Neighborhood housing rent index construction and spatial discontinuity: a machine learning approach

Yao-pei Wang， Shanghai University of Finance and Economics[1]

Yong Tu, National University of Singapore[2]

Abstract:

We integrate a hedonic housing rent model (econometric approach) into a state-space model (reinforcement machine learning approach). We adopt the kalman filter and smoother recursive algorithm and the expectation maximization algorithm (statistical estimation methods) to estimate the proposed state-space housing rent hedonic model. The method is applied to the Singapore public open rental housing market to construct housing rent indexes. Compared with the conventional econometric methods in index construction, the proposed model has three advantages. Firstly, a state-space modeling approach technically allows us to construct neighborhood level housing rent indexes through a reinforcement learning process regardless the sample size in a neighborhood. Secondly, the expectation maximization algorithm effectively enhances the robustness of maximum likelihood estimation for a dataset being repleted with unobservable information, for example, fewer or zero transactions in certain time periods. Thirdly Kalman filter and smoother recursive algorithm optimizes the estimates by capturing all information (before and after a time point) to predict a housing rent at a time point. This helps reduce the bias caused by sticky rents. The paper empirically proves that the proposed model outperforms other types of index models in prediction accuracy, hence produces more accurate housnig rent indexes at neighborhood level.

Accurately constructing neighborhood housing rent indexes are impotant in real estate valuation, real estate investment returns and risk analyses. This is because the spatial patterns of housing price distribution may change over time, which is resulted from urban developments. To illustrate it, we apply K-shape clustering algorithm in unsupervised machine learning literature to the neighborhood housing rent indexes to analyze the dynamic patterns of the spatial distribution of housing rents. We find the spatial discontinuity of housing rent dynamics. The housing rent indexes in some spatially disconnected neighborhoods appear to have similar dynamic pattern, while different dynamic patterns are found in some spatially adjacent neighbothoods.

---

[1] School of Public Economics and Administration， Shanghai University of Finance and Economics， Shanghai， 200433， China

[2] Department of Real Estate, School of Design and Environment, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore.

Key words : *housing rent Index; Price dynamic; State-Space Model; K-shape Clustering*

JEL Classification: *C43 ; E30; C32; C38*

Acknowledgement: the authors would like to thank the NUS Academic Research Fund R – 297-000-136-115.

1. Introduction

A housing price index construction method derived from the conventional econometric approaches often claims that it can capture certain information (Sun et al, 2005; Tu et al, 2007) or correct certain biases, such as, non-constant quality bias (Wu et al, 2014), recall bias (Crone, 2001) or the vacancy component of nonresponse bias (Ambrose et al, 2015). Thus, the proposed method can produce a more accurate index number. This is not always true as a new method often creates a new problem after correcting an old one (Eichholtz et al, 2012).

The classic hedonic housing price model conceptualizes a quality-constant index (Wyngarden, 1927). The disapproval is some hedonic variables, especially those fine-grained spatial and temporal information, are often unmeasurable. Repeat-sales method (Shiller,1991) reduces the impacts of omitted hedonic variables, but giving rise to the sample selection bias (Meese and Wallace, 1997) and the inconsistent estimates of indexes because some temporal variations are left in the residuals (Giannetti Antoine, 2018). The locally weighted quantile model, which is an advanced spatial statistical model, can demonstrate the impacts of hidden information by showing the distributional changes of housing prices overtime (McMillen, 2014). However, the application of the model to constructing neighborbood housing indexes is questioned due to the reduction of sample size at neighborhood level and the uncaptured spatial and temporal information which differentiates neighborhoods. Furthermore, the exsiting housing price index construction methods are mainly derived from economietric approaches, requiring both random sample with an acceptable sample size and the satisfcations of model assumptions. The machine learning approach neither requires excessive hypothesis tests under a sample deistribution assumption nor the minimum sample size. The approach makes no assumptions about the causal relationships among variables and the estimated model is less sensitive to an institutional environment on predictability (Grimmer,2015). These advantages motivate us to ask how a machine learing approach may help improve the accuracy of neighborhood housing price index construction.

The current housing rent index construction methods have addressed but are unable to effectively correct the bias arising from the sticky rent. Eichholtz et al. (2012) studied the sticky rent problem and identified a temporal aggregation bias in housing rent index construction using the re-contract sample established by checking on the rent contract dates. However, the method may also give rise to the sample selection bias. Ambrose et al. (2015) compared the move-in date and the date of the first rental payment to distinguish new tenants from renewal tenants to resolve the sticky rent problem. Re-contract rents can be higher (a tenant may accept a higher rent if he is unwilling to more) or lower (a landloard may offer a lower rent if he wants to keep a good tenant) than a market rent. This may cause biases in rent index construction. Physically, we may see excessive spikes or bumps in an index series. Differentiating new rents from recontract rents is practically difficult and time consuming. How can we mitigate the impacts of sticky rents?

It is observed that housing market structure evolves over time. Literature has attempted to identify housing market spatial pattern at a time point (Keskin and Watkins 2017) or to explore the driving forces behind the pattern changes (Jeanty, Partridge, Irwin 2010). Little research has been done to uncover the dynamics of a housing market spatial structure. This is important because the Neil Smith's rent-gap hypothesis (Boulay 2012) points out that investment in real estate will only be made if a rent gap exists over time. One of the advantages in machine learning approaches is to identify patterns. This motivates us to adopt maching learning clustering method to demonstrate the dynamic patterns of neighborhood housing rent indexes.

To answer the above questions, this paper extends the work of Ren, Fox and Bruce (2017) in two ways. Firstly we integrate a hedonic housing rent model (econometrics) into a state-space model (reinforcement machine learning). Secondly, we adopt and clearly explain how to apply the kalman filter and smoother recursive (KR) algorithm (Kalman 1960; Rauch et al., 1965) and the expectation maximization (EM) algorithm (Holmes, 2012) to statistically improve the estimation of a state-space housing rent hedonic model (SSH_KREM model). Using this model, we construct housing rent indexes at pre-defined neighborhood level. The SSH_KREM model is applied to the Singapore public open rental housing market, where neighborhoods are highly homogeneous and only fine-grained spatial temporal information can differentiate them, yet they are often unmeasurable.

A state-space model originates from the reinforcement machine learning literature (Akaike 1976). It was first used in the field of control engineering (Kalman 1960). In the recent years, it is used for modeling housing prices (Schulz and Werwatz 2004; Zhang, Hudson and Manahov 2015; Ren, Fox and Bruce 2017). "State" represents a time series (such as a series of housing rent indexes) in a "Space" or a "system", (it is a neighborhood in this paper). The state at a time point (a housing rent in this paper) is determined by a set of variables associated with the "Space". A state-space model estimates and predicts the "State" at any time point through a reinforcement machine learning process within the "Space" regardless of the sample size and sample distribution. This technically makes neighborhood housing rent index construction feasible.

The Singapore public open rental housing market at neighborhood level are typically a thin market. In certain time periods, we may get fewer or even zero transactions. To minimize the impacts of sample size on estimation, we adopt the expectation maximization algorithm. The EM statistical estimation algorithm can effectively enhances the robustness of maximum likelihood estimation for a dataset being repleted with scarce observations or missing values.

In a rental housing dataset, a housing unit is typically associated with both new rents and recontract rents. The latter gives rise to an estimation bias called as sticky rent problem. The Kalman filter and smoother recursive algorithm, firstly used in the Apollo spacecraft's navigation computer (Kalman ,1960), can optimize the estimates by capturing all information (before and after a time point) to predict a “State” (a housing rent in this paper) at a time point. The estimation approach, to certain degree, reconciles the difference between new rents and recontrcat rents, mitigating the impacts of sticky rent problem on rent prediction.

At last, the paper applies a K-shape clustering algorithm introduced in machine learning literature to the neighborhood housing rent indexes to analyze the dynamics of the spatial distribution of neighborhood housing rents. Based on it, we classifiy the neighborhoods into groups within which the neighbrohoods share similar pattern of housing price dynamics, while across the groups, there are larger variations in the patterns of housing rent dynamics. This exercise provides insightful information to both the public authority and housing investors to estimate the investment returns and risk analyses.

The main findings are that the SSH_KREM housing rent indexes outperform the conventional rent indexes because the model captures more unobserved or unmeasured information as well as use more information embedded in a dataset as compared to a conventional model. We also find the spatial discontinuity of housing rent dynamics. The housing rent indexes in some spatially disconnected neighborhoods appear to have similar dynamic pattern, while different dynamic patterns are found in some spatially adjacent neighbothoods. The findings have meanful implications to real estate valuation, real estate investment returns and risk analyses.

The next section summarizes the related literature. The data and methodology are introduced in Section 3. Empirical findings are reported in Section 4. Section 5 concludes.

2. Literature Review

Three streams of housing price or rent index construction methods are reviewed. This is followed by a review of state-space modelling approach, Klaman filter and smoother recursive and expectation maximixation algorithms.

One stream of housing price index construction literature follows the concept of hedonic housing price theory. Rosen (1974) proposed the theory of hedonic housing price and illustrated how housing prices change over time after controlling the qualify changes of a housing unit. The weaknesses are that the hedonic housing price theory cannot determine the true functional form and we are not able to obtain a full set of hedonic characteristics. In addition, a hedonic housing price model often assumes that a hedonic coefficient is constant across time, space and housing units. Meese and

Wallance (1994) extended a hedonic housing price model to a locally weighted hedonic regression, which allowed varying coefficients. They proved that the model could better predict housing prices. McMillen (2012) proposed a matching estimator, which was applied to the Singapore residential and commercial housing markets by Deng et al. (2012, 2014). Brunauer et al. (2013) presented a multilevel structured additive regression model by leveraging the hierarchical structure of neighborhood attributes and house-level hedonics. These methods have improved the prediction accuracy, but are still unable to resolve the problems of uncaptured variables and the problem of functional forms.

The second stream of literature follows the repeat-sales modeling approach, proposed by Bailey et al. (1973). A repeat-sales housing price index construction method may, to certain degree, avoid the impacts of omitted variables but give rise to the problem of sample selection bias. Besides, the approach assumes that the quality of housing attributes are constant over time. However, housing quality changes with the age and the renovations of a housing unit. Case and Shiller (1987, 1989) extended a repeat-sales model to a weighted repeat-sales model, attempting to correct heteroscedastic errors. Gatzlaff and Haurin (1997) developed a repeat-sales model to reduce the baises caused by sale frequencies. Shiller (1991) and Goetzmann and Peng (2002) proposed an arithmetic average repeat-sales estimator to replace the original geometric average estimator. The method is adopted by Standard and Poors to produce the S&P/Case–Shiller Home Price Index. Nagaraja et al (2011) proposed an autoregressive repeat-sales method using all sales' information but without engaging any hedonic information. Repeat-sales method typically requires a relatively large sample including sufficient number of repeated transactions, which constrains its application to neighborhood housing price index construction.

The third stream of literature adopts the spatio-temporal modeling approach. Holly et. al. (2010) applied a Pesaran common correlated effect estimator to a panel dataset. They explored the interactions between geographical proximity and unobserved common factors. Baltagi et al. (2014) proved that the method produced robust findings and MÁRCIO et al. (2016) further used Bayesian estimation method to improve the model estimation. However, these models have an inherent problem that we don't know how time and space interacts to determine a housing price, thus, the relation function between time and space is often assumed. A space has two dimentions but can have unlimited directionality, such as, from north to south or from west to east, or any direction in between, while time is unidimensional and can only move in one direction which is "FORWARD". Thus the interactions between time and space are too complicated to be modelled.

Housing rent index construction typically adopts the methods used in housing price index constructions. Wheaton et al. (1994) identified a pure rent and pointed out that furniture, water and electricity bills differentiated an observed rent from its pure rent. Margo (1996) argued that the houses with furniture charge significant higher rents than the houses without furniture. Webb and Fisher (1996) estimated efficient rent defined as the annual-equivalent cash flows of the present value of all cash flows that are explicitly identified in the lease contracts. Eichholtz et al. (2012) addressed temporal aggregation bias and sticky rent problems using the re-contract rental housing transaction sample established by checking the contract dates. However, the method may also give rise to sample selection bias. Ambrose et al. (2015) compared the move-in date and the date of the first rental payment to distinguish new tenants from renewed tenants to solve sticky rent problem. An et al. (2016) engaged a panel dataset to construct commercial real estate rent indexes by adopting an autoregressive model with age adjustment. Hu et al. (2019) incorporated a machine-learning method into the hedonic housing price approach to predict rents but addressed little on what and why we can benefit from the method in rent index construction.

A state-space model belongs to the family of reinforcement learning in machine learning literature. It is different from the supervised and unsupervised learnings (Chart A_1 in the Appd) that don't help build indexes. Unsupervised learning helps get better sensory cognition but don't produce coefficients, while supervised learning, such as, random forest, can estimate coefficients but without economic meanings. A state-space model has increasingly attracted the attentions from urban researchers. Schulz and Werwatz (2004) developed a state-space model guided by asset pricing theory to estimate the investment returns of single-family houses. Zhang, Hudson and Manahov (2015) proposed a relative valuation approach to quantify a bubble in housing by incorporating the housing user cost into a state-space model. Ren, Fox and Bruce (2017) constructed a hybrid model of a state-space model and a hedonic model to provide more accurate predictions of housing prices. Tao et al. (2018) adopted it to analyze the dynamic determinants of a time series. Katagiri (2018) used it to investigate the developments of housing price synchronization across the countries.

The present paper enriches the literature in two ways. It proposes a SSH_KREM model and applies it to construct neighborhood rent indexes. The model outperforms the conventional econometric housing rent index contruction models. It applies K-shape clustering method to the constructed neighborhood housing rent indexes and explores the spatial dynamics of neighborhood rental housing market.

3.Data Collection and Research Design

3.1 Data Collection

The working dataset is drawn from the SRX that is the publically available real estate transaction database in Singapore. The distance information is self-calculated using GIS.The descriptive statistics and variables are given in Table 1. The “HDB” stands for the Housing Development Board in Singapore. This is the Singapore public housing authority providing fully subsidized rental housing to low income

Singaporean households, partially subsidized owner occupied housing to Singaporeans or permanent residents and non subsidized rental housing to all through the HDB open rental housing market.

TABLE 1.a
Descriptive Statistics of the Singapore HDB open rental housing transactions between 1st May 2006 and 30th April 2018

Variable	Obs	Mean	Std.Dev.	Min	Max
Pub_Rent (S$)	186,740	2165.241	441.366	250	10000
Size (m^2)	186,740	94.42846	28.22823	8	1636
Floor (storey)	186,740	8.136639	4.894714	1	80
Age (year)	186,740	25.51478	10.57679	0	52
Dis_Bus_Interchange (m)	186,740	1292.333	773.3795	54.51022	4222.508
Dis_CBDCentral (m)	186,740	11293.3	4592.27	578.9364	20060.56
Dis_Hospital (m)	186,740	5406.633	3598.582	29.13114	14423.77
Dis_MRT (m)	186,740	613.0877	385.9335	21.93404	2133.387
Dis_Park (m)	186,740	356.6549	243.757	4.72818	1396.32
Dis_Shoppingmall (m)	186,740	802.4818	463.5725	5.340025	2579.894
Rental transacts (times)	186,740	4.146	2.719	1	24

Source: SRX and the geographic information are calculated by author by GIS

TABLE 1.b
Defination of variables in TABLE 1.a

Variable	Defination
Pub_Rent	Monthly rent of a housing unit in Singapore dollar (S$).
Size	The size of a housing unit (m^2).
Floor	The floor level of a housing unit locates at (storey).
Age	The age of a housing unit at a transaction date (Year).
Dis_Bus_Interchange	Distance to nearest bus interchange (m).
Dis_CBDCentral	Distance to the CBD (m).
Dis_Hospital	Distance to the nearst hospital (m).
Dis_MRT	Distance to the nearest Mass rail transit station (m).
Dis_Park	Distance to the nearest park (m).
Dis_Shoppingmall	Distance to the nearest shopping mall (m).
Rental transacts(times)	How many time a housing unit has been transacted during the period (tmes).

It is noted that in the HDB open rental housing market, a rental contract is pro-landlord and the rental period of each contract should be longer than 6 monthes. A rental contract is typically renewed yearly. Sticky rent problem is expected to be serious. In a HDB housing estate, housing units and neighborhoods are highly homogenous, often differentiated by fine-grained spatial factors after a location is controlled.

In this paper, we construct the HDB neighborhood housing rent indexes. We divide the HDB housing estatets into 123 neighborhoods using the first three digits of a postcode. The neighborhood sample size varies from 64 to 10,680 (Table A in the Appendix Table A gives more details). We then apply a SSH-KREM model (see section 3.2) to estimate the state variable for each neighbornood, generating a series of neighborhood indexes.

3.2 Research Design

The estimation procedure of the proposed SSH_KREM model is introduced first. This is followed by an introduction of the K-shape clustering method. At last we present nine randomly selected neighborhoods which are used to demonstrate the empirical results.

3.2.1 A state-space hedonic model with kalman filter and smoother and expectation maximization algorithms (SSH_KREM model)

The proposed state-space hedonic housing rent model with the kalman filter and smoother recursive (KR) algorithm and the expectation maximization (EM) algorithm (SSH_KREM model) is an extention of Ren, Fox and Bruce (2017). We firstly incorporate a hedonic housing rent model into a state-space model. The advantage of adopting reinforcement machine learning approach is discussed. We secondly adopt the KR and EM algorithms to estimate the model as well as carefully justify the advantages of uing these algorithms in constructing housing rent indexes.

Assuming that a SSH_KREM model is applied to a set of rental housing transactions in neighborhood i to construct neighborhood housing rent indexes.

In the first step, a hedonic model is estimated using equation 3.2.1. $y_{i,j,t}$ indicates a housing rent for housing unit *j* in neighborhood *i* at time *t*, *t* = 1,2,3,…,*T*. *T* is equal to 156, representing 156 months between May 2006 and April 2018. The exogenous hedonic attributes are represented by a vector of $Z_{n,j,t}$. N indicates the total number of hedonic attributes and n indicates a hedonic attribute. η is residual, following the conventional assumetions in a linear hedonic model and β is a vector of coefficients. Let $\tilde{y}_{i,j,t}$ indicate $\sum_{n=1}^{N} \beta_{i,n} Z_{n,j,t}$.

$$y_{i,j,t} = \sum_{n=1}^{N} \beta_{i,n} Z_{n,j,t} + \eta_{i,j,t} \tag{3.2.1}$$

In the second step, we estimate the proposed state-space model. It contains *the Transition Equation (3.2.2)* which is the training equation and *the Measurement Equation (3.2.3)* which is the reinforcement learning equation. The model has two assumptions, a) the mean value of the initial state, $\tilde{x}_{0,i}$, is an input parameter with the default value of zero; b) in any time, the residual terms $\varepsilon_{t,i}$ and $v_{t,i,l}$ are independent of each other, and they are not related to the initial state $\tilde{x}_{0,i}$. In both equations, we also assume the residuals are normally distributed. But in machine learning approach, this assumption will not affect the iteration to converge. $\tilde{x}_{t,i}$ in Transition Equation 3.2.2 is a latent variable or a hidden state.

The difference between the observed, $y_{i,j,t}$, and the estimated, $\tilde{y}_{i,j,t}$, is $y_{i,j,t}$-$\tilde{y}_{i,j,t}$, which contains all the information in a space (neighborhood i), and, is used as the dependent variable in *Measurement Equation* 3.2.3 . Through a reinforcement deep learning process using all the information in the space, the equation helps improve the estimation of hidden state $\tilde{x}_{t,i}$, generated by the training equation of *Transition Equation* 3.2.2. Then, the improved estimate of hidden state $\tilde{x}_{t,i}$ is brought back to Transition equation (3.2.2.) to get the hidden state of $\tilde{x}_{t+1,i}$. The process continutes till we estimate all hidden states, obtaining a series of housing rent indexes for neighborhood i.

$$\tilde{x}_{t,i} = a_i \tilde{x}_{t-1,i} + \varepsilon_{t,i}, \qquad \varepsilon_{t,i} \sim \mathcal{N}(0, {\sigma_i}^2) \tag{3.2.2}$$

$$y_{t,i,j} - \tilde{y}_{t,i,j} = b_t \tilde{x}_{t,i} + v_{t,i,j}, \qquad v_{t,i,j} \sim \mathcal{N}(0, {R_i}^2) \tag{3.2.3}$$

The estimated states, $\tilde{x}_{t,i}$, t=1,2,…T, directly capture the dynamics of housing rents n neighborhood i, giving rise to neighborhood housing rent indexes. Therefore, a state-space hedonic housing rent model allows us to construct neighborhood housing rent indexes regardless the sample size and distribution, overcoming the drawbacks of econometric rent index construction methods.

The calibration of equations 3.2.2 and 3.2.3 involves the estimation of the hyperparameters of $b_t \; and \; R_i$ in *Measurement Equation* and a_i, σ_i in *Transition Equation*, represented by Φ and the hidden state $\tilde{x}_{t,i}, t = 1, 2 \ldots T$, for neighborhood i. The reinforcement learning process is made possible using the Kalman filter and smoother algorithm (KR) and the expectation maximization algorithm (EM). The EM estimation involves a two-step's iteration process. In the first step of each EM iteration, the KR algorithm is used to estimate the hidden state $\tilde{x}_{t,i}$. KR is also an iteration process till $\tilde{x}_{t,i}$ converges. The estimation moves to the second step of the EM estimation to get an estimate of Φ. Then, the iteration goes back the first step, till Φ converges as well as $\tilde{x}_{t,i}$ converges, for t=1,2,…T (see Table A_2 in Appd).

Through the iteration process, the EM is able to produce robust maximum likelihood estimates for the data with unobserved samples. This can happen if there are few housing rental transactions in certain months in a neighborhood. This is also the advantage of using KM algorithm.

KR algothrim includes two estimation procedures: the Kalman filter (KF) and kalman smoother (KS). KF gives the optimal (lowest mean square error) estimate of the hidden state $\tilde{x}_{t,i}$, using the observed data up to time t. The KS also gives the optimal (lowest mean square error) estimate of the hidden state $\tilde{x}_{t,i}$, but using all information before and after time t. When both Kalman filter and smoother estimations converges, the hidden state $\tilde{x}_{t,i}$ is estimated. It is an optimal estimate and orthogonal to $v_{t,i,j}$ and $\varepsilon_{t,i}$.

KR algorithm helps mitigate the estimation bias caused by sticky rents. For example, if a rental housing unit has two new rents and a few recontract rents between the two new rents, the KS algorithm will use both two new rents and all recontract rents to predict a rent at a time point. To certain degree, it reduces the bias caused by recontract rents (sticky rents).

As for model diagnosis, K-fold cross-validation (CV) is often used to justify the model fit and to compare and select a best model among several machine learning models. CV is easy to be understood and implemented and has a lower bias than other methods (Geisser, 1993; James, 2013)

At last, we put the estimated states $\tilde{x}_{t,i}$ into the hedonic regression (3.2.1) to predict a SSH_KREM housing rent $y_{i,j,t}^{pred}$:

$$y_{i,j,t}^{pred} = \sum_{n=1}^{N} \beta_{i,n} Z_{n,j,t} + \tilde{x}_{t,i} \qquad (3.2.4)$$

In summary, the proposed SSH_KREM model has three advantages in housing rent index construction. A state-space process directly estimates rent index series for a neighborhood regardless of the sameple size and sample distribution in the neighborhood. The EM algorithm ensures that the estimation is robust in the presence of missing observations (when market is thin). The KR algorithm produces optimal estimates and a rent is predicted using full set of information before and after a time point t, mitigating the impacts of sticky rent.

A disadvantage of a SSH_KREM model is that it cannot catch the missing information at unit level, for example, the impacts of furniture or renovation on a rent. Besides, any problems associated with missing data in a variable or data inaccuracy are not solved by the model.

3.3 Identifying the dynamics of rental housing market structure: k-shape clustering

K-shape clustering method is a noval algorithm in time series clustering. It replies on a scalable iterative refinement procedure to create well separate clusters. The time series in the same cluster have similar shapes while, across the clusters, they have different shapes. In order to compare the shapes of two time series when clustering, K-shape adopts a normalized version of the cross-correlation statistical measure to compute cluster centroids, then to update the members of each cluster using the centroids. This is the scalable iterative refinement procedure (Yang et al, 2017), which is illustrated by eq. 3.3.1

$$\mathrm{SBD}(\vec{m},\vec{n}) = 1 - \max\left(\frac{CC_w(\vec{m},\vec{n})}{\sqrt{D_0(\vec{m},\vec{m}) * D_0(\vec{n},\vec{n})}}\right) \qquad (3.3.1)$$

Where

$$CC_w(\vec{m},\vec{n}) = D_{w-T}(\vec{m},\vec{n}) \quad w \in (1,2\,, \ldots\ldots,\ \ 2T-1)$$

And

$$D_k(\vec{m},\vec{n}) = \begin{cases} \sum_{l=1}^{T-k} m_{l+k} \cdot n_l\,, k \geq 0 \\ D_{-k}(\vec{n},\vec{m}), k < 0 \end{cases}$$

Eq. 3.3.1 gives the shape based distance (SBD) between two time series of $\vec{m} = (m_1, \ldots, m_T)$ and $\vec{n} = (n_1, \ldots, n_T)$. Its values fall between -2 and 2, with 0 indicating perfect similarity between two time series. It is used to update the cluster memberships.

Cross-correlation is a measure of similarity for time-lagged signals that is widely used for signal and image processing. $D_k(\vec{m},\vec{n})$ demonstrates the original explanation function of Cross-correlation measurement. $CC_w(\vec{m},\vec{n})$, measures the shape similarity of the two time series, $D_0(\vec{m},\vec{m})$ is the matric norm of time series $\vec{m}$, which is used to nomalised the cross-correlation. A position "w" is found when $\frac{CC_w(\vec{m},\vec{n})}{\sqrt{D_0(\vec{m},\vec{m}) * D_0(\vec{n},\vec{n})}}$ is maximized, which means we move the time series of $\vec{m}$ when time point w has the same position as time point t in $\vec{n}$, the shapes of $\vec{m},\vec{n}$ is most similar. This calculate help us solve the misorientation in time series.

Applying K-shape algorithm to neighborhood housing rent indexes, we can identify the neighborhoods which rent indexes have similar dynamic patterns, indicating the renal housing markets in these neighborhoods have experienced similar transformations. We can also identify the neighborhoods which have had different dynamic patterns.

The methodology presented in this section is illustrated by Chart A_2 in the appendix.

3.4 Identifying neighborhoods

In the Singapore HDB open rental hosuing market, each building has its own 6-digit postcode. The buildings are grouped into a neighborhood if their first three digits of postcodes are the same. The definition ensures that each neighborhood is geographically enclosed. We obtain 123 neighborhhods (or postal sectors). The size of a neighborhood is approximately 1 square kilometer in diameter on average. The full smaple is then divided into 123 rental housing transaction subsamples between May 2006 and Aprirl 2018.

We applied the SSH_KREM model to each of 123 neighborhhods to consrtuct neighborhhod indexes. To illustate the results, we choose 9 neighborhoods from 9 different HDB Towns in Singapore as shown in Figure 1.

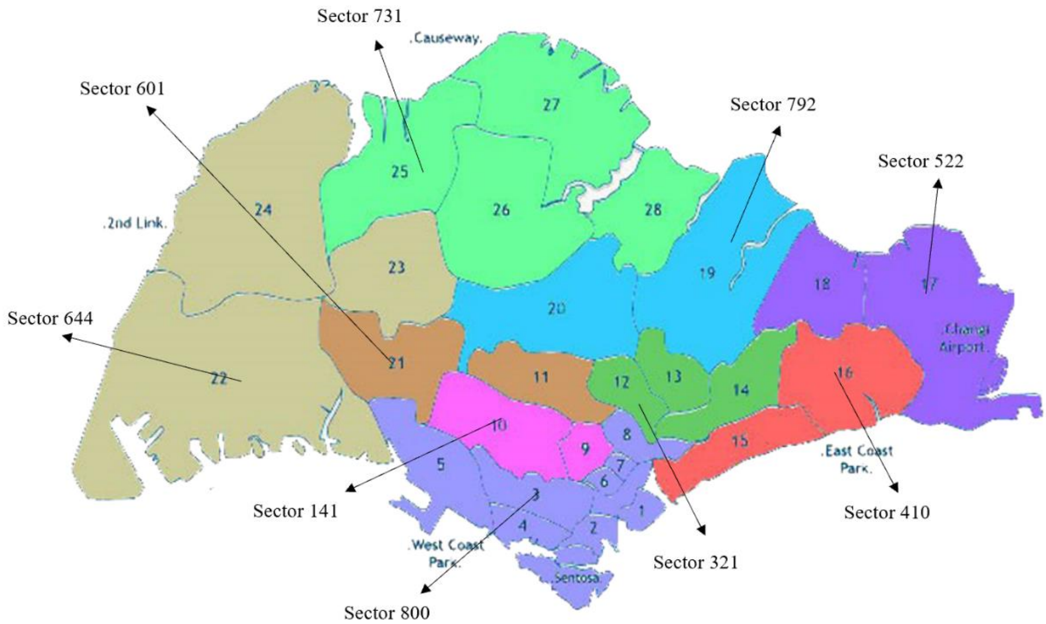


FIG. 1. The selected 9 neighborhoods (postal sectors) in Singapore.

4. Empirical Results

We firstly compare the model fits and predictability using a hedonic model, a Locally weighted regression, a state-space model and the proposed SSH_KREM model in Section 4.1. We then present the national and neighborhood indexes for the Singapore HDB open rental housing market (Section 4.2). Finally, we demonstrate the results of k-shape clustering in Section 4.3

4.1 The Evaluation of SSH_KREM Model

Theoretically, a SSH_KREM model is free from selection bias, sticky problem and partly solves the problem caused by unobserved samples. Thus, it should outperform the exsiting econometrically based index construction models in model fits and predictability. To empirically prove it, a hedonic linear regression, a locally weighted regression and a state-space model without EM and KR algorithms are chosen to compare with the SSH_KREM model.

More specifically, a hedonic regression is treated as the benchmark. A locally weighted regression (LWR) is a noval spatial model which allows us to estimate varying coefficients across space . A boostrap mothed is sued to decide the optimal bandwide of LWR and the Generalized Linear Model (GLM) is used to estimate the model. A state-space model is estimated by the quasi-Newton method.

Table 2 reports the K-fold cross-validation (k = 5) results of the estimated models. The CV errors are measured by “Root Mean Squared Error (RMSE)” and “Mean Absolute Error (MAE)”. MAE is generally smaller than RMSE. It is noted that using the square in RMSE makes the measure substantially larger although the errors may be small and acceptable, and, MAE is the most natural diagnostic measure (Willmott and Matsuura 1995).

TABLE 2

Model Comparisons by Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)

		(1)	(2)	(3)	(4)	(5)
		Hedonic	Hedonic+time	LWR	State-Space	SS_KREM
N_1	RMSE	397.75	379.18	395.13	326.91	309.92
	MAE	263.79	287.04	285.83	223.14	213.90
N_2	RMSE	355.52	427.61	341.99	273.29	251.03
	MAE	269.51	257.61	244.13	202.60	186.63
N_3	RMSE	368.66	342.44	357.37	279.76	262.07
	MAE	286.69	252.06	251.73	212.03	197.68
N_4	RMSE	341.74	548.23	346.00	244.04	210.01
	MAE	354.32	266.62	263.43	187.15	158.82
N_5	RMSE	344.30	605.25	323.23	273.25	248.13
	MAE	276.94	401.70	240.31	209.92	183.88
N_6	RMSE	325.50	275.48	315.90	249.73	217.08
	MAE	260.18	205.00	245.49	190.88	166.79
N_7	RMSE	292.82	291.13	293.41	220.98	208.45
	MAE	233.04	192.62	227.75	162.54	151.07
N_8	RMSE	291.42	262.36	282.32	190.49	165.79
	MAE	241.88	241.88	209.49	140.61	122.93
N_9	RMSE	419.18	423.23	407.79	369.59	353.98
	MAE	327.27	321.87	311.42	274.72	257.71

Note:

1. Column_1 “Hedonic” means just use hedonic attributes like age, floor, size;
2. Column_2: “Hedonic+time” means hedonic attributes and monthly time dummies;

3. Column_3: "LWR" means locally weighted regression;
4. Column_4: "State-space" means state-space model without EM and KR algorithm;
5. And N_1, N_2… N_9 represent 9 neighborhoods. N_1 indicate "Sector 141", "Sector 321", "Sector 410", "Sector 522", "Sector 601", "Sector 644", "Sector 731", "Sector 792", "Sector 800"(see Fig 1).

Table 2 shows that the CV errors of SSH_KREM is consistently the lowest in both RMSE and MAE measures. Generally speaking, SSH_KREM model reduces more than 30% prediction error when comparing it with the hedonic regression. The KR and EM algorithms contribute 10% reduction in prediction errors. SSH_KREM model performs well in retrieving both spatial and temporal information from the exsting sample, so it outperforms the other methods.

To examinate if the state variable captures any hidden information. We add the state variables $\{\tilde{x}_{t,i}\}$ into a hedonic model. Table 3b shows the estimations of state-space hedonic housing rent models in 9 neighborhoods. Comparing the results in Table 3a and 3b, the adjusted R^2 across 9 neighborhoods are dramatically improved. On average, R^2 rises from around 0.3 to around 0.6. Furthermore, the Standard Error of Mean (SEM) of each coefficient is reduced, which proves that the residuals are significantly decreased after adding the state variables. The findings are consistent for the rest of the 123 neighborhoods and the results are available at request.

TABLE 3a

Classic hedonic housing rent regressions in the 9 neighborhoods

	(1) N_1	(2) N_2	(3) N_3	(4) N_4	(5) N_5	(6) N_6	(7) N_7	(8) N_8	(9) N_9
Age	-10.527***	-0.758	17.976***	-8.443***	-5.083	16.429***	-12.862***	-10.572**	-5.177***
	(1.363)	(1.573)	(3.754)	(2.669)	(3.941)	(5.595)	(2.265)	(4.075)	(1.906)
Floor	4.145**	6.886***	1.915	8.799*	2.331	-1.326	6.499**	-0.220	7.072***
	(1.858)	(2.126)	(2.871)	(4.886)	(5.072)	(2.747)	(3.070)	(1.765)	(1.900)
Size_Sqm	13.301***	9.260***	8.778***	2.275**	10.309***	6.604***	6.969***	7.719***	13.386***
	(0.634)	(0.661)	(0.778)	(1.050)	(0.894)	(1.009)	(0.765)	(1.025)	(0.974)
N	772	557	720	271	230	421	602	355	602
Adjusted R^2	0.601	0.246	0.185	0.102	0.479	0.109	0.224	0.171	0.239

Note: the dependent variable is housing rent. N_1 to N_9 are defined in Table 2.

TABLE 3.2

SSH_KREM hedonic housing rent regressions in the 9 neighborhoods

	(1) N_1	(2) N_2	(3) N_3	(4) N_4	(5) N_5	(6) N_6	(7) N_7	(8) N_8	(9) N_9
Age	-10.616***	-0.890	23.41***	-10.317***	-0.847	18.986***	-9.774***	-15.891***	-4.939***
	(1.218)	(1.228)	(2.449)	(1.935)	(2.561)	(2.811)	(1.694)	(2.683)	(1.535)
Floor	4.321***	5.243***	1.996	6.113*	13.764**	-3.178	10.602**	0.169	8.547***
	(1.601)	(1.767)	(2.469)	(3.603)	(3.362)	(2.375)	(2.511)	(1.440)	(1.721)

Size_Sqm	13.072***	9.475***	8.673***	4.353***	9.078***	6.129***	7.428***	7.962***	13.386***
	(0.570)	(0.559)	(0.657)	(0.744)	(0 .628)	(0.884)	(0.619)	(0.766)	(0.974)
State variable	0.955***	1.009***	0.972***	0.971***	0.981***	0.982***	0.977***	0.982***	0.971***
	(0.042)	(0.043)	(0.036)	(0.045)	(0.068)	(0.042)	(0.040)	(0.035)	(0.061)
N	772	557	720	271	230	421	602	355	602
Adjusted R^2	0.758	0.619	0.589	0.668	0.729	0.606	0.703	0.736	0.463

Note: "state variable" is the state series $\tilde{x}_{t,i}$. the dependent variable is housing rent. the dependent variable is housing rent. N_1 to N_9 are defined in Table 2.

We further test the sensitivity of the SSH_KREM model estimation to sample size. Table 4a shows the Mean Absolute Percentage Error(MAPE) of the CV errors across 9 neighborhoods, MAPE doesn't not change significantly as sample size. In order to validate the test, we use a single neighborhood (sector 141 in Fig 1) and stepwisely drop 20% samlple randomly, the changes in MAPE are very small and can be neglected (Table 4b)

TABLE 4a

SSH_KREM model sample size sensitive analysis: across neighborhoods

	(1) N_1	(2) N_2	(3) N_3	(4) N_4	(5) N_5	(6) N_6	(7) N_7	(8) N_8	(9) N_9
MAPE	0.091	0.078	0.078	0.067	0.078	0.069	0.075	0.060	0.094
Obs	772	557	720	271	230	421	602	355	602

TABLE 4b

SSH_KREM model sample size sensitive analysis: in a single neighborhood

		(1)	(2)	(3)	(4)	(5)
		100%	80%	60%	40%	20%
N_1	MAPE	0.091	0.088	0.094	0.098	0.097
	Obs	772	617	463	308	154

In summery, SSH_KREM model retrieves the hidden information from the existing dataset through its reinforment leaning process and beats the hedonic regression, locally weighted regression and state-space regression in prediction. The performance of SSH_KREM model is robust with sample size.

4.2 The SSH_KREM Housing Rent Indexes

We demonstrate the estimated SSH_KREM housing rent indexes at both national and neighborhood levels in the Singapore HDB open rental housing market. Then we demonstrate a few applications of SSH_KREM model.

4.2.1 National HDB Housing Rent Indexes in Singapore

The national HDB housing rent indexes are estimated by weighted averaging the 123 neighborhood indexes. The weight is the monthly frequency of rental housing transactions in each neighborhood. The SSH_KREM national indexes are compared with the repeat sales indexes and the hedonic indexes (Fig 2). The index baseline is March 2009.

It is noted that we didn't include locally weighted housing rent indexes. This is because LWS model can't estimate temporal coeffients. To derive indexes, we need to repeatedly run a regression at each time point. This also explains why the existing LWS indexes are all yearly based indexes.

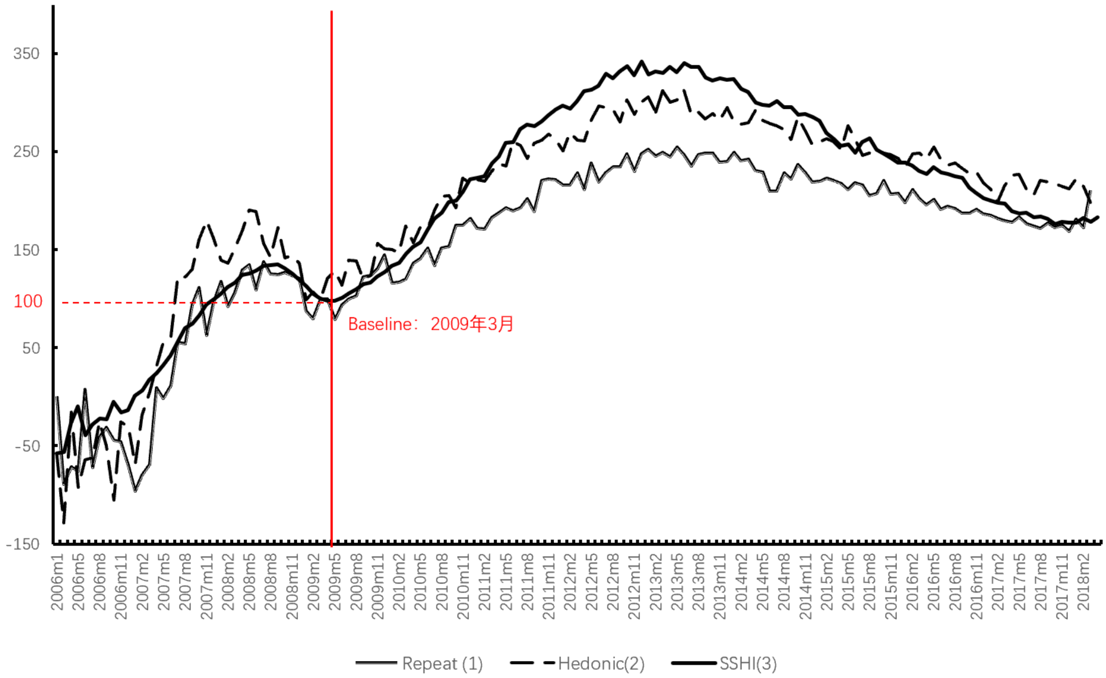


FIG. 2. Housing Rent Indexes: the repeat sales indexes (Repeat), hedonic method (Hedonic) and SSH_KREM indexes

The FIG. 2 Shows that the three indexes series have similar trend between January 2006 and April 2018. But SSH_KREM national index series is smoother than both repeat sales and hedonic indexes. This is because the KR algorism reduces the impacts of sticky rents.

4.2.2 Neighborhood Housing Rent Indexes

The Figures 3.1-3.9 show the neighborhood housing rent indexes of 9 neighborhoods with 95% confidence interval. The figures illustrate housing rent dynamics vary significantly across space.

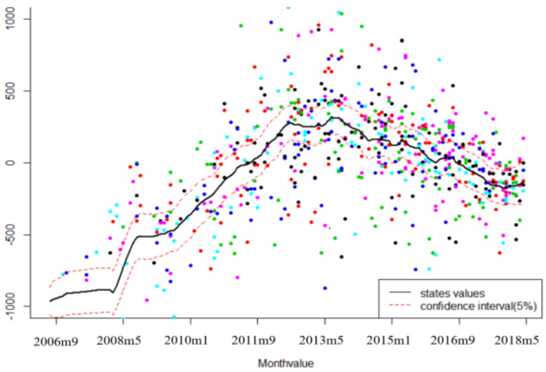


FIG. 3.1. The housing rent indexes in N_1

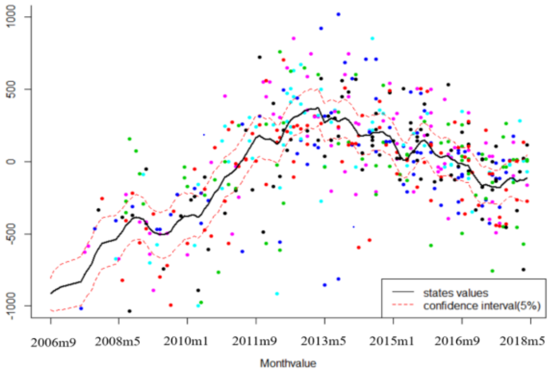


FIG. 3.2. The housing rent indexes in N_2

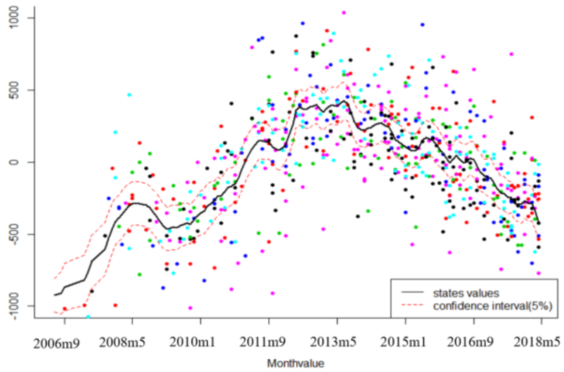


FIG. 3.3. The housing rent indexes in N_3

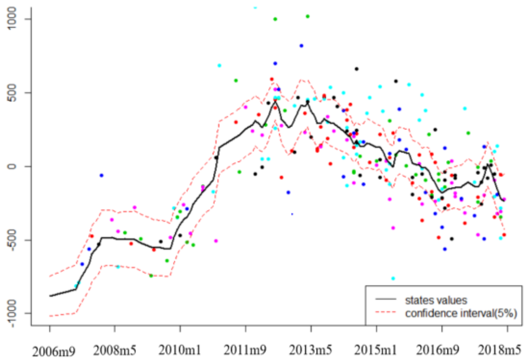


FIG. 3.4. The housing rent indexes in N_4

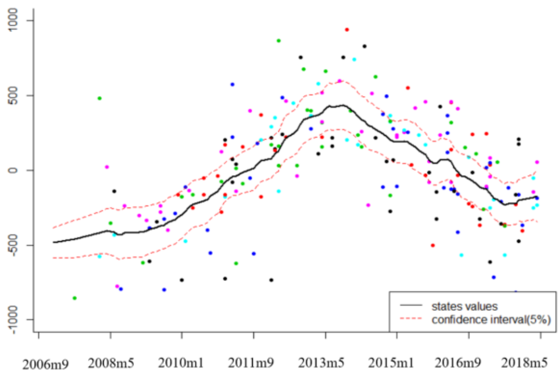


FIG. 3.5. The housing rent indexes in N_5

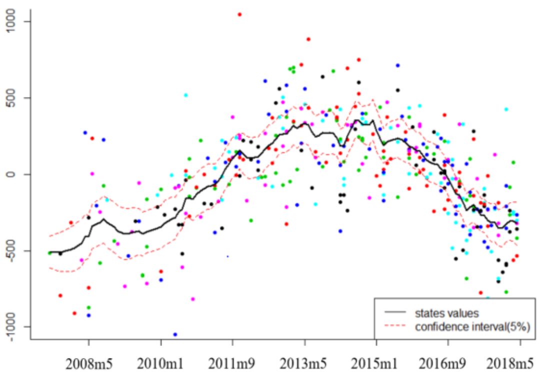


FIG. 3.6. The housing rent indexes in N_6

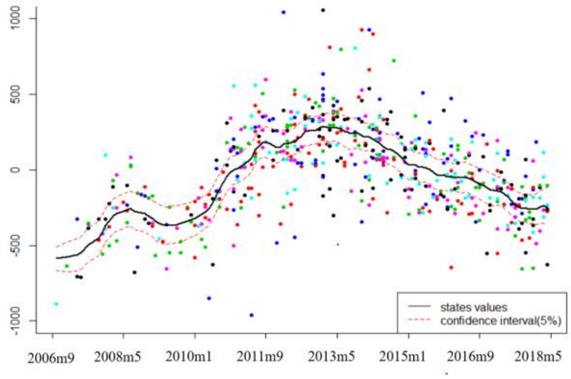


FIG. 3.7. The housing rent indexes in N_7

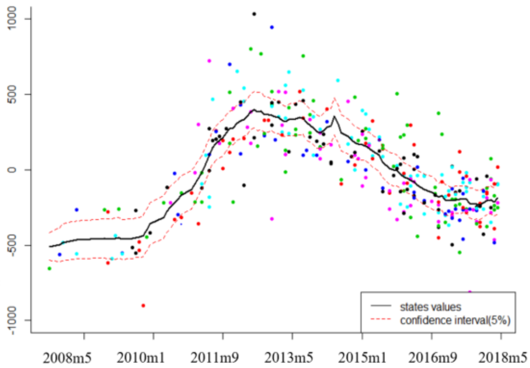


FIG. 3.8. The housing rent indexes in N_8

FIG. 3.9. The housing rent indexes in N_9

4.2.3 The applications of SSH_KREM model

Housing rent is "sticky" as it often doesn't change during a contract period. The SSH_KREM model can predict a rent for each housing unit at any time point. To illustrate it, we choose a housing unit located at "407B FERNVALE ROAD #14-09, Sengkang, Singapore".

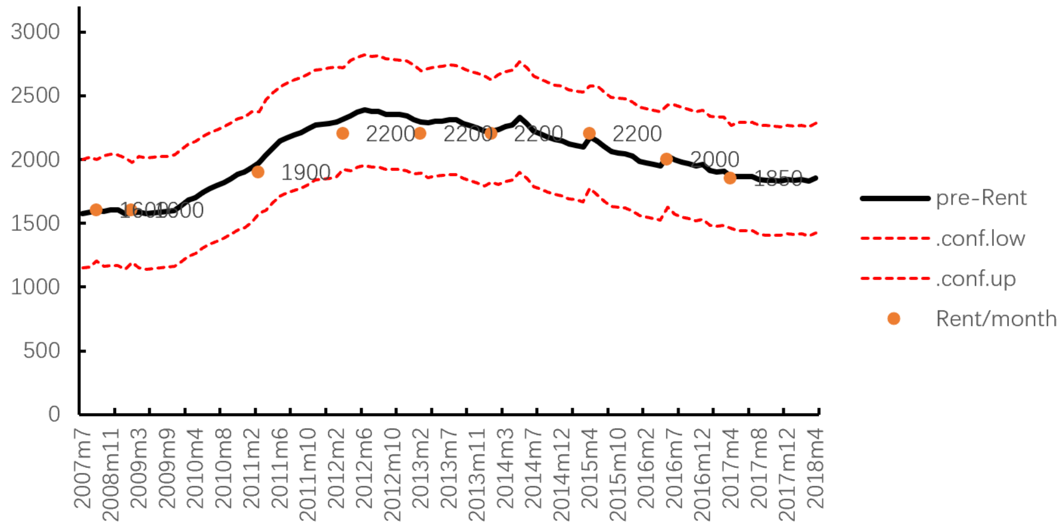


FIG. 4. Predicted rent series for a housing unit and the 95% predictive confidence interval

The unit locates in N_8 (Sector 792 in Fig 1). Between 1st July 2007 and 30th April 2018, the unit was transacted 10 times. We obtain 10 transacted rents (the dots in Fig 4), but we cannot identify which are new rents and which are recontract rents. Using the SSH_KREM model estimated for this neighborhood, we predict the 10 rents (see the line "pre-Rent" and its 95% confidence intervals).

The SSH_KREM model can forecast future rents. FIG. 5 shows the state variable forecasts for the next 5 months in N_8 (Sector 792 in Fig 1). However, we observe the 95% confidence interval becomes much wider during the forecast period.

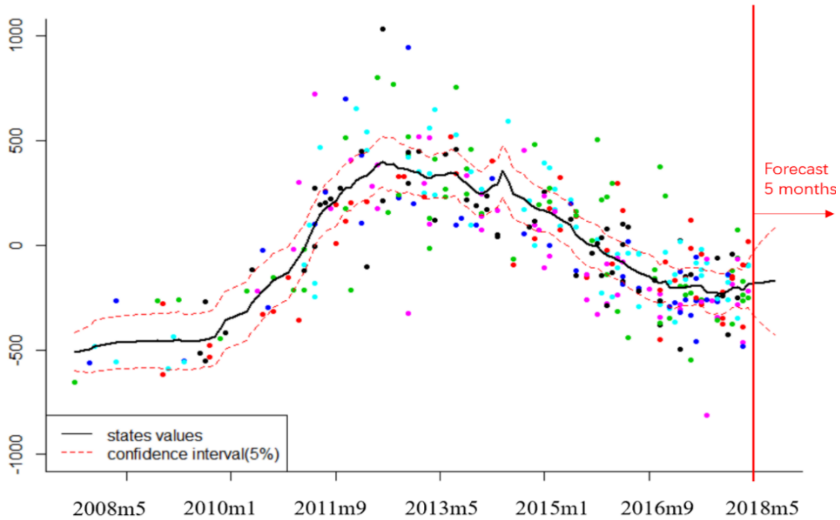


FIG. 5. Five months' state variable forecast N_8 (Sector 792 in Fig 1)

4.3 The dynamic patterns of the Singapore HDB open rental housing Market

The K-shape clustering method can help identify which neighborhoods have more stable investment returns or if the rents across some beighborhoods may share the similar dynamic patterns. The housing units locating in the different neighborhoods but sharing similar dynamic pattern, may provide insightful information for investors. The K-shape clustering method is applied to 123 neighborhhod indexes. Six clusters are formed (Fig 6). K-shape cluatering method is sensitive to missing values, we use mean interpolation method to fill in the missing index numvers in each time series. It is noted that missing index numbers in a index series is caused by zero transactions in that month. The interpolation index numbers take account of 15% of all.

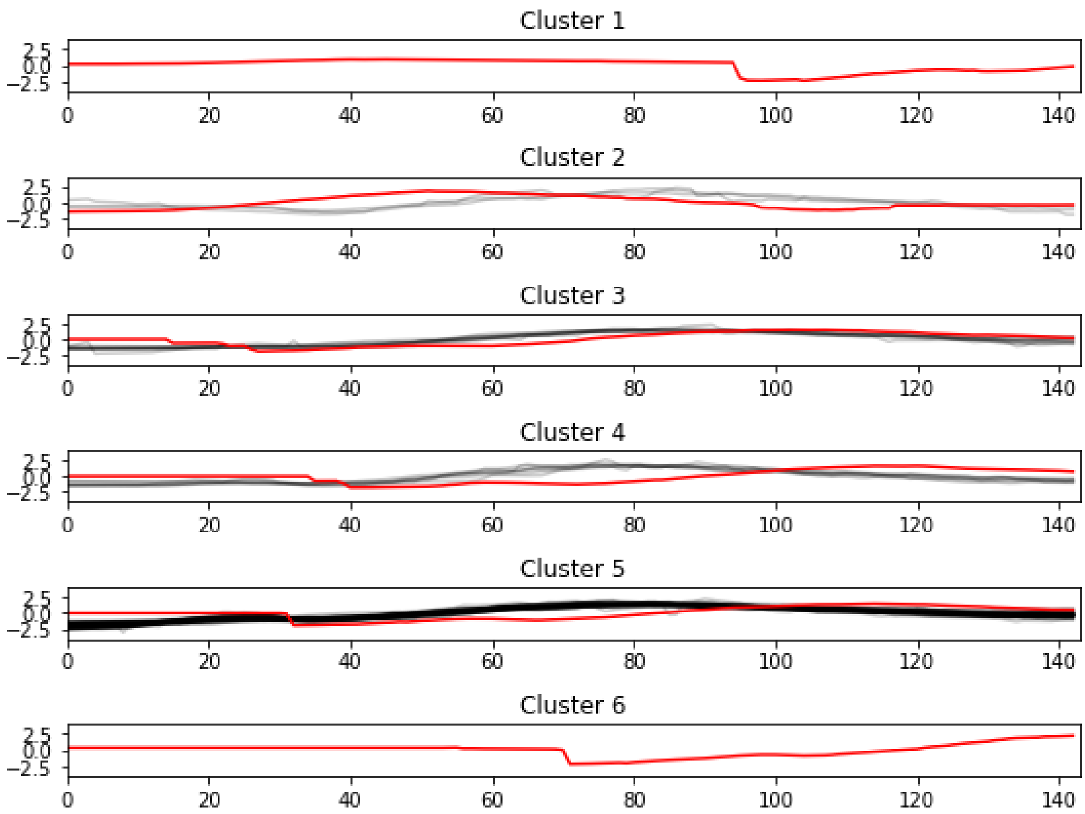


FIG. 6. The trends of 6 clusters in the Singapore HDB open rental housing market

AS we mentioned before,when processing K-shape clustering, every time series is normalized between -2 to 2. Fig 6 shows the normalized SSH_KREM rental indexes. The Red lines are the cluster centroids of the six clusters and each gray line is the normalized SSH_KREM rental index which belongs to the certain cluster.

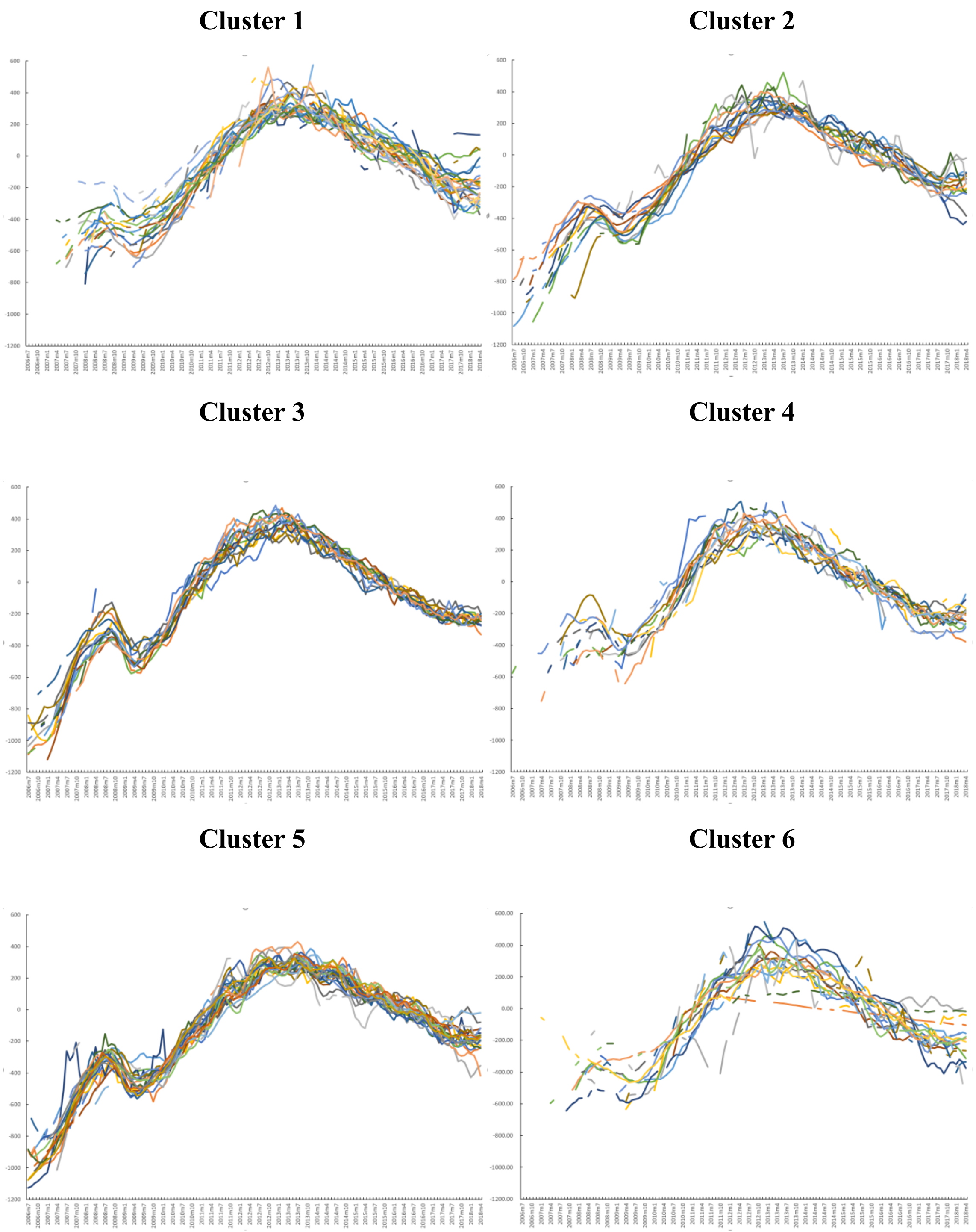


FIG. 7. The SSH_KREM neighborhood housing rent indexes by clusters

FIG. 7 gives the 123 neighborhoods K-shape clustering results. The 6 clusters indicate six dynamics patterns. We further analyze the attributes of the neighborhoods and find that the housing units in cluster 1 mainly locate in the central area, which explain why they share similar patterns. However, the neighborhoods in other clusters are not necessarily bounded in one geopgraphical area.

5. Conclusion

We present a machine learning method to construct housing rent indexes at a fine-scale geographical unit, which outperforms the selected exisiting econometric models. In particular, the proposed state-space model generates more precise rental indexes than other models and it is also the first state space model that is specifically designed for constructing housing rent indexes.

The proposed SSH_KREM model uses a reinforcement learning approach to capture hidden spatial and temporal information from the hedonic residuals, which increase model prediction accuracy. The prediction accuracy won't change with the reduction in sample size. This enables us to construct housing rent indexes at neighborhood level. the adoption of KR algorithm helps correct the bias caused by sticky rents and the EM algorithm produces robust likelihood estimates avoiding the problem caused by unobserved information, such as fewer or zero transactions in certain months at neighborhood level. Moreover, K-shape clustering method in unsupervised learning literature can improve our understanding on the dynamics of housing market structure.

Appendix

TABLE A_1

The sample sizes of the 123 neighborhoods in the HDB open rental housing market

Postal Sector	Sample Size	Postal Sector	Sample Size	Postal Sector	Sample Size	Postal Sector	Sample Size	Postal Sector	Sample Size
651	64	793	191	644	421	643	1,063	541	2,193
551	65	734	201	681	425	642	1,097	542	2,261
653	65	754	202	590	426	210	1,121	521	2,333
212	66	562	204	732	435	610	1,147	470	2,567
564	69	165	205	531	443	544	1,148	440	2,708
211	75	683	221	164	446	350	1,155	140	3,449
463	77	523	222	753	467	461	1,185	550	4,159
315	79	601	230	430	491	822	1,238	600	4,343
602	82	524	234	810	511	641	1,303	670	4,765
656	83	312	235	431	514	380	1,316	570	4,977
528	86	505	244	752	522	270	1,324	680	5,661
794	87	381	252	390	541	370	1,342	510	5,733
735	91	311	261	901	551	271	1,349	310	6,146
563	92	152	263	321	557	400	1,377	650	6,414
323	93	391	267	731	602	900	1,378	640	7,142
322	102	522	271	800	602	821	1,439	120	7,442
533	110	162	293	200	665	330	1,453	530	8,049
603	112	791	307	180	707	100	1,527	730	8,152
604	118	260	317	410	720	150	1,551	460	8,237
101	130	190	355	130	731	320	1,700	760	9,246
526	139	792	355	151	746	540	1,770	560	10,055
525	144	142	366	751	762	543	1,781	520	10,680
462	146	682	373	141	772	820	1,803		
532	173	360	384	824	800	160	1,972		
684	185	163	411	823	947	750	2,085		

Note: If The neighborhood sample size outnumber 1500, we use High Performance Computer(HPC) to run programs. Because it needs more than 45GB CPU memory and will cost more than 24 hours to converge.

Table A_2 The iteration process of the KREM algorithms

To estimate the hyperparameters (Φ) and the State series X={$\tilde{x}_{t,i}, t = 1,2 \dots T$ } in a state-space model for neighborhoopd i, two iterative processes are needed. The EM estimation for Φ and $\tilde{x}_{t,i}$ is completed by an iteration process. In each iteration in the EM, the KR estimation for a State $\tilde{x}_{t,i}$ is completed by another iteration process.

Mathematically, in each iteration of an EM algorithm, the following equation is maximized.

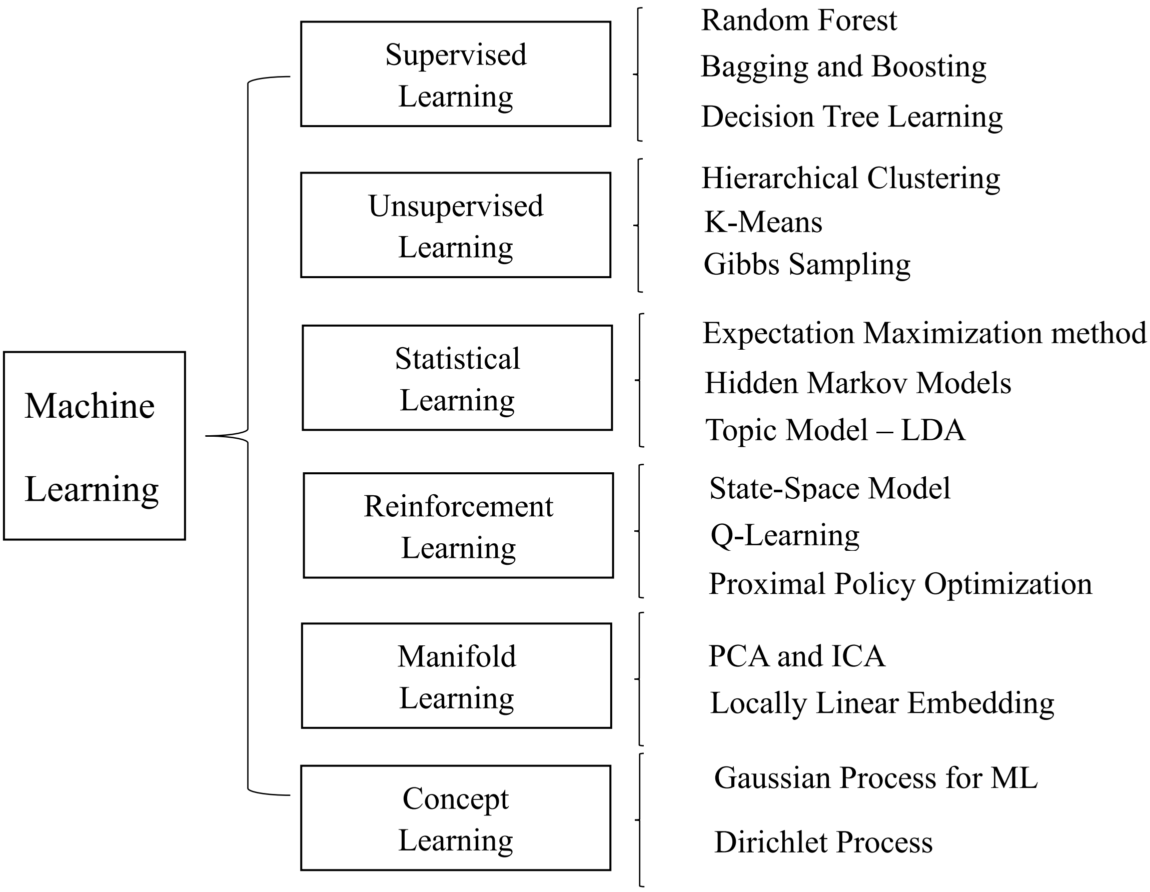
$$\widehat{\Phi}_2 = arg \max_{\Phi} \quad E_{X|\widehat{\Phi}_1}[\log L(\Phi|Y = y_1^T, X)]$$

Where: y_1^T means the expectation values of the observable inputs conditioned on full time information. This value is calculated by KR algorithm. $arg \max_{\Phi}()$ is a function means the a certain value set of Φ which makes the function value in () is maximum.
The iterating process is as below:

a) Set an initial set of hyperparameters, $\widehat{\Phi}_1$;
b) At E step,
    a. using the state-space model for the hidden state X and hyperparameter $\widehat{\Phi}_1$ to calculate the expected values of State X, conditioned on all the data, y_1^T. The calculation is based on the iteration process using the Kalman filter and smoother recursive algorithm, this gives the x_t^T output;
    b. calculating the expected values of State X (or Y if there are missing using both *Transition Equation* (3.2.2) and *Measurement Equation* (3.2.3) of state **X** (or **Y** if there are missing observations in Y) that appear in your expected log-likelihood function, like $\mathrm{E}[X|Y = y_1^T, \widehat{\Phi}_1]$and $\mathrm{E}[\mathrm{g}(X)|Y = y_1^T, \widehat{\Phi}_1]$.

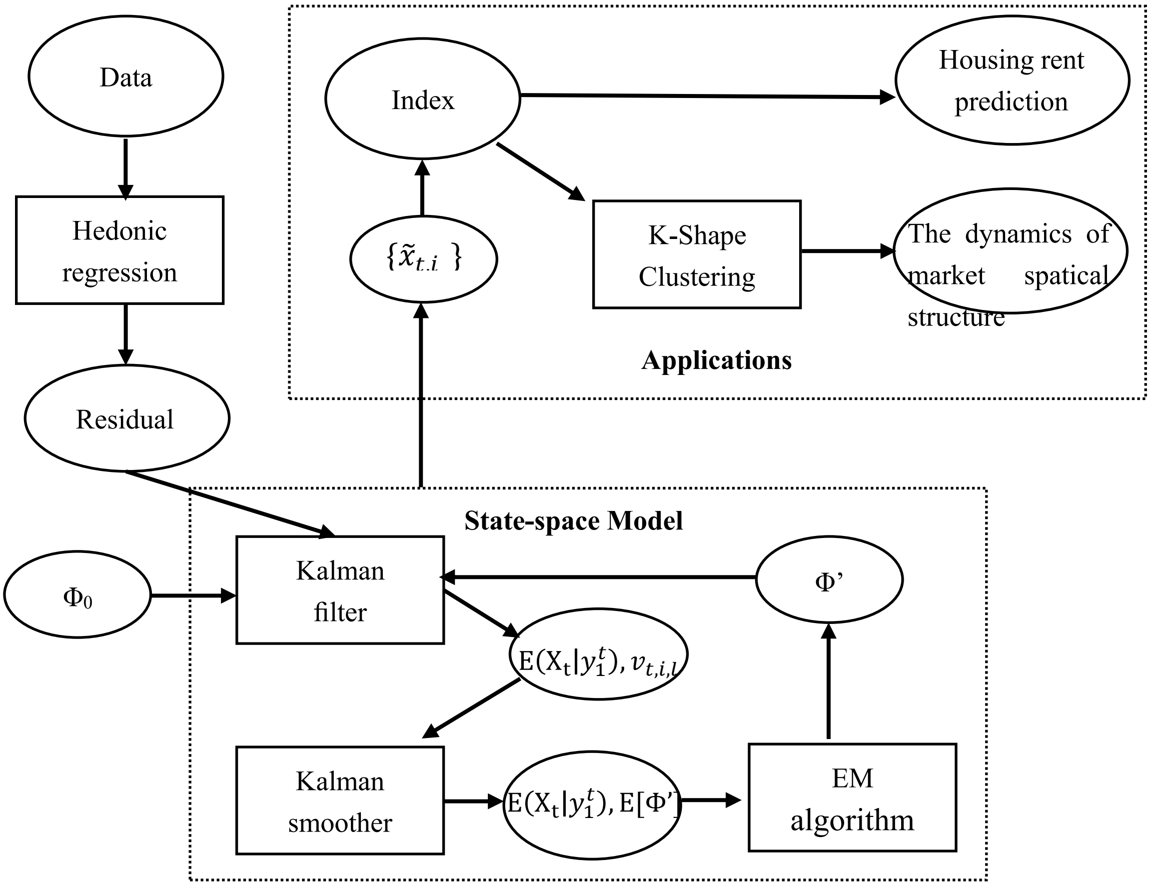
c) At M step, put those $\mathrm{E}[X|Y = y_1^T, \widehat{\Phi}_1]$and $\mathrm{E}[\mathrm{g}(X)|Y = y_1^T, \widehat{\Phi}_1]$ into your expected log-likelihood function in place of **X** (and g(X)) and maximize with respect to Φ. This gives you Φ_2.
d) Repeat the E and M steps until the log likelihood $E_{X|\widehat{\Phi}_1}[\log L(\Phi|Y = y_1^T, X)$ converges.

Chart A_1. The classifications of machine learning approaches



Source: Writer summeried form “An Introduction to Statistical Learning”(James, 2013)

Chart A_2 The research design framework



REFERENCES

Ambrose, B. W., Coulson, N. E., & Yoshida, J. (2013). The repeat rent index. *Review of Economics and Statistics*, 97(5) , 939–950.

An, X., Deng, Y., Fisher, J. D., & Hu, M. R. (2016). Commercial real estate rental index: a dynamic panel data model estimation. *Real Estate Economics, 44*(2), 378–410.

Bailey, M., Muth, R., & Nourse, H. (1963). A regression method for real estate price index construction. *Publications of the American Statistical Association*, 58(304), 933-942.

Baltagi, B. H. , & Li, J. . (2013). Further evidence on the spatio-temporal model of house prices in the united states. *Journal of Applied Econometrics* 29(3):515-522.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. 1998. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18:7411-7425.

Boulay, G. . (2012). Real estate market and urban transformations: spatio-temporal analysis of house price increase in the centre of marseille (1996-2010). Articulo - *Journal of Urban Research*, 9(9).

Brunauer, W., Lang, S., & Umlauf, N. (2013). Modeling house prices using multilevel structured additive regression. *STATISTICAL MODELLING -LETCHWORTH-, 13*(2), 95-123.

Case, K. E. , & Shiller, R. J. . (1987). Prices of single family homes since 1970: new indexes for four cities. *England Economic Reviews*, 5(Sep), 45-56.

Case, K. E. , & Shiller, R. J. . (1989). The efficiency of the market for single family homes. *America Economic Reviews*. 79, 125–137.

Crone, Theodore M. , L. I. Nakamura , and R. Voith . Measuring American rents: a revisionist history.Working Papers (2001).

Eichholtz, P. , Straetmans, S. , & Theebe, M. . (2012). The amsterdam rent index: the housing market and the economy, 1550–1850. *Journal of Housing Economics*, 21(4).

Gatzlaff, D. H., & Haurin, D. R. .(1997). Sample selection bias and repeat-sales index estimates. *Journal of Real Estate Finance & Economics, 14*(1-2), 33-50.

Geisser, S. . (1993). Bayesian interim analysis of censored exponential observations. Statistics and Probability Letters, 18(2), 163-168.

Giannetti, A. (2018). Does temporal aggregation explain the persistence of the s&p/case-shiller indices? evidence from a longitudinal specification. *Real Estate Economics*, 46(3).

Goetzmann, W. N., & Peng, L. (2002). The bias of the rsr estimator and the accuracy of some alternatives. *Real Estate Economics, 30*(1), 13–39.

Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *PS: Political Science & Politics*, 48(1), 80-83.

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

James, G. , Witten, D. , Hastie, T. , & Tibshirani, R. . (2013). An Introduction to Statistical Learning. Springer New York.

Jeanty, P. W. , Partridge, M. , & Irwin, E. . (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Regional Science and Urban Economics, 40(5), 343-352.

Kalman RE. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME--Journal of Basic Engineering*, 82:35-45.

Keskin, Berna; Watkins, Craig.(2017). Defining spatial housing submarkets: Exploring the case for expert delineated boundaries. *Urban Studies.* 54(6):1446-1462.

Laurini, M. P. (2017). A continuous spatio-temporal model for house prices in the USA. The Annals of Regional Science, 58(1), 235-269.

Margo, R. A. (1996). The rental price of housing in new york city, 1830-1860. *Journal of Economic History, 56*(3), 605-625.

Mcmillen, D. P. (2012). Repeat sales as a matching estimator. *Real Estate Economics, 40*(4), 745–773

McMillen, D. P. (2014). Local quantile house price indices. Journal of Urban Economics.

McMillen, Daniel P., and Chihiro Shimizu. "Decompositions of Spatially Varying Quantile Distribution Estimates: The Rise and Fall of Tokyo House Prices ". working paper, 2017.

Meese, R., & Wallace, N. (1994). Testing the present value relation for housing prices: Should I leave my house in San Francisco?. *Journal of urban Economics*, 35(3), 245-266.

Katagiri, M. (2018). House Price Synchronization and Financial Openness: A Dynamic Factor Model Approach. International Monetary Fund.

Rauch, H. E., Striebel, C. T., & Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *Aiaa Student Journal American Institute of Aeronautics & Astronautics, 38*(38), 1445-1450.

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

Schulz, R., & Werwatz, A. (2004). A state space model for berlin house prices: estimation and economic interpretation. *Journal of Real Estate Finance & Economics, 28*(1), 37-57.

Shiller, R. J. (1991). Arithmetic repeat sales price estimators *. *Cowles Foundation Discussion Papers*, 1(1), 110-126.

Tao, M., Zhou, Z., & Abdulhai, B. (2015). Nonlinear multivariate time–space threshold vector error correction model for short term traffic state prediction. *Transportation Research Part B*, 76, 27-47.

Titsias, M. K., Holmes, C. C., & Yau, C. (2016). Statistical inference in hidden markov models using k-segment constraints. *Journal of the American Statistical Association, 111*(513), 200-215.

Ren, You; Emily Y B. Fox and Andrew Bruce.(2017). Clustering correlated, sparse data streams to estimate a localized housing price index. *The Annals of Applied Statistics* 11(2):808-839

Wallace, N. E., & Meese, R. A. (1997). The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics*, 14(1-2), 51-73.

Webb, R. B., & Fisher, J. D. (1996). Development of an effective rent (lease) index for the chicago cbd. *Journal of Urban Economics, 39*(1), 1-19.

Wheaton, W. C., & Torto, R. G. (1994). Office rent indices and their behavior over time. *Journal of Urban Economics*, 35(35), 121-139.

Willmott, C. J., & Matsuura, K. (1995). Smart interpolation of annually averaged air temperature in the United States. *Journal of Applied Meteorology*, 34, 2577–2586.

Wu, J., Deng, Y., & Liu, H. (2014). House price index construction in the nascent housing market: the case of china. *Journal of Real Estate Finance & Economics, 48*(3), 522-545.

Wyngarden, H. (1927). An Index of Local Real Estate Prices. *Michigan Business Studies*, 1(2).

Zhang, H., Hudson, R., Metcalf, H., & Manahov, V. (2015). Identification of house price bubbles using user cost in a state space model. *Applied Economics*, 47(56), 6088-6101.