High Frequency, High Fidelity: Constructing Daily Housing Price Indexes

Shuya Yang[1], Daniel Melser[1], Farshid Vahid[1], Silvio Contessi[2]

[1]Department of Econometrics and Business Statistics, Monash University
[2]Department of Banking and Finance, Monash University

July 11, 2022
Australian Conference of Economists 2022

Why a daily house price index is necessary

Housing price movements have been poorly measured in comparison with other asset classes;

Housing futures derivatives demands for more timely information on housing price movements (Bokhari and Geltner, 2012; Shiller, 2015; Fabozzi et al., 2020);

Lower-frequency housing price indexes impose challenges on information delivery effectiveness due to temporal aggregation bias (Calhoun et al., 1995; Melser and Hill, 2019).
$\Rightarrow$ During housing booms, prices can routinely move by 10% per quarter or 3% per month. (Corelogic, 2022)

Why a daily house price index is necessary

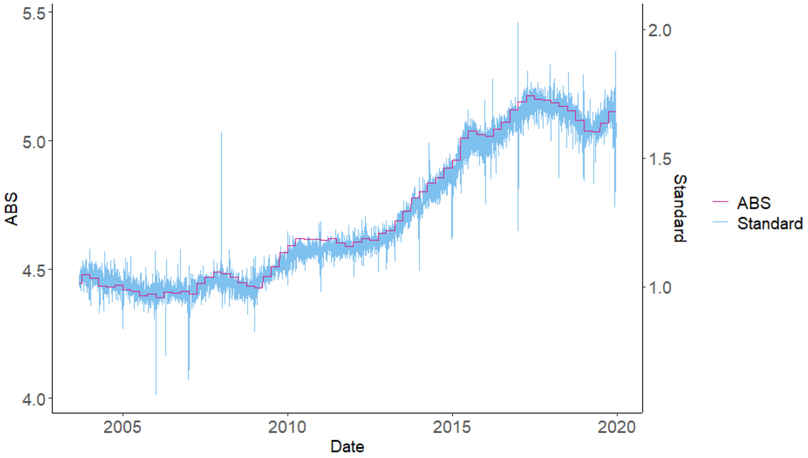


Figure 1: The comparison between the housing price index from the ABS and the estimated housing price indexes. Daily housing price indexes contains more information. (Sources: ABS; NSW Valuer General)

Objectives

1. Empirical Objective: Build a daily house price index that is of a higher frequency than currently available indexes. It will provides more timely information about the housing market dynamics.
2. Methodological Objective: Address the challenges imposed by estimating daily price index, for example, scarce and noisy data.

The repeat sales model

$$\ln\left(\frac{p_{it}}{p_{is}}\right) = \delta_t - \delta_s + \nu_{it}, \tag{1}$$

where

- $\ln p_{it}$ and $\ln p_{is}$ denote log price of property i at time t and time s, respectively, $t > s = 1, 2, \ldots, T-1$;
- δ_t and δ_s is the value of the logarithm of the market trend at time t and s.

The rationale for introducing the Flexible Regularized Repeat Sales Method

Equation (1) can be rewritten as

$$\mathbf{\Delta p} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\nu}, \tag{2}$$

where

$\boldsymbol{\delta}$ is a $T \times 1$ vector;

$\mathbf{Z}$ is an $N \times T$ matrix that records the transaction time for N pairs of repeat sales transactions.

The raw repeat sales index is based on $\hat{\boldsymbol{\delta}}$ (the least square estimates of $\boldsymbol{\delta}$).

As the frequency level increases, there are:

- more time dummies ($\times 4, 12, 52, 365$) $\Rightarrow$ lower efficiency
- more periods with scarce data $\Rightarrow$ measurement errors
- more rapid changes of prices $\Rightarrow$ more volatility

Existing methods in high-frequency housing price index

I **Two-step approach** first estimates $\hat{\delta}$, and then uses

- Frequency Conversion Approach (Bokhari and Geltner, 2012; Bourassa and Hoesli, 2017);
- Kalman Filter (Bollerslev et al., 2016).

II **One-step approach** re-parameterises the housing price index by a semi-parametric functional form

$$\boldsymbol{\delta} = \mathbf{H}\boldsymbol{\beta},$$

where **H** can be Fourier bases (McMillen and Dombrow, 2001; McMillen, 2003).

In this study, we introduce the use of:

- Spline bases, and
- Wavelet bases.

The Flexible Regularized Repeat Sales method

With a basis matrix **H**, we rewrite Equation 2 as:

$$\boldsymbol{\Delta}\mathbf{p} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\nu} = \mathbf{ZH}\boldsymbol{\beta} + \mathbf{e}, \tag{3}$$

where **H** is a basis matrix of dimensions $T \times J$. J denotes the size of the basis. $J = T$ for wavelet basis and $J = \frac{T}{2}$ for spline basis. This means that the basis will have a minimum wavelength of 2, allowing the capture of day-to-day movements.

The model is estimated by solving the following optimization problem,

$$\min_{\boldsymbol{\beta}} \frac{1}{2N} ||\boldsymbol{\Delta}\mathbf{p} - \mathbf{ZH}\boldsymbol{\beta}||_2^2 + \lambda \sum_{j=1}^{J} \left[\frac{(1-\alpha)}{2} \|\beta_j\|_2^2 + \alpha \|\beta_j\|_1 \right], \qquad \lambda \geq 0 \tag{4}$$

The estimated log house price index is

$$\hat{\boldsymbol{\delta}} = \mathbf{H}\hat{\boldsymbol{\beta}}. \tag{5}$$

The estimated daily housing price index (Fourier)

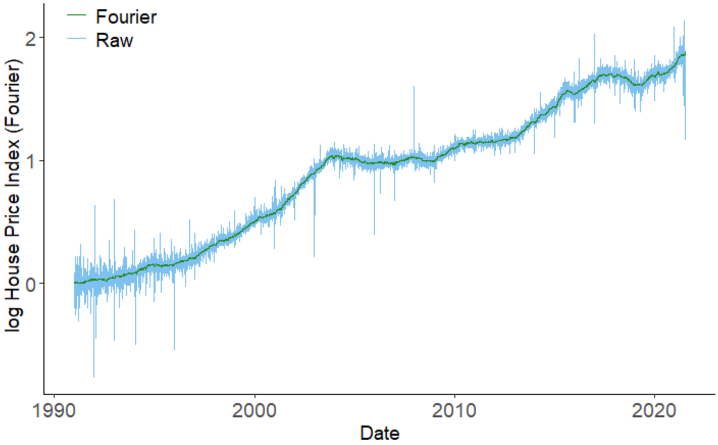


Figure 2: The comparison between FRRS Fourier housing price index and raw housing price index of Sydney (Source: NSW Valuer General)

The estimated daily housing price index (B-Spline)

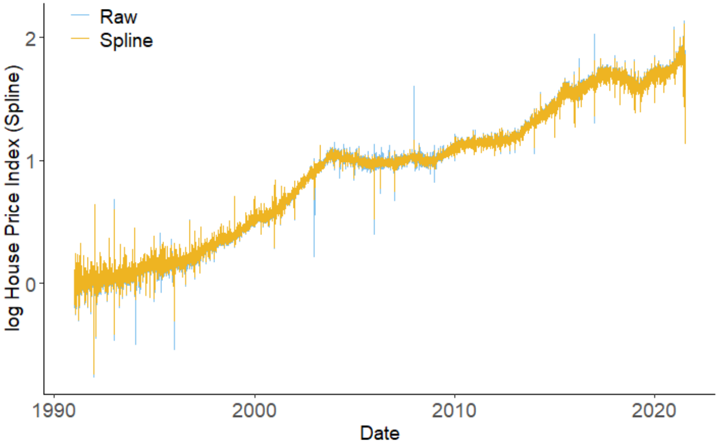


Figure 3: The comparison between FRRS B-Spline housing price index and raw housing price index of Sydney (Source: NSW Valuer General)

The estimated daily housing price index (Wavelet)

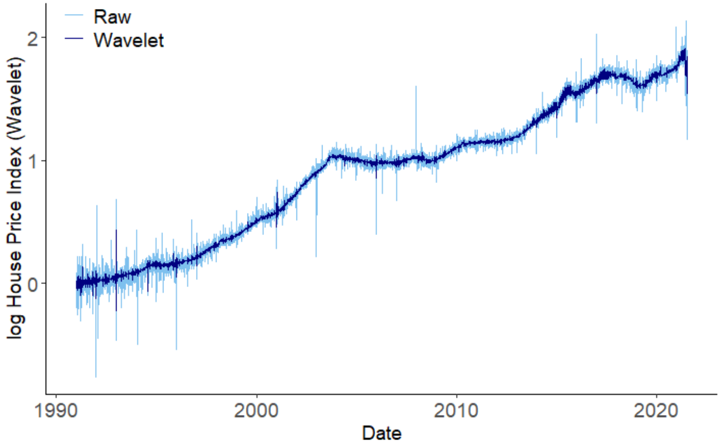


Figure 4: The comparison between FRRS Wavelet housing price index and raw housing price index of Sydney (Source: NSW Valuer General)

The Daily House Price Index provides localized smoothing feature

com2000

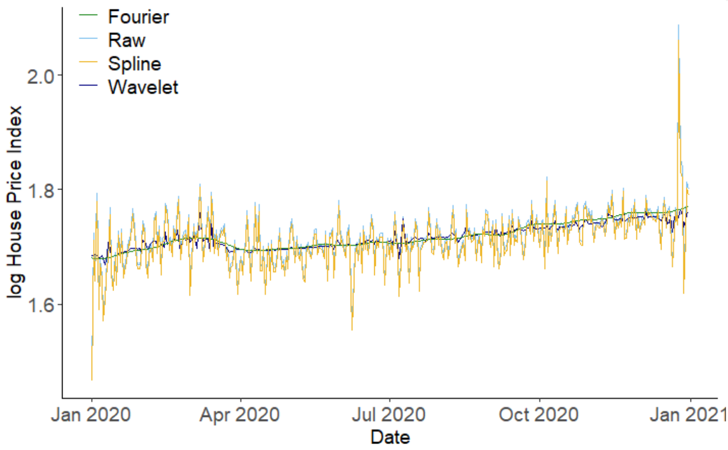


Figure 5: Comparison between methods in a subsample period in year 2020 (Source: NSW Valuer General)

Comparison of the stability of the methods

Table 1: K-Fold Cross Validation Results (K=60)

MSE	Raw RS	Fourier	Spline	Wavelet
Mean	0.0519	0.0512	0.0531	**0.0511**
Std. Dev.	0.0014	0.0014	0.0061	0.0014
MAE	Raw RS	Fourier	Spline	Wavelet
Mean	0.1553	0.15305	0.1561	**0.15300**
Std. Dev.	0.0019	0.0020	0.0067	0.0020

Comparison of the stability of the methods

We perform test if the information criteria are significantly different from each other: $H_0 : IC_{M1} = IC_{M2}$ vs $H_1 : IC_{M1} \neq IC_{M2}$ (Devore and Berk, 2012; Hill et al., 2020)

Table 2: P–Values of the Hypothesis Test of the Cross Validation Results

Model2 (M2)	Fourier		Spline		Wavelet	
Model1 (M1)	MAE	MSE	MAE	MSE	MAE	MSE
Raw	0.00	0.00	0.20	0.06	0.00	0.00
Fourier	—	—	0.00	0.01	0.44	0.34
Spline	—	—	—	—	0.00	0.01

Conclusion and Contribution

This study proposes and implements the Flexible Regularized Repeat Sales method that

- considers both localized effects and long term trends
- uses one-step estimation
- provides more accurate prediction of house price changes

The FRRS daily house price indexes for Sydney, Australia

- provide timely information.
- convey more information about abrupt changes that reflect short term market dynamics.

Potential Applications

The high frequency house price index should be of interest for policymakers and investors: policy implication, event study;

The FRRS method will also be useful when there are few transactions such that there is sparse data environment:

- price indexes for assets that are traded infrequently, for example, small-cap stocks, junk bonds and alternative assets;
- price indexes reconstruction for infrequently recorded assets, for example, securities traded outside the exchange hours, historical stock prices

Thank you!

The Fourier basis matrix

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0.71 & 0.71 & 1 & 0 & 0.71 & -0.71 & 0 & -1 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\ 0.71 & -0.71 & -1 & 0 & 0.71 & 0.71 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ -0.71 & -0.71 & 1 & 0 & -0.71 & 0.71 & 0 & -1 \\ -1 & 0 & 0 & -1 & 1 & 0 & 0 & 1 \\ -0.71 & 0.71 & -1 & 0 & -0.71 & -1.71 & 0 & -1 \end{bmatrix}. \tag{6}$$

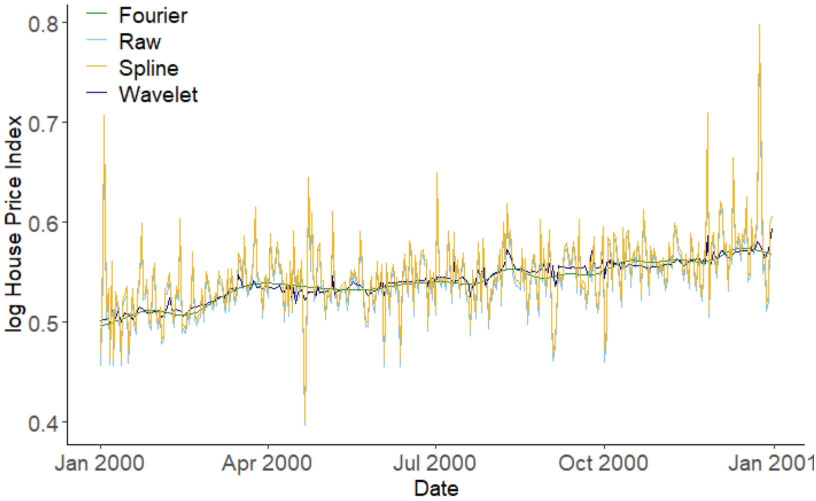
The B-spline basis matrix for $T = 8$ periods

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.05 & 0.54 & 0.36 & 0.04 & 0 & 0 & 0 & 0 \\ 0 & 0.11 & 0.58 & 0.32 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.24 & 0.65 & 0.11 & 0 & 0 & 0 \\ 0 & 0 & 0.02 & 0.48 & 0.47 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0.11 & 0.59 & 0.30 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.18 & 0.60 & 0.22 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

The Daubechies D4 wavelet basis matrix for $T = 8$ periods

$$\boldsymbol{D} = \begin{bmatrix} -0.13 & 0 & 0 & 0 & -0.17 & 0 & -0.06 & 0.35 \\ -0.22 & 0 & 0 & 0 & -0.05 & 0 & 0.24 & 0.35 \\ 0.84 & -0.13 & 0 & 0 & -0.14 & 0 & 0.37 & 0.35 \\ -0.48 & -0.22 & 0 & 0 & -0.17 & 0 & 0.55 & 0.35 \\ 0 & 0.84 & -0.13 & 0 & 0.35 & -0.17 & 0.06 & 0.35 \\ 0 & -0.48 & -0.22 & 0 & 0.73 & -0.05 & -0.24 & 0.35 \\ 0 & 0 & 0.84 & -0.13 & -0.05 & -0.14 & -0.37 & 0.35 \\ 0 & 0 & -0.48 & -0.22 & -0.51 & -0.17 & -0.55 & 0.35 \end{bmatrix}. \tag{8}$$

Appendix: The Daily House Price Index provides localized smoothing feature ▸ com2020



Appendix: Elastic Net Estimation Results

Table 3: Elastic Net Estimation Results

		Fourier	Spline	Wavelet
α=0	λ	0.0165	13.2355	0.0643
	$\lvert\hat{\beta}_{+-}\rvert$	802	11154	11189
α=1	λ	0.0002	0.0000	0.0005
	$\lvert\hat{\beta}_{+-}\rvert$	431	10713	2088
$\alpha_{optimal}$	λ	0.0008	0.0000	0.0005
	$\lvert\hat{\beta}_{+-}\rvert$	428	10713	2086

Note: 1. $\alpha_{optimal}$ provides lowest MSE among possible α values between 0 and 1 with an incremental increase of 0.1. The optimal value of α for Fourier is 0.3 and it is 1 and 0.9 for spline and wavelet method, respectively.

2. $|\hat{\beta}_{+-}|$ records the number of non-zero estimated values of