J Real Estate Finan Econ (2015) 50:289–306
DOI 10.1007/s11146-014-9493-9

Intercity Information Diffusion and Price Discovery in Housing Markets: Evidence from Google Searches

Jing Wu · Yongheng Deng

Published online: 25 January 2015

Abstract We provide an innovative measure of information flow in Chinese housing markets based on search records from the Internet search engine Google. The measure depicts a substantial flow of house-price related information from national "superstar" cities, such as Beijing and Shanghai, and regional "star" cities, such as Tianjin and Chongqing, to other "normal" cities. The empirical results based on Granger causality test and turning point detection analysis both suggest that such information diffusion is a key factor that influences the intercity house price discovery process in the short run. The "superstar" and "star" cities lead the country in terms newly-built house prices changes in the sample period between 2006 and 2011.

Keywords Information diffusion · Price discovery · House price · Google searches

Introduction

The phenomenon of information diffusion and price discovery in asset markets has been well-documented in numerous studies. Typically, all of the information available about an asset is not immediately and simultaneously reflected in its price in all markets because of market inefficiencies. Instead, market price formation firstly occurs in a particular market, before such price signals are transmitted from this leading market to other markets, significantly affecting price formation in those lagging markets (Grossman and Stiglitz 1976; Hong and Stein 2007).

J. Wu
Hang Lung Center for Real Estate and Department of Construction Management, Tsinghua University, Beijing, China
e-mail: ireswujing@tsinghua.edu.cn

Y. Deng (✉)
Institute of Real Estate Studies, NUS Business School, and School of Design and Environment, National University of Singapore, Singapore, Singapore
e-mail: ydeng@nus.edu.sg

In the housing market, the prior literature has provided extensive empirical evidence on the interdependence of house prices across different markets. However, the pattern of information spillover and its effects can only be examined indirectly via certain tests because of the lack of an accurate measure of information flow. In this paper, we seek to fill this gap by providing a direct measure of the direction and density of intercity information flow in housing markets, by analyzing records of web search queries (Da, Engelberg, and Gao 2011; Bank, Larch, and Peter 2011). Using the search query data provided by "Google Trends", we built an "Information Flow Index", which measures the propensity of Google users to focus on a certain city when they are searching for house-price related information on the web. This unique method reveals the pattern of information dissemination across various housing markets.

Using the Google Trends data for mainland China between 2004 and 2011 as an example, our information flow index suggests that, while house price information in most "normal" cities does not spread beyond their provinces, a few major cities have substantial influence at the regional or even national level. A significant amount of house price information flows from three national "superstar" cities- Beijing, Shanghai and Shenzhen- to almost all other provinces. There are also several regional "star" cities, such as Tianjin, Chongqing and Wuhan, which mainly attract the attention of market participants in nearby provinces.

The empirical analysis based on the newly-built house price series in 35 major cities also suggests that, the pattern of information spillover is significantly correlated with the spatial pattern of intercity house price discovery. First, the Granger causality in short-run house price changes is more likely to exist if the information is transmitted from a leading city to a laggard city. As a consequence, house price changes in "superstar" cities such as Beijing Granger-cause house price changes in most of the other cities. Second, we follow Ferreira and Gyourko (2011) and identify the key time points during the recent housing boom during 2009–2010 in each city. On average, the whole process of the boom in "superstar" cities is about 3–4 months before the "star" cities, and 6–8 months before the "normal" cities.

The key contribution of this paper is to provide a novel and intuitive measure of information flow in the housing market. This measure enables a detailed description of the spatial pattern of the dissemination of house price information for the first time. It also demonstrates the role of information as a factor in intercity house price discovery. This research contributes to the growing literature on house price discovery, and is among one of the first empirical studies on the pattern of intercity house price discovery in mainland China.

The paper proceeds as follows. Literature Review section provides a brief review of research on information diffusion and price discovery in the housing market. The Pattern of Intercity Information Diffusion in Chinese Housing Markets section explains how the Information Flow Index was built and describes the key features of the pattern of information diffusion in Chinese housing markets based on the index. Empirical Results for Intercity House Price Discovery section provides two empirical examples of intercity price discovery in China, and discusses how they are related to the information spillover pattern. Concluding Remarks section concludes the paper.

Literature Review

The diffusion of house price changes across different markets (or different segments in a market) has long been of interest in the housing literature. Hitherto, researchers have focused on two aspects of the phenomenon. The first is the relationship between securitized (public) and unsecuritized (private) real estate markets. In general, most research points out that the price/return changes in the securitized sector lead changes in the unsecuritized sector (Giliberto 1990; Gyourko and Keim 1992; Barkham and Geltner 1995; Yunus, Hansz, and Kennedy 2012), although there are a few counterexamples, such as Tuluca, Myer, and Webb (2000). The second aspect of the phenomenon are spatial patterns of diffusion. The results suggest that house price changes can diffuse between contiguous areas (Clapp, Doldo, and Tirtiroglu 1995; Pollakowski and Ray 1997; Holly, Pesaran, and Yamagata 2011), or from certain "core" countries/cities/neighborhoods to others (Meen 1999; Oikarinen 2004; Bandt, Barhoumi, and Bruneau 2010). Recently, researchers expanded the scope to include other related issues. For example, the diffusion of price changes has been shown to exist between the land and housing markets (Ooi and Lee 2004; Chau et al. 2010), the public and private housing sectors (Ong and Sing 2002), various quality tiers (Ho, Ma, and Haurin 2008), and spot and presale housing markets (Wong, Chau, and Yiu 2007).

The literature also provides several possible explanations for the cross-market interdependence of house prices, where information is expected to play an important role, especially in the short run. The logic dates back to the work of Grossman and Stiglitz (1976), which later became the theory of "gradual information flow" (Hong and Stein 1999, 2007). Housing markets exhibit different levels of efficiency – some markets/segments react faster to newly available information than others and house prices adjust more quickly, because they have more experienced participants, more frequent transactions, or lower information costs (Case and Shiller 1989; Gyourko and Keim 1992; Clapp, Doldo, and Tirtiroglu 1995; Oikarinen 2004; Deng and Quigley 2008; Ferreira and Gyourko 2012). By contrast, some markets react relatively slowly to new information. One way to ameliorate this delayed response for the participants in these markets is to learn from leading markets via price signals. By this means, information will spread from the leading to the lagging markets, and facilitate price adjustment in the latter.

However, the absence of a direct and reliable measure of information flow means that the pattern and effect of information diffusion in the housing market can only be indirectly investigated. One approach is to learn from the financial literature and view an unexpected component in asset returns/prices in the leading markets as a signal of "news". A recent example is Chau et al.'s (2010) use of the unexpected outcome of land auctions in Hong Kong as a proxy of "new information" of the land market, which has been shown to have a significant effect on house prices. Other studies have created proxies for information cost. For instance, Clapp, Doldo, and Tirtiroglu (1995) use population density as a proxy for information cost and conclude that house prices change more quickly in cities that are denser compared to those that are less dense. Nevertheless, these indirect analyses have been unable to provide a detailed picture of information diffusion, or direct evidence of its effect on the price discovery process.

Research in the field of financial offers more ideas for measuring information flow. While some studies have analyzed news from traditional media outlets, such as the *Wall*

Street Journal or *Dow Jones Newswire* (Tetlock 2007, 2010; Tetlock, Saar-Tsechansky, and Macskassy 2008; Fang and Peress 2009), recent studies have begun using instruments that rely on new channels of information, such as web search engines. Da, Engelberg, and Gao (2011) were the first to suggest using the volume of Google searches as an information indicator in financial markets, and their empirical research suggests that this indicator can predict stock prices in the short run. Bank, Larch, and Peter (2011), Mondria and Wu (2011) and Dzielinski (2012) used a similar proxy measure based on various topics in the financial markets. We follow this strategy in this paper, and fill the gap in the literature on housing by providing a direct measure of information flow in housing markets, using mainland China as our example. Our study focuses on the spatial pattern of intercity information spillovers, which has seldom been analyzed in the previous financial literature but is especially important for housing markets.

The Pattern of Intercity Information Diffusion in Chinese Housing Markets

Internet Search Engines as an Emerging Information Channel in China

Like many other large economies, the Internet is playing an increasingly important role in China's economic and social development. After continuous and rapid growth, China has the largest number of Internet users, relative to all other countries. The China Internet Network Information Center (CNNIC)[1] reported that there were 513.1 million Internet users in China at the end of 2011, making up 38.07 % of the population,[2] with each user spending 18.7 h on the Internet per week on average. The Internet is especially popular among educated and young people, who are also the largest group of potential home buyers. The CNNIC report pointed out that more than 95 % of graduates and some 60 % of individuals aged 20–39 used the Internet in 2011.

Among the different ways of using the Internet, a web search engine is one of the most important tools for Chinese Internet users. 79.4 % of the respondents to CNNIC's 2011 survey listed "*searching for information* via *web search engines*" as one of their major activities on the Internet,[3] second only to "*instant messaging*" (80.9 %), among the 18 options. This implies that over 400 (513.1 * 79.4 % = 407.4) million Chinese people are searching for information on the Internet. While it is too early to say whether (or to what extent) web search engines will eventually replace traditional media, they have already become an important information dissemination channel in China. In addition, since house prices are a key issue of concern in China today, it is reasonable to expect web search engines to play an important role in the diffusion and spillover of house price information.

[1] CNNIC is a non-profit organization with support from China's Ministry of Information Industry and the Chinese Academy of Sciences. Since 1997, CNNIC has been publishing the "Statistical Report on Internet Development in China" semi-annually. The latest version of the report was published in January 2012 and is available on CNNIC's official website (www.cnnic.cn).

[2] Estimates of the number of Internet users in China by other institutes, such as the National Bureau of Statistics of China and the World Bank, are very close to CNNIC's figures.

[3] This figure has ranged between 70 and 80 % in CNNIC's annual survey over the last 5 years.

Compared to traditional media, such as newspapers or television, web search engines are especially helpful for understanding the pattern of information diffusion in the housing market for at least two reasons. First, it is almost impossible (or at least very costly) to accurately identify how many people have acquired certain pieces of information from newspapers or television, where these people are located, and when they obtained such information, all of which are essential for understanding the effect of information diffusion (Engelberg and Parsons 2011). In contrast, technical details of all web queries, including the originating IP addresses and date/time, are automatically recorded on a search engine's servers, and can be used to accurately measure the spatial and temporal distribution of users' queries for certain information, thus providing a detailed picture of the flow of information. Second, information disseminated on the web via search engines is target-oriented: while individuals may obtain information from newspaper articles or television programs simply by chance, even if they are not interested at all, the opposite is true for web searches. In most cases, individuals search for a certain piece of information only if they need it. For instance, most queries for "*house price*" should come from (potential) sellers and buyers of housing units, developers, brokers, market analysts, or policy makers, all of whom are participants in the housing market. Hence, information diffused via the Internet is more likely to affect the following dynamics in housing markets.[4]

Fortunately several major internet search companies have been collecting and providing web search statistics. We use the data provided by "Google Trends" (www.google.com/trends/) to build our measure of information flow in Chinese housing markets.[5] "Google Trends" is a free service provided by Google since May 2006, and (currently) covers query records from January 2004 to the present. For any given term, "Google Trends" can report its "Search Volume Index", which measures how often this term has been searched for in Google in a particular region and period.[6] This is the raw input for our information flow measure.[7]

[4] There are two potential problems. First, the pattern of information diffusion revealed via web search records may over-sample inexperienced market participants, since professional participants are more likely to use channels such as market analysis reports or the business media, instead of web search engines. However, given that buyers in the current Chinese housing market largely consist of inexperienced households, we believe the potential sampling problem will not substantially affect our results. Second, a measure of web searches based on web search records may underestimate the density of information flow from some of the largest cites with very developed housing markets. For example, cities such as Beijing and Shanghai are the focus of well-known professional property websites (e.g., www.soufun.com), which Internet users may visit without searching the broader Internet via search engines. Moreover, for mega cities like Beijing, people may search directly for information on house price changes in their districts (e.g., "*Haidian*" + "*house price*", instead of "*Beijing*" + "*house price*"). We acknowledge this as a limitation of the information flow index and leave it for future research.

[5] The leading web search engine in China, Baidu, also provides a similar service named "Baidu Index" (index.baidu.com). However, in most provinces, this index only date back to 2008. In addition, its calculation formula is opaque. Therefore in this study, we choose to rely on "Google Trends".

[6] More technical details on the "Search Volume Index" are available on the "About Google Trends" webpage (www.google.com/intl/en/trends/about.html); or see discussions in Da, Engelberg and Gao (2011).

[7] In the original work by Da, Engelberg and Gao (2011), web search frequency is adopted as a measure of "attention". Here, we treat it a proxy of information flow, with the assumption that Google users will browse at least some of the search results (websites) and hence get some related information.

The National Level Information Flow Index

We start with aggregate-level analysis by measuring the degree of influence of each city's house price information at the national level. This indicator can allow us to: (1) identify the most influential cities in Chinese housing markets, which will be the emphasis in the next analysis; and (2) test the reliability of this index by comparing it with measures based on other channels of information diffusion (Da, Engelberg, and Gao 2011).

A city's National Information Flow Index (*NIFI*) is defined as the propensity of all Google users in mainland China to focus on this city when they are searching for house-price related information. More specifically, for city i its index ($NIFI_i$) is calculated as the volume of searches of the combination of this city's name and the keyword of "*house price*" (*fang jia* in Mandarin)[8] from mainland China during the sample period, normalized by the total volume of searches for "*house price*" only in the same interval.[9] Thus, this index reflects the degree of relative importance of each city's house price information at the level of the overall country, while cities with higher scores in this index can be perceived to be more influential.

We calculate the index for all 287 cities in China during the sample period of 2004–2011.[10] The results are depicted on the map in Fig. 1. Instead of a time series for each city, we only have an aggregated index for the entire period. The average and standard deviation of the index are 0.18 and 0.69 %, respectively. However, these statistics mask a high level of heterogeneity in the influence of the various cities. For example, house price information in Shanghai and Beijing attracts much attention. The NIFI in these two cities reaches 7.5 and 6.5 %, which means that, on average, one of every 13 search requests for house price information from Chinese Google users is explicitly restricted to Shanghai, while one of every 15 searches only focuses on Beijing. At the other end, the index is lower than 1.0 % for 267 cities, which implies that the influence of their house price information is almost negligible at the national level. In 244 of these 267 cities, the index is reported to be very close to 0.0 %. Between these two extremes, 18 cities have NIFIs between 1.0 and 2.5 %, and could be expected to have a non-negligible influence at the national level. Several large cities, such as Shenzhen, Tianjin, Chongqing, Chengdu and Wuhan, appear in this group.

To verify whether this index can effectively capture information flow, we calculate another indicator based on a traditional information diffusion channel, newspapers. So far, we still cannot find an equivalent to the *Wall Street Journal* in China which dominates the business media at the national level. Hence, we choose to combine multiple sources as Engelberg and Parsons (2011). Based on the Genius Database (www.genius.com.cn), we count the total number of articles which have both each

[8] In this paper, we use the keyword "*house price*" (*fang jia*). The evidence from "Google Trends" suggests that this word is used much more frequently by Chinese users than other options. The number of search requests for these terms- "*price of house*" (*zhu fang jia ge*), "*real estate price*" (*fang di chan jia ge*) and "*building price*" (*fang wu jia ge*) – is only 1.5, 2.0 and 3.5 % respectively of the number of searches on "*house price*" (*fang jia*).

[9] We explicitly exclude the term "*hotel*" (*jiu dian* or *bin guan*) in our search, since in Mandarin, the term "*hotel rate*" is also "*fang jia*".

[10] In early 2010, Google closed its business in mainland China; after that, users in China could still access the Google server in Hong Kong.

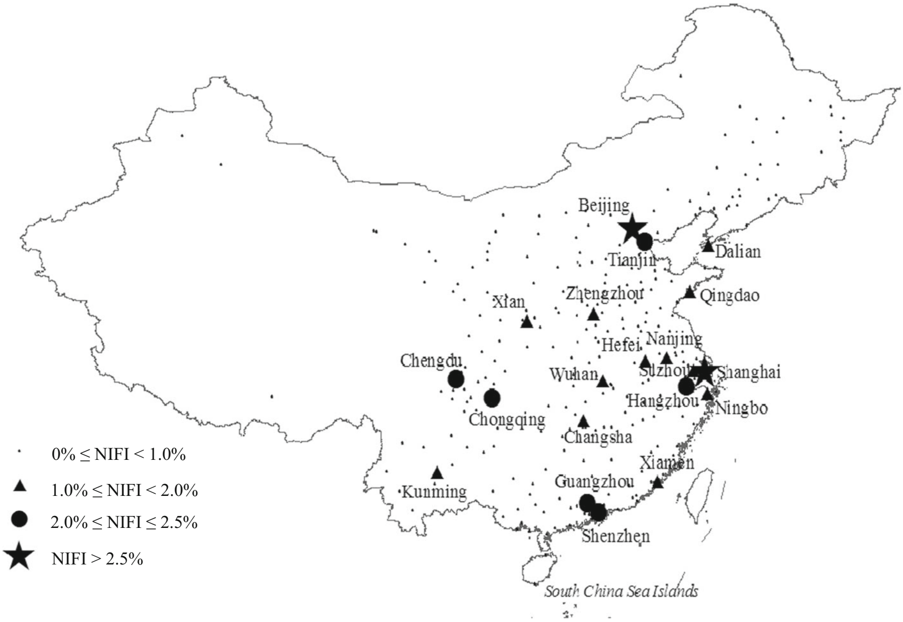

Fig. 1 National level information flow index

city's name and the keyword "*house price*" from more than 30 national Chinese newspapers in the period 2004 to 2011, normalized by the number of articles with the keyword "*house price*" only, and again for each city, we only calculate the aggregated index for the whole sample period. For our 287 sample cities, the correlation coefficient between the city-level "newspaper index" and the NIFI is 0.935. Therefore, at least at the aggregate level, the index based on "Google Trends" is consistent with the pattern revealed in traditional channels.

The Provincial Level Index and the Spatial Pattern of Information Diffusion

The national-level index suggests that 20 cities have substantial influence in the national housing market. A more important question is the spatial pattern through which information diffuses. Even with exactly the same NIFI, search requests may only come from users within the city, nearby contiguous areas, or from all around the country, which would result in totally different effects in housing markets. Therefore, we further decompose the national level index to the provincial level so that we can estimate each city's influence beyond its own province.[11]

The basic logic of the provincial Information Flow Index (PIFI) is consistent with the national indicator. In particular, the index capturing the flow of information from city i to province j ($PIFI_{i,j}$) is defined as the number of Google search requests with the combination of city i's name and the term "*house price*" from province j, normalized by

[11] Ideally, a city-level index would reflect the spatial pattern of information diffusion more clearly than the provincial-level indicator, especially in describing information flows within a province. However, the total number of Google search requests for "*house price*" is not large enough to report in many cities currently, leaving us unable to calculate city-level index.

the total volume of searches for the term "*house price*" from province *j* in the same time period. For each of the 287 cities, we obtain 30 provincial level indexes, including one for the city's original province.[12] A higher $PIFI_{i,j}$ value indicates a denser flow of house price information from city *i* to province *j*.

Again, the results suggest a large divergence between the cities.[13] For 258 of the 287 cities, the provincial level indexes are lower than 1.0 % in all provinces other than the cities' own provinces, suggesting that for most "normal" cities, house price information hardly diffuses beyond their province. This leaves only 19 cities with a considerable level of influence beyond their local province, which is our major interest.

The summary statistics of the provincial indexes for these 19 cities are listed in Table 1. Not surprisingly, all these cities have the highest index values in their own provinces. The other 29 provinces are divided into two groups to investigate the spatial pattern in more detail: provinces with any part of their jurisdictions within 500 km of the target city are defined as nearby provinces within the same region, and all other provinces defined to be outside the region.

According to the results, these 19 cities can be grouped into two tiers. The first tier, or the national "superstars", includes Beijing, Shanghai and Shenzhen, whose house price information have nationwide influence. As an extreme case, the PIFI for Beijing is quite high in all 8 provinces that are within a 500 km radius, and it also reaches at least 1.0 % in 20 of the 21 provinces beyond that distance (Fig. 2a). As a whole, the average PIFI in all 29 provinces is 3.59 %. This indicates that housing market participants in almost all provinces are watching changes in housing prices in Beijing. The situation is similar for Shanghai (Fig. 2b), with the average PIFIs in all 29 provinces being 2.45 %. In comparison, the influence of Shenzhen is weaker (Fig. 2c). It mainly influences the east and middle regions (especially the southern part), and the density of the information flows (i.e., the magnitude of PIFIs) is lower than Beijing and Shanghai's.

The second tier cities includes the other 16 cities, which are "regional stars". In general, these cities are very influential at the regional level, but their influence quickly declines with distance, hardly reaching provinces outside their region. A typical example is Tianjin (Fig. 2d). Its house price information is sought out across most provinces in northern China, with PIFIs reaching 1.0 % in 7 of the 8 provinces within 500 km, but only to a limited extent in other regions. Similar cases include Chongqing (Fig. 2e) and Chengdu in the southwestern region, Wuhan in the central region, Xian in the northwestern region, Guangzhou in the southern region (Fig. 2f), Nanjing and Hangzhou in the eastern region, and Dalian and Shenyang in the northeastern region. The influence of other cities like Hefei, Suzhou, Xiamen and Fuzhou are even smaller and concentrated in a few adjacent provinces.

These results lead to two questions. The first is why such a pattern exists. In particular, the existence of the three nationwide "superstars" is especially striking in a huge country like China: why would people in provinces such as Yunnan (southwest) or Xinjiang (northwest) keenly observe house prices in Beijing, which is over 2000 km

[12] There are 31 provinces (including 4 municipalities and 5 autonomous region) in mainland China. The Xizang (Tibet) Autonomous Region is not included in the following analysis because the volume of Google search requests for "*house price*" from that region is too small to report.

[13] The detailed results are available on request.

Table 1 Summary statistics of the provincial level information flow index (PIFI)

City	Local province	All other provinces			Other provinces within 500 km			Other provinces beyond 500 km		
		Total number of provinces	Number of provinces with PIFI above/in 1.0 %	Average of PIFI	Total number of provinces	Number of provinces with PIFI above/in 1.0 %	Average of PIFI	Total number of provinces	Number of provinces with PIFI above/in 1.0 %	Average of PIFI
Beijing	22.00 %	29	28	3.59 %	8	8	5.75 %	21	20	2.76 %
Shanghai	34.00 %	29	27	2.45 %	6	6	3.92 %	23	21	2.07 %
Shenzhen	16.00 %	29	19	0.86 %	5	4	1.30 %	24	15	0.77 %
Chongqing	45.50 %	29	11	0.74 %	7	5	1.50 %	22	6	0.50 %
Chengdu	35.50 %	29	10	0.72 %	7	5	1.29 %	22	5	0.55 %
Tianjin	50.00 %	29	10	0.60 %	8	7	1.44 %	21	3	0.29 %
Wuhan	30.50 %	29	6	0.41 %	8	3	0.69 %	21	3	0.31 %
Xian	34.00 %	29	5	0.50 %	8	2	0.81 %	21	3	0.38 %
Guangzhou	13.00 %	29	5	0.43 %	5	3	0.90 %	24	2	0.33 %
Nanjing	16.00 %	29	4	0.41 %	8	3	0.81 %	21	1	0.26 %
Hangzhou	20.50 %	29	3	0.34 %	7	2	0.64 %	22	1	0.25 %
Dalian	20.00 %	29	3	0.33 %	6	2	0.75 %	23	1	0.22 %
Qingdao	14.00 %	29	3	0.28 %	7	1	0.50 %	22	2	0.20 %
Changsha	28.50 %	29	2	0.24 %	6	1	0.42 %	23	1	0.20 %
Hefei	25.00 %	29	2	0.17 %	9	2	0.39 %	20	0	0.08 %
Shenyang	15.50 %	29	2	0.14 %	4	2	0.63 %	25	0	0.06 %
Suzhou	11.50 %	29	1	0.17 %	6	1	0.50 %	23	0	0.09 %
Xiamen	19.50 %	29	1	0.16 %	4	1	0.50 %	25	0	0.10 %
Fuzhou	16.50 %	29	1	0.07 %	4	1	0.38 %	25	0	0.02 %

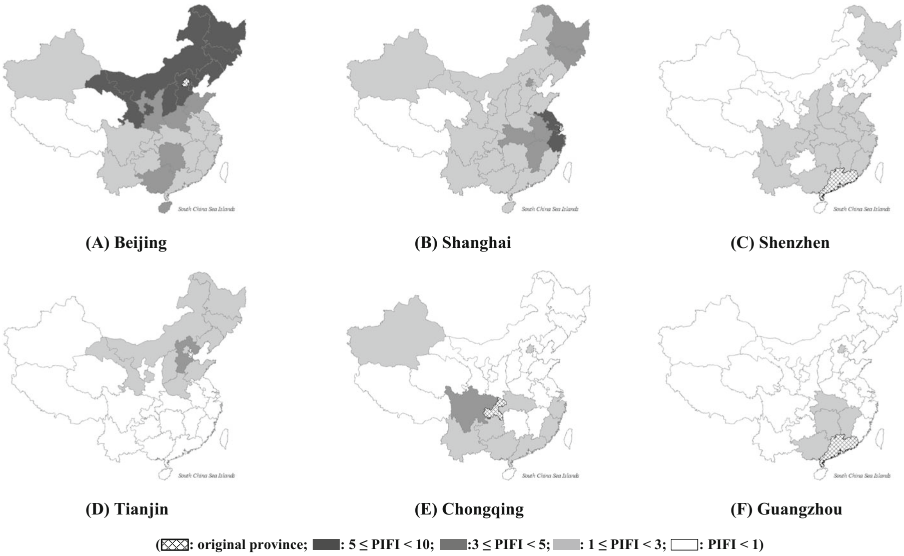

Fig. 2 Provincial level information flow indexes of select cities (a) Beijing (b) Shanghai (c) Shenzhen (d) Tianjin (e) Chongqing (f) Guangzhou

away? Moreover, a preliminary international comparison suggests this phenomenon may be unique to China. We apply a similar calculation method to the United States, and calculate both the national and state level information flow indexes for major cities such as New York and Los Angeles. However, we do not find any evidence for the existence of "superstars" whose house price information spills over to the majority of states[14]: for New York City, besides the state of New York, its home price information attracts limited attention from nearby states only, such as New Jersey, Pennsylvania and Massachusetts, but not any other states. Meanwhile, only people in California search online for information on Los Angeles' home prices.

One explanation, as suggested by Deng, Gyourko, and Wu (2012), is that the Chinese housing market is characterized by a strong national trend where prices tend to move the same direction across most markets in a given period, reflecting the strong influence of the national effect because of shifts in the macroeconomic environment, market sentiment, or the central government's policies. City-specific effects, by contrast, are found to be far less important in determining local house price change. This implies that understanding such a national trend is an important task for housing market participants to predict local house price changes, and a most feasible way to learn such information is from the "superstar" and "star" cities.[15]

The second question is the determinant of a specific city's influence: why do cities like Beijing, instead of some others, seize the top position in the hierarchy? A detailed

[14] We adopt the key word "*home price*" in the US analysis. The results based on other key words like "*house price*" are consistent.

[15] As a preliminary test of this explanation, we calculate provincial level indexes for house price information in Hong Kong. The results suggest that few people in mainland China (even in Guangdong Province, which is very close to Hong Kong) search for information about house prices in Hong Kong. This finding is consistent with the fact that the movement of house prices in Hong Kong does not share the same common trend with cities in mainland China (mainly because the central government's policies do not apply to Hong Kong).

investigation of this issue is beyond the scope of this research, but the statistics listed in Table 2 provide some preliminary evidence. According to the latest available statistics in 2009, the "superstar" cities have the most developed real estate industry (i.e., the highest share of employment in the real estate industry), the most active housing market, and the highest population density, followed by the regional "stars", which is consistent with the attributes of a more efficient market, as suggested in the literature. Therefore, new information should be captured by changes in house prices in these markets first, and then attract the attention of market participants in other cities.[16]

Empirical Results for Intercity House Price Discovery

After the analysis of the intercity house price information diffusion pattern, we now turn to the other side of the story and investigate the pattern of intercity price discovery in Chinese housing markets.

Data

The following analysis uses the constant quality price indexes of newly-built housing units in 35 major Chinese cities[17] from January 2006 to December 2011 provided by Tsinghua University (Wu, Deng, and Liu 2014). This index is calculated based on the full sample of micro-level data of newly-built housing transactions, while the conventional hedonic model is applied to control for potential quality changes. Hence, this index can be expected to reflect short-run house price changes more accurately than official price indexes without quality control.[18]

One major limitation of the data is that the sample period only covers 6 years, which is shorter than most studies that studied intercity price correlations. However, it is the only interval during which a reliable house price indicator is available in China. The following two factors can at least partially mitigate this limitation. First, the Chinese housing market has experienced large and frequent house price fluctuations in these 6 years (as an example, Fig. 3 depicts the annual growth rate calculated based on the aggregated index of these 35 cities). As a result, the sample period indicates a large variance in short-run house price changes, which is the focus of this analysis. Second, given the large number of transactions, we are able to use the monthly house price series in the analysis.[19] In addition, we choose to adopt two different approaches to achieve a robust result of the intercity price discovery pattern.

[16] Another possible explanation is that more people from other provinces plan to purchase housing units in these "superstar" or "star" cities, and thus search for related information in advance. The last column in Table 2 calculates the average proportion of home buyers from other provinces, and the pattern is consistent with this explanation. But, the differences between various tiers are only marginally significant.

[17] The 35 major cities include all the 3 "superstar" cities defined above, 15 of the 16 "star" cities with Suzhou as the only exemption, and 17 of the 268 "normal" cities. So far the constant quality house price indicator is not available for other cities.

[18] More details about this index and its comparison with the official house price indicators are reported in Wu, Deng and Liu (2014).

[19] The total volume of newly-built housing units transacted in these 35 cities in the sample period reaches 8.39 million, or 3330 units per city per month on average.

Table 2 Major features of cities in three tiers

	Percentage of urban employments in real estate industry	Per capita annual housing transaction volume (yuan)	Population density (population / sq.km.)	Percentage of home buyers from other provinces
Average of "Superstar" Cities	4.433 %	1.330	3221	22.283 %
Average of "Star" Cities	2.175 %	0.855	792	15.100 %
Average of "Normal" Cities	0.099 %	0.186	411	10.403 %
T Test Stat. for the Difference between "Superstar" and "Star"	3.53***	2.48**	4.99***	1.53
T Test Stat. for the Difference between "Star" and "Normal"	5.97***	11.61***	3.78***	1.28

Source: Authors' calculations based on statistics published by National Bureau of Statistics and Ministry of Housing and Urban–rural Development of China

Granger Causality Test

The first method is the standard Granger causality test (Granger 1969) which is widely applied in most existing studies of inter-market house price relationships. For city i and j, the model is estimated as:

$$\text{dlog}\left(PRICE_{i,t}\right) = \sum_{m=1}^{l} \alpha_m \cdot \text{dlog}\left(PRICE_{i,t-m}\right) + \sum_{m=1}^{l} \beta_m \cdot \text{dlog}\left(PRICE_{j,t-m}\right) + \varepsilon_t \quad (1)$$

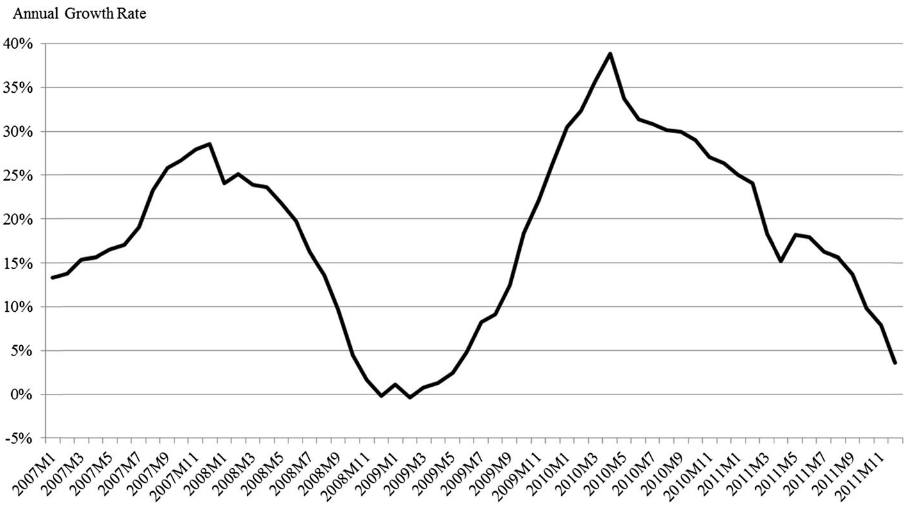

Fig. 3 Annual growth rate of aggregated house price index in 35 major cities

where the variable of dlog($PRICE_{i,t}$) is the log difference of city i's house price index in period t, which is stationary in all 35 cities during the sample period according to the conventional Augmented Dickey Fuller (ADF) test. The lag structure (l) is determined by the Akaike information criteria (AIC), with the maximum lag allowed of 6 months. Then if the null hypothesis that $\beta_1 = \ldots\ \beta_l = 0$ is rejected by the conventional F test, the price change in city j is expected to Granger-cause the price change in city i in the short run.[20]

The above procedures are applied to each of the 1190 (35*34) pairs between the 35 major cities. Granger causality is found to be significant at the 95 % or higher level in 466 pairs. Table 3 lists the distribution of these 466 pairs by the leading cities, which provide some evidences for the strong relationship between the price discovery pattern revealed by the Granger causality test and the information diffusion pattern discussed before. In general, both the national "superstars" and regional "stars" affect many cities, especially the former group. House price changes in Beijing can Granger- cause house price changes in 28 of the 34 cities in the short run, with the number for Shanghai also reaching 24; in other words, house price changes in these two cities can affect most of the other major cities around the country. The influence of Shenzhen is comparably weaker- it only leads 15 cities. As for the regional "stars", house price changes in each of them can, on average, Granger-cause house prices changes in about 14.8 other cities, which is fewer than the "superstars" (22.3 cities on average), but still higher than the 17 "normal" cities (10.4 cities on average), with both difference significance at the 90 % or higher confidence level.

As a more direct test of the correlation between information diffusion and house price discovery, we find causality in 72 of the 110 pairs, or 65.45 %, with significant information flow (i.e., the PIFI from the leading city to the province of the lagged city reaches at least 1.0 %). In contrast, causality only exists for 36.48 % of the other 1080 pairs (i.e., 394 pairs), which is lower than the proportion in the other group (the difference is significant at the 99 % confidence level). This also explains why the house price changes in the "superstars" and "stars" can affect much more cities than the "normal" cities.

Detection of Key Time Points in the Recent Housing Booming

The second method aims to reveal the price discovery pattern more intuitively. Following the estimation strategy developed by Bai (1997), Card, Mas, and Rothstein (2008), Ferreira and Gyourko (2011), we seek to identify the key time points in the recent housing booming during 2009 and 2010, which allow us to compare the whole process of this boom in different cities.

For each city, we define three key time points in the boom process. The first point is when the market witnessed the first signal of recovery from the "recession" during the financial crisis, $T_{i,recover}$. This point is defined as the month with the lowest annual house price growth rate during 2008 to 2010. The second point is the end of the boom

[20] The standard Granger causality test augmented with error correction terms, or the VEC approach, is suggested in some studies if the price levels (in log term) in the two cities are cointegrated. However, in this analysis, the sample period is too short to test for cointegration and the focus of the analysis is just the short-run house price dynamics. Accordingly the standard Granger causality test without error correction term is adopted.

Table 3 Number of cities with granger causality

	City	Number		City	Number
National "Superstars"	Beijing	28	"Normal" Cities	Ningbo	22
	Shanghai	24		Guiyang	18
	Shenzhen	15		Nanning	14
	Average	22.3		Xining	14
Regional "Stars"	Chengdu	21		Haikou	12
	Nanjing	21		Changchun	11
	Chongqing	20		Harbin	11
	Hangzhou	19		Yinchuan	11
	Xiamen	18		Nanchang	10
	Fuzhou	17		Lanzhou	10
	Tianjin	16		Urumqi	10
	Xian	15		Kunming	9
	Qingdao	15		Shijiazhuang	8
	Dalian	13		Hohhot	7
	Wuhan	11		Zhengzhou	7
	Changsha	11		Jinan	2
	Shenyang	11		Taiyuan	1
	Guangzhou	9		*Average*	10.4
	Hefei	5	–	–	–
	Average	14.8	–	–	–

period, $T_{i,end}$, which is defined as the month with the highest annual house price growth rate after $T_{i,recover}$.

The third point is the start of the boom in the housing market, $T_{i,booming}$, which is more difficult to identify. We adopt the method by Ferreira and Gyourko (2011) to detect the structural breakpoint during the period between $T_{i,recover}$ and $T_{i,end}$. We estimate the following question for all potential structural breakpoints ($T^*_{i,booming}$) for each city i and month t:

$$PG_{i,t} = a + d_i 1\left[T_{i,t} \geq T^*_{i,booming}\right] + \varepsilon_{i,t}, \text{ for } T_{i,recover} < T^*_{i,booming} < T_{i,end} \quad (2)$$

where $PG_{i,t}$ is the annual house price growth rate, d_i estimates the importance of the potential break, $T_{i,t}$ is a quarter, and $T^*_{i,booming}$ is the location of the potential structural break. The breakpoint point, $T_{i,booming}$, is defined as the month which maximizes the R^2 of this equation.

Figure 4 depicts the results in Beijing as an example. First, as shown in Fig. 4b, it can be seen that the housing market started to recover in December 2008 ($T_{Beijing,recover}$), and that the boom process ended in April 2010 ($T_{Beijing,end}$). During this period, Eq. (2) reaches its maximum R^2 value in October 2009 as shown in Fig. 4c, while the coefficient d_i depicted in Fig. 4d is also significantly larger than 0 in that month. It is thus identified as the starting point of the boom period.

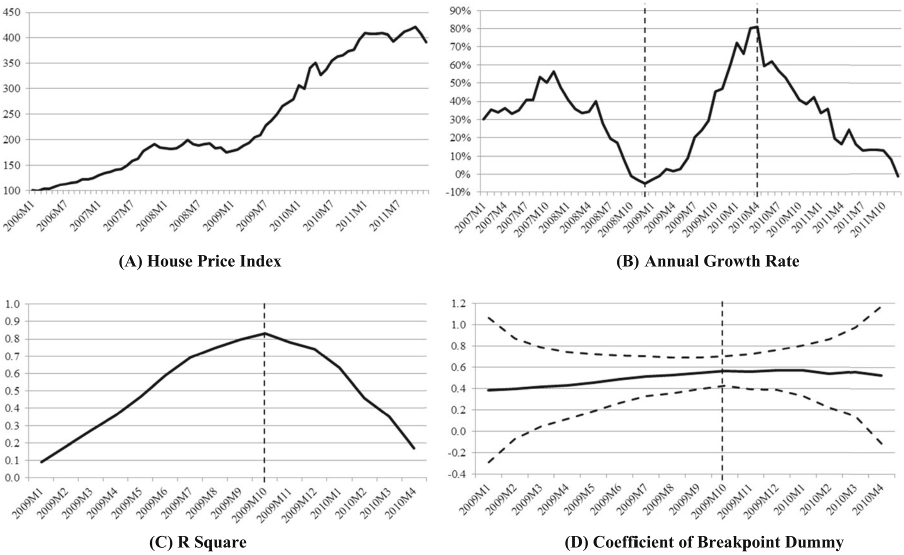

Fig. 4 Detecting key time points during the housing boom in Beijing (**a**) House price index (**b**) Annual growth rate (**c**) R square (**d**) Coefficient of breakpoint dummy

Following these procedures, Table 4 lists the three key time points for all 35 cities. In the cities of Harbin, Yinchuan and Xining, the coefficient d_i is not significantly larger than 0 at the 90 % confidence level when the R^2 reaches its maximum value, which suggests that there is no boom in these cities by definition.

Again, the results show a clear lead-lag relationship between the three tiers of cities defined in last section. In general, house prices in the three "superstars" began rising again at the end of 2008. This was almost the same time as the announcement of the Chinese government's stimulus package. Then the housing boom started around October 2009, and house price growth rate reached its peak and then started to decline in May 2010, immediately after the central government issued intervention policies in the housing market. Comparably, the process in the "star" cities lags by about 3 to 4 months on average. According to the results, a representative "star" city resumed house price growth in the first quarter of 2009, began its boom around the end of 2009 and the beginning of 2010, and reached the peak in about mid-2010. The time points in the "normal" cities are even later and lag by about 3 more months after the regional "stars".

According to the above tests, the spatial pattern of intercity house price information flow and of house price discovery are highly correlated. This suggests that information diffusion is at least one of the major factors in determining intercity house price discovery pattern, especially in the short run. Further quantitative analysis from the temporal perspective may provide more evidence for such causality relationship, although it is not feasible at present due to data constraints.[21]

[21] As a preliminary attempt, we collect the quarterly series of all provinces' *PIFI* on Beijing's house prices, which is the only available continuous time series of *PIFI* so far. The empirical tests suggest that, on one hand, a sharp change in house prices (either an increase or decrease) in Beijing will lead to house price information flows to other cities in the following one to three quarters (i.e., higher *PIFI*). On the other hand, such information will immediately affect house prices in the lagging cities. However, a more definite conclusion requires evidence from more cities, and we leave this for future research.

Table 4 Key time points in the housing boom in 2009–2010

City	Recover	Boom	End	City	Recover	Boom	End
A. National "Superstars"				C. Normal Cities			
Shenzhen	2008M10	2009M10	2010M04	Ningbo	2009M1	2009M10	2010M1
Shanghai	2008M12	2009M09	2010M04	Nanchang	2009M1	2009M10	2011M6
Beijing	2008M12	2009M10	2010M04	Zhengzhou	2009M2	2010M9	2011M6
Average	2008M11	2009M10	2010M04	Haikou	2009M2	2009M12	2010M2
B. Regional "Stars"				Changchun	2009M3	2010M3	2010M4
Guangzhou	2008M11	2009M9	2010M2	Nanning	2009M3	2009M11	2010M4
Chengdu	2008M12	2009M10	2010M2	Shijiazhuang	2009M4	2010M10	2011M5
Tianjin	2008M12	2009M10	2010M6	Taiyuan	2009M5	2010M2	2010M5
Wuhan	2008M12	2009M11	2010M5	Guiyang	2009M5	2009M12	2010M5
Xiamen	2009M2	2009M10	2010M4	Kunming	2009M6	2011M7	2011M12
Fuzhou	2009M2	2010M1	2010M12	Lanzhou	2009M6	2011M4	2011M5
Hangzhou	2009M3	2009M10	2010M4	Harbin	2009M8	–	2011M4
Changsha	2009M3	2010M3	2011M2	Jinan	2009M9	2010M1	2010M9
Chongqing	2009M4	2009M11	2010M4	Yinchuan	2009M9	–	2010M7
Nanjing	2009M4	2009M11	2010M6	Hohhot	2010M1	2011M1	2011M3
Dalian	2009M4	2010M3	2010M4	Urumqi	2010M4	2010M7	2011M2
Qingdao	2009M5	2009M11	2010M10	Xining	2010M7	–	2011M7
Hefei	2009M5	2010M1	2010M5	*Average*	2009M7	2010M5	2010M11
Xian	2009M5	2010M2	2010M10				
Shenyang	2009M5	2010M4	2010M12				
Average	2009M3	2010M1	2010M7				

Concluding Remarks

In this paper, we suggest an innovative measure of information flow in Chinese housing markets based on Google search records. The search request records depict a substantial flow of information about house prices from national "superstar" cities such as Beijing and Shanghai and regional "star" cities like Tianjin and Chongqing to other "normal" cities. The results also suggest that such information diffusion is at least one of the major factors that influence intercity house price discovery in the short run.

Obviously these results highlight the fact that cities like Beijing and Shanghai should be the major target for investors, market analysts, and researchers if they want to better understand China's housing market, and for policy makers, who want to formulate more effective housing policies. An even more important implication is that house price dynamics in "superstar" cities will have substantial externalities in the national level – a bubble in Beijing may generate misleading signals to market participants in other cities and quickly spread around the country. Unfortunately, these cities have been shown to be vulnerable to exceptional house price surges (Gyourko, Mayer, and Sinai 2013), while some unique institutional factors in China further dampen such vulnerability (Deng et al. 2014). Therefore, policy makers should pay particular attention to any potential mispricing in these core cities.

We believe this research can also serve for much broader and in-depth future works on information issues in the housing market. In this paper, we do not separate market participants' rational learning process from (irrational) herding or positive feedback behaviors, which can be a task for future research. Similarly, the analysis of variance during the booming and recession periods is also interesting. Moreover, while this paper mainly focus on intercity information flow, we can also build an index based on searches for house price information from local market participants. Possible topics related include how certain events (announcements of new intervention policies, land auctions, or a listing of new complexes, etc.) affect market participants' attentions, and whether the volume of Google searches can predict house price change or transaction volume.

Acknowledgments We thank the referee and the editor, Kelvin Wong, Yuichiro Kawaguchi and participants in the 2012 Asia-Pacific Real Estate Research Symposium, the 2012 GCREC Annual Conference and the 2012 AREUEA-AsRES Joint Conference for helpful comments. The authors thank the Institute of Real Estate Studies at National University of Singapore for financial support. Wu also thanks the National Natural Science Foundation of China for financial support (No. 71003060 & 71373006).

References

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics, 79*(4), 551–563.

Bandt, O., Barhoumi, K., & Bruneau, C. (2010). The international transmission of house price shocks. In O. Bandt et al. (Eds.), *Housing markets in Europe: A macroeconomics perspective*. Berlin: Springer-Verlag Berlin Heidelberg.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management, 25*(3), 239–264.

Barkham, R., & Geltner, D. (1995). Price discovery in American and British property markets. *Real Estate Economics, 23*(1), 21–44.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the dynamics of segregation. *Quarterly Journal of Economics, 123*(1), 177–218.

Case, K., & Shiller, R. (1989). The efficiency of the market for single family homes. *American Economic Review, 70*(1), 125–137.

Chau, K. W., Wong, S. K., Yiu, C. Y., et al. (2010). Do unexpected land auction outcome bring new information to the real estate market? *Journal of Real Estate Finance and Economics, 40*(4), 480–496.

Clapp, J., Doldo, W., & Tirtiroglu, D. (1995). Imperfect information and investor inferences from housing price dynamics. *Real Estate Economics, 23*(3), 239–269.

Da, Z., Engelberg, J., & Gao, P. J. (2011). In search of attention. *Journal of Finance, 66*(5), 1461–1499.

Deng, Y., & Quigley, J. (2008). Index revision, house price risk, and the market for house price derivatives. *Journal of Real Estate Finance and Economics, 37*(3), 191–209.

Deng, Y., Gyourko, J., & Wu, J. (2012). Land and house price measurement in China. In A. Heath, F. Packer, & C. Windsor (Eds.), *Property Markets and Financial Stability*. Bank of International Settlement and Reserve Bank of Australia.

Deng, Y., Morck, R., Wu, J., & Yeung, B. (2014). China's pseudo-monetary policy. *Review of Finance*, forthcoming.

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters, 9*(3), 167–175.

Engelberg, J., & Parsons, C. (2011). The causal impact of media in financial markets. *Journal of Finance, 66*(1), 67–97.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance, 64*(5), 2023–2052.

Ferreira, F., & Gyourko, J. (2011). Anatomy of the beginning of the housing boom: U.S. neighborhoods and metropolitan areas. NBER Working Paper #W17374.

Ferreira, F., & Gyourko, J. (2012). Heterogeneity in neighborhood-level price growth in the United States, 1993–2009. *American Economic Review, 102*(3), 134–140.
Giliberto, M. (1990). Equity real estate investment trusts and real estate returns. *Journal of Real Estate Research, 5*(2), 259–264.
Granger, C. (1969). Investigating causal relationship by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438.
Grossman, S., & Stiglitz, J. (1976). Information and competitive price system. *American Economic Review, 66*(2), 246–253.
Gyourko, J., & Keim, D. (1992). What does the stock market tell us about real estate returns? *AREUEA Journal, 20*(3), 457–485.
Gyourko, J., Mayer, C., & Sinai, T. (2013). Superstar cities. *American Economic Journal: Economic Policy, 5*(4), 167–199.
Ho, L. S., Ma, M., & Haurin, D. (2008). Domino effects within a housing market: the transmission of house price changes across quality tiers. *Journal of Real Estate Finance and Economics, 37*(4), 299–316.
Holly, S., Pesaran, H., & Yamagata, T. (2011). A spatio-temporal model of house prices in the USA. *Journal of Econometrics, 158*(1), 160–173.
Hong, H., & Stein, J. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance, 54*(6), 109–128.
Hong, H., & Stein, J. (2007). Disagreement and the stock market. *Journal of Economic Perspectives, 21*(2), 109–128.
Meen, G. (1999). Regional house prices and the ripple effect: a new interpretation. *Housing Studies, 14*(6), 733–753.
Mondria, J., & Wu, T. (2011). Asymmetric attention and stock returns. Working paper (SSRN 1772821).
Oikarinen, E. (2004). The diffusion of house price movements from center to surrounding areas. *Journal of Housing Research, 15*(1), 3–28.
Ong, S. E., & Sing, T. F. (2002). Price discovery between private and public markets. *Urban Studies, 39*(1), 57–67.
Ooi, J., & Lee, S. (2004). Price discovery between residential land and housing market. *Journal of Housing Research, 15*(2), 95–112.
Pollakowski, H., & Ray, T. (1997). House price diffusion patterns at different aggregation levels: an examination of housing market efficiency. *Journal of Housing Research, 8*(1), 107–124.
Tetlock, P. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance, 62*(3), 1139–1168.
Tetlock, P. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies, 23*(9), 3520–3557.
Tetlock, P., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance, 63*(3), 1437–1467.
Tuluca, S., Myer, N., & Webb, J. (2000). Dynamics of private and public real estate markets. *Journal of Real Estate Finance and Economics, 21*(3), 279–296.
Wong, S. K., Chau, K. W., & Yiu, C. Y. (2007). Volatility transmission in the real estate spot and forward markets. *Journal of Real Estate Finance and Economics, 35*(3), 281–293.
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 and Economics, 48*(3), 522–545.
Yunus, N., Hansz, A., & Kennedy, P. (2012). Dynamic interaction between private and public real estate markets: some international evidence. *Journal of Real Estate Finance and Economics, 45*(4), 1021–1040.