Journal of Housing Economics 50 (2020) 101715

Contents lists available at ScienceDirect

Journal of Housing Economics

journal homepage: www.elsevier.com/locate/jhec

House price index based on online listing information: The case of China

Xiaodan Wang, Keyang Li, Jing Wu*

Hang Lung Center for Real Estate, Department of Construction Management, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Online listing information
House price index
China

JEL classification codes:
R31
R38
C40
C81

ABSTRACT

While a timely, accurate house price index with broad coverage is of significant importance in housing market research and analysis, the lack of reliable raw data sources remains a major constraint in the house price index construction in nascent housing markets such as China. In this study, we introduce online listing information as an innovative data source for house price index construction, using China's housing resale markets as an example. Compared with alternative data sources, such as the officially-registered transaction information of housing resales, our analysis shows that online listing data provide a better trade-off between accuracy, reliability, and feasibility, especially after the resolution of potential replicated and/or manipulated data issues using our proposed procedures. Based on the cleaned online listing information, we calculate the first housing resale price indices covering almost all (274) Chinese cities. In particular, for around 200 relatively smaller cities, the index provides the first regular house price indicator, which shows a significant divergence in house price dynamics between different tiers of cities. We also briefly discuss the potential extensions of the listing price index, including the daily house price index and the housing rental price index.

1. Introduction

Construction of a high-quality house price index has become increasingly important due to sharp house price fluctuations in major economies. A timely, accurate house price index with broad coverage plays an essential role in decision making for both market participants and policymakers, as well as in housing market analysis and academic researches. Several recent studies have also emphasized that monitoring house prices is especially important but challenging in nascent housing markets such as China (Ciarlone, 2015; Guo et al., 2014; Wu et al., 2014, 2016).

Existing literature on house price index constructions mostly focuses on the methodological perspective. Besides the conventional hedonic (Kain and Quigley, 1970) and repeat sales (Case and Shiller, 1987) methods, recent studies also introduce several innovative concepts, such as the matched (Francke, 2010; Guo et al., 2014; Jiang et al., 2015; McMillen, 2014), quantile regression (Liao and Wang, 2012; McMillen and Thorsnes, 2006), and semi-parametric (Hill and Scholz, 2017; Karato et al., 2015; Zhu et al., 2019) methods. However, much less efforts have been made from the raw data source perspective. Currently, most existing house price indices rely on transaction or appraisal data, both of which have been revealed to expose to potential problems (Deng et al., 2018; Silver, 2016). In particular, in nascent housing markets such as China, the lack of reliable price information has become the key obstacle to further improving regular house price monitoring.

In this study, we introduce online listing information as an innovative data source to construct house price indices in China's housing resale market. Currently, most listing information for housing resales in urban China is posted online and can be automatically and freely collected, and thus offers significantly better accessibility compared to transaction information. Nevertheless, online listing information still poses its own problems in the context of nascent housing markets, especially in terms of replication and manipulation of data. Therefore, instead of directly introducing the procedures developed by Anenberg and Laufer (2017) based on the multiple listing services (MLS) system in the U.S., we develop a set of procedures to carefully remedy these potential problems before using the listing information in the house price index construction. The empirical evidence from 16 sample cities suggests that, the price index based on the cleaned online listing information can well reflect both the long-term trend and short-term fluctuations of actual transaction prices. In particular, it is more reliable than the resale transaction price information officially registered in local housing authorities, or even the existing official house price indicator in China.

The high data accessibility of the online listing information makes it possible to extend the scope of regular house price monitoring from major cities to most of the cities in China. Based on cleaned online

* Corresponding author.
E-mail address: ireswujing@tsinghua.edu.cn (J. Wu).

https://doi.org/10.1016/j.jhe.2020.101715
Received 12 May 2018; Received in revised form 4 July 2020; Accepted 18 July 2020
Available online 22 July 2020
1051-1377/

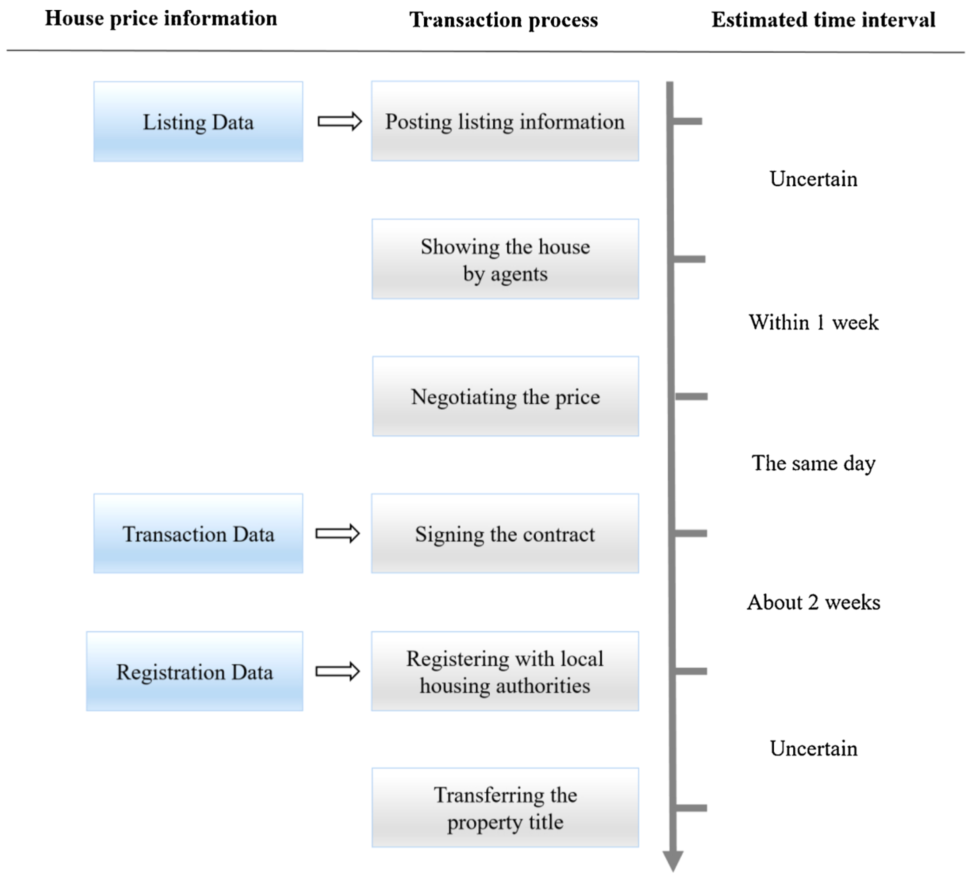

Fig. 1. Typical housing resale transaction process in China.

listing information, we develop the first set of house price indices covering the housing resale markets in almost all (274) cities across the country between 2015 and 2019. In particular, for around 200 3rd or 4th tier cities, the index provides the first indicator which is able to trace their monthly house price dynamics regularly. It thus can serve as the cornerstone of high-quality analysis and research of these increasingly important markets. We find significant divergence in house price dynamics between different tiers of cities; in particular, both the extraordinary house price surges and housing market recessions concentrate in some smaller cities. Meanwhile, we note that house prices in smaller cities near megacities tend to follow the megacities' trends. According to the results, the existing house price indices focusing on major cities only may lead to a biased understanding of the overall housing market conditions in China, a large country with substantial regional disparities.

Additionally, we briefly discuss the potential extensions of the online listing information in house price index constructions. First, while most existing house price indices are on a monthly or quarterly basis, a growing body of literature highlights that a high-frequency house price index is of great importance (Anenberg and Laufer, 2017; Bollerslev et al., 2016; Deng et al., 2017). We provide an example of a daily listing price index for housing resales in Beijing, which can provide rich details in short-term house price dynamics, and thus can benefit not only housing market analysis but also investment decision making. Second, the procedures developed in this study also apply to the collection and cleaning of online listing information of housing rental transactions, facilitating the construction of housing rental price index. We make a preliminary attempt by calculating a set of monthly rental price indices in 35 major cities in 2015–2019. It can reflect changes in living costs and, or even more importantly, help better identify the potential divergence between house prices and the fundamental factors (Davis et al., 2008; Gallin, 2008; Sinai and Souleles, 2005).

Although a few previous studies, such as Anenberg and Laufer (2017), have raised the idea of applying listing information in house price monitoring, this study contributes to the literature by exploring nascent housing markets. First, and most importantly, unlike the case of Anenberg and Laufer (2017) in the U.S., where the well-developed MLS platform provides high-quality listing information, the major challenge for an immature housing resale market like China comes from the poor quality of the original listing information. Therefore, this study strives to develop a cleaning process to identify replicated and manipulated listing information before adopting such information in house price monitoring. Second, in mature housing markets such as the U.S., the listing price mainly serves as complementary information to the already-comprehensive housing market indicators. By contrast, in a nascent market like China, the listing price may be the major, or even only, feasible house price information source for a large number of relatively smaller cities. As a most important contribution, our new index provides the first available house price measure that can be regularly updated for over 200 relatively smaller Chinese cities.

This paper is organized as follows. Section 2 provides background information of China's housing resale market and discusses the feasibility of adopting listing data as a new data source for house price monitoring. Appropriate methodologies for data collection, cleaning, and index construction are introduced in Section 3, especially focusing on how to deal with replicated and manipulated listing data. Section 4 provides empirical evidence on the reliability of the listing price index, by comparing the listing price indices with the indices based on actual resale transactions in 16 sample cities. Section 5 presents the listing price index for housing resales in 274 cities, while Section 6 further discusses the potential extensions in two aspects. The last section concludes the paper.

2. Features of online listing information in China's housing resale market

Since the housing reform of the late 1990s, the housing market in urban China has been dominated by sales in the new construction sector, in which newly-built dwelling units are sold by developers to households (Wu et al., 2014). However, the housing resale markets comprising transactions between households have developed rapidly in recent years when the housing stock has increased and the market for newly-built dwelling units has slowed.

Fig. 1 summarizes the typical process of a housing resale transaction in urban China. There are two noteworthy differences compared with the arrangements in the U.S. or most other developed economies. First, in China, brokers do not serve as the agents of sellers or buyers, but provide intermediary services to match sellers and potential buyers. Second, the MLS system currently does not exist in China. Thus, brokers and households cannot share information on listing units via MLS; instead, brokers need to circulate listing information to attract potential buyers.

Specifically, a seller would first contact one or more brokerage companies to provide information of his/her dwelling unit for sale. The brokerage companies would then list the unit on various public websites in order to attract potential buyers. There are two major types of such websites: the brokerage companies' own official websites such as Homelink (lianjia.com) and Centaline (www.centanet.com), and a large number of national-, regional-, or city-level online housing platforms such as Fang (fang.com) and 58 City (www.58.com). The marginal cost of circulating such information online is generally low; hence, typically, brokers would choose to post the listing information on multiple websites and/or re-post the information on the same websites multiple times.[1] In most cases, the listing information would include major hedonic attributes and a brief description of the unit (including the complex where the unit is located), a few photographs, and, most importantly, the listing/asking price of the unit. In contrast, the specific address (i.e., building and unit number within the complex) is typically omitted. The seller can freely change the listing price on the websites via the brokers until the unit is sold. The websites do not have an obligation to verify the validity and accuracy of the posted information.

A potential buyer, attracted by an online posting, could contact the broker for more details and to arrange a house viewing. Post viewing, if the potential buyer remains interested in the unit, the broker could facilitate a meeting between both parties within a few days (or even on the same day in the case of a hot housing market), over which the buyer and seller could negotiate the price. A transaction would take place if the buyer and seller agree on the price and other transaction terms. In most cases, the broker will immediately prepare the contract, including the (actual) transaction price of the unit. After contract signing, it typically takes a fortnight to register the transaction in the local housing authority officially. One noteworthy phenomenon is that a large share of buyers/sellers choose to register a lower, false transaction price to evade the substantial tax on housing resale transactions,[2] and the gap between the actual and the registered prices could be influenced by factors such as tax codes and buyers' financing demands (Agarwal et al., 2020; Dai and Xu, 2018). The transaction completes after the buyer transfers the payment and obtains the keys from the seller.

Three types of price information are generated in the above process: the listing price (both the initial listing price and its following adjustments); the actual transaction price recorded on the contract (which is private information); and the registered (and typically fake) transaction price recorded in government systems. Most existing housing resale price indices in China are based on the registered price information provided by the local housing authorities. However, this is obviously not an appropriate data source for price index construction, not only because it typically lags by a few weeks, but also because, as discussed above, it is typically manipulated for tax evasion purposes and does not reflect actual transaction prices. The actual transaction price information recorded on the contracts could overcome these two problems, but its accessibility is a major challenge as it is private information shared between the transacting parties and the brokerage company only. Hence, it is not feasible to rely on this type of actual transaction price information for regular house price monitoring, especially for a multiple-city or national-level index.

Online listing information provides a better trade-off between the accuracy, timeliness, and accessibility of housing resale price information. Most importantly, the nature of the online listing information determines that it is released and circulated as public information and, hence, can be easily accessed and collected from public websites with web crawling tools. Additionally, it should be more reliable than the registered transaction price information because the tax evasion incentive does not exist at this stage. In addition, online listing information typically includes both the listing price and detailed hedonic attributes of the dwelling unit, which enables the hedonic method of index construction.

Admittedly, the listing price is not identical to the actual transaction price (i.e., the equilibrium price) for most housing resales. However, we believe that listing price information can still reflect housing market conditions reasonably for the following two reasons. First, because listing prices could significantly affect the search-and-match process of housing resales, sellers are typically very careful in both setting the initial listing prices and making subsequent adjustments (Knight, 2002; Merlo et al., 2015). Second, existing literature has provided strong evidence of the high correlation between listing prices and transaction prices. Listing prices not only influence transaction prices (Bucchianeri and Minson, 2013; Northcraft and Neale, 1987), but also contain information that could improve predictions of transaction prices (Horowitz, 1992). Through Granger causality tests on quarterly data from Baton Rouge, Louisiana, between 1985 and 1992, Knight et al. (1994) concluded that listing prices are leading indicators of transaction prices. The recent work by Shimizu et al. (2016) indicated that house price indices based on listing prices and transaction prices are highly comparable. We will also provide similar empirical evidence through our empirical analysis in Section 4.

Nonetheless, it is important to keep in mind that online listing information also poses potential problems, two of which are especially important. First, a single listing unit may appear more than once online. As mentioned above, the seller might contact more than one brokerage company in an attempt to expand influence; a broker might also choose to post the listing information on multiple websites or multiple times on the same website in order to attract more potential buyers. If such replicated data were disproportionally concentrated on certain types of dwelling units, it would lead to bias in the house price indices constructed. Second, the problem of manipulated data also exists in online listing information, although not for tax-evasion purposes. Some brokers might concoct highly eye-catching listing information (e.g., units with listing prices well below market levels) to entice potential buyers to contact them. If such forged listing information existed, it would surely lead to bias in the index constructed. Therefore, the online listing

[1] Theoretically, sellers could choose to post the listing information on the second type of websites themselves, despite working with one or more brokerage companies. However, most sellers tend to rely entirely on brokerage companies in circulating listing information.

[2] For example, Beijing imposes a deed tax of 1% for units less than 90 sq. m, and 1.5% for units over 90 sq. m. If a unit is resold within 5 years of purchase, an income tax equaling 20% of the appreciation is applied. Moreover, if a unit is resold within 2 years, an additional business tax equaling 5.6% of the transaction price is also applied. Thus, transaction participants have strong incentives to report a lower registration price to evade tax. Agarwal et al. (2020) found that the registration prices were substantially lower than the corresponding actual prices for about 95% of housing resale transactions in a major anonymous city, and on average, the gap between the registration and actual prices reached about one third of the actual price.

information collected cannot be directly used for index construction. Instead, careful data cleaning is needed to deal with the potential problems of replicated and manipulated data. This is also the major difference between our procedures based on a nascent market like China and that developed by Anenberg and Laufer (2017) in the U.S., as in the latter case, the well-developed MLS platform could provide high-quality original listing information.

3. Data collection, cleaning, and index construction

3.1. Listing data collection

With the help of GXD, a leading real estate data company in China, we introduce state-of-the-art web crawling techniques to automatically collect online listing information of housing resale units in mainland China. Specifically, we are currently able to monitor more than 500 websites, including brokerage companies' official websites and housing information platforms as described in Section 2, covering 274 of the 298 cities on or above the prefectural level in mainland China. The remaining 24 cities are located mainly in western provinces such as Gansu and Ningxia, where the housing resale markets remain underdeveloped.[3] It is noteworthy that, besides the 70 major cities covered in the "NBSC 70-City Home Price Index" ("NBSC 70-City Index" hereafter) officially released by the National Bureau of Statistics of China (NBSC), so far there is no regular house price index for most of the remaining 204 cities.

We aim to include information on all units that are still listed on the market (i.e., units whose owners are willing to sell but have not achieved in a transaction). We cannot directly identify whether a listing unit has been sold because brokers do not necessarily withdraw or label the units as sold in a timely manner. However, considering the negligible online posting cost, we can reasonably assume that the brokers will keep posting the unit's information on websites until the unit has been sold. In other words, a unit will be considered as still on the market and, thus, included in our sample if there is newly posted information on the unit. For this purpose, our program automatically browses each website every 15 min and collects all newly posted information on listing units. Obviously, one listing unit will appear as multiple records in the original dataset and potentially lead to the replicated problem as mentioned above. We then merge records of an identical unit in the data cleaning process.

Based on the above approach, we collected a total of over 1.7 billion listing records from 274 cities between January 2015 and September 2019, or about 30.6 million records per month on average. The sample volume greatly varies by city, from less than 100 thousand in small cities, to over 50 million in megacities such as Beijing and Shanghai. For each record, we collect its listing price, listing date, and major hedonic attributes including the name of the housing complex[4]; unit size in floor area; number of rooms including bedrooms, living rooms, and washrooms; number of levels; major direction the dwelling unit is facing; and standard of interior decoration.

3.2. Identifying replicated and manipulated data

As emphasized in Section 2, replicated and manipulated data are two major potential problems in online listing information. Hence, identifying such data is a key task in the data cleaning process before index construction.

For the replicated data issue, if information on one specific listing unit (i.e., listing records with identical hedonic attributes) appears more than once within a reporting period,[5] we choose to merge all these records into one observation in this reporting period in the working dataset. If the listing prices of these records differ within the reporting period (e.g., the seller adjusts the listing price), we compute the simple average of these recorded prices as the listing price of this unit in that reporting period. Therefore, a listing unit will only appear once in one reporting period in the working dataset for house price index construction.

A remaining challenge is that the broker may, intentionally or unintentionally, slightly change the attributes when posting a unit on the same or different websites. For example, the unit size might be reported as 120.05 sq. m. when posted on one website but be rounded to 120 sq. m. on another website. In such cases, the listing records will be identified as separate units and will not be merged by the above process even if they appear within the same reporting period. In order to solve this problem, we further evaluate the dissimilarity between any two listing records within the same housing complex during the same reporting period. Specifically, we follow the strategy developed by Guo et al. (2014), and define the degree of dissimilarity between two listing records as the Manhattan distance:

$$\Delta_{ij} = \sum |\hat{\beta}_m \times (x_{i,m} - x_{j,m})| \tag{1}$$

where Δ_{ij} measures the degree of dissimilarity between listing records i and j; $x_{i,\ m}$ and $x_{j,\ m}$ are the same hedonic attribute, m, of these two records; the coefficient $\hat{\beta}_m$ represents the relative importance of the attribute, which equals to the hedonic value of the corresponding attribute estimated via the hedonic model of the city in the given reporting period. Accordingly, a hedonic attribute will be endowed with a higher weight when calculating the dissimilarity if it has a larger impact on house prices.

The schematic diagram of the representative probability density of Δ is shown in Fig. 2. If there are no replicated listing records, the probability should normally be distributed due to the law of large numbers. However, some pairs with high similarity (or low dissimilarity) would also exist because of the replicated listings issue. Therefore, there will be a turning point (TP) in the probability density curve, as shown in Fig. 2. We take this turning point as the threshold, and a pair of listing records will be merged if the calculated dissimilarity between them is smaller than the threshold. For example, in Fig. 2, the listing records of i and j will be identified as derived from one identical housing unit and be merged. In contrast, the listing records of i and k, or j and k, will be identified as belonging to different listing units.

For the manipulated data issue, we follow the strategy of Kontrimas and Verikas (2006) and Morano and Tajani (2014) and adopt the conventional method of outlier detection to identify listing units whose listing prices abnormally diverge from the market level. Specifically, for each listing unit, we use all other units listed in the same reporting period to estimate the hedonic model and use this hedonic model to impute the fair value of the unit. Following the standard suggested by Morano and Tajani (2014), we define units whose listing prices fall beyond three standard deviations from the corresponding fair values as outliers. These outliers will be excluded from the working dataset for the following house price index construction.

We use the period between January 2015 and December 2017 in the capital city, Beijing, as an example to present the outcomes of online listing data collection and cleaning. During these 36 months, we collected an original dataset containing 29.19 million online listing records. Following the above data cleaning procedures, we finally achieve

[3] According to the latest available data from the 2010 Population Census, the 274 cities accounted for 94.1% of the urban population in China. The share in total housing stock or total resale housing transaction volume should be even higher, although the accurate statistics are not available.

[4] We will follow the hedonic method improved by Wu et al. (2014), in which complex dummies are adopted to control for complex-level attributes. Thus, at the complex level, we only need to identify the complex name.

[5] The reporting period is determined by the frequency of the price index to be constructed, such as month, week, or day.

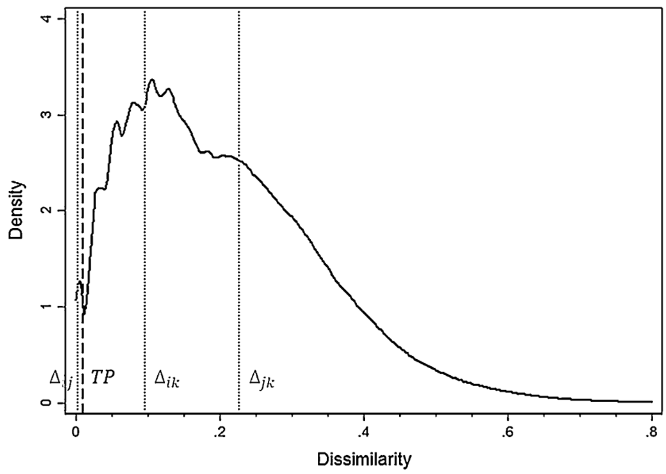

Fig. 2. Schematic diagram to identify replicated data.

in a cleaned working dataset with an average of 72,725 listing units per month if we adopt monthly reporting periods, or an average of 14,436 units per day at the daily basis. As a rough estimate, according to the statistics of the local housing authority, during the three-year sample period there were totally 586,475 dwelling units sold in the resale market in Beijing, or 535 units per day on average. Based on the statistics from a leading brokerage company in Beijing, the average time-on-market for resale units was 58.7 days during this period. Therefore, on average we could expect 31,404 (535 * 58.7) units listed on the market per day. Accordingly, it is reasonable to assert that our sample represents a significant share of the resale market.

Table 1 provides further evidence of the necessity of the data cleaning process. In Column (1), we include all the 29.19 million original records. After controlling for the complex fixed effects, unit-level hedonic attributes, and year-month fixed effects, we further introduce the dummy for the identified replicated records (with daily reporting period). Controlling for other factors, the replicated records' listing prices are 0.65% higher and statistically significant at 1% level. One possible explanation is that brokers have higher incentives to post a unit on multiple websites or multiple times if the unit is listed at a higher price. In Column (2), we include the 16.63 million unique records (i.e., after merging the replicated records), and introduce the dummy for the identified manipulated records. Controlling for other factors, the prices of the identified manipulated listing records are 23.05% lower, which is consistent with the pattern that brokers intentionally falsify some listing records with extremely low prices to seduce potential buyers. According to these results, the existence of replicated and manipulated listing records, especially the latter, would lead to bias in the house price index constructed, thereby highlighting the importance of the data cleaning process.

Table 1
Price difference associated with identified replicated and manipulated data.

Variable	(1)	(2)
	ln(unit listing price)	
Identified as replicated data	0.0064***	
	(0.0001)	
Identified as manipulated data		−0.2620***
		(0.0003)
Year-month FE	YES	YES
Complex FE	YES	YES
Hedonic Attributes	YES	YES
Observations	29,184,729	16,631,489
R-squared	0.908	0.899

Note: Robust standard errors clustered at complex level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

3.3. Price index construction

Based on the online listing information collected and cleaned via the approach discussed above, we further calculate the house price index for specific housing resale markets. Considering that it is less feasible to identify repeat sales, we choose to calculate the resale price indices using the hedonic method improved by Wu et al. (2014); that is:

$$HP_{ijt} = \alpha + \beta X_{it} + \gamma C_j + \delta_t D_t + \varepsilon_{ijt} \tag{2}$$

$$HPI_t = \exp(D_t) \times 100 \tag{3}$$

where: HPI_t refers to the resale price indices; HP_{ijt} is the logarithm of listing price per sq. m. of unit i in complex j at time t; X_{it} and C_j represent the unit-level housing characteristics and complex-level dummy (equals 1 for complex j and 0 otherwise), respectively; D_t is the time dummy variable (equals 1 in period t and 0 otherwise); and ε_{ijt} is the independent and identically distributed (i.i.d.) error term.

4. Evidence on the reliability of listing price index

Before turning to the applications, we provide empirical evidence on the reliability of the listing price index. With the help of a leading brokerage company in China, we can access full-sample micro-level actual transaction data of resale units sold by this brokerage company between January 2015 and December 2017 in 16 major cities.[6] We firstly calculate the constant-quality monthly price index based on these actual transaction data using the hedonic method shown in Eqs. (2) and (3). As discussed in Section 2, we believe the actual transaction information can most accurately reflect price changes in the resale market and adopt this index as the benchmark. Then, as our major interest, we calculate the monthly listing price index in these 16 cities based on the data cleaning and index construction procedures described in Section 3. As a comparison, we also include two existing house price indicators in the same set of cities, including the resale market sub-index of the official “NBSC 70-City Index”, and the monthly average registration prices for resale units as released by local housing authorities. In Fig. 3, we plot the four index series in all these cities, with January 2017 as the basic period. Generally, the graphical evidence shows that the listing price indices could well trace the fluctuation of actual transaction price indices, compared with the NBSC indices and the average registration prices.

In order to provide more conclusive evidence, we then adopt the following two aspects of quantitative analysis for all the 16 cities. First, in order to test the consistency in the long-term trend, in Panel A of Table 2 we calculate the cumulative growth rate between January 2015 and December 2017 for each of the four price indices. As the benchmark, the average cumulative growth rate of the actual price index of the 16 cities was 47.8% during the sample period (or a monthly compound average growth rate of 1.3%). The average cumulative growth rate of the listing price index also reached 45.1% during the same period, which is not significantly different from the growth rate of the actual transaction price index.[7] By contrast, the corresponding cumulative growth rate was only 18.9% for the “NBSC 70-City Index”, or 27.8% for the average registration price indicator, both of which were significantly lower than the actual transaction price index.

[6] These 16 cities are Beijing, Changsha, Chengdu, Chongqing, Dalian, Guangzhou, Hangzhou, Hefei, Jinan, Nanjing, Qingdao, Shanghai, Shenyang, Shenzhen, Wuhan, and Xiamen. The actual transaction sample cover the whole period between 2015 and 2017 in Beijing, Chengdu, Hangzhou, Nanjing, Qingdao and Shanghai. In the other 11 cities, the actual transaction sample only started between January 2015 and December 2016. We adopt January 2017 as the basic period in the following index comparison, because we have the four indices after that for all the 16 cities.

[7] We also depict the cumulative growth rates for all these 16 figures in Fig. A1 in the appendix. Besides the consistency in the average value, the growth rates are also highly consistent in most of these 16 cities.

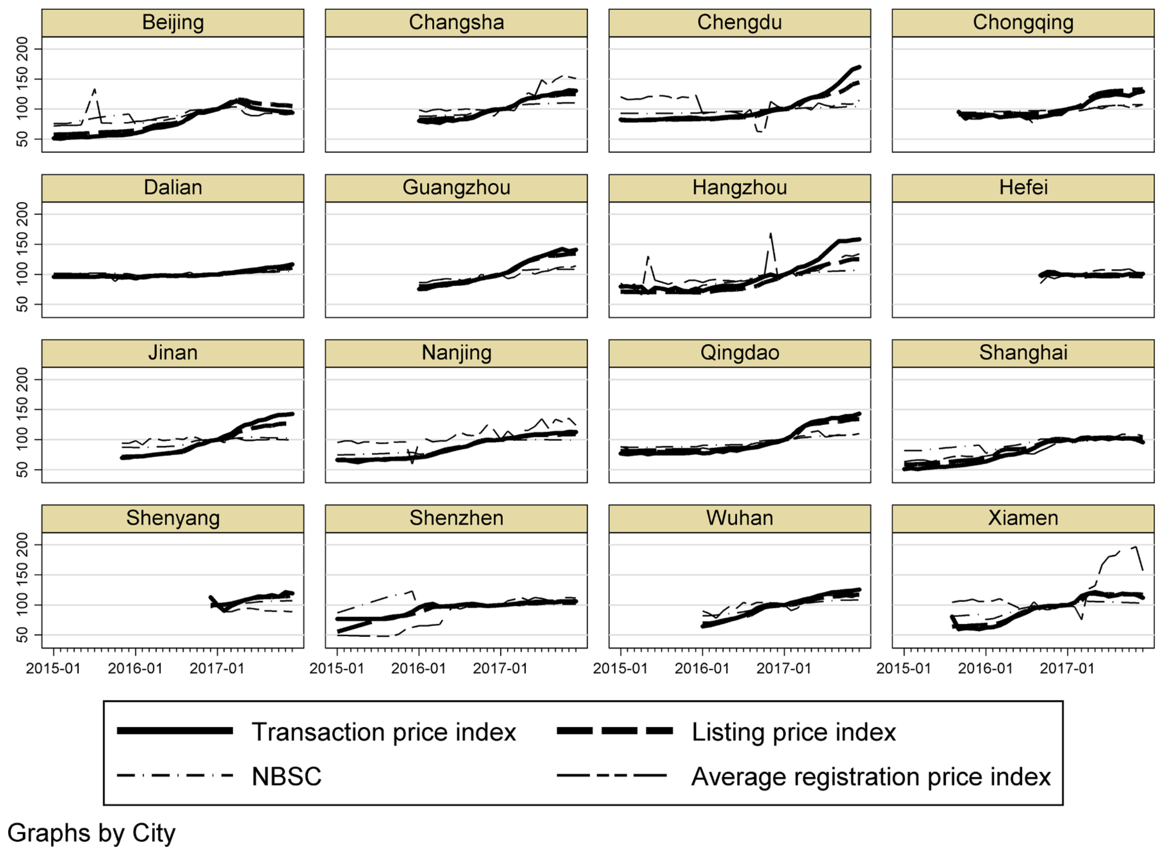

Fig. 3. Four house price indices in 16 sample cities (January 2015 to December 2017; Benchmark: January 2017).

Table 2
Comparison between four resale house price indicators (16 sample cities; January 2015 to December 2017).

Panel A: Test on the consistency in long-term trend

Variable	Mean	Std. Dev.	Min	Max	T-test
Actual transaction price grow rate (%)	47.801	22.934	2.540	72.500	–
Listing price grow rate (%)	45.141	19.541	0.178	64.857	−2.660
Resale sub-index of NBSC 70-City grow rate (%)	18.940	10.218	−0.1320	37.138	−28.861***
Average registration price grow rate (%)	27.906	21.646	−7.254	80.640	−19.895**

Panel B: Test on the consistency in short-term dynamics

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	D.ln(actual transaction price index)					
D.ln(listing price index)	0.878***			0.915***		
	(0.066)			(0.083)		
D.ln(average registration price)		−0.001			−0.002	
		(0.028)			(0.024)	
D.ln(NBSC 70-City Index)			0.158			0.053
			(0.154)			(0.153)
City FE	NO	NO	NO	YES	YES	YES
Year-month FE	NO	NO	NO	YES	YES	YES
Observations	434	434	434	434	434	434
R-squared	0.295	0.000	0.014	0.389	0.206	0.207

Note: Robust standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Second, in order to measure the consistency in short-term dynamics, in Panel B we calculate the correlation coefficient between the actual transaction price index and each of the other three indicators. As the most important finding, the results in Column (1) show that the correlation coefficient between our listing price growth rate and the actual transaction price growth rate reaches as high as 0.878, which is also statistically significant. Instead, there is almost no correlation between the growth rates of the registration price and the actual transaction price (Column 2), while the correlation coefficient between the growth rates of the official NBSC index and the actual transaction price is only 0.158 and statistically not significant (Column 3). In the latter three columns, we further introduce the city fixed effect and monthly fixed effect to calculate the conditional correlation coefficients, and the pattern remains almost unchanged.[8]

[8] Another potential bias of the listing price index may come from the so-called "bidding wars" as suggested by Han and Strange (2014); that is, a seller

As a summary, the empirical results based on the 16 sample cities provide reliable evidence that the listing price index can well reflect the pattern of the actual transaction price in China's resale market, from both the long-term trend and short-term dynamics perspectives. By contrast, consistent with the findings by Agarwal et al. (2020) and Dai and Xu (2018), due to the tax evasion behaviors, the price information officially registered in local housing authorities substantially deviates from the actual housing resale price changes. The official "NBSC 70-City Index" is also less consistent with the actual transaction price index compared with our listing index. This finding echoes the conclusions by Fang et al. (2016), Wu et al. (2014) and Wu et al. (2016) that the official house price indices in China, such as the "NBSC 70-City Index", still suffer from substantial bias due to problems in both underlying data and compiling methodology. Therefore, although the listing price index may still be different from the actual transaction price index, the listing price performs better than the officially-registered transaction information and provides a valuable data source for house price index construction.

5. House price indices covering all resale markets in mainland China

Due to data accessibility constraints in China, the existing price indices based on housing transactions have hitherto concentrated on major cities. In particular, the only official multiple-city index for housing resales (i.e., the sub-index of the NBSC 70-City Index) covers 70 large and medium cities on a monthly basis. However, according to the latest available statistics from the 2010 Population Census, these 70 major cities only accounted for about 50% of the urban population and 40% of the urban housing stock by floor area, with the rest in the over 200 relatively smaller cities.[9] Even more importantly, as highlighted by several recent studies such as Fang et al. (2016) and Liu and Xiong (2018), these smaller cities are associated with higher market risks and are more likely to witness a substantial house price correction in the near future. Therefore, extending the scope of regular housing market monitoring to these relatively smaller cities is also of great importance, especially from the policy perspective.

Based on the online listing information collected and cleaned via the approach described in Section 3, we are able to develop the first national-wide constant-quality price index, which can regularly trace housing resale price dynamics in 274 cities around the country. For each city, the monthly index covers the period between January 2015 and September 2019 in this study, with January 2015 as the basic period, while it is also feasible to keep the index regularly updated afterwards. Fig. 4 depicts the compound average monthly growth rates of the resale price indices during the sample period. It clearly reveals a substantial divergence in housing market conditions between various cities, with the compound average monthly growth rates ranging from −0.62% to 2.31% during the 57 months.

As a most remarkable contribution, for over 200 relatively smaller cities beyond the scope of the official "NBSC 70-City Index", this listing price index provides a first regular constant-quality price indicator for their housing resale markets. While it is infeasible to go through all these 200 cities in this paper, in Fig. 5 we plot the index series of three representative groups of cities as the example. Panel A highlights the strong relationship between a megacity (Beijing) and its nearby smaller cities (Langfang and Baoding, both in Hebei province). While the capital city of Beijing witnessed a doubling in resale price index in 2015–2016, the nearby cities such as Baoding experienced a similar house price surge during the same period. The housing market fluctuation was even more remarkable in another nearby city, Langfang: the resale price index increased by over 170% in 2015–2016, but then dropped by about one-third in the following two years. Panel B includes three cities in Middle China, namely, Nanyang (in Henan province), Anqing (in Anhui province), and Yichun (in Jiangxi province). In all these three cities, the resale price indices increased by over 50% in 2017–2018. Thus, the housing market booming in China does not only concentrate in large, coastal cities, but also exist in at least some of the smaller, inland cities. However, there were also some housing resale markets suffering from a recession during the sample period. As shown in Panel C, in Fuxin, Panjin (both in Liaoning province), and Qitaihe (in Heilongjiang province), all of which locate in the northeast region, the resale price index kept decreasing in all these five years. Market analysts and academic researchers can apply various kinds of quantitative analysis methods to these index series and shed more lights on these relatively smaller housing markets, which have been rapidly developed but greatly ignored in previous studies.

Based on the city-level indices, we further calculate the aggregated indices at the national level. Following Fang et al. (2016), we adopt the urban population reported in the 2010 Population Census as the weight.[10] There are three monthly time series in Fig. 6, with the first covering all of the 274 cities, the second including only the 70 cities covered by NBSC, and the last covering the 204 cities other than the 70 cities. As a most remarkable finding, these three indices point out that the understanding of the overall conditions of Chinese housing resale markets would be biased if the house price index only covers major cities. Regarding to the sample period of 57 months, for all the 274 cities, the compound average growth rate of the aggregated house price index was 0.82% per month, but would increase to 0.96% if we only consider the 70 major cities. Additionally, although the overall trends are similar across the sample period, individual indices may follow different paths during specific months. For example, at the beginning of 2015, house prices increased only in the 70 major cities, and then followed by the smaller cities. All these results further highlight the importance of completed coverage for a national-level index.

In summary, for a large country with significant regional variations like China, a comprehensive understanding of the housing market conditions of the entire country, instead of only a few major cities, is of great importance for both investors and policymakers. Such demand highlights the value of online listing information, which enables regular house price monitoring for almost every city in the country, especially the relatively smaller cities.

6. Potential extensions of the listing information applications

Besides the national-level monthly listing price indices for housing resales, in this section we briefly discuss two other potential applications of the online listing information in house price monitoring.

(*footnote continued*)
intentionally sets a low listing price, which attracts multiple buyers to compete for the house and results in a transaction price above the listing price. Based on the micro-level dataset, we find that the share of such "bidding wars" is around 3.6% from 2015 to 2017 in these 16 cities, which is quite low and similar to the share in the U.S. between 1986 and 1994 (Han and Strange, 2014). More importantly, following Han and Strange (2014), we analyze the relationship between the bidding war share and housing market conditions at the city-level using the same specification as Han and Strange (2014), and find no statistically significant correlations. In sum, the bidding wars are rare and stable, hence, unlikely to bias our listing price index. The regression results mentioned above are available upon request.

[9] The statistics of urban population and floor area come from the 2010 Population Census. We also use the average listing housing resale price in 2015 to infer the market value of housing stock. In 274 cities, the over 200 relatively smaller cities accounted for 34.5% of the urban housing stock by total market value.

[10] We also try adopting the imputed total market value of housing stock in 2015 as the weight, and plot the aggregated monthly indices in Fig. A2.

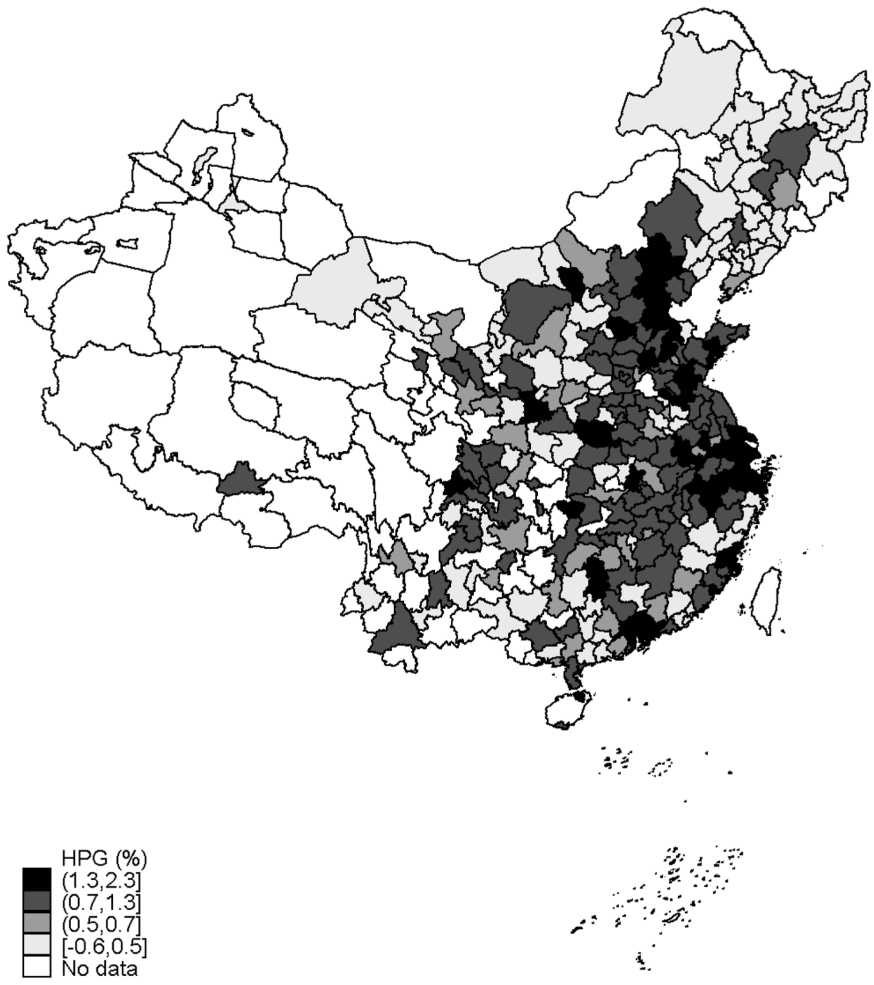

Fig. 4. Compound average monthly growth rate of housing resale price indices in 274 cities (January 2015 to September 2019).

6.1. Daily house price index

Most existing house price indices are on monthly or quarterly basis, but a growing body of literature highlights that a high-frequency (e.g., daily) house price index is also of great importance. The high-frequency house price index has been adopted to investigate the relationship between house price and related financial asset price (Anenberg and Laufer, 2017; Deng et al., 2017), to predict the subsequent house price changes (Bollerslev et al., 2016), and in academic researches on the pattern of housing market dynamics (Deng et al., 2018). Hitherto, only a few attempts have been made in China to calculate a daily house price index, covering only one or two of the most important cities (Deng et al., 2018).

The timeliness and richness of online listing information make it particularly suitable for constructing a daily house price index. In general, the hedonic model adopted here is consistent with Eqs. (2) and (3). The only difference is that, following Bollerslev et al. (2016) and Deng et al. (2018), we further adopt the standard Kalman filter to smooth the original daily house price indices obtained from the hedonic method. To be specific, let $P_{i,t}^*$ denote the price index in city i at time t based on the hedonic method, which can be decomposed into two parts:

$$\ln P_{i,t}^* = \ln P_{i,t} + w_{i,t} \tag{4}$$

where $P_{i,\ t}$ is the true price index and $w_{i,\ t}$ represents serially uncorrelated random errors. Assuming that the logarithm of the true price follows a random walk with drift, we have:

$$r_{i,t} = \Delta \ln P_{i,t} = u_i + v_{i,t} \tag{5}$$

$$r_{i,t}^* = \Delta \ln P_{i,t}^* = r_{i,t} + w_{i,t} - w_{i,t-1} = u_i + v_{i,t} + w_{i,t} - w_{i,t-1} \tag{6}$$

where $v_{i,\ t}$ represents random errors, $w_{i,\ t}$ and $v_{i,\ t}$ are i.i.d. with normal distributions and are mutually uncorrelated. The true price series $r_{i,\ t}$ can be obtained by the Kalman filter.

As a first attempt, Fig. A3 depicts the daily listing price index (Panel A) in Beijing between January 1, 2015 and December 31, 2017, along with the corresponding monthly price index (Panel B). Compared with the monthly index series, the daily index provides much more details on subtler market fluctuations, although the potential applications of the daily price information are beyond the scope of the current paper. It is noteworthy that, based on online listing information, we are also able to construct daily house price indices for a large number of other Chinese cities.

6.2. Rental price index

The online listing information also applies to the monitoring of housing rental prices. Currently, most rental transactions in urban China are not officially registered in local housing authorities, making it even more challenging to collect information on rental transaction prices. Thus, so far, there is no official multiple-city housing rental price index in China. However, at least in the major cities with relatively more matured housing rental markets, landlords and brokers also typically circulate listing information for the rental units on websites. Therefore, it is feasible to adopt the same procedures described in Section 3 to collect and clean the online rental listing information and calculate the constant-quality rental price index.

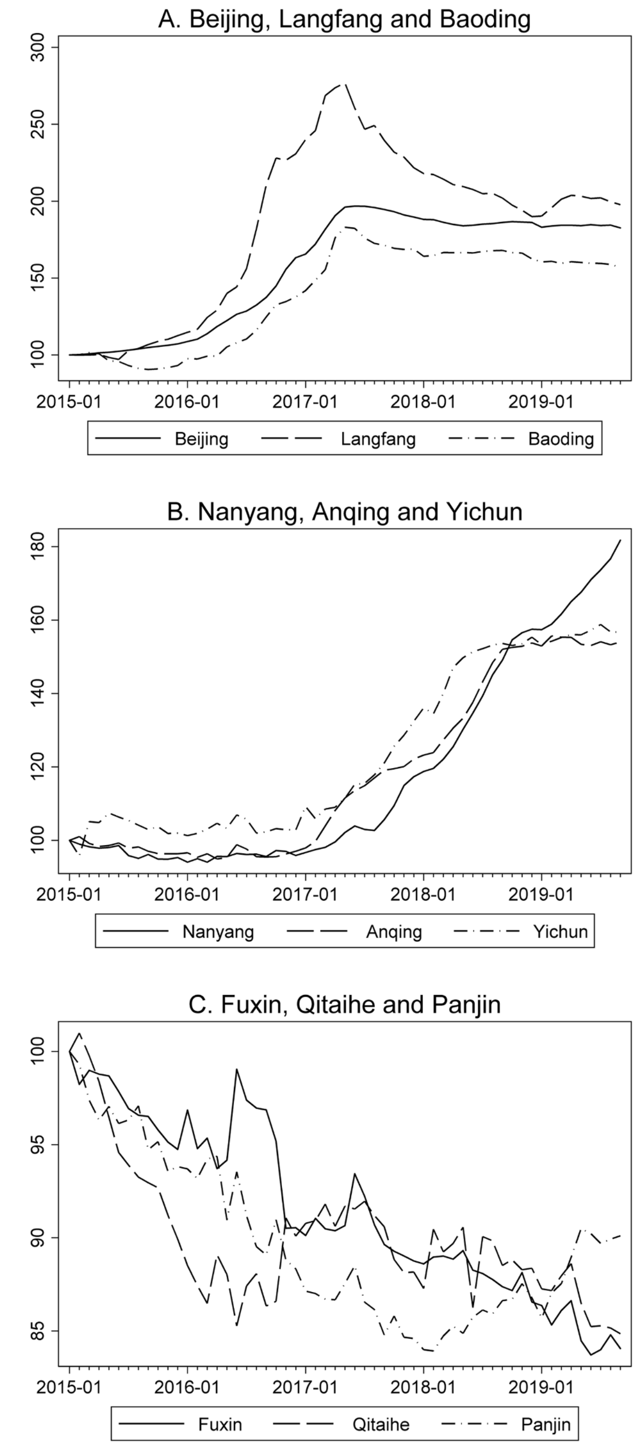

Fig. 5. Monthly listing price indices for three representative groups of cities (January 2015 to September 2019).

As one example, we adopt the online listing information to calculate the monthly rental price index for each of the 35 major cities[11] between January 2015 and September 2019. Among various potential applications of the rental price index, here we highlight its usage in bubble gauge: the comparison between the resale and rental price indices can provide preliminary but valuable information on the potential mispricing in the housing market. In Fig. A4, we depict the compound average growth rates of both resale and rental price indices for each of the 35 major cities during the sample period. It is clear that, for almost all these major cities, the resale price grew faster than the rental price during the sample period. For instance, the compound average monthly growth rate of the resale price index reached 1.53% in Shenzhen, substantially higher than the corresponding growth rate of 0.59% for the rental price index. Although more conclusive judgment still relies on stricter empirical analysis, the pattern shown in Fig. A4 at least implies that the divergence between house prices and the fundamental factors is still expanding during recent years in major Chinese cities.

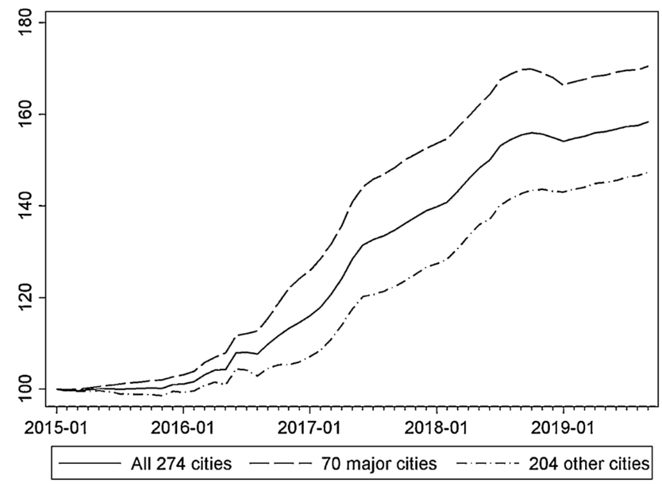

Fig. 6. Aggregated housing resale price indices (January 2015 to September 2019; weighted by urban population).

7. Conclusions

In this study, we introduce online listing information as an innovative data source for house price index construction, taking China's housing resale market as an example. As well as its inherent advantage of high data accessibility, our results show that online listing information can also accurately reflect the changes in actual transaction prices in the housing resale markets, especially after the potential replicated and manipulated data problems are adequately solved based on our proposed approach. This new data source can thus greatly benefit the construction of a timely, accurate house price index with broad coverage, especially for relatively smaller cities.

We highlight the remarkable advantages of high accessibility of the online listing information in house price index construction, and provide the first set of house price indices covering almost all cities in China. In particular, for around 200 3rd or 4th tier cities, the index provides the first indicator regularly tracing their monthly house price dynamics, and thus can significantly benefit the market analysis and research in these markets. We also point out a significant divergence in house price dynamics between different tiers of cities. According to the results, the existing national-level house price indices based on major cities only might lead to a biased understanding of the overall housing market conditions in China. Additionally, we briefly discuss the potential extensions in housing market analysis based on the listing information. First, we provide an example of a daily listing price index in Beijing, which can provide rich details in short term house price dynamics. Second, the listing price information can also be used to trace housing rental price changes, which enriches the monitoring of the housing market, and makes it possible to compare dynamics of resale prices and rental prices.

[11] These 35 cities are Beijing, Tianjin, Shijiazhuang, Taiyuan, Hohhot, Shenyang, Dalian, Changchun, Harbin, Shanghai, Nanjing, Hangzhou, Ningbo, Hefei, Fuzhou, Xiamen, Nanchang, Jinan, Qingdao, Zhengzhou, Wuhan, Changsha, Guangzhou, Shenzhen, Nanning, Haikou, Chongqing, Chengdu, Guiyang, Kunming, Xi'an, Lanzhou, Xining, Yinchuan, and Urumqi. The rental markets are relatively more active in these major cities.

The listing price index developed in this study can serve as a basis for several future researches. One the one hand, researchers can make efforts to further improve the listing price index; in particular, a more comprehensive understanding on the relationship between listing and transaction price indices is called for. On the other hand, the listing price indices can facilitate empirical analyses in several aspects. For instance, researchers can gauge the real estate market risks in the small cities using the listing price index with broad coverage, while the daily house price index can be used for prediction analysis in the financial market. Besides, scholars can also apply the rental price index to identify the divergence between house prices and fundamental factors.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 91,546,113 and 71,874,093] and Tsinghua University Initiative Scientific Research Program.

Appendix

Figs. A1–A4.

CRediT authorship contribution statement

Xiaodan Wang: Conceptualization, Methodology, Formal analysis, Investigation, Conceptualization, Writing - original draft, Writing - review & editing. **Keyang Li:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Jing Wu:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

We appreciate the comments from the editor and the anonymous referee, and seminar participants at the 2017 AsRES and GCREC Joint Annual Conference, the 2017 CAREA Annual Conference, and the 2018 APRES Symposium.

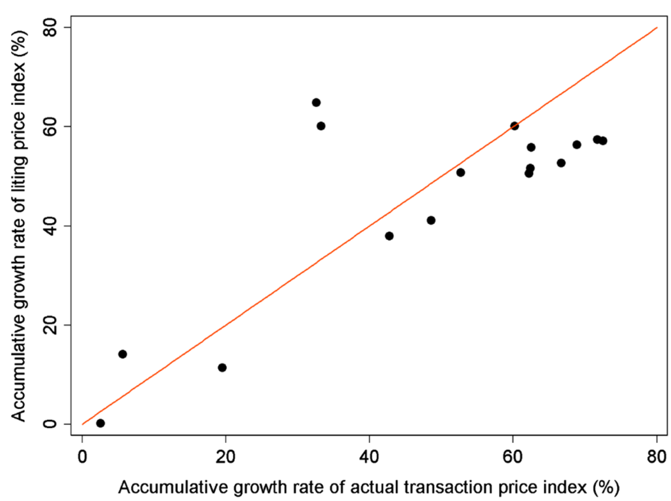

Fig. A1. Cumulative growth rates of actual transaction price index and listing price index (16 sample cities; January 2015 to December 2017).

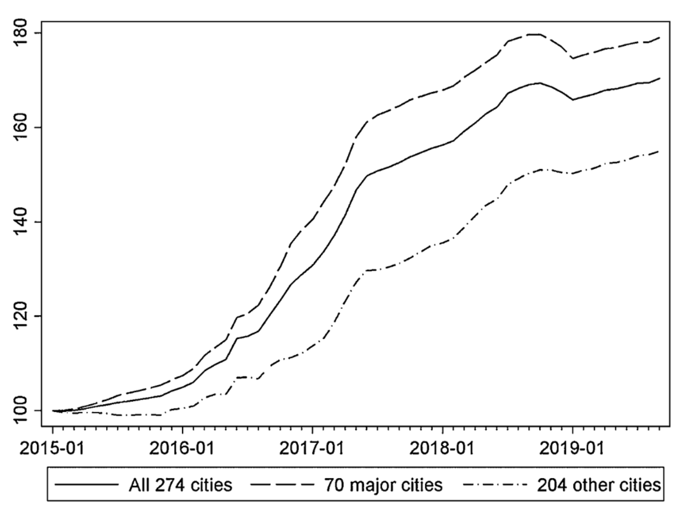

Fig. A2. Aggregated housing resale price indices (January 2015 to September 2019; weighted by market value of housing stock).

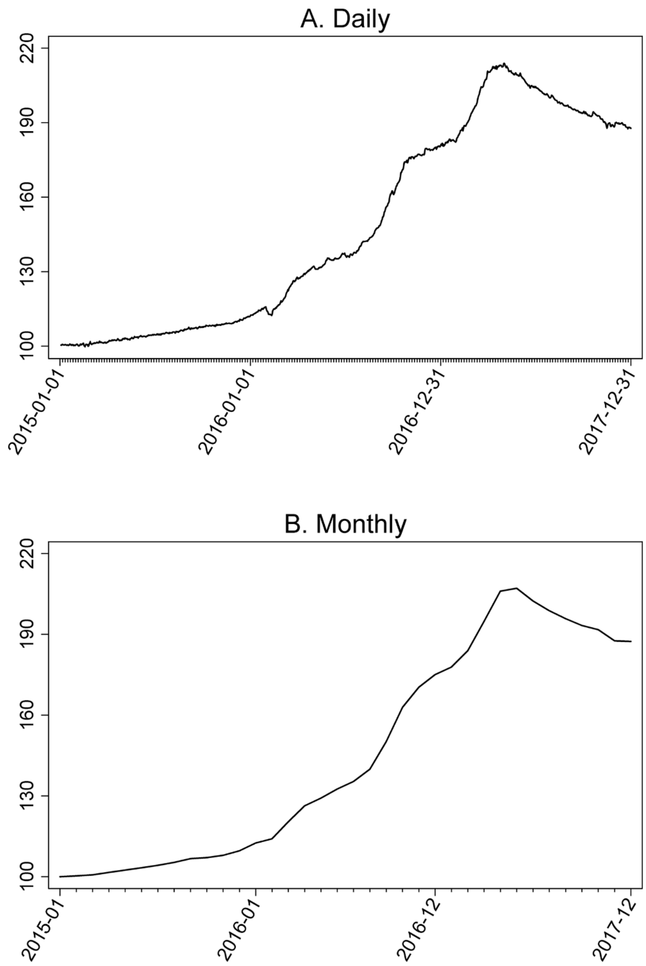

Fig. A3. Daily and monthly house price indices in Beijing (January 2015 to December 2017).

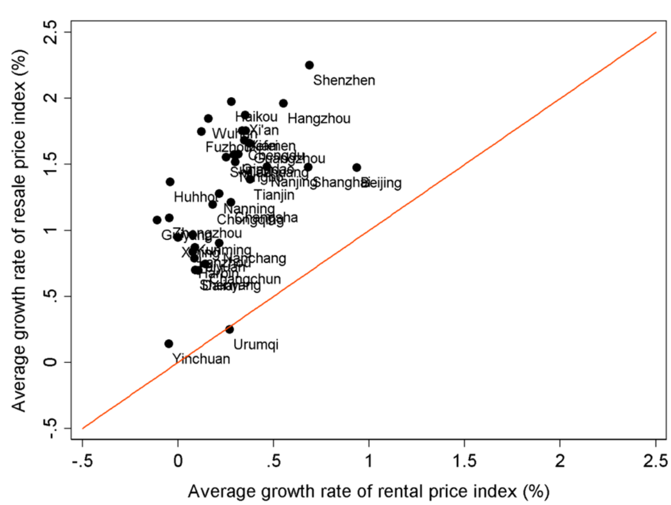

Fig. A4. Comparison of the monthly average growth rates of resale and rental price indices (35 major cities; January 2015 to September 2019).

References

Agarwal, S., Li, K.Y., Qin, Y., Wu, J., Yan, J.B, 2020. Tax evasion, capital gains taxes, and the housing market. J. Public Econ. 188, 104222. https://doi.org/10.1016/j.jpubeco.2020.104222.

Anenberg, E., Laufer, S., 2017. A more timely house price index. Rev. Econ. Stat. 99 (4), 722–734. https://doi.org/10.1162/REST_a_00634.

Bollerslev, T., Patton, A.J., Wang, W, 2016. Daily house price indices: construction, modeling, and longer-run predictions. J. Appl. Econom. 31 (6), 1005–1025. https://doi.org/10.1002/jae.2471.

Bucchianeri, G.W., Minson, J.A, 2013. A homeowner's dilemma: anchoring in residential real estate transactions. J. Econ. Behav. Organ. 89, 76–92. https://doi.org/10.1016/j.jebo.2013.01.010.

Case, K., Shiller, R., 1987 Sep Sep. Prices of single-family homes since 1970: new indexes for four cities. N. Engl. Econ. Rev. 45–56.

Ciarlone, A., 2015. House price cycles in emerging economies. Stud. Econ. Finance 32 (1), 17–52. https://doi.org/10.1108/sef-11-2013-0170.

Davis, M.A., Lehnert, A., Martin, R.F, 2008. The rent-price ratio for the aggregate stock of owner-occupied housing. Rev. Income Wealth 54 (2), 279–284. https://doi.org/10.1111/j.1475-4991.2008.00274.x.

Dai, Y.K., Xu, Y.F., 2018. Regulations enlarge inequality: evidence from Yin and Yang contracts on Beijing's housing market. Work. Pap.

Deng, Y., Girardin, E., Joyeux, R, 2018. Fundamentals and the volatility of real estate prices in China: a sequential modelling strategy. China Econ. Rev. 48, 205–222. https://doi.org/10.1016/j.chieco.2016.10.011.

Deng, Y., Girardin, E., Joyeux, R., Shi, S.P, 2017. Did bubbles migrate from the stock to the housing market in China between 2005 and 2010? Public Econ. Rev. 22 (3), 276–292. https://doi.org/10.1111/1468-0106.12230.

Fang, H., Gu, Q., Xiong, W., Zhou, L.A, 2016. Demystifying the Chinese housing boom. NBER Macroecon. Annu. 30 (1), 105–166. https://doi.org/10.1086/685953.

Francke, M.K., 2010. Repeat sales index for thin markets: a structural time series approach. J. Real Estate Finance Econ. 41 (1), 24–52. https://doi.org/10.1007/s11146-009-9203-1.

Gallin, J., 2008. The long-run relationship between house prices and rents. Real Estate Econ. 36, 635–658. https://doi.org/10.1111/j.1540-6229.2008.00225.x.

Guo, X.Y., Zheng, S.Q., Geltner, D., Liu, H.Y, 2014. A new approach for constructing home price indices: the pseudo repeat sales model and its application in China. J. Hous. Econ. 25, 20–38. https://doi.org/10.1016/j.jhe.2014.01.005.

Han, L., Strange, W.C., 2014. Bidding wars for houses. Real Estate Econ. 42 (1), 1–32. https://doi.org/10.1111/reec.12015.

Hill, R.J., Scholz, M., 2017. Can geospatial data improve house price indexes? A hedonic imputation approach with splines. Rev. Income Wealth. https://doi.org/10.1111/roiw.12303.

Horowitz, J.L., 1992. The role of the list price in housing markets: theory and an econometric model. J. Appl. Econ. 7 (2), 115–129. https://doi.org/10.1002/jae.3950070202.

Jiang, L., Phillips, P.C.B., Yu, J, 2015. New methodology for constructing real estate price indices applied to the Singapore residential market. J. Bank. Finance 61, 121–131. https://doi.org/10.1016/j.jbankfin.2015.08.026.

Kain, J.F., Quigley, J.M., 1970. Measuring the value of housing quality. J. Am. Stat. Assoc. 65 (330), 532–548. https://doi.org/10.1080/01621459.1970.10481102.

Karato, K., Movshuk, O., Shimizu, C, 2015. Semiparametric model of hedonic housing prices in Japan. Asian Econ. J. 29 (4), 325–345. https://doi.org/10.1111/asej.12077.

Knight, J.R., Sirmans, C.F., Turnbull, G.K, 1994. List price signaling and buyer behavior in the housing market. J. Real Estate Finance Econ. 9 (3), 177–192. https://doi.org/10.1007/bf01099271.

Knight, J.R, 2002. Listing price, time on market, and ultimate selling price: causes and effects of listing price changes. Real Estate Econ. 30 (2), 213–237. https://doi.org/10.1111/1540-6229.00038.

Kontrimas, V., Verikas, A., 2006. Tracking of doubtful real estate transactions by outlier detection methods: a comparative study. Inf. Technol. Control 35 (2), 94–105.

Liao, W.-.C., Wang, X, 2012. Hedonic house prices and spatial quantile regression. J. Hous. Econ. 21 (1), 16–27. https://doi.org/10.1016/j.jhe.2011.11.001.

Liu, C., Xiong, W., 2018. China's real estate market. Work. Pap.

McMillen, D., Thorsnes, P., 2006. Housing renovations and the quantile repeat-sales price index. Real Estate Econ. 34 (4), 567–584. https://doi.org/10.1111/j.1540-6229.2006.00179.x.

McMillen, D, 2014. Local Quantile House Price Indices. University of Illinois, Mimeo.

Merlo, A., Ortalo-Magné, F., Rust, J, 2015. The home selling problem: theory and evidence. Int. Econ. Rev. (Philadelphia) 56 (2), 457–484. https://doi.org/10.1111/iere.12111.

Morano, P., Tajani, F., 2014. Least median of squares regression and minimum volume ellipsoid estimator for outliers detection in housing appraisal. Int. J. Bus. Intell. Data Min. 9 (2), 91–111. https://doi.org/10.1504/IJBIDM.2014.065074.

Northcraft, G.B., Neale, M.A, 1987. Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions. Organ. Behav. Hum. Decis. Process. 39 (1), 84–97. https://doi.org/10.1016/0749-5978(87)90046-x.

Shimizu, C., Nishimura, K.G., Watanabe, T, 2016. House prices at different stages of the buying/selling process. Reg. Sci. Urban. Econ. 59, 37–53. https://doi.org/10.1016/j.regsciurbeco.2016.04.001.

Silver, M, 2016. Real-estate price indexes. Availability, importance, and new developments. Reality, data, and space. Int. J. Stat. Geogr. 7 (1), 4–25.

Sinai, T., Souleles, N.S., 2005. Owner-occupied housing as a hedge against rent risk. Q. J. Econ. 120 (2), 763–798. https://doi.org/10.1093/qje/120.2.763.

Wu, J., Deng, Y.H., Liu, H.Y, 2014. House price index construction in the nascent housing market: the case of China. J. Real Estate Finance Econ. 48 (3), 522–545. https://doi.org/10.1007/s11146-013-9416-1.

Wu, J., Gyourko, J., Deng, Y.H, 2016. Evaluating the risk of Chinese housing markets: what we know and what we need to know. China Econ. Rev. 39, 91–114. https://doi.org/10.1016/j.chieco.2016.03.008.

Zhu, E.W., Wu, J., Liu, H.Y., Li, X.D, 2019. Within-city spatial distribution, heterogeneity and diffusion of house price: evidence from a spatiotemporal index for Beijing. Real Estate Econ. https://doi.org/10.1111/1540-6229.12293.