition 1, for given variances, the posterior of θ is $\mathcal{N}(\hat{\theta}, \Sigma_\theta)$ with $\hat{\theta}$ and Σ_θ as in (9). In particular $\hat{\mu}_t$ and Σ_{tu} coincide with the fixed-interval smoothed mean and covariance of the level produced by the Kalman filter and the Rauch–Tung–Striebel smoother applied to the state-space form with state $(\mu_t, \alpha, \beta, \bar{\beta})$, in which the static parameters are constant states with diffuse initial conditions.*

Proof. The log posterior is $-\frac{1}{2}J(\theta)$ plus a constant, a quadratic form in θ ; the posterior is therefore Gaussian, with mode equal to mean $\hat{\theta}$ and covariance equal to the inverse Hessian $(X^\top W X)^{-1}$. The Kalman filter and the fixed-interval smoother compute the conditional mean and covariance of the state given all observations for the same joint Gaussian distribution [16, ch. 4–5]; the diffuse initialisation corresponds to the flat priors of Definition 1. Both procedures therefore return the same moments. $\square$

The normal matrix is symmetric positive definite once the reference district fixes the level of α ; with 1235 observations and 160 unknowns it is 160×160 . One Cholesky factorisation gives $\hat{\theta}$ and, column by column, Σ_θ ; the full estimation takes under a second in an interpreted language with no numerical library. Missing months need no imputation: a month without observations contributes only its level increments.

Algorithm 1 Variance components (three passes, bounded)

- 1: Initialise $\sigma_A = \sigma_B = 0.06$, $\sigma_\mu = 0.01$, $\tau = 0.10$; set $\sigma_z^2 = \sum_k (n_k - 1)s_k^2 / \sum_k (n_k - 1)$ from the anchor rows (default 0.20).
 - 2: **for** pass = 1, 2, 3 **do**
 - 3: Solve (9) for $\hat{\theta}$, Σ_θ .
 - 4: $\sigma_r \leftarrow \max(0.01, [\frac{1}{|\Omega_r|} \sum_{\Omega_r} \tilde{w}_d \hat{\varepsilon}_{dt}^2]^{1/2})$ for each regime r .
 - 5: On months with at least 10 districts, $m_t \leftarrow$ mean of $y_{dt} - \hat{\alpha}_d + \hat{\beta}_d \mathbf{1}_A$; for consecutive such months, $\sigma_\mu^2 \leftarrow \widehat{\text{Var}}(m_{t+1} - m_t) - \sigma_{\max}^2(1/n_t + 1/n_{t+1})$, clipped to $[0.004, 0.03]^2$.
 - 6: $\tau^2 \leftarrow \widehat{\text{Var}}(\hat{\beta}_d) - \text{Var}(\hat{\beta}_d)$ over districts observed in both regimes, clipped to $[0.03, 0.30]^2$.
 - 7: **end for**
-

5.2 Variance components

The five standard deviations are estimated by a deterministic method-of-moments iteration (Algorithm 1), so that a build is a pure function of its data. Reading σ_μ from the smoothed path itself would understate it and feed back into more smoothing; the estimator uses the cross-sectional mean of the de-measured observations on dense months instead, from which the sampling noise is subtracted.

On the data the iteration gives $\sigma_A = 0.022$ and $\sigma_B = 0.034$ (log units), $\sigma_\mu = 0.012$ per month, $\tau = 0.131$ and $\sigma_z = 0.127$. The estimated monthly volatility of the city level lies inside its bounds of 0.004 and 0.03; the sensitivity of every result to it is reported in Section 7. The bands are conditional on these estimates: they do not include the uncertainty of the variance components.

6 Results

6.1 The common break

The common break is $\hat{\beta} = 0.205$ in logarithms (standard error 0.028), that is +22.8% with a 95% interval of +16.2% to +29.7%. Without anchors, identification rests on the random-walk prior alone and the estimate is +21.4% (+11.3% to +32.4%); with the first anchor sample only it is +23.3% (+16.2% to +30.9%). The anchors move the estimate by 1.4 points because they agree with the prior: the market move the model fits over the anchors' spans averages -0.7% (pair-weighted), against -0.9% observed on the buildings. Two independent routes, a flat-market prior through the gap and the same buildings read on both sides, give the same answer.

The 98 building pairs taken alone imply a larger district step, 25.2% on the same pages, because they over-represent three central districts whose own offsets are above the median (Khlong Toei, Pathum Wan, Huai Khwang). The model weights districts by stock, not by where archived buildings happen to be.

6.2 District heterogeneity

Figure 4 and Table 3 report the offset of each district observed in both regimes. The median offset is +21.5%, the stock-weighted mean +26.1%, and the range runs from -2.5% (Rat Burana) to +67.5% (Bang Kho Laem); the estimated dispersion around the common break is $\tau = 0.131$. A single common correction would leave errors of this size in district histories, which is why the model carries one offset per district. The district observed in regime A only (Samphanthawong) takes the common break, with a standard error of 13.0 points.

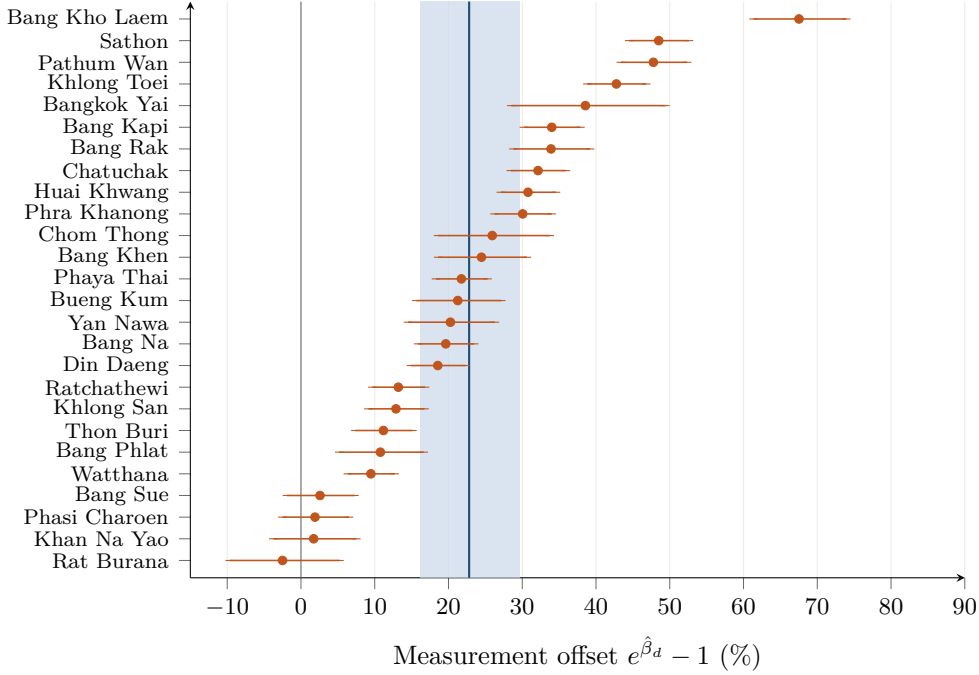


Figure 4: Estimated measurement offset of each district observed in both regimes, with 95% intervals. The vertical line and the shaded band are the common break $\hat{\beta}$ and its 95% interval.

6.3 The index

Figure 5 shows the index and Table 4 selected months. The index is essentially flat over seven years: 102.2 in June 2019, 103.3 in December 2022, 100 in January 2024 and 99.5 in September 2026, with a high of 104.0 (2022-11) and a low of 98.2 (2026-02). The half-width of the band has a median of 1.44% on the 40 months with at least ten districts, and a median of 3.4% (maximum 4.3%) on the 51 months before the break, where the level rests on the estimated offset.

Within each era the model and the chained index agree. From January 2024 to June 2025 the index moves -0.45% against -0.57% for the chain; from October 2023 to November 2025, -2.51% against -1.61% ; from the fourth quarter of 2019 to the fourth quarter of 2022, $+0.61\%$ against $+1.68\%$ (on four to eight districts). The difference between the two constructions is the level across the gap: the chain bridged it on six districts with a $+19.8\%$ step, the model's level moves by -2.8% from 2022-12 to the first chart page (2024-09).

The published city figure can be set beside the index within each period: from 2019-06 to 2022-12 it moves $+0.3\%$ and the index $+1.0\%$; from 2024-09 to 2026-09 it moves -10.0% and the index -0.9% . The matched-district construction removes composition effects.

7 Robustness and backtests

Table 5 summarises the tests. Each is computed by the reference implementation on the same data, in memory.

Recovery. A $+40\%$ break injected into a clean synthetic panel is recovered at $+42.3\%$, with the truth inside the interval and a band that covers the true path in 95.2% of months; at a nominal 95% the band is close to its nominal level. Injected into the real regime-B data, where no true break exists, the same break is recovered at $+40.8\%$.

Table 3: District offsets $e^{\hat{\beta}_d} - 1$ (%) and their standard errors (log points), districts observed in both regimes.

District	Offset	s.e.	District	Offset	s.e.
Rat Burana	−2.5	3.9	Phaya Thai	+21.7	1.5
Khan Na Yao	+1.7	2.8	Bang Khen	+24.5	2.5
Phasi Charoen	+1.9	2.3	Chom Thong	+25.9	3.1
Bang Sue	+2.6	2.3	Phra Khanong	+30.1	1.5
Watthana	+9.5	1.5	Huai Khwang	+30.8	1.5
Bang Phlat	+10.8	2.6	Chatuchak	+32.1	1.4
Thon Buri	+11.2	1.8	Bang Rak	+33.9	2.0
Khlong San	+12.9	1.7	Bang Kapi	+34.0	1.5
Ratchathewi	+13.2	1.6	Bangkok Yai	+38.6	3.8
Din Daeng	+18.6	1.6	Khlong Toei	+42.8	1.4
Bang Na	+19.6	1.6	Pathum Wan	+47.8	1.5
Yan Nawa	+20.3	2.5	Sathon	+48.5	1.4
Buang Kum	+21.3	2.4	Bang Kho Laem	+67.5	1.9

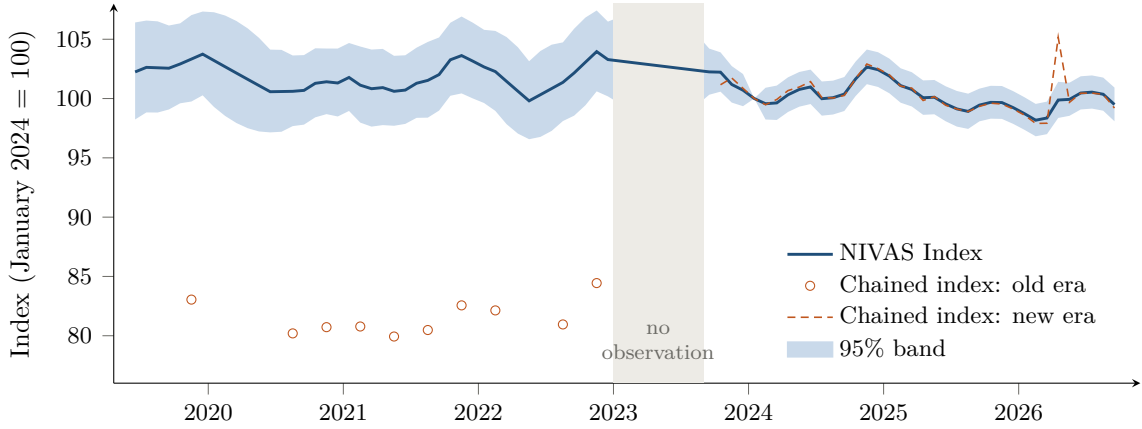


Figure 5: The NIVAS Index with its 95% band, and the matched-pair chained index that preceded it. The chain bridged the 2023 gap on six districts and inherited the measurement break; the model removes it. Within each era the two constructions move together, apart from the one-month move of April 2026, which the model down-weights.

Stability. Six more months of data move published history by at most 2.13%, twelve months by at most 2.19%. Removing any one district keeps the break within 21.2–23.9%. Fixing the city’s monthly volatility at three times its lower bound moves the break by less than one point and widens the old-era band from 1.1% to 1.7% (one standard deviation).

What carries the estimate. The chart-era data pin the district levels: without the archived chart pages the break is +26% (interval 16–36%), and the dense regime-B era is what fixes the offsets. The anchors, in contrast, refine the estimate but do not drive it.

8 Limitations

The common break rests on an assumption. Proposition 1 is a statement about the data, not about the estimator: no estimator can extract the common break from district series that stop and restart around a gap. The random-walk prior fixes the expected market move through the gap at zero; the anchors replace that assumption by the stability

Table 4: Selected months of the index (January 2024 = 100), with the 95% band, the number of districts observed and the matched-pair chained index (quarterly before 2023).

Month	Index	95% band	Districts	Chained	
2019-06	102.2	[98.2, 106.4]	1	—	
2019-12	103.7	[100.3, 107.3]	2	83.0	
2020-12	101.3	[98.5, 104.2]	12	80.7	
2021-12	103.2	[100.0, 106.4]	2	82.6	
2022-12	103.3	[100.2, 106.5]	4	84.4	
2023-04	102.8	[98.8, 107.0]	0	—	no observation
2023-09	102.2	[100.3, 104.2]	6	—	
2023-10	102.2	[100.6, 103.9]	10	101.2	
2024-01	100.0	—	18	100.0	base
2024-06	101.0	[99.5, 102.4]	28	101.4	
2024-12	102.4	[101.0, 103.9]	30	102.5	
2025-06	99.5	[98.1, 101.0]	24	99.4	
2025-11	99.7	[98.3, 101.1]	42	99.5	
2026-09	99.5	[98.1, 100.9]	42	99.2	

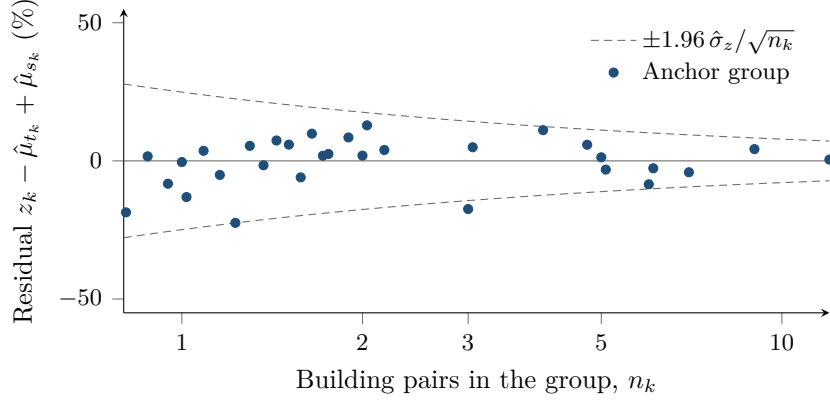


Figure 6: Anchor residuals against group size. The funnel is the 95% range implied by the estimated per-pair dispersion $\hat{\sigma}_z = 0.127$; the residuals are consistent with it.

of the building measure (Assumption 1). Both give about +21.4% and +22.8% respectively. Neither measures the market between January and August 2023, for which no page form exists.

Early history. Before 2023 the index rests on 122 district-months across 27 districts, often one or two per month. The band reflects this, and the early level is best quoted with its band.

Static district effects. A district drifting against the city over several years is treated as noise around a fixed α_d . The observation variance absorbs the drift, but residuals are serially correlated and the band is approximate.

Anchor sample. The 98 pairs come from 57 buildings concentrated in central districts, and the second sample re-uses buildings of the first at other dates. The model treats each pair as independent, which gives the anchors somewhat more precision than the data support; the independent information is closer to 57 buildings than to 98 pairs. The archive

Table 5: Backtests and sensitivity analyses.

Test	Result
Synthetic market (20 districts, 84 months, monthly s.d. 0.8%), $\times 1.4$ rescale from September 2023, 8-month gap, old era thinned to one district-month in three	Break recovered at +42.3% [35.1, 50.0]; path error at most 3.8% (old era 3.0%); band covers the truth in 95.2% of months
Real regime-B data (September 2023 to September 2026, no true break), $\times 1.4$ from January 2025, gap May–December 2024	Break recovered at +40.8% [29.1, 53.5]; path within 1.7% of the clean fit
Revision: refit without the last 6 / 12 months	History moves at most 2.13% / 2.19% (median 0.57% / 0.69%); break moves by at most 0.7 point
Leave one district out (45 fits)	Median of the per-district maximum change 0.16%; worst Bang Na 3.25% (2022-05, a month with 2 districts); break within 21.2–23.9%
Drop the live source	Maximum change 2.16%; break +22.7%
Drop the archived chart source	Break +26% [16, 36]: the chart era identifies the offsets
Drop the anchors	Maximum change 1.25% (2020-08); break +21.4%
Second anchor sample (54 to 98 pairs)	Break +23.3% to +22.8%; interval width 14.7 to 13.5 points; σ_z 0.141 to 0.127
σ_μ fixed at 0.004 / 0.007 / 0.012 per month	Break +22.1 / +22.4 / +22.8%; old-era level s.d. 1.1 / 1.4 / 1.7%
Anchors, observed against fitted (39 groups)	Standardised residual s.d. 0.87, consistent with $\sigma_z / \sqrt{n_k}$

holds no further building with a copy in each era.

Plug-in variances. The band conditions on the estimated variance components and ignores their uncertainty; the synthetic coverage of 95.2% suggests that this is not material at present.

Future breaks and coverage. A further change in the data can be handled as one more offset per district in (1). 5 districts have no history and enter the index only from their first reading.

9 Governance and versioning

Buils. A build is identified by its formula, its weights and a hash of its input data and anchors. Buils, their monthly rows and their district offsets are append-only: a new reading produces a new build, and an earlier build is never edited. Every published series names its build.

Revisions. Each build re-estimates the whole history, because new months refine the level and the variance components. A move of a district’s figure in one month that the next month reverses is down-weighted, never deleted: the observation stays in the fit with a reduced weight, so a later month can revise the month before. The measured revision when six months of new data are added is at most 2.13% (median 0.57%) of a published

month, and at most 2.19% (median 0.69%) for twelve months. A change of method that would move any published month by more than 0.5% is held for review before release.

The break is stored with each build. The common break and the district offsets are re-estimated with the rest of the history at each build and stored with it, so every published value traces to its estimate. A new source regime or a change of method is stated here, never a silent change.

Publication rules. Every published month carries its band and its district count. Months with no observation are drawn dashed and labelled as model-only; months before the break carry their band and a note that their level rests on the estimated offset; months on fewer than five districts are marked thin.

Data hygiene and independence. The index uses only published aggregates; no listing, lister or listing text is stored. Proprietary data held by NIVAS never enters this index; a series built on it would be a separate dataset and a separate build.

10 Conclusion

A break in published data mid-history need not end a price index, nor be hidden inside it. The NIVAS Index treats the break as a measurement offset per district, estimates the common part from a stated prior and from independent same-building evidence, and publishes every month with a band that shows what the data can and cannot support. On Bangkok condominium asking prices, the method attributes a +22.8% step in the published district averages to the measurement, not to the market, and yields an index that is essentially flat from 2019 to 2026. The same machinery admits a further regime or a second district-level source as one more set of rows in a single linear system.

Disclaimer

This paper is for general information purposes only. It describes a statistical benchmark of asking prices. It does not constitute investment advice, an appraisal of any property, or a recommendation to buy or sell any asset, and it should not be relied upon for accounting, legal, tax or lending decisions. Figures are estimates with stated uncertainty and are subject to revision under the policy of Section 9.

References

- [1] Hayden Adams, Noah Zinsmeister, and Dan Robinson. Uniswap v2 core. Technical whitepaper, Uniswap, March 2020. URL <https://app.uniswap.org/whitepaper.pdf>.
- [2] Hayden Adams, Noah Zinsmeister, Moody Salem, River Keefer, and Dan Robinson. Uniswap v3 core. Technical whitepaper, Uniswap, March 2021. URL <https://app.uniswap.org/whitepaper-v3.pdf>.
- [3] Guillermo Angeris and Tarun Chitra. Improved price oracles: Constant function market makers. In *Proceedings of the 2nd ACM Conference on Advances in Financial*

- Technologies (AFT '20)*, pages 80–91. Association for Computing Machinery, 2020. doi: 10.1145/3419614.3423251.
- [4] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998. doi: 10.2307/2998540.
 - [5] Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003. doi: 10.1002/jae.659.
 - [6] Martin J. Bailey, Richard F. Muth, and Hugh O. Nourse. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942, 1963. doi: 10.1080/01621459.1963.10480679.
 - [7] George E. P. Box and George C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349):70–79, 1975. doi: 10.1080/01621459.1975.10480264.
 - [8] Lorenz Breidenbach, Christian Cachin, Benedict Chan, Alex Coventry, Steve Ellis, Ari Juels, Farinaz Koushanfar, Andrew Miller, Brendan Magauran, Daniel Moroz, Sergey Nazarov, Alexandru Topliceanu, Florian Tramèr, and Fan Zhang. Chainlink 2.0: Next steps in the evolution of decentralized oracle networks. Whitepaper v1.0, Chainlink Labs, April 2021. URL <https://research.chain.link/whitepaper-v2.pdf>.
 - [9] Charles A. Calhoun. OFHEO house price indexes: HPI technical description. Technical report, Office of Federal Housing Enterprise Oversight, Washington, DC, March 1996. URL <https://www.fhfa.gov/research/papers/house-price-indexes-hpi-technical-description>. Hosted by the Federal Housing Finance Agency as *House Price Indexes: HPI Technical Description*.
 - [10] Karl E. Case and Robert J. Shiller. Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*, pages 45–56, September 1987. URL <https://ideas.repec.org/a/fip/fedbne/y1987isepp45-56.html>. Federal Reserve Bank of Boston; also NBER Working Paper 2393.
 - [11] Karl E. Case and Robert J. Shiller. The efficiency of the market for single-family homes. *American Economic Review*, 79(1):125–137, 1989. URL <https://ideas.repec.org/a/aea/aecrev/v79y1989i1p125-37.html>.
 - [12] Chainlink. Offchain reporting. Protocol documentation. URL <https://docs.chain.link/architecture-overview/off-chain-reporting>. Accessed 11 October 2026.
 - [13] Antonio G. Chessa. A new methodology for processing scanner data in the Dutch CPI. *Eurostat Review on National Accounts and Macroeconomic Indicators (EURONA)*, (1/2016):49–69, 2016. URL <https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-gp-16-001>. Eurostat, Luxembourg, cat. no. KS-GP-16-001.
 - [14] Andrew T. Court. Hedonic price indexes with automotive examples. In *The Dynamics of Automobile Demand*, pages 99–117. General Motors Corporation, New York, 1939. URL https://allengoodman.wayne.edu/Research/PUBS/Court_Hedonic.pdf.
 - [15] Jan de Haan and Frances Krsinich. Scanner data and the treatment of quality change in nonrevisable price indexes. *Journal of Business & Economic Statistics*, 32(3):341–358, 2014. doi: 10.1080/07350015.2014.880059.

- [16] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*. Number 38 in Oxford Statistical Science Series. Oxford University Press, 2nd edition, 2012. doi: 10.1093/acprof:oso/9780199641178.001.0001.
- [17] Michael Egorov. StableSwap — efficient mechanism for stablecoin liquidity. Whitepaper, Curve Finance, November 2019. URL <https://classic.curve.finance/files/stableswap-paper.pdf>.
- [18] Eurostat, ILO, IMF, OECD, UNECE, and World Bank. *Handbook on Residential Property Prices Indices (RPPIs)*. Publications Office of the European Union, Luxembourg, 2013. doi: 10.2785/34007. Cat. no. KS-RA-12-022-EN.
- [19] Zvi Griliches. Hedonic price indexes for automobiles: An econometric analysis of quality change. In *The Price Statistics of the Federal Government*, number 73 in NBER General Series, pages 173–196. National Bureau of Economic Research, 1961. URL <https://www.nber.org/chapters/c6492>. Staff Paper 3, Price Statistics Review Committee.
- [20] Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989. doi: 10.1017/CBO9781107049994.
- [21] Robert J. Hill. Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, 27(5):879–914, 2013. doi: 10.1111/j.1467-6419.2012.00731.x.
- [22] Robert J. Hodrick and Edward C. Prescott. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997. doi: 10.2307/2953682.
- [23] Lorraine Ivancic, W. Erwin Diewert, and Kevin J. Fox. Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1):24–35, 2011. doi: 10.1016/j.jeconom.2010.09.003.
- [24] William Stanley Jevons. On the variation of prices and the value of the currency since 1782. *Journal of the Statistical Society of London*, 28(2):294–320, 1865. doi: 10.2307/2338419.
- [25] Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. doi: 10.1115/1.3662552.
- [26] Frances Krsinich. The FEWS index: Fixed effects with a window splice. *Journal of Official Statistics*, 32(2):375–404, 2016. doi: 10.1515/jos-2016-0021.
- [27] Peter C. B. Phillips and Sainan Jin. Business cycles, trend elimination, and the HP filter. *International Economic Review*, 62(2):469–520, 2021. doi: 10.1111/iere.12494.
- [28] Pyth Data Association. Pyth network: A first-party financial oracle. Whitepaper v2.0, Pyth Data Association, September 2023. URL <https://pyth.network/whitepaper>.
- [29] Herbert E. Rauch, F. Tung, and Charlotte T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965. doi: 10.2514/3.3166.
- [30] Sky Protocol. OSM (Oracle Security Module). Protocol documentation (formerly MakerDAO). URL <https://developers.skyeco.com/protocol/core/osm/>. Accessed 11 October 2026; contract source <https://github.com/makerdao/osm>.

- [31] S&P Dow Jones Indices. S&P Cotality Case-Shiller home price indices methodology. Index methodology, S&P Dow Jones Indices, 2025. URL <https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-cotality-cs-home-price-indices.pdf>. Formerly *S&P CoreLogic Case-Shiller Home Price Indices Methodology*.
- [32] James H. Stock and Mark W. Watson. New indexes of coincident and leading economic indicators. In Olivier J. Blanchard and Stanley Fischer, editors, *NBER Macroeconomics Annual 1989*, volume 4, pages 351–394. MIT Press, 1989. doi: 10.1086/654119.
- [33] James H. Stock and Mark W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002. doi: 10.1198/073500102317351921.
- [34] Howard L. Weinert. Efficient computation for Whittaker–Henderson smoothing. *Computational Statistics & Data Analysis*, 52(2):959–974, 2007. doi: 10.1016/j.csda.2006.11.038.
- [35] Edmund T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1923. doi: 10.1017/S0013091500077853.

A Parameters of formula v1c

Parameter	Setting	Role
Regimes	text form $\rightarrow A$; archived chart, live $\rightarrow B$	B is the published scale
Base month t_0	January 2024	First January of the dense era
σ_μ	start 0.01; estimated; bounds $[0.004, 0.03]$	Monthly s.d. of the city level
σ_A, σ_B	start 0.06; estimated per regime; at least 0.01	Observation noise
τ	start 0.10; estimated; bounds $[0.03, 0.30]$	Dispersion of district offsets
σ_z	pooled s.d. of the anchor rows; default 0.20	Per-pair noise of an anchor
Passes	3	Method-of-moments iterations
Dense month	at least 10 districts	Calibrates σ_μ
Band	$z = 1.96$	95%
Thin month	fewer than 5 districts	Publication flag

B Aggregate anchors

Each row aggregates the building pairs read in a text-era month s_k and a chart-era month t_k : their number n_k , the mean log change of the building figure z_k (shown in %), its standard deviation, the change fitted by the model, and the sample. No building identity is published. Totals: 98 pairs on 57 buildings; mean -0.9% , pooled s.d. 0.122.

s_k	t_k	n_k	z_k (%)	s.d. (log)	Fitted (%)	Sample
2019-12	2025-04	1	-4.0	—	-3.6	1
2020-09	2024-11	2	+3.9	0.128	+1.9	1
2020-08	2025-05	6	-9.0	0.224	-0.5	1
2020-11	2024-11	7	-2.9	0.073	+1.2	1
2020-10	2025-02	1	-13.3	—	-0.2	1
2020-11	2025-05	9	+3.0	0.084	-1.3	1
2020-11	2025-08	3	-19.9	0.326	-2.5	1
2020-11	2025-10	1	-10.0	—	-1.7	1
2021-02	2024-11	5	+2.8	0.059	+1.5	1
2021-01	2025-03	1	+2.7	—	-1.0	1
2021-01	2025-05	4	+9.5	0.093	-1.7	1
2021-01	2025-08	1	-1.2	—	-2.9	1
2021-05	2024-11	2	+14.9	0.008	+2.0	1
2021-06	2025-05	1	-5.7	—	-0.6	1
2021-06	2025-09	1	-19.9	—	-1.3	1
2021-08	2024-12	1	-21.6	—	+0.9	1
2021-07	2025-08	2	+6.1	0.085	-2.4	1
2021-12	2025-03	1	+1.0	—	-2.3	1
2021-11	2025-05	1	+2.0	—	-3.5	1
2022-01	2024-11	1	-1.0	—	+0.0	1
2022-01	2025-05	1	-4.1	—	-2.5	1
2022-08	2024-12	1	-2.1	—	+1.0	1
2022-08	2025-09	1	+5.5	—	-1.9	1
2019-12	2025-06	1	-0.9	—	-4.1	2
2020-09	2024-11	1	+7.8	—	+1.9	2
2020-09	2025-03	1	-0.5	—	+0.1	2
2020-08	2025-05	3	+4.4	0.042	-0.5	2
2020-09	2025-09	1	-7.2	—	-1.2	2
2020-11	2024-11	5	-1.9	0.148	+1.2	2
2020-12	2025-03	1	-0.7	—	-0.5	2
2020-11	2025-05	12	-0.8	0.089	-1.3	2
2020-11	2025-08	6	-5.2	0.105	-2.5	2
2021-01	2024-11	2	+4.8	0.163	+0.9	2
2021-02	2025-05	5	+4.8	0.139	-1.0	2
2021-02	2025-08	2	+0.3	0.068	-2.2	2
2021-06	2025-04	1	+9.2	—	-0.6	2
2021-05	2025-08	1	-18.7	—	-1.7	2
2022-01	2025-04	1	-0.7	—	-2.6	2
2022-08	2025-09	1	+5.5	—	-1.9	2

C Reproduction

The reference implementation is a Python module that uses only the standard library. Given the district series and the anchor file, it rebuilds the index, the backtests and every number and figure of this paper from 1235 district-month points. The estimation is deterministic: the same data give the same index to the last digit.