

A High-Frequency Hedonic Housing Price Index for Quebec: Weekly Measurement with Hierarchical Pooling

Simon-Pierre Boucher*

August 2026

Abstract

We construct quality-adjusted *weekly* housing price indexes for Quebec from roughly 745,000 geocoded residential transactions (2021–2026), published for the province, its 17 administrative regions, and ten major municipalities, by property type. The method combines a pooled hedonic regression with week-by-type time effects (stage 1) and a hierarchical state-space model that shrinks thin geography–type cells toward their parent trends (stage 2). A repeat-sales benchmark, a downsampling experiment, and a composition-shock simulation validate the design. Weekly measurement is feasible far below conventional liquidity thresholds: with only 25 transactions per week the smoothed path deviates from the full-sample path by under 0.8% RMSE. All series carry calibrated uncertainty bands, reliability grades, and a real-time (one-sided) variant that never uses future data.

1 Introduction

House price indexes are published monthly or quarterly almost everywhere, yet the underlying transactions arrive continuously. The obstacle to weekly measurement is not conceptual but statistical: weekly cells are thin, and naive weekly averages confound price change with composition change. The composition problem is classical (??); the thinness problem is acute at weekly frequency for any geography below a large metropolitan area. ? demonstrate that timely house-price measurement is feasible with appropriate pooling; we push frequency to the week and geography to the municipality for a single Canadian province.

Our contribution is a transparent two-stage architecture. Stage 1 is a single pooled hedonic regression over the full sample in the time-dummy tradition (??): structural characteristics, fine location fixed effects, and week $\times$ type effects that constitute the provincial weekly price paths. Stage 2 treats each sub-provincial cell’s weekly deviation from its parent path as a latent random walk observed through the cell’s weekly mean residual, estimated by Kalman filtering (??). The observation variance scales inversely with the week’s transaction count, so liquid weeks speak for themselves while thin weeks shrink toward the parent trend — an empirical-Bayes resolution of the thinness problem in the spirit of hierarchical modeling.

2 Data

The input is a registry extract of 745,119 residential transactions in 1,100+ Quebec municipalities from January 2021 through July 2026, with price, date, address, coordinates, property type

*Contact: contact@spboucher.ai. Code and data pipeline: QWHPI platform repository.

(single-family, condominium, plex, undetermined), floor area, year built, building type, and the municipal assessment value. Coordinates are spatially joined to the official SDA 1/20,000 administrative boundaries (100% within-polygon match; 99.9% agreement with the raw city text field). The *undetermined* type (14.4% of records, 95%+ missing on all structural characteristics) never enters any published index.

Screens are economically motivated and fully tabulated: transfers priced outside [0.2, 5.0] times the municipal assessment (predominantly non-arm’s-length family transfers and estate settlements), conveyances above \$10M (portfolio scale), and duplicate registrations — 3.5% of rows in total, each flagged with a reason and never silently deleted. Implausible characteristics (floor area outside [20, 2000] m², year built after the sale year) are nulled with indicators, not dropped. The research sample is 626,127 transactions across all 291 weeks.

3 Method

3.1 Stage 1: pooled hedonic

For transaction i in week $t(i)$, type $k(i)$, location $\ell(i)$:

$$\log P_i = \beta_1 \log \text{FA}_i + \beta_2 \mathbb{I}[\text{FA miss}]_i + f(\text{age}_i) + \gamma_{b(i)} + \alpha_{\ell(i)} + \delta_{t(i),k(i)} + \varepsilon_i, \quad (1)$$

with age bins $f(\cdot)$, building-type effects γ , forward-sortation-area fixed effects α (the postal-code level; finer than municipality), and week $\times$ type effects $\delta_{t,k}$. Missing floor areas are imputed by conditional medians within type $\times$ municipality and always flagged; no future price information enters any imputation. The assessment value is excluded from the headline model to avoid valuation feedback (it appears only in a robustness specification and in the screens). Estimation is by high-dimensional fixed-effects least squares on the full pooled sample ($R^2 = 0.60$); the paths $\delta_{\cdot,k}$, rebased so the 2021 average equals 100, are the provincial indexes per type. Independent weekly regressions are never run.

3.2 Stage 2: hierarchical weekly state

Let $e_i = \log P_i - \widehat{\log P_i}$ be stage-1 residuals. For a cell c (region $\times$ type, then municipality $\times$ type) define the weekly mean $\bar{e}_{c,t}$ over $n_{c,t}$ transactions. The latent deviation $d_{c,t}$ of the cell from its parent’s path evolves as

$$\bar{e}_{c,t} = d_{c,t} + \epsilon_{c,t}, \quad \epsilon_{c,t} \sim N(0, \sigma_c^2/n_{c,t}), \quad d_{c,t} = d_{c,t-1} + \eta_{c,t}, \quad \eta_{c,t} \sim N(0, \tau_c^2), \quad (2)$$

with σ_c^2 the cell’s residual variance and τ_c^2 estimated by maximum likelihood via the prediction-error decomposition. Zero-transaction weeks are pure predictions — the weekly grid is never broken. Municipality deviations are estimated relative to the smoothed regional path, so the published municipal index is $\delta_{t,k} + \hat{d}_t^{\text{region}} + \hat{d}_t^{\text{city}}$. The Kalman gain provides an interpretable shrinkage weight; the posterior variance combines with the delta-method variance of $\delta_{t,k}$ into the published 95% band. Two variants ship for every series: the raw weekly estimate and a strictly one-sided filtered path (*no lookahead*) for real-time use; a two-sided smoother is retained for research. The “all types” series is a fixed-weight combination of the three type indexes with full-sample transaction shares, in the stratification tradition of ?.

4 Validation

Repeat sales. A Bailey–Muth–Nourse index (??) on same-unit pairs (unit signature = address, floor area, year built; 90-day minimum gap; 63k+ pairs) tracks the hedonic indexes’ relative paths closely: Montréal condominiums end 8.8% below the provincial condominium index by repeat sales versus 8.2% by the hierarchical model. Hedonic levels sit a few percent above repeat sales, consistent with the well-documented depreciation/renovation bias of repeat-sales measures (??). Notably, a city-specific time-dummy regression *overstates* Montréal’s divergence (−14%) — pooled coefficients with local deviations are closer to the composition-free benchmark than fully local estimation.

Downsampling. Using Montréal condominiums (~145/week) as a laboratory, we thin each week to a target count and compare against the full-sample path (20 replications): RMSE rises from 0.32% (100/wk) to 0.64% (50/wk), 0.80% (25/wk), 1.15% (15/wk), 1.36% (10/wk) and 1.63% (5/wk); 95% CI coverage stays above 0.97 down to 25/week and degrades to ~0.83 at 10/week. These results anchor the published A–E reliability grades and the decision to publish (with warnings) rather than suppress thin cells.

Composition shock. Removing 70% of below-median floor-area single-family sales for 26 weeks moves the raw median by +11.7% and the hedonic index by −0.7%: the index measures prices, not mix.

Seasonality. Annual harmonics on weekly log-changes are jointly insignificant for all tested series ($p > 0.36$); all series are therefore published NSA.

5 Findings

Between the 2021 base and July 2026 the provincial all-type index rises to ~156 (+56%), with the 2022 spring boom, the 2022–23 rate-hike correction, and a sustained 2024–26 expansion clearly timed at weekly resolution. Regional heterogeneity is large: Montréal condominiums appreciate roughly 27% while several regions exceed 70%. The weekly index dates turning points 4–8 weeks earlier than the same methodology aggregated to monthly frequency. The assessment-gap module (sale price / municipal assessment) tracks the mechanical lag of triennial assessment rolls and is published as a separate concept, never mixed with the price index.

6 Conclusion

Weekly, quality-adjusted, sub-provincial price measurement is practical with transparent econometrics: one pooled hedonic regression plus one small state-space model per cell, running end-to-end in under a minute on a laptop. Honest uncertainty — confidence bands, reliability grades, effective sample sizes, partial-week flags, and preserved vintages — is what makes publication of thin cells defensible.