Architecture and the Built environment

#04
2017

The Spatial Dimension of House Prices

Yunlong Gong

The Spatial Dimension of House Prices

Yunlong Gong
Delft University of Technology, Faculty of Architecture and the Built Environment,
OTB - Research for the Built Environment

Architecture
and the
Built environment

abe.tudelft.nl

Design: Sirene Ontwerpers, Rotterdam

ISBN 978-94-92516-51-0
ISSN 2212-3202

The Spatial Dimension of House Prices

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 8 mei 2017 om 15:00uur
door Yunlong GONG
Master of Science in Land Resource Management, CUMT, China
geboren te Shandong, China

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. P. J. Boelhouwer
promotor: Prof. dr. J. de Haan

Samenstelling promotiecommissie bestaat uit

Rector Magnificus,	voorzitter
Prof. dr. P. J. Boelhouwer,	promotor
Prof. dr. J. de Haan,	promotor

Onafhankelijke leden

Prof. dr. E. Buyst,	Katholieke Universiteit Leuven
Prof. dr. P. M. A. Eichholtz,	Universiteit Maastricht
Prof. dr. A. J. van der Vlist	Rijksuniversiteit Groningen
Prof. dr. J. Rouwendal	Vrije Universiteit Amsterdam
Prof. emer. dr. A. Kleinknecht,	Technische Universiteit Delft
Prof. dr. M. G. Elsinga,	Technische Universiteit Delft, reservelid

Contents

Preface

After a short while of happiness when I learned that I was going to pursue a PhD in the Netherlands, an overwhelming feeling of fear surged through me. I am afraid of not surviving my PhD journey. Fortunately, with the help and support of many people, I finally make it possible.

I am very grateful to my promotor Peter Boelhouwer and Jan de Haan for giving me this opportunity to study in a different culture. Thanks very much for your constant patience and encouragement, which builds up my confidence gradually. Thanks for your trust in me, for the freedom you give to me in research, and for your valuable feedback and suggestions. All these have trained me to be a qualified independent researcher, which will benefit my whole career. I would also like to extend my gratitude to my committee members for their efforts in assessing my work.

Thanks are given to my dear colleagues and fellow PhDs for creating the comfortable and pleasant environment which makes my work and life much enjoyable. Special thanks go out to Harry B. and Marietta for helping me out with the international comparison of housing affordability, to Gust for your company in every cycling tour around South Holland, and to Qi for the tasty food you made to us. Kees, Joris, Marja, Harry H., Sylvia, Alfred and Cor, I really appreciate your remarks whenever I presented my work to you. Wenjing, Juan, Xin, Job, Rosa, KyungHo, Vitnarae, and Queena thanks very much for the nice conversations and inspiring discussion which broadens my horizons. Furthermore, I want to express my gratitude to Jinxing Hu for sharing the research data with me, to Gavin Wood and Evert Meijers for commenting on my working papers, to Henri de Groot and Jan Rouwendal for freely including me into the master course "Regional and Urban Economics" and for your stimulating discussion on my work.

Thanks also go to the support staff of OTB for their great assistance and support, which makes the things much easier. Special thanks go to Elianne for arranging almost everything around the defence ceremony, and to Dirk for your assistance in publication of my work. Besides, I would like to thank Véro Crickx and Ken Arroyo Ohori for helping me with the layout of this dissertation.

Although being in a country with different culture, I have felt at home with the company of a circle of Chinese friends. Among them, thank you especially to

Changgong, Jingtang, Kaihua, Guangliang, Yuanhao, Xinchao, Zhijie, Liguo, Xiangming and Zhiyong.... I will never forget the funny and joyful drinking moments.

I want to express my gratitude to China Scholarship Council (SCS) for financing my study in the Netherlands.

Finally, thank you to my family for your endless love, patience and support, which enables me to fully concentrate on my PhD research.

Summary

The economic reform in China, launched in the late 1970s, gradually promotes the free mobility of capital and labour between rural and urban areas, and between cities. The following housing market reform in the late 1990s thoroughly terminates the socialist allocation of housing and introduces market forces into the housing sector. Such institutional shifts have profound effects on the evolution of the Chinese interurban housing market. Yet, little is known about the spatial behaviour of house prices across cities in the post-reform era. How do the housing markets of different cities organise across space? What is the relationship between the house price dynamics of different cities? To answer these questions, this research performs economic and econometric analysis of the spatial dimension of the Chinese interurban housing market. In addition, this research also concerns the construction of a reliable house price index in the presence of spatial heterogeneity and dependence in the urban housing market of China. A reliable house price index is essential to the analysis of house price dynamic behaviour. However, owing to the data problem, this part is conducted based on the housing market of a Dutch city.

This research discovers the spatial regularities of house prices across Chinese prefecture cities in an economic common area and investigates the underlying formation process. It reveals an uneven distribution of house prices across cities, with those large and/or higher-tier cities and their neighbours having significantly higher house prices. Such an uneven pattern of house prices demonstrates the agglomeration spillovers in the interurban housing market. Two forms of spillovers are empirically examined. The first is the urban hierarchy distance effect, which is related to the position of a city in a hierarchical urban system. In general, the distance penalty of higher-tier urban centres is confirmed, that is, all else being equal, the further away a city is from the higher-tier city, the lower the house price. The second form of spillovers relates to a city's position in a city network system, in which no hierarchical structure is imposed. In such a situation, the spillovers arise from the interaction with neighbouring cities and it is found that a city that has larger neighbours tends to have higher house prices. These two forms of spillovers are somewhat correlated with each other because a higher-tier city is always associated with a larger urban size.

It is argued that the spillovers in the interurban housing market work through two channels: the productivity and amenity channel. First, because of the economies of agglomeration, a location that has good access to large urban concentrations is likely to

enjoy some productivity advantages and thus can bear higher house prices. Second, a location that is surrounded by large urban concentrations can easily get access to some unique amenities that need a large market potential to survive; households value such access and thus bid up the house price there. However, it seems that the role that the productivity channel plays is much more important than the role of the amenity channel.

In addition to the static distribution of house prices across space, this research also concerns the time series behaviour of house price dynamics across Chinese cities. Geography plays an important role in explaining the cross-city differences of house price dynamics. For the housing markets of major cities across the whole of China, the cluster analysis generally uncovers two relatively homogeneous groups, within which the house price growth series share a similar dynamic pattern. One cluster contains mainly the cities in the undeveloped central, western and northeast China, whereas the other is composed of the most important economic centres in eastern China. However, the spatial segregation of housing markets is more likely to occur in the most recent period. In the early period before 2010, the house price dynamics of cities are much more homogeneous.

The similarities and/or dissimilarities among house price dynamics of different cities indicate the complicated interrelationships between each of the markets. This research further examines various spatial interrelationships between the housing markets of an economic common area in south China. The spatial causal relationships between housing markets are first tested by the Granger causality test. The results reveal a complicated pattern, but it can be tentatively confirmed that house price changes in the developed eastern-central markets 'cause' the house price dynamics in the less-developed western markets. Then a spatial-temporal model is built to depict the diffusion pattern of house prices between markets. In general, a shock given to the house price of a certain market gradually spreads to its neighbouring cities. However, the interurban housing market can hardly remain an equilibrium relationship in the long-run, that is, it tends to be divergent.

The last part of this research concerns the treatment of spatial effects in the hedonic house price model as well as its influence on the construction of hedonic imputation indexes, which measure the pure house price changes over time. It is argued that the value of a dwelling can be split into the value of the land and the value of the structure, and that the value of the location characteristics of a dwelling is capitalised into the price of the land. Thus, land prices can be expected to vary significantly across space. Indeed, the mixed geographically weighted regression framework adopted in this research, which allows the shadow price of structure to be constant across space and the implicit price of land to be property-based, is found to be superior to, in terms of model prediction, those models that restrict the spatial variation of land prices. Nevertheless, the Fisher imputation house price index based on the most sophisticated

model is almost identical to those based on the simple specifications. The land and structure price indexes, on the other hand, are sensitive to the treatment of location in land prices.

This research underlines market forces in the operation of Chinese interurban housing markets in the post-reform era, and contributes to the understanding of spatial dimension of house prices, not only in China, but also in other market-oriented economies.

Samenvatting

De economische hervorming in China, die eind jaren zeventig van start ging, bevordert geleidelijk de vrije mobiliteit van kapitaal en arbeid tussen landelijke en stedelijke gebieden en tussen steden onderling. Met de daaropvolgende hervorming van de woningmarkt eind jaren negentig kwam er definitief een einde aan de socialistische toewijzing van woningen en werd er marktwerking in de woningmarkt geïntroduceerd. Dergelijke institutionele veranderingen hebben ingrijpende gevolgen gehad voor de ontwikkeling van de interstedelijke huizenmarkt in China. Toch is onze kennis van het ruimtelijke gedrag van huizenprijzen in de steden in de periode na de hervormingen beperkt. Hoe zijn de huizenmarkten van verschillende steden in de ruimte georganiseerd? Wat is de relatie tussen de dynamiek in de huizenprijzen in verschillende steden? Om antwoord te geven op die vragen zijn voor dit onderzoek economische en econometrische analyses uitgevoerd van de ruimtelijke dimensie van de Chinese interstedelijke huizenmarkt. Daarnaast is er onderzoek gedaan naar het opzetten van een betrouwbare index voor huizenprijzen met inachtneming van de ruimtelijke heterogeniteit en onderlinge afhankelijkheid binnen de stedelijke huizenmarkt in China. Een betrouwbare huizenprijzenindex is essentieel om het dynamische gedrag van huizenprijzen te kunnen analyseren. Als gevolg van dit gebrek aan data is dit onderdeel uitgevoerd op basis van de huizenmarkt in een stad in Nederland.

Dit onderzoek heeft de ruimtelijke regelmatigheden van de huizenprijzen in steden in verschillende Chinese prefecturen in een economisch samenhangend gebied aan het licht gebracht. Daarnaast is het achterliggende prijsvormingsproces onderzocht. Daaruit blijkt dat de huizenprijzen onregelmatig over de steden zijn verdeeld, met significant hogere prijzen in grote en/of meer ontwikkelde steden en de steden daaromheen. Zo'n onregelmatige verdeling van de huizenprijzen wijst op de overloopeffecten van agglomeraties in de interstedelijke huizenmarkt. Twee soorten overloopeffecten zijn empirisch onderzocht. De eerste is het effect van afstand binnen de stedelijke hiërarchie, dat samenhangt met de positie van een stad binnen een hiërarchisch stedelijk systeem. Over het algemeen ondervinden steden een negatief effect als zij zich op grotere afstand van meer ontwikkelde stedelijke centra bevinden: hoe verder een stad van een meer ontwikkelde stad gelegen is, hoe lager de huizenprijzen (onder verder gelijke omstandigheden). De tweede vorm van overloopeffect heeft betrekking op de positie van een stad binnen een netwerk van steden waarop geen hiërarchie van toepassing is. In die situatie zijn de

overloopeffecten het gevolg van interactie met omringende steden. Het is gebleken dat steden met grotere buursteden veelal hogere huizenprijzen kennen. Deze twee vormen van overloopeffect vertonen enige mate van correlatie, aangezien een meer ontwikkelde stad ook altijd groter zal zijn.

Er wordt betoogd dat de overloopeffecten in de interstedelijke huizenmarkt via twee kanalen tot uiting komen: productiviteit en voorzieningen. In de eerste plaats is het zo dat een locatie die goede toegang biedt tot grote stedelijke concentraties vanwege de economische voordelen van agglomeratie meestal productiviteitsvoordelen biedt en dus hogere huizenprijzen kan dragen. Ten tweede biedt een locatie die is omringd door grote stedelijke concentraties goede toegang tot unieke voorzieningen die een groot marktpotentieel nodig hebben om te kunnen overleven. Doordat die voorzieningen door huishoudens worden gewaardeerd, stijgen de lokale huizenprijzen. Het lijkt er echter wel op dat productiviteit een veel grotere rol speelt dan voorzieningen.

Naast de statische distributie van huizenprijzen in de ruimte is voor dit onderzoek ook gekeken naar het tijdreeksgedrag van de huizenprijzendynamiek tussen Chinese steden. Geografie speelt een belangrijke rol in de verschillen in de dynamiek van huizenprijzen tussen steden. Als er een clusteranalyse wordt uitgevoerd voor de huizenmarkten in grote steden in heel China, komen daarbij twee relatief homogene groepen naar voren waarbinnen de stijging van de huizenprijzen een vergelijkbaar dynamisch patroon vertoont. Eén cluster bestaat hoofdzakelijk uit de steden in het onontwikkelde midden, westen en noordoosten van China, terwijl de andere bestaat uit de belangrijkste economische centra in oostelijk China. De ruimtelijke segregatie van de huizenmarkten heeft echter voornamelijk in het recente verleden plaatsgevonden. Tijdens de periode vlak voor 2010 was de dynamiek van huizenprijzen in steden veel homogener.

Uit de overeenkomsten en/of verschillen tussen de dynamiek in huizenprijzen in verschillende steden blijkt hoe complex de onderlinge verbanden tussen de diverse markten zijn. Er is nader onderzoek gedaan naar verschillende ruimtelijke verbanden binnen de huizenmarkt in een economisch samenhangend gebied in Zuid-China. De ruimtelijk-causale verbanden tussen de huizenmarkten zijn eerst getoetst met behulp van de Granger-causaliteitstoets. Daarmee werd een complex patroon zichtbaar, maar voorlopig lijkt bevestigd dat de veranderingen in huizenprijzen in de ontwikkelde markten in het oosten en het midden van het land de 'oorzaak' zijn voor de dynamiek van de huizenprijzen in de minder ontwikkelde markten in het westen. Vervolgens is er een ruimtelijk-temporeel model opgezet om het verspreidingspatroon van de huizenprijzen tussen markten zichtbaar te maken. Over het algemeen zal een schok voor de huizenprijzen in een bepaalde markt zich geleidelijk uitbreiden naar de omringende steden. De interstedelijke huizenmarkt is echter nauwelijks in staat om op lange termijn in evenwicht te blijven en is vaak divergent.

Het laatste deel van dit onderzoek betreft de omgang met ruimtelijke effecten in het hedonische model voor huizenprijzen en de invloed daarvan op de ontwikkeling van 'hedonische toewijzingsindexen' die de zuivere veranderingen aan de huizenprijzen in de tijd weergeven. Er wordt betoogd dat de waarde van een woning kan worden onderverdeel in de waarde van de grond en de waarde van het gebouw, en dat de waarde van de omgevingskenmerken van een woning in de prijs van de grond wordt verdisconteerd. Daardoor kunnen de grondprijzen in de ruimte aanzienlijk variëren. Het geografisch gemengde gewogen regressiekader dat voor dit onderzoek is gebruikt, maakt het mogelijk om de schaduwprijs van een bouwwerk constant te houden in de ruimte en de impliciete grondprijs op gebouwen te baseren. Het heeft een superieur voorspellend vermogen in vergelijking met modellen die de ruimtelijke variatie van grondprijzen beperken. Toch blijkt de Fisher toewijzingsindex voor huizenprijzen op basis van het meest geavanceerde model vrijwel identiek te zijn aan die gebaseerd op de eenvoudigste specificaties. De indices voor grond- en gebouwenprijzen zijn daarentegen gevoelig voor de manier waarop bij grondprijzen rekening wordt gehouden met de locatie.

Dit onderzoek onderstreept de marktwerking binnen de Chinese interstedelijke huizenmarkten in de periode na de hervormingen en levert een bijdrage aan ons inzicht in de ruimtelijke dimensie van huizenprijzen, niet alleen in China maar ook in andere markteconomieën.

1 Introduction

§ 1.1 Motivation

China has been undergoing significant social and economic structural changes since launching its policy of economic reform and opening up in 1978. This has involved a transformation from a centrally planned economy, where there is no role for the market, to a market-oriented economy in which market principles play a major role. During the last four decades, great achievements have been made in terms of economic growth and social well-being. To name a few indicators: the Gross Domestic Product (GDP) of the country increased from USD 189.65 billion in 1980 to USD 10.866 trillion in 2015, positioning China as the second largest economy in the world, with an average annual growth rate over 10%.[1] Meanwhile, poverty levels have greatly improved. The poverty headcount ratio at USD 1.90 a day (2011 PPP) has decreased dramatically, from 42.15% in 1981 to 10.68% in 2013. The rapid economic growth, combined with the reform of the *Hukou* registration system, has also accelerated the migration flow from rural areas to urban areas.[2] The population living in urban China in 2015 reached 763 million, making the urbanisation level of 55.61%, almost three times that in 1980.

With the rapid growth of the urban population, the welfare-based public housing provision system founded in the central planning era could no longer meet the increasing housing demand of urban residents. Thus, in 1994, comprehensive housing reforms were implemented, aiming to privatize the public housing sector and promote a housing allocation system based on market principles. The milestone of housing reform occurred in 1998, when the government completely suspended the traditional housing allocation system, making the housing market the only way to access housing services (Wang et al. 2012). The emergence of the private urban housing market spurred both housing transactions and prices. In 1998, the housing area traded on the

1 All the data in this paragraph, including GDP, poverty level and urban population, was collected from the World Bank.

2 The *Hukou* (household registration) system in China was initially designed as a mechanism for monitoring population movements in the early 1950s. Subsequently, it became a strong tool to restrain rural-urban migration and labour mobility between cities.

market was approximately 108 million square metres on an average transaction price of 1854 yuan/m^2. These two figures were nearly ten and three times higher in 2014, soaring to 1.05 billion square metres and 5933 yuan/m^2, respectively.[3]

At the regional level, rapid economic development has been accompanied by increasing inequality. Soon after the launch of the economic reforms, some coastal regions, Guangdong and Zhejiang in Eastern China, for example, grew quickly, due to the influx of foreign direct investment (FDI), advanced technologies and equipment, and favourable policies of the central government. The 'core' position of these regions in the national economy was further enhanced through a self-reinforcing process (Anderson 2012, p.127), shaping a core-periphery economic structure in China. In 1980, the regional gross product of Eastern China accounted for 43.69% of total GDP in China, while in 2014 this ratio increased to 51.16%, reflecting the polarization of economic activities .[4]

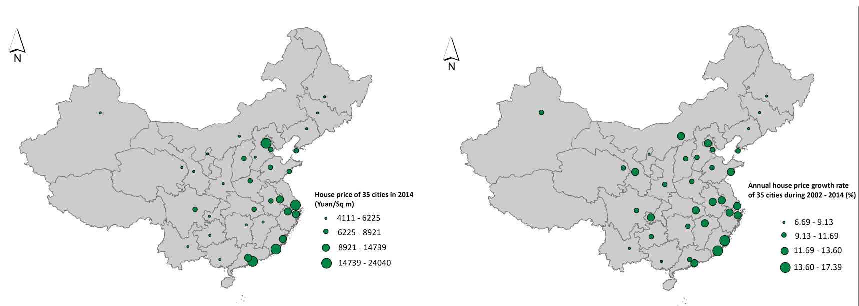


FIGURE 1.1 Spatial distribution of house prices and house price growth rates in 35 cities

Reflecting the distribution of economic activities, the inequality in the cost of housing between regions is also striking. In 2014, the average sale price in 35 main cities in mainland China was approximately 8599 yuan/m^2, with the standard error also high, at 4651 yuan/m^2, making the coefficient of variance 0.54, thus indicating a high degree of heterogeneity across this city-level housing market. The left panel of Figure 1.1 shows the spatial distribution of average house prices. It is apparent that the prices in the coastal cities of Eastern China are generally greater than the prices of inland cities. However, the picture of house price dynamics is a little different. From 2002 to 2014, the rapid growth in house prices, on average 11.38% per year, seems to be a

3 The housing data used in this section was collected from *China Statistical Yearbook*. Note that the average house prices are calculated without controlling for housing quality.

4 The data for regional economic indicators was collected from the Statistical Yearbook of China and the provinces. Eastern China includes Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan.

national phenomenon and there is very little variance between the annual growth rates in different cities; the coefficient of variance is only 0.18, much lower than that of the house price level. Perhaps the most prominent spatial pattern of house price growth rate is that the northeastern cities experienced the lowest price appreciation during the period 2002-2014.

This dissertation is fundamentally concerned with the spatial patterns of house prices and their dynamics across cities in China. Although literature on the Chinese housing market has been emerging in recent years, little is known about the spatial interaction of regional housing markets. The following four chapters will be dedicated to responding to questions concerning the emerging market: Why is there a core-periphery structure in the distribution of interurban house prices? To what extent are the house price developments across cities similar? How do house price dynamics in one city affect the house price changes in other cities?

The investigation of the spatial dimension of the Chinese housing market has been always hampered by the quality of the data, especially when analysing house price dynamics. This situation has inspired the pursuit of research to construct house price indexes that reflect the house price changes as accurately as possible. In line with a key theme of this study, particular attention has been paid to the influence of the spatial dimension on constructing hedonic imputation house price indexes. Since access to detailed housing transaction records in China is rather restrictive, the analysis in Chapter 6 is based on one housing market in the Netherlands. Using this information, the chapter provides some useful guidance for the construction of house price indexes for Chinese markets.

The remaining sections of this chapter are organised as follows. Section 1.2 provides some background on the formation of the urban private housing market in China, followed by Section 1.3, which briefly reviews the literature on the spatial dimension of house prices. Several questions relating to this research are presented in Section 1.4, while Section 1.5 outlines the structure of the dissertation and briefly introduces the main content of each chapter.

§ 1.2 The emergence of an urban private housing market in China

After the foundation of the People's Republic of China in 1949, the country was characterised by a rural-urban dual structure. In urban China, the government gradually established a welfare-oriented housing system through the socialist transformation of private housing and the construction of new public housing. By the late 1970s, the private housing market had almost been eliminated (Huang 2004). In this public rental system, housing was allocated through a work unit-employee linkage.

The rent charged was highly subsidised and the quality of the housing one received relied on a set of non-monetary criteria, such as job rank, job seniority and household size. While this system was beneficial to tenants, it created a huge financial burden and thus restrained housing supply. In 1978, the average housing area consumed per capita in urban China was only 3.6 square metres, which was even below the consumption level in 1949 (Fang et al. 2016).

As China embarked on a transition from a centrally planned economy to a market economy with Chinese characteristics in 1978,[5] the reform of the housing system was also placed on the agenda, aiming to increase the housing supply. A big step was made in 1988, when land transactions were legally permitted through a constitutional amendment; however, it should be noted that since land in China belongs to the State, only a land lease right can be transacted on the market.[6] This reform established the legal basis for private housing construction and the creation of an urban housing market. The housing provision system in urban China subsequently entered a dual structure period, with housing provided either by the public or the private sectors. By 1993, approximately 40% of urban households still resided in state-owned housing (Wang 2011); however, in 1994, the State Council of China introduced a policy to guide work units to sell the public housing to occupying tenants, which significantly accelerated the public housing privatization process.

The dual structure of housing provision lasted until 1998, when the welfare-oriented housing system was officially discontinued and the market mechanism took control in the allocation of housing. However, government was still involved in two types of housing catering for low-income households. The first was 'cheap rental housing' (*lianzhu fang*), aiming to assist extremely low-income households. The second was 'affordable housing' (*jingji shiyong fang*), which is similar to normal owner-occupied housing except that the construction of affordable housing is subsidised and the price controlled by the government, and with certain restrictions being placed on transactions. Following these reforms, the rate of home ownership soared to over 80% by 2002 (Wang et al. 2012).

After the establishment of the private urban housing market in the late 1990s, recent years have witnessed the emergence of studies on the operation of this new market. Most of the studies are dedicated to exploring the determinants of house prices and to testing whether there is a house price bubble. The house price fundamentals found in the housing markets of Western countries also play a role in these emerging housing markets. For example, at the national level, monetary policy is thought of as a key

5 The market economy with Chinese characteristics, also known as a socialist market economy, is an economic model which emphasizes the dominance of the state-owned sector but the importance of the market mechanism in the economy.

6 For residential land, the private housing developer can buy use rights for a period of 70 years.

factor driving national house price changes (Xu and Chen 2012; Zhang et al. 2012). In the cross-city housing market, differences in the population of cities, income levels, the level of air pollution, amenities and supply conditions have been found to be responsible for the disparities between city house prices (Zheng et al. 2010; Li and Chand 2013; Zheng et al. 2014). Within cities such as Beijing, the monocentric city model can still predict the urban structure to some extent, although a polycentric structure has been emerging in recent years (Zheng and Kahn 2008; Qin and Han 2013). Nevertheless, little is known about the spatial pattern of Chinese interurban house prices and their dynamics. The aim of this dissertation is to address this deficit.

§ 1.3 The spatial dimension of house prices

§ 1.3.1 Spatial distribution of interurban house prices

'Location, location, location' – the famous motto of the real estate market – underlines how closely property price is related to property position. Since locations are inherently differentiated in terms of both natural endowments and human activities, it is not a surprise to see certain structures that regulate the distribution of house prices across space. Within an urban area, a common regulatory effect is that, all other things being equal, house price will have a negative relationship with the distance to the central business district (CBD), as depicted by the well-known monocentric city model (Alonso 1964; Mills 1967; Muth 1969). Such a house price gradient has been empirically established in many cities across the world, such as Chicago, Berlin, and Beijing. Of course, a modern city is more than a monocentric city and will be more likely to be characterised by a polycentric structure, in which subcentres and important urban nodes such as hospitals and universities also play a role in shaping the house price pattern (Heikkila et al. 1989; Waddell et al. 1993). Whether in a monocentric or polycentric city, the house price structure reflects a trade-off between a household's desire for space and the commuting cost to those urban centres.

Beyond the intra-urban market, how do the house prices differ between urban areas? Note that the focus is on the aggregate house price measured for an entire urban area.[7] Before examining the cross-city house price structure, one first needs to determine how the characteristics of a city affect aggregate house prices. In a general spatial equilibrium framework, where the marginal consumer is indifferent across cities, the effect of a city on its house prices should be equivalent to the combined effect that the city has on productivity (wages) and amenity (quality of life) (Roback 1982; Glaeser et

7 For the purpose of house price comparison between cities, an ideal house price measure is the price for an imaginary standard house in a standard location in the city. However, such a measure is rarely available. A commonly used alternative measure is the average price of all houses within the city.

al. 2001). In this regard, the spatial pattern of cross-city house prices is tied to how the city's location affects its productivity and amenity.

A framework used to describe the city's location as well as its effect on productivity and amenity is central place theory, in which cities form an urban hierarchy (Christaller 1933). In the hierarchical system, the higher-tier cities enjoy a productivity and amenity premium not found in the lowest-tier cities. The higher the city is in the hierarchy, the larger the premium. The productivity premium is thought to come from various sources. For example, firms in higher-tier cities are more productive because they can economise on transportation costs in relation to delivering goods and providing services. In addition, the frequent exchange of new ideas in higher-tier cities also benefits productivity. The amenity advantage is mainly due to economies of scale, with some higher-order amenities, such as exotic restaurants, luxury shops and specialised healthcare facilities, only present in higher-tier cities because they need a large market to survive. While the lowest-tier cities do not possess productivity and amenity premiums, they may share in the advantages of these higher-tier cities. The extent to which this can happen depends on their proximity to these higher-tier cities. Therefore, in a hierarchical urban system, house price is expected to decrease with greater distance from higher-tier cities.

In addition, the productivity/amenity advantages/disadvantages of a city might not only depend on its hierarchical position as discussed above, that is, distance to higher-tier cities, but also on its position in the city network, that is, its connection with neighbouring cities generally. The latter view treats the urban system as a network of cities where each interacts with all the others, whether they have a higher-rank, a lower-rank or the same rank (Capello 2000; Boix and Trullén 2007). In the network system, a city's productivity advantage relates to aggregate and undifferentiated market potential measured by population or income within a broader region, as suggested by New Economic Geography (Fujita et al. 1999; Partridge et al. 2009). In general, greater proximity to larger markets tends to raise factor prices. The amenity advantage of a city in the network relates to the concept of 'borrowed size', whereby a city can maintain a higher level of amenities than its own size indicates through borrowing size from the cities within the network; at the same time, the cities which offer the support simultaneously have access to these amenities and thus perform better than when they are isolated (Meijers and Burger 2015; Meijers et al. 2016). Note that the position in the hierarchy and the position in the city network are not entirely independent, because a higher-tier urban centre always yields larger market potential and higher degree of amenity spillover, but they are complementary to each other in explaining the spatial pattern of cross-city house prices (Partridge et al. 2009).

The house prices of cities are far more complicated than a theoretical model can predict, with elements such as a 'bubble' that cannot be explained by fundamentals. The spatial pattern of bubbles, therefore, partly contributes to the spatial distribution of house prices. Bubbles are usually not uniformly distributed across space and it is common for bubbles to exist in some cities but not in others. Since housing buyers are not perfectly informed and not rational, they tend to revise their beliefs concerning housing markets through information gained from other agents, a process which can be thought of as 'social learning' or 'social dynamics'(Burnside et al. 2016). In doing so, the optimistic attitudes in bubble markets can easily spread to neighbouring markets and drive up house prices there as well. Such spillovers can result in the spatial clustering of interurban house prices and can be modelled in empirical analysis using spatial econometrics (Fingleton 2008).

§ 1.3.2 Spatial dimension of house price dynamics

The house price dynamics of cities are driven by city-specific demand and supply shifters. If the housing supply is elastic and the market is efficient, then, in the long-run, the house price dynamics of a city will only reflect the changes in construction costs (including the cost of land) of that city. However, this is not the whole story in reality, where the housing supply is always constrained by topographical and planning factors and housing markets cannot clear immediately. Thus, it is the interaction of demand and supply shifters that determines the tendency of house prices to change. In addition, common national factors, such as monetary policy and business cycles, are also important determinants of house price dynamics.

Like the house price level, the house price dynamics of cities also have a spatial dimension. In general, cities that have a close geographical proximity tend to be exposed to similar demand and supply shifters, a similar interaction structure and a similar response to common factors, and hence their house price dynamics are closer to each other than to more distant cities. Clustering homogeneous markets can aid in discovering the spatial pattern of house price dynamics on a larger scale and in identifying sub-national markets. For example, Abraham et al. (1994) revealed three groups of US metropolitan housing markets, namely the West Coast, East Coast and Central US. This clustering logic is also the basis of many regional analyses of house price dynamics. For example, many UK studies have been carried out on the level of Standard Statistical Regions, and their underlying assumption is that the house price dynamics within the region are virtually identical (MacDonald and Taylor 1993; Alexander and Barrow 1994; Holly et al. 2011). However, these regions, designed for administrative purposes, might not completely correspond to the homogeneous market aggregation. At least, for the commercial housing market in the UK, aggregation according to administrative boundaries is not a good solution (Jackson 2002). Thus, in housing analysis, one should be very careful in choosing the appropriate spatial scale.

Another component of the spatial dimension of house price dynamics is the spatial interrelationships between markets. An important hypothesis related to this issue is that the relationship between markets will be stable in the long-run, although in the short-run, prices might be quite different (Meen 1996). This hypothesis was proposed due to the observation that North/South house price differences in the UK widened in the 1980s and then narrowed in the 1990s (Giussani and Hadjimatheou 1991). The UK housing economists also adopted a related concept of the ripple effect, whereby house prices first rise in the southeast and then spread to the rest of the country. What are the mechanisms or behavioural reasons behind these phenomena? Meen (1999) offered five explanations: migration, equity transfer, spatial patterns in the determinants of house prices, spatial arbitrage and coefficient heterogeneity of regional house price models. Of these explanations, the latter two would be the most plausible.

Although the long-run convergence and ripple effect hypotheses originate from empirical observations, the statistical evidence has failed to reach a consensus. While some studies, such as Meen (1996) and Cook (2003), present evidence favouring these hypotheses in the UK, others cast doubt on them (Drake 1995; Abbott and Vita 2013). These hypotheses have also been tested in other markets outside the UK, such as those of Ireland, Sweden, Australia, South Africa and Malaysia, and again the evidence is mixed. The lack of consensus can partly be attributed to the confusion of long-run convergence and the ripple effect. While some studies consider long-run convergence as a cointegration relationship which states that the house prices of different regions are tied together in the long-run through an equilibrium relationship (e.g.,MacDonald and Taylor 1993), others argue that, to ensure the convergence, certain constraints should be imposed on the long-run equilibrium relationship (Abbott and Vita 2013). In other words, cointegration is necessary for convergence, but not sufficient. Some studies think of the ripple effect as Granger causality, which merely describes a relationship in which house price changes in certain markets lead house price changes in other markets (Stevenson 2004). Others emphasize a transmission pattern from leading markets to lagged markets, whereby shocks should first spread to nearby areas, with areas further away taking a longer time to respond (Ashworth and Park 1997). Nevertheless, almost all studies agree that, in the short-run, house price changes in one market can spread to other markets, which is generally defined as a diffusion effect (Pollakowski and Ray 1997).

The long-run convergence discussed above does not imply that house prices are equalized across cities. However, there is another stream of studies focusing on the equalization of city-level house prices, which indicate that properties in areas with lower initial house prices will grow faster in price than those in higher initial price areas. The origins of these studies lie in economic growth theory (e.g.,Solow 1956; Swan 1956). Subject to diminishing returns in capital accumulation, growth theory predicts that economies with different initial conditions will ultimately (absolutely) converge to the same steady-state level of income, with poor economies gradually catching up with the leaders. If one applies this theory to the economy of cities, it is natural to conjecture that per capita income of different cities will ultimately converge, which further leads to the convergence of house prices across cities. However, Kim and Rous (2012) found little evidence of overall convergence among US state and metropolitan housing markets, instead revealing a few 'convergence clubs'. Within each club, the house price disparities between the markets diminish over time, while at the same time, the house price difference between the markets of different clubs might increase. In addition to the US market, the phenomenon of club convergence has also been documented in the UK and Spanish housing markets (Montagnoli and Nagayasu 2015; Blanco et al. 2016). Sometimes the house price disparity between markets diminishes over time after controlling for local characteristics, known as conditional convergence. In such a case, house prices of different markets converge towards some permanent disparity relationships that are determined by the heterogeneity in city-specific house price determinants. An example of conditional convergence can be found in a study by Gyourko and Voith (1992), who revealed that higher priced metropolitan areas in the US tend to have lower appreciation rates after controlling for a local fixed effect and a time-varying national effect.

§ 1.3.3 Spatial dimension in house price index construction

The quality-adjusted house price indexes that measure pure temporal house price changes are usually constructed by two methods: the repeat sales model and the hedonic price model. The repeat sales model is interested in price *changes* and has been applied to houses sold at least twice during the study period, which omits many single sales and is prone to sample selection bias (Wang and Zorn 1997). While the repeat sales method satisfactorily controls for housing qualities, especially for the location characteristics, if one is interested in the *level* of house prices and the shadow prices of housing characteristics, the repeat sales model does not work. The hedonic price model is a desirable alternative, which assumes that the price of a dwelling can be recovered by a set of housing characteristics. When constructing house price indexes, three methods can be employed: time-dummy methods, imputation methods and characteristics methods (Hill 2013).[8]

The challenges in applying hedonic price models to the construction of a house price index concern specifying the correct functional form and choosing the appropriate housing characteristics, with the latter being the larger issue. In general, there are two

8 The time-dummy method estimates a pooled hedonic house price model with time dummies for different periods; the time dummies can be directly used to construct the price index. The imputation method estimates a separate hedonic house price model for each of the periods and imputes the prices of dwellings for each period using the estimated shadow prices. Standard price index formulas, such as Laspeyres and Paasche formulas, are then applied to the imputed prices. The characteristics method is very similar to the imputation method. The key difference is that the characteristics method constructs a hypothetical dwelling and the price index is built on the imputed prices of this hypothetical dwelling.

groups of house price characteristics. Firstly, there are the physical characteristics, with the most common variables used in the literature being floor area, land area, age, number of bedrooms and bathrooms, garage, swimming pool, fireplace and air conditioning (Sirmans et al. 2006). Secondly, there are the location-related characteristics, such as distance to city centre, distance to parks, and the quality of the local school.

Owing to reasons such as data availability, it is impossible to include all the variables that might exert an influence on house prices into the hedonic model. The omission of location variables in particular is likely to cause spatial dependence, the ignorance of which in the hedonic model will yield inconsistent estimates of parameters, which consequently affects the construction of a house price index. To address these problems, spatial econometric models, such as the spatial autoregressive model (SAR), which incorporates the weighted average house price of neighbouring cities as a predictor, and the spatial error model (SEM), which directly models the spatial correlation structure of error terms, have been introduced to the hedonic house price framework (e.g.,Can 1992; Can and Megbolugbe 1997). The spatial-augmented hedonic model can then be used to improve the calculation of hedonic house price indexes. Some examples can be found in Hill et al. (2009), Dorsey et al. (2010), Pace et al. (1998) and Tu et al. (2004).

It is widely recognized that the value of a dwelling is comprised of two components: the value of the land on which the structure sits and the value of the structure. Some researchers have been interested in separate land price indexes and structure price indexes, because it is very plausible that these two indexes evolve differently over time. However, estimating the land and structure price indexes is not easy for markets where there are no explicit land transactions. Two methods have been proposed to separately estimate land and structure prices from home sales: the residual approach and the hedonic approach. Both of these approaches assume that the house value can be split into a reproducible structure component and an unreproducible land component which capitalizes the value associated with location. The residual approach derives land value from the difference between property value and the replacement cost of the same structures after accounting for depreciation (Davis and Heathcote 2007; Davis and Palumbo 2008). The hedonic approach, in contrast, simultaneously estimates the value of the structure and the land in a hedonic framework, where the land price refers to the marginal implicit price per unit of land plot (Kuminoff and Pope 2013; de Groot et al. 2015).

However, in practice, the residual approach is more commonly used because the hedonic approach suffers from omitted variable bias. The houses located in a better neighbourhood with higher land values tend to have nicer physical structural characteristics that cannot be readily observed, and therefore the estimated land value will be confounded with the value of unobservable physical characteristics.

Nevertheless, the hedonic approach has one virtue: if estimated correctly, the hedonic principle of value seems more consistent with the notion of market value (Kuminoff and Pope 2013). Therefore, if one wants to take advantage of the hedonic approach, it is necessary to seek a solution to the omitted variable bias. In particular, the spatial dimension of land values requires better treatment. A common treatment assumes land prices to be constant within the neighbourhood, recognizing the fact that houses within the neighbourhood are exposed to the same local public goods and amenities. However, it can be argued that this treatment might be too crude and that the price of land plots might vary significantly even within one neighbourhood. Imagining a neighbourhood alongside a lake, it is very likely that the land plots near the lake have higher prices than the land plots further away. Therefore, the spatial dimension of land prices associated with location should be treated more concisely when constructing the price index of land, structure and houses.

§ 1.4 Research questions

The objective of this dissertation is to draw a comprehensive picture of house price behaviour in the spatial dimension. It is mainly concerned with three aspects involving two spatial scales. On the interurban scale, the spatial distribution of interurban house prices and the spatial relationships of interurban house price dynamics are the main topics. On the intra-urban scale, the focus is on the spatial correlation and heterogeneity of property prices. In Chapters 2 to 5, a great deal of attention is paid to interurban housing markets in China. In Chapter 6, the focus is on the intra-city housing market in the Netherlands.

The spatial distribution of interurban house prices is intensively dealt with in Chapters 2 and 3. The key questions that need to be answered are:

What is the spatial distribution of house prices across cities? How can that pattern be explained? What role does location play in shaping the interurban house price pattern? (Chapters 2 and 3)

Chapter 2 attempts to answer these questions from the point of view of the hierarchical urban system, where the top-tier cities provide the entire range of urban products and the lower-tier cities only offer a few. The specific questions relating to this chapter are:

Can an interurban house price gradient, whereby house prices decrease when moving away from the core cities to periphery cities, be observed in the urban hierarchical system? If yes, how can we explain this pattern? Which factors can it be attributed to? (Chapter 2)

Chapter 3 argues that the modern urban system is characterized by a city network paradigm, which nests the possibilities of both hierarchical and non-hierarchical structures (Capello 2000). Therefore, this chapter looks into the spatial distribution of house prices from the perspective of city network externalities. The related questions are:

Do cross-city spillovers, which mean that the house price of one market depends on the market conditions of neighbouring markets, contribute to explaining the spatial clustering pattern of interurban house prices? If so, is city network externality one of the channels that generate such spillovers? (Chapter 3)

Chapters 4 and 5 are concerned with the spatial dimension of interurban house price dynamics. The general questions related to this are:

Are house price dynamics across cities different from each other or are they homogeneous? What are the long-run and short-run relationships between them? (Chapter 4 and 5)

Chapter 4 investigates the national interurban housing market and focuses on the overall heterogeneous (or homogeneous) house price dynamics across cities and structural changes across different sub-periods. The main sub-questions in this chapter are:

Can city house price dynamics be divided into a few homogeneous clusters within which the cities have similar house price growth trajectories? Are there structural changes such that the cluster memberships are not consistent across different periods? Can geography play a role in explaining the cluster structure? (Chapter 4)

Chapter 5 pays more attention to the relative relationships between housing market dynamics, such as the leading-lag relationships, and, long-run and short-run relationships. These aspects are reflected in the questions:

Is there any leading-lag relationship across the city housing markets such that the historical house price information in one market can be used to predict the current house prices in other markets? Has a long-run equilibrium relationship been maintained such that the markets will not deviate from each other? Is there a distinct house price diffusion pattern in the short-run such that shocks to one particular market gradually propagate to other markets? (Chapter 5)

Chapter 6 deals with the construction of a house price index, which measures the house price development of a city. Particular attention is paid to the impact of spatial characteristics on the house price index. The questions related to this chapter are:

How can the house price index be decomposed into a land price index and a structure price index? Does better treatment of location benefit the construction of a house price *index?* (Chapter 6)

§ 1.5 Introduction to chapters

Each of the chapters following this introduction responds to the corresponding questions raised in Section 1.4 above. Figure 1.2 below outlines the structure of the dissertation and the theories and/or methodologies used in each chapter. A detailed introduction to each chapter can also be found below.

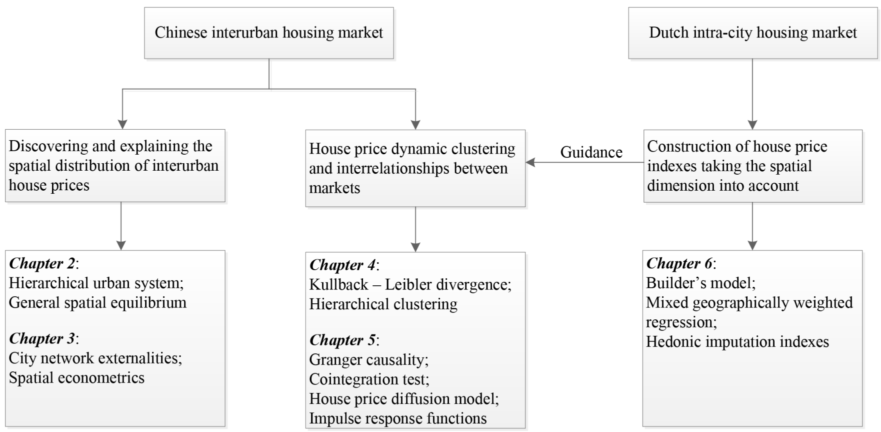


FIGURE 1.2 Outline of the chapters

Chapters 2 and 3 are mainly concerned with the spatial distribution of interurban house prices within the urban system of the Pan-Yangtze River Delta (PYRD) in Eastern China, which includes 42 cities. A panel data set was compiled from various sources for these two chapters. Chapter 2 treats this urban system as a hierarchical urban system, in which one city is deemed to be the top-tier city, three cities to be second-tier cities and all the other cities to be third-tier (lowest-tier) cities. Based on Central Place Theory, which asserts that higher-tier cities will be more productive and produce more urban functions than lower-tier cities (Partridge et al. 2009), the general spatial equilibrium model of Rosen-Roback (Rosen 1979; Roback 1982) demonstrates that the further a city is from the higher-tier cities, the lower the house prices in that city. This negative interurban house price gradient is shaped by two channels: a 'productive component', whereby the more distant cities receive less agglomeration spillovers from higher-tier cities, and an 'amenity component', whereby it is more costly for the more peripheral cities to gain access to higher-order amenities. The interurban house price gradients in relation to higher-tier cities are then empirically estimated in terms of the urban system of PYRD, and they are further decomposed into the productivity component and amenity component so that the relative contribution of these two components can be assessed.

Chapter 3 understands the urban system through the paradigm of a city network system, within which a city can 'borrow size' from neighbouring cities, allowing that city to achieve better performance in terms of productivity and amenity than is indicated by its size (Alonso 1973; Meijers and Burger 2015). Such city network externalities will generate some cross-city spillovers, such that having good access to larger neighbouring markets tends to increase house prices. Based on the urban system of PYRD, the city network externalities in the housing market are empirically modelled using the methods of spatial econometrics, in which the spatial interaction structure is captured by a spatial weight matrix. Specifically, the spatial lag of X model (SLX), which includes the spatial lags of independent variables, and the spatial Durbin error model (SDEM), which captures the spatial lag information of both independent variables and error terms, are employed. These two models can reveal the relationship between one city's house price and the urban size of neighbouring cities, which carries information about city network spillovers. In general, the cross-city spillovers of housing markets – that is, the house price of a city being dependent on the housing market conditions of neighbouring cities – may be raised not only by city network externality, but also by other channels, such as yardstick competitions (Brady 2014).[9] In this sense, another two common approaches, the spatial autoregressive model (SAR), which incorporates the spatial lag of dependent variable, and the spatial Durbin model (SDM), which includes the spatial lags of both dependent and independent variables, are also estimated. However, these two methods are hard to be theoretically justified, and thus suffer from the identification problem (Gibbons and Overman 2012).

Chapters 4 and 5 both deal with the spatial dimension of house price dynamics but with different focuses. Chapter 4 is dedicated to the overall clustering patterns of house price dynamics across the whole country based on some similarity measures. Specifically, it attempts to group the housing markets of 34 major Chinese cities –which are either municipalities directly controlled by the central government, capitals of provinces or vital economic centres –into a few clusters according to the house price appreciation trajectories from 2005 July to 2016 June. The data are extracted from the 'Price Indices of Newly Constructed Residential Buildings in 35/70 Large- and Medium-sized Cities', in which the quality changes have been controlled for to some extent, published monthly by the National Bureau of Statistics of China (NBSC). Before performing the cluster analysis, a measure that reflects the degree of similarity between housing markets must be defined, such as the Euclidean distance. This chapter, being different from the literature, adopts a distribution-based dissimilarity measure – Kullback-Leibler (*KL*) divergence (Kullback 1968), which has been applied in

9 Yardstick competition in housing markets simply means that market participants in one market compete with participants in the neighbouring markets such that the house price formation processes in these markets are correlated with each other.

machine learning and environmental studies but not in housing analysis. The *KL* divergence has a probability meaning and thus can allow one to make inferences, while Euclidean distance does not. The homogeneous clusters are then obtained using the hierarchical agglomerative clustering method, which has been extensively used in the literature on the homogeneous grouping of commercial markets. Considering the changing conditions of Chinese housing markets, structural changes are also tested to see whether or not the cluster membership is consistent throughout the period. To do so, the sample period is split into three sub-periods and the cluster analysis is performed on each sub-sample. Furthermore, this chapter closely examines the effectiveness of two commonly used classification schemes in describing the interurban housing market structure in China – the geographical demarcation system defined by NBSC and the city-tier system published by various institutes.

In order to carefully examine the relative relationships between housing markets, Chapter 5 concentrates on the housing markets of ten vital cities in a common economic area in South China – the Pan-Pearl River Delta (Pan-PRD), which includes cities from developed Eastern China and less developed Central and Western China. The NBSC monthly price indexes are also used, covering the period from June 2005 to May 2015. The first question regarding the relative relationship between housing markets concerns whether the house price change information in some markets leads the house prices changes in other markets. The leading-lag relationships of housing markets are examined using the Toda-Yamamoto Granger causality test (Toda and Yamamoto 1995). Compared to the standard Granger causality procedure, the Toda-Yamamoto procedure is more flexible and powerful. Subsequently, the long-run equilibrium relationship between housing markets is investigated. If the house price ratio between two markets moves around a constant level in the long-run, the two housing markets are then considered to be convergent. The long-run convergence properties between pairwise housing markets are investigated using the Engle-Granger cointegration test, with certain restrictions imposed on the cointegration space. Finally, a house price diffusion model that considers both the long-run and short-run spatial relationships is built. In this diffusion model, the house price growth of a city at time depends not only on its own lagged price changes, but also on the lagged price changes of its neighbours and on the long-run equilibrium relationship with neighbours. In particular, the house price information of neighbours is synthesized using a spatial weight matrix. This model is a variant of that of Holly et al (2011), and, combined with the General Impulse Response Function (GIRF), presents a full picture of the house price behaviour between markets.

Chapter 6 switches the focus from the interurban housing market to an intra-urban housing market, and is concerned with the construction of a house price index, which is the input for the analysis of interurban house price dynamics. Since access to housing transactions in the Chinese market is not available, this chapter is based on a small city in the Netherlands. The chapter starts with a 'builder's model', which decomposes the value of a dwelling into the value of the structure and the value of the land. The land component is of particular interest because location characteristics are mainly capitalized intoland values. As such, land prices are expected to vary significantly within the whole market, even within the neighbourhood, whereas the implicit price of structural characteristics will be the same across space. To capture these features, this chapter applies a mixed geographically weighted regression (MGWR) model, which models the land prices in a nonparametric fashion and the structural prices in a parametric fashion. Specifically, the nonparametric part of the MGWR model assumes that the land price of a location depends on neighbouring land prices, whereby both the spatial dependence and spatial heterogeneity of land prices can be properly dealt with. Another two restrictive models are also estimated, with one assuming that land prices are fixed across the city-wide markets and the other assuming that land prices vary across neighbourhoods. The performance of these three models is then comprehensively assessed. Most importantly, various hedonic imputation house price indexes, land price indexes and structural price indexes are compiled based on the estimates of these three models and the differences between indexes generated by different models are investigated. To simplify the treatment of land component, this analysis mainly uses the sales of single-family dwellings. For the apartments, some special treatment is needed to extract the land component. However, the essence of the model is indifferent to the dwelling types. In this regard, the model presented in this chapter is still enlightening about the construction of separate price indexes for Chinese markets.

References

Abbott, A., & Vita, G. D. (2013). Testing for long-run convergence across regional house prices in the UK: A pairwise approach. Applied Economics, 45(10), 1227-1238.

Abraham, J. M., Goetzmann, W. N., & Wachter, S. M. (1994). Homogeneous groupings of metropolitan housing markets. Journal of Housing Economics, 3(3), 186-206.

Alexander, C., & Barrow, M. (1994). Seasonality and cointegration of regional house prices in the UK. Urban Studies, 31(10), 1667-1689.

Alonso, W. (1964). Location and land use: toward a general theory of land rent. Cambridge MA: Harvard University Press.

Alonso, W. (1973). Urban zero population growth. Daedalus, 102(4), 191-206.

Anderson, W. P. (2012). Economic Geography. Abingdon,Oxfordshire: Routledge.

Ashworth, J., & Park, S. C. (1997). Modeling regional house prices in the UK. Scottish Journal of Political Economy, 44(3), 225-246.

Blanco, F., Martín, V., & Vazquez, G. (2016). Regional house pice convergence in

Spanish during the housing boom. Urban Studies, 53(4), 775-798.
Boix, R., & Trullén, J. (2007). Knowledge, networks of cities and growth in regional urban systems. Papers in Regional Science, 86(4), 551-574.
Brady, R. R. (2014). The spatial diffusion of regional housing prices across U.S. states. Regional Science and Urban Economics, 46, 150-166.
Burnside, C., Eichenbaum, M., & Rebelo, S. (2016). Understanding booms and busts in housing markets. Journal of Political Economy, forthcoming.
Can, A. (1992). Specification and estimation of hedonic housing price mdoels. Regional Science and Urban Economics, 22(3), 453-474.
Can, A., & Megbolugbe, I. (1997). Spatial dependence and house price index construction. Journal of Real Estate Finance and Economics, 14(1-2), 203-222.
Capello, R. (2000). The city network paradigm: Measuring urban network externalities. Urban Studies, 37(11), 1925-1945.
Christaller, W. (1933). Central places in Southern Germany. (translated by C.W.Baskin,1966). London: Prentice Hall.
Cook, S. (2003). The convergence of regional house prices in the UK. Urban Studies, 40(11), 2285-2294.
Davis, M. A., & Heathcote, J. (2007). The price and quantity of residential land in the United States. Journal of Monetary Economics, 54(8), 2595-2620.
Davis, M. A., & Palumbo, M. G. (2008). The price of residential land in large US cities. Journal of Urban Economics, 63(1), 352-384.
de Groot, H. L. F., Marlet, G., Teulings, C., & Vermeulen, W. (2015). Cities and the Urban Land Premium. Cheltenham: Eaward Elgar.
Dorsey, R. E., Hu, H., Mayer, W. J., & Wang, H. (2010). Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle. Journal of Housing Economics, 19(2), 75-93.
Drake, L. (1995). Testing for convergence between UK regional house prices. Regional Studies, 29(4), 357-366.
Fang, H., Gu, Q., Xiong, W., & Zhou, L. (2016). Demystifying the Chinese housing boom. In M. Eichenbaum, & J. Parker (Eds.), NBER Macroeconomics Annual 2015 (Vol. 30, pp. 105-166). Chicago: University of Chicago Press.
Fingleton, B. (2008). Housing supply, housing demand, and affordability. Urban Studies, 45(8), 1545-1563.
Fujita, M., Krugman, P., & Mori, T. (1999). On the evolution of hierarchical urban systems. European Economic Review, 43(2), 209-251.
Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? Journal of Regional Science, 52(2), 172-191.
Giussani, B., & Hadjimatheou, G. (1991). Modeling regional house prices in the United Kingdom. Papers in Regional Science, 70(2), 201-219.
Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. Journal of Economic Geography, 1(1), 27-50.
Gyourko, J., & Voith, R. (1992). Local market and national components in house price

appreciation. Journal of Urban Economics, 32(1), 52-69.
Heikkila, E., Gordon, P., Kim, J. I., Peiser, R. B., Richardson, H. W., & Dale-Johnson, D. (1989). What happened to the CBD-distance gradient? Land values in a policentric city. Environment and Planning A, 21(2), 221-232.
Hill, R. J. (2013). Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy. Journal of Economic surveys, 27(5), 879-914.
Hill, R. J., Melser, D., & Syed, I. (2009). Measuring a boom and bust: The Sydney housing market 2001 - 2006. Journal of Housing Economics, 18(3), 193-205.
Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of Urban Economics, 69(1), 2-23.
Huang, Y. (2004). The road to homeownership: a longitudinal analysis of tenure transition in urban China (1949-94). International Journal of Urban and Regional Research, 28(4), 774-795.
Jackson, C. (2002). Classifying local retail property markets on the basis of rental growth rates. Urban Studies, 39(8), 1417-1438.
Kim, Y. S., & Rous, J. J. (2012). House price convergence: Evidence from US state and metropolitan area panels. Journal of Housing Economics, 21(2), 169-186.
Kullback, S. (1968). Information theory and statistics. New York: Dover Publications.
Kuminoff, N. V., & Pope, J. C. (2013). The value of residential land and structures during the great housig boom and bust. Land Economics, 89(1), 1-29.
Li, Q., & Chand, S. (2013). House prices and market fundamentals in urban China. Habitat International, 40, 148-153.
MacDonald, R., & Taylor, M. P. (1993). Regional house prices in Britain: long-run relationships and short-run dynamics. Scottish Journal of Political Economy, 40(1), 43-55.
Meen, G. (1996). Spatial aggregation, spatial dependence and predictability in the UK housing market. Housing Studies, 11(3), 345-372.
Meen, G. (1999). Regional house prices and the ripple effect: A new interpretation. Housing Studies, 14(6), 733-753.
Meijers, E. J., & Burger, M. J. (2015). Stretching the concept of 'borrowed size'. Urban Studies, Forthcoming.
Meijers, E. J., Burger, M. J., & Hoogerbrugge, M. M. (2016). Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe. Papers in Regional Science, 95(1), 181-199.
Mills, E. S. (1967). An aggregation model of resource allocation in a metropolitan area. American Economic Review, 57(2), 197-210.
Montagnoli, A., & Nagayasu, J. (2015). UK house price convergence clubs and spillovers. Journal of Housing Economics, 30, 50-58.
Muth, R. F. (1969). Cities and housing: the spatial pattern of urban residential land use. Chicago: University of Chicago Press.
Pace, R. K., Barry, R., Clapp, J. M., & Rodriquez, M. (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. Journal of Real Estate Finance

and Economics, 17(1), 15-33.
Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. Journal of International Economics, 78(1), 126-140.
Pollakowski, H. O., & Ray, T. S. (1997). Housing price diffusion patterns at different aggregation levels: An examination of housing market efficiency. Journal of Housing Research, 8(1), 107-124.
Qin, B., & Han, S. S. (2013). Emerging polycentricity in Beijing: Evidence from housing price variations,2001-05. Urban Studies, 50(15), 2006-2023.
Roback, J. (1982). Wages, rents, and the quality of life. Journal of Political Economy, 90(6), 1257-1278.
Rosen, S. (1979). Wage-based indexes of urban quality of life. In P. Mieszkowski, & M. Straszheim (Eds.), Current Issues in Urban Economics. Baltimore: The Johns Hopkins University Press.
Sirmans, G. S., MacDonald, L., Macpherson, D. A., & Zietz, E. N. (2006). The value of housing characteristics: A meta analysis. The Journal of Real Estate Finance and Economics, 33(3), 215-240.
Solow, R. M. (1956). A contribution to the theory of economic growth. Quarterly Journal of Economics, 70(1), 65-94.
Stevenson, S. (2004). House price diffusion and inter-regional and cross-border house price dynamics. Journal of Property Research, 21(4), 301-320.
Swan, T. W. (1956). Economic growth and capital accumulation. Economic Record, 32(2), 334 - 361.
Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66, 225-250.
Tu, Y., Yu, S.-M., & Sun, H. (2004). Transaction-Based Office Price Indexes: A Spatiotemporal Modeling Approach. Real Estate Economics, 32(2), 297-328.
Waddell, P., Berry, B. J. L., & Hoch, I. (1993). Residential property values in a multinodal urban area: New evidence on the implicit price of location. Journal of Real Estate Finance and Economics, 7(2), 117-141.
Wang, F. T., & Zorn, P. M. (1997). Estimating house price growth with repeat sales data: What's the aim of the game? Journal of Housing Economics, 6(2), 93-118.
Wang, S.-Y. (2011). State misallocation and housing prices: Theory and evidence from China. American Economic Review, 101(5), 2081-2107.
Wang, Y., Shao, L., Murie, A., & Cheng, J. (2012). The maturation of the neo-liberal housing market in urban China. Housing Studies, 27(3), 343-359.
Xu, X. E., & Chen, T. (2012). The effect of monetary policy on real estate price growth in China. Pacific-Basin Finance Journal, 20(1), 62-77.
Zhang, Y., Hua, X., & Zhao, L. (2012). Exploring determinants of housing prices: A case study of Chinese experience in 1999 - 2010. Economic Modelling, 29(6), 2349-2361.
Zheng, S., Cao, J., Kahn, M. E., & Sun, C. (2014). Real estate valuation and

cross-boundary air pollution externalities: Evidence from Chinese cities. Journal of Real Estate Finance and Economics, 48(3), 398-414.
Zheng, S., & Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. Journal of Urban Economics, 63(2), 743-757.
Zheng, S., Kahn, M. E., & Liu, H. (2010). Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. Regional Science and Urban Economics, 40(1), 1-10.

2 Interurban house price gradient: Effect of urban hierarchy distance on house prices

Published as: Gong, Y., Boelhouwer, P., de Haan, J., 2016. Interurban house price gradient: Effect of urban hierarchy distance on house prices. Urban Studies, 53(15):3317-3335.

Abstract: This paper applies a general spatial equilibrium model to investigate the effect that distance within urban hierarchy can have on interurban house prices. Our spatial model predicts a negative price gradient towards higher-tier cities, which can be decomposed into a 'productivity component' and an 'amenity component', representing respectively the effect of wage differences and households' valuation of access to higher-order services. The theoretical findings are tested on data for the hierarchical urban system of the Pan-Yangtze River Delta in China. Both central and subcentral cities are shown to impose statistically significant distance penalties on interurban house prices, even after we control for amenities and characteristics that are generally considered to be the determinants of house prices. According to the empirical decomposition, the negative house price gradients are largely accounted for by the productivity component.

Keywords: China, distance effect, house price gradient, spatial equilibrium, urban hierarchy

§ 2.1 Introduction

House prices vary significantly across areas. For example, the average house price for the U.S. metropolitan statistical areas (MSAs) in the top price decile in 2000 was about 290,000 dollars, almost four times the average for the MSAs in the bottom decile and more than twice the mean value[1]. Likewise, developing economies such as China exhibit a huge house price differential between cities. According to the "100 city house price index report", the 90th percentile price for 95 prefecture cities (municipalities) in

1 Details on mean house price values are available in the paper by Gyourko et al. (2010).

December 2010 was more than three times higher than the 10th percentile price[2].

The remarkable variation in house price across cities is usually attributed to differences in socio-economic conditions and amenities.For instance, cities with a warm winter or cool summer are always expensive (e.g., Rappaport 2007). In addition, man-made amenities, notably school quality and crime rate, have a significant effect on house prices (e.g.Gyourko and Tracy 1991)[3]. Topographical constraints and legal regulations, which determine the housing supply, can also affect house prices (e.g., Malpezzi 1996). Furthermore, as we argue here, the relative location of a city should be taken into account when explaining interurban house price patterns. This paper demonstrates a persistent spatial pattern whereby differences in house prices tend to increase as the location shifts from the core city to peripheral cities.

The effect of location on the price of inner-city land (or houses) has been widely investigated since the pioneering work of Alonso (1964), Mills (1967) and Muth (1969). Their work predicted a negative effect of distance on price when moving away from the Central Business District (CBD). Empirical evidence for negative gradients of population density, house prices or land values has been found in studies of Chicago, Berlin, Stockholm, Beijing and the southern part of West Norway (McMillen 1996; Söderberg and Janssen 2001; Osland et al. 2007; Zheng and Kahn 2008; Ahlfeldt 2011). The pattern of house prices in a modern polycentric city is much more complicated, however. There, location is also shaped by proximity to subcentres and other important nodes like universities, hospitals and parks (Heikkila et al. 1989; Waddell et al. 1993; Qin and Han 2013).

In an interurban context where cities form a hierarchy[4], it is no surprise to find that house prices in top-tier cities tend to be the highest, whereas lower prices characterize the lowest-order cities in the hinterland. Yet relevant studies on the effect of a city's location on house prices are largely absent; to our knowledge, only two have been published. de Bruyne and van Hove (2013) developed a theoretical model to explain, from the perspective of commuters, how access to a core municipality will affect house prices. The underlying premise is that commuters have to compensate for their loss in leisure time and for the cost of the journey to work by economizing on housing expenditure. Using municipal-level data for Belgium, they found solid evidence supporting their hypothesis: good access to economic centres (capital city or provincial

2 This index is published monthly by the China Index Academy. Source:http://fdc.fang.com/index/BaiChengIndex.html.

3 Climatological or environmental conditions are natural (pure) amenities because they are non-produced and have no explicit prices. On the other hand, amenities like government services are considered man-made (non-pure), as they can be priced in terms of taxes.

4 According to Central Place Theory, a city in the top tier of an urban hierarchy provides the entire range of urban products, whereas lower-order cities provide fewer products (Fujita et al. 1999).

capitals) will increase house prices.

Commuting between core and peripheral cities might not be realistic in some countries. In this regard, Partridge et al. (2009) present a general analytical framework that combines the spatial general equilibrium framework of Roback (1982) and Central Place Theory. They state that location characteristics (access to higher-tier centres) will enhance a firm's profitability and households' utility respectively by providing access to greater markets and unique consumer services like exotic restaurants, renowned museums and specialized healthcare facilities. Spatial differences in house prices are thus outcomes of the location responses of firms and households to the urban hierarchy. Data for rural and urban counties in the U.S. shows that estimated incremental distance penalties for remoteness from the combined tiers of the urban hierarchy are about 12% to 17%[5].

In the study underlying this paper, we systematically investigated how the location of a city – i.e., distance to higher tiers within the urban hierarchy – would affect house prices by applying a general spatial equilibrium framework analogous to that of Partridge et al. (2009). This framework predicts a negative interurban house price gradient with respect to higher-tier cities. The price gradient can be decomposed into a 'productivity component', which represents the effects of wage difference caused by agglomeration spillovers, and an 'amenity component', which reflects households' valuation for access to higher-tier consumer services. We used aggregate data from China's Pan-Yangtze River Delta, where a housing market has emerged and matured since the housing system reform was launched in 1998. With that data, a series of interurban house price gradients were estimated and empirically decomposed after controlling for city-specific amenities and characteristics.

The contribution we intend to make with this paper is twofold. First, we test the penalties imposed by distance within the urban hierarchy on house prices in developing countries where the spatial pattern of interurban house prices has been largely understudied. Second, we attempt to decompose the house price gradient rather than wage (growth) differentials, the latter having been analysed previously by Partridge et al.(2010).

§ 2.2 Related literature

5 The housing cost gradient of Partridge et al. (2009), which is based on incremental distance, differs from the common house price gradient in intra-city studies. We use an example to explain incremental distances. Consider an urban hierarchy with three tiers: the third-tier (lowest-level) city is 100 km from the nearest second-tier city and 300 km from the nearest first-tier (highest-level) city. The incremental distance with respect to the first-tier city is thus 200 km (300-100).

Explaining house price differences across markets has long been a concern, as amply demonstrated in the literature[6]. Ozanne and Thibodeau (1983) developed an implicit demand and supply model of metropolitan housing markets. The markets were divided into rental and homeowner sectors, which are linked by tenure choice and the urban land market. Reduced equations for house prices and rents were then estimated using seemingly unrelated regression (SUR) method based on a data set of MSAs in the USA. Among other independent variables, they considered median income, number of households, demographic characteristics, tax, construction cost, price of land and other consumer goods, as well as geographic features and government restrictions on land supply. Surprisingly, they found that two variables, namely income and number of households, significantly affect rents but not house prices. Coastal location, as a proxy for topographical land use restrictions, had no influence on house prices either. Potepan (1996) further extended the framework of Ozanne and Thibodeau (1983) to include housing service, housing capital and urban land markets, of which the first two are linked through user-cost relationships. In contrast, their reduced-form estimates based on data for MSAs confirm the significant effect of income and population (change) on house prices. Amenities, such as climate and quality of public services, were also shown to influence house prices. Using a provincial panel data set for China, Li and Chand (2013) also found that income level and the ratio of impending marriage population to total population have a significant effect on house prices.

While most of the studies include income and population as independent variables, the spatial general equilibrium framework (Roback 1982; Glaeser et al. 2006) clearly justifies the endogeneity between wage, population and house prices. This framework accommodates migration across markets to equalize the inter-urban utility level. Accordingly, price differences between cities are considered as compensating differentials that compensate for city amenities. The implicit prices of amenities can further be used to calculate a quality-of-life index. Gyourko and Tracy (1991) regressed housing expenditure on a set of pure amenities such as climate and environmental indicators and a set of non-pure amenities such as education, safety and healthcare. In general, they found that those amenities, as a group, significantly affect housing expenditure in the USA. Similarly, Rappaport (2007) provides evidence from the U.S. market that counties with warmer winters and cooler summers enjoy higher growth in house prices. Not surprisingly, amenities are also highly valued in Chinese housing markets. For example, green space and beach access have a positive relationship with house prices, while air pollution, measured as particulate matter (PM), affects house

6 House prices at different levels of spatial aggregation are influenced by different attributes. For the attribute of property, house price determinants usually fall into three categories: structural, locational and neighbourhood characteristics (e.g. Ottensmann et al. 2008; Qin and Han 2013). House price determinants at the national level usually include income, interest rate, population and construction cost (e.g. Drake 1993; Meen 2002). In this paper, we pay particular attention to the regional determinants. Although there is a large volume of literature using time-series analysis, here we are mainly interested in cross-sectional studies.

prices negatively (Zheng et al. 2010). Moreover, cross-boundary pollution flows, referring to pollutants carried by wind from other cities, also have a negative effect on a specific city's house price (Zheng et al. 2014).

On the supply side the price of raw land and construction cost are the two main factors, even playing a bigger role in explaining house prices in more developed cities (Li and Chand 2013). Conditions such as topographical features and regulation constraints, which may be directly or indirectly correlated with land prices and construction cost, can also affect house prices. When facing a demand shock, cities with a relatively elastic supply will experience a modest house price increase because of the unfettered new supply. On the other hand, house prices must rise dramatically in cities with an inelastic supply (Glaeser et al. 2006). Malpezzi (1996) investigated the relationship between the regulatory environment, as measured by a series of rent controls and zoning plans, and housing markets in the USA and found that regulation raises rents and house prices but lowers homeownership rates. The finding that greater regulatory restrictiveness will increase house prices or foster a larger house price growth in a booming period is further confirmed by Ihlanfeldt (2007) and Huang and Tang (2012), among others.

More recently, some studies have considered the spatial dimension of house price determinants. A few authors have investigated the role that the relative location of a city within the urban hierarchy plays in forming house prices, assuming that central cities that have larger market potential and higher consumer amenities will have a positive effect on nearby cities' house prices (Partridge et al. 2009; de Bruyne and van Hove 2013). Our study will contribute to this stream of research by investigating the effect of distance within the urban hierarchy on interurban house prices in an emerging market – China.

§ 2.3 Theoretical framework

Our theoretical framework follows the spatial general equilibrium model of Roback (1982), which has been extensively used by Beeson and Eberts (1989) and Partridge et al. (2010). To perform our analysis, we made several assumptions. Both capital and labour can move freely across cities, thereby allowing individuals to select their residential location within a particular city and to choose between different cities. However, the option of living in one city and working in another is ruled out. Further, land is fixed in each city but can be freely changed between uses.

Households maximize utility subject to a budget constraint by choosing amounts of traded composite goods (x) and housing production (h^c), given the bundle of urban amenities (s) and location characteristics (D):

$$\max U(x, h^c; s, D) \quad \text{subject to} \quad w = x + ph^c \tag{1}$$

where w represents the wage that makes up the bulk of the household budget; p denotes the house price while the price of composite goods is determined by international markets and set to unity. Urban amenities (s) include pure amenities (e.g., weather) and man-made amenities (e.g., government services). Location characteristics (D) are a set of distances to cities of different (higher) tiers; thereby, these characteristics capture what we call the urban hierarchy distance effect. The indirect utility function (V) can be derived from equation and must be equal to c across cities in spatial equilibrium:

$$V(w, p; s, D) = c. \tag{2}$$

The indirect utility function has the usual properties, $V_w = \partial V/\partial w > 0$ and $V_p < 0$. If urban amenities have a positive effect on utility, then $V_s > 0$.

Suppose that land is the only input of housing production according to a constant-return-to-scale production function: $H = h(l^h; s, D)$. The problem for a representative housing developer is to minimize costs subject to the production function. We can consider the unit cost function since the production function is the constant return to scale and the equilibrium condition is:

$$G(r; s, D) = p(s, D) \tag{3}$$

where r represents the land price. As usual, the unit cost function is increasing in factor prices, so $G_r = l^h/H > 0$. If urban amenities (s) provide net productivity advantages to housing developers, then $G_s < 0$; otherwise $G_s > 0$.

Following the tradition of Rosen (1979), households are viewed as self-producers of composite goods. The assumption of self-production ensures that land is not a factor of production. That is, $X = f(N^x; s, D)$, where N^x is the labour for producing composite goods. Assuming that f is also a constant-return-to-scale function, then the unit cost must be equal to the product price in equilibrium:

$$C(w; s, D) = 1. \tag{4}$$

The unit cost function $C(\cdot)$ has properties similar to $G(\cdot)$, $C_w = N^x/X > 0$ and $C_s < 0$ if urban amenities have positive productivity effects. Note that the labour and land market clearing conditions require that $N = N^x$ and $L = l^h$, where N and L denote a city's total amount of labour and land, respectively.

Now we turn to the effect that distances in the urban hierarchy (D) exert on a household's utility and production costs. As central place theory suggests, higher-tier cities can provide all the functions of lower-tier cities as well as a higher level of service or product (with a higher demand threshold) that is first available at the higher tier. Therefore, agglomeration economies are expected to be greatest in the highest-tier cities in the hierarchical urban systems (Partridge et al. 2009). Owing to agglomeration spillovers or benefits (e.g., of knowledge), a city that is close to the higher-tier city usually has higher productivity advantages in both housing production and composite goods; thus, $G_D > 0$ and $C_D > 0$. Households in that city also benefit from their access to higher-tier cities, which can offer unique higher-order cultural, recreational and consumer services, so that $V_D < 0$.

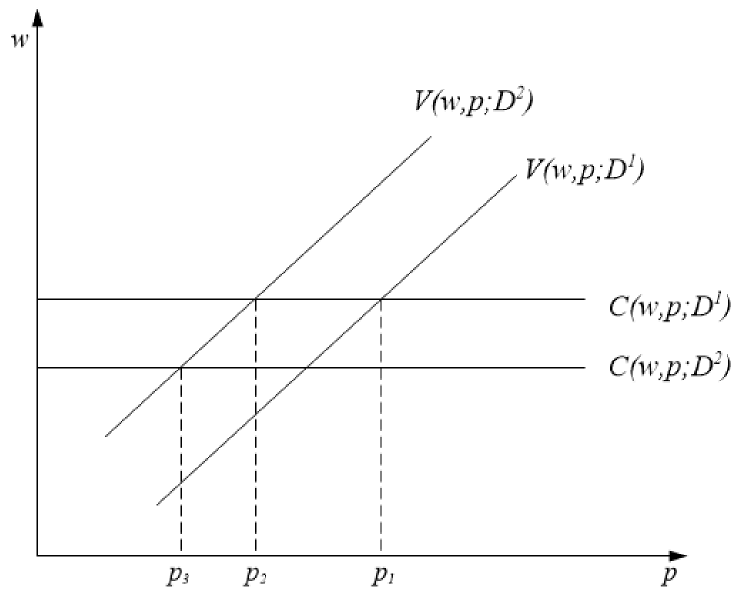


FIGURE 2.1 Illustration of distance effects on equilibrium wages and house prices.

Holding within-area amenities(*s*) constant across cities, the combination of Equations (2), (3) and (4) can be solved simultaneously for *w*, *p* and *r* in terms of location characteristics (*D*). Figure 2.1 illustrates the effect of greater remoteness from a higher-tier city – say, the central city in the highest tier of an urban hierarchy. The upward-sloping line represents the iso-utility curve where higher house prices require higher wages to keep utility constant. Since land is not a factor for composite goods production, the iso-cost curve of a composite good is represented by the horizontal line. Suppose that city 2 is more distant from the central city than city 1. Then the lower degree of access to the central city moves the iso-utility curve leftward, reducing house prices, and shifts the iso-cost curve downward, reducing both wages and house prices. The total decrease in house price due to remoteness from the central city is $p_1 - p_3$.

Totally differentiating the equations (2), (3) and (4) and solving for dw/dD, dp/dD and

dr/dD (assuming $ds = 0$), we obtain:

$$\frac{dw}{dD} = -\frac{C_D}{C_w} < 0, \quad \frac{dp}{dD} = \frac{C_D V_w}{V_p C_w} - \frac{V_D}{V_p} < 0,$$
$$\frac{dr}{dD} = \frac{1}{G_r}\left[\frac{C_D V_w}{V_p C_w} - \frac{V_D}{V_p} - G_D\right] < 0 \tag{5}$$

Rearranging dp/dD and using Roy's identity, the negative house price gradient can be decomposed into two components:

$$\frac{dp}{dD} = \frac{1}{h^c}\frac{dw}{dD} + \left(-\frac{V_D}{V_p}\right). \tag{6}$$

The first term in Equation (6) is the effect on p due to the changes in wages, i.e., the productivity advantages ($(dp/dD)^C$). The second term is the amount of house price required to compensate for remoteness from the higher-order services offered by the central city, i.e., the effect of consumer amenities ($(dp/dD)^V$). These two terms correspond to $p_2 - p_3$ and $p_1 - p_2$ in Figure 2.1, respectively. Again, land is not a factor of production of composite goods. Therefore, the value of consumer amenities will be completely capitalized in house prices. The contribution of the amenity component to the total house price gradient can be derived from Equation (6):

$$\left(\frac{dp}{dD}\right)^V = \frac{dp}{dD} - \frac{1}{h^c}\frac{dw}{dD} \quad \text{or} \quad \left(\frac{d\log p}{dD}\right)^V = \frac{d\log p}{dD} - \frac{1}{k}\frac{d\log w}{dD}, \tag{7}$$

where $k = h^c p/w$ is the share of the consumer's budget spent on housing. The magnitude of the productivity component can be derived in a similar way.

§ 2.4 Hierarchical urban system and empirical data

§ 2.4.1 Hierarchical urban system of interest

Prior to introducing the readers to the hierarchical urban system covered in this study, we offer some background on the administrative arrangement of Chinese urban areas. A typical prefecture city, or a municipality directly under the central government (municipality for short), usually consists of districts and counties (or county-level cities). The 'city proper' (*shiqu*) of the prefecture city is made up of the districts (Ding 2013) [7]. The hierarchical urban system mentioned in this paper pertains to the city proper of prefecture cities and municipalities.

7 It should be noted that the city proper in China is a smaller subset of the administrative area of a prefecture city. It is made up of city districts, the boundaries of which are determined by legal and administrative criteria. Distinct from common usage, the 'city proper' of a prefecture city in China is overbounded and usually encompasses urban, suburban and rural areas.

The empirical grounds for the study refer to the hierarchical urban system of the Pan-Yangtze River Delta (PYRD). The area comprises one municipality (Shanghai) and three provinces (Jiangsu, Zhejiang and Anhui), including 42 cities, with a land area of 350,000 km^2 and a population of 215 million in 2010 (see Figure 2.2). With the *hukou* restriction on labour mobility being phased out in the transition to a market economy, a more liberal labour market has emerged. People can freely migrate to cities that offer higher real wages or better urban amenities. For example, the population of Shanghai increased by 43% from 2000 to 2010. Furthermore, urbanites tend to live and work in the same city because of cultural traditions, the expense of commuting and so on. Given these features, the PYRD constitutes a natural experimental setting for our theoretical analysis.

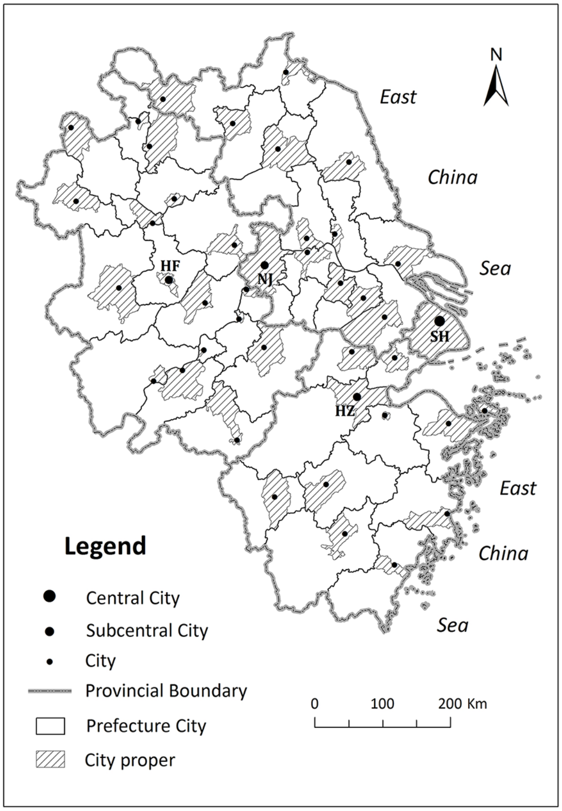


FIGURE 2.2 Hierarchical urban system of Pan-Yangtze River Delta

Accompanying the rapid economic growth and liberalization of the labour market, the increasing urban population has been accommodated in a modern, market-oriented housing sector since the housing reform of 1998. Three types of housing are provided to meet the demand of different income groups: commercial housing, government-supported affordable housing (*Jingji shiyong fang*) and government-subsidized rental housing (*Lianzu fang*) (Wang et al. 2012). The commercial sector is market-oriented. At present, it comprises the majority of the units, even though affordable housing has been encouraged and supported by governments in recent years.

TABLE 2.1 Economic development of central and subcentral cities, 2010

city	Working population (10^4 Person)	Overall rank	Rank in corresponding province	GDP (billion Yuan)	Overall rank	Rank in corresponding province
Shanghai	716.74	1		1697.16	1	
Nanjing	259.18	3	1	451.52	3	1
Hangzhou	333.58	2	1	474.08	2	1
Hefei	114.28	8	1	192.04	8	1

In theory, a higher-tier city should have a relatively large market and provide higher-order services and products for lower-tier cities. In China, urban development and the spatial layout of cities are usually guided by the upper-level governments' plans. Therefore, the Outline of National Urban System Planning (2005-2020) offers a good perspective from which to define the urban hierarchy in the PYRD area. The Outline identifies a three-tiered urban system. Shanghai, planned to be the nationwide central city, is undoubtedly the only highest-tier city. Nanjing, Hangzhou and Hefei, as local-central cities and the capitals of Jiangsu, Zhejiang and Anhui province respectively, comprise the second (subcentral) tier. They offer higher-order functions and services for third-tier cities. Note that Ningbo, designated as a local-central city, is excluded from the list of subcentral cities. Because Ningbo lies very close to Hangzhou; its influence on third-tier cities can be easily overshadowed by Hangzhou [8]. The evidence presented in Table 2.1 suggests that the inclusion of Hefei as a subcentral city is a bit controversial; its economic indicators, namely the working population and GDP, are not in the top rank. However, considering its leading position in Anhui province, it is reasonable to define Hefei as one of the local centres.

§ 2.4.2 Model specification and data

We use a set of panel data for 42 cities (41 prefecture cities and one municipality) spanning the period from 2006 to 2010. Thus, we have 210 annual observations. According to the theoretical model, the full specification of the pooled cross-sectional

8 The correlation between distance to Hangzhou and to Ningbo is 0.922. The estimated subcentral city gradient with or without Ningbo as subcentral city, using the Semi-log/Log-log model shown in Table 2, does not differ much (-0.0701 without Ningbo against -0.0744 with Ningbo).

model can be expressed as:

$$\begin{aligned} Hprice_{it} &= \alpha^H + \mathbf{Location}_i\boldsymbol{\beta}^H + \mathbf{Amenity}_{it}\gamma^H + \mathbf{Time}_t\boldsymbol{\lambda}^H + \epsilon_{it}^H \\ Wage_{it} &= \alpha^I + \mathbf{Location}_i\boldsymbol{\beta}^I + \mathbf{Amenity}_{it}\gamma^I + \mathbf{Time}_t\boldsymbol{\lambda}^I + \epsilon_{it}^I \end{aligned} \quad (8)$$

where $Hprice_{it}$ and $Wage_{it}$ represent the house prices and wages of city i at time t, respectively; the vector $\mathbf{Location}_i$ contains a set of distance measures to central and subcentral cities in the urban hierarchy and probably remains constant over time; the vector $\mathbf{Amenity}_{it}$ includes the measures of city amenities and characteristics; and the vector $\mathbf{Time}_t$ contains dummy variables that control for time effects. The house price equation in model (8) is our main focus.

The primary data sources for this paper are the city-level or province-level statistical yearbooks as well as China City Statistical Yearbooks. Here, the notion of house price refers to the average sale price of newly sold residential buildings per square metre of floor space in the city proper [9]. This measure includes both finished and pre-sale housing [10].One drawback of this measure is that it does not control for housing quality. Nonetheless, to our knowledge, this is the only aggregate measure of house price that can cover all of the cities in this analysis. The wage level is approximated by annual average wages of employees working in state-owned, collective-owned non-private sectors. The wage data is gathered from China City Statistical Yearbooks.

Both geographical distance and travel time are used to measure the accessibility of a city to higher-tier cities. Geographical distance is the straight line distance between the CBD of two cities, while travel time means the least amount of driving time extracted from Google Maps in December 2012. By the same approach, geographical distance and travel time to the nearest subcentral city are constructed. Our measure here differs from the *incremental* distance (see note 5 for a detailed explanation) of Partridge et al. (2009). Under the assumption of *incremental* distance, a third-tier city, say city *i*, that is proximal to the central city will no longer be influenced by the subcentral city. In contrast, the distance measures applied here contain the influence of both the central city and the subcentral city on city *i*.

A set of variables are chosen as proxies for city amenities and characteristics. The main climate variable is the winter temperature, specifically the average temperature of December, January and February. The summer temperature is excluded, as it does not vary much across our study area. The environmental indicator is the annual amount of

9 A few cities only have sale data incorporating all kinds of buildings: commercial, residential and some other types. However, according to data from other cities, residential buildings account for the great majority of total sales. Therefore, the average price of residential buildings in these cities is a corrected average price for all buildings; the correction coefficient is determined by the nearby cities.

10 In the pre-sale housing (*qifang*) market, purchasers buy houses that are not yet completed from the developers in the form of a forward contract. Since the Chinese housing market is experiencing rapid growth, the pre-sale housing sector enjoys a very large share of the market.

industrial smoke and dust emissions per GDP. Smoke and dust are the two major components of particulate matter, which is an important aspect of quality of life. This measure reflects the intensity of particulate matter emissions. Higher emission intensities usually indicate a higher share of the polluting sector in the industrial composition, which will make the city less pleasant to live in. We also create the dummy variable 'coastal city' to measure the living comfort of a city. It takes the value of 1 if the city proper borders an ocean but the value of 0 otherwise. The man-made amenities we consider are healthcare and education conditions, the most important aspects of quality of life in a city. They are approximated by the ratio of students to teachers and the number of physicians per thousand inhabitants. Finally, the variable 'arable land per capita in 2004' is incorporated as a proxy for planning and regulation constraints. To ensure grain security, the central government has drawn a 'red line' minimum for arable land at 120 million hectares in the whole country. In this regard, a city with less arable land will probably face more strict planning and regulation constraints, which will consequently push up the house price but limit its population growth (Glaeser et al. 2006). Note that the spatial context of winter temperature, smoke and dust emissions, and arable land per capita does not pertain to the city proper but covers the whole prefecture city (including counties or county-level cities) [11].

§ 2.5 Estimating interurban house price gradients

A set of interurban house price gradients were estimated to investigate the distance penalties of central and subcentral cities. First, a parsimonious model that only considers the effect of a central city was estimated based on three distance-decay forms. Second, the augmented models that contain both central and subcentral cities were used to detect the house price pattern in a polycentric urban system. Third, after controlling for city amenities and characteristics, the house price gradients of central and subcentral cities were re-estimated.

§ 2.5.1 Central-city house price gradient

Specifying the functional form is an important issue in empirical analysis. In order to choose the 'best' model specification, we considered three distance-decay forms of parsimonious models: linear (Level-Level), semi-log (Log-Level) and log-log (Log-Log). In addition, we included two regional dummy variables to control for the provincial fixed effects of Jiangsu and Anhui, such as natural resource availability and policy difference.

11 The descriptive statistics of all variables are available on request.

The estimated central-city house price gradients are reported in Table 2.2. *t* statistics were produced on the basis of standard errors clustered by city that are robust to correlation between error terms of the same city and heteroskedasticity over time. Except for the linear functional form where the negative coefficients of distance measures are not significant, both the semi-log and log-log form detected the highly significant distance penalties of the central city (Shanghai) on house prices in other cities. These penalties are in accordance with our theoretical findings as well as findings in U.S. housing markets. To our surprise, geographic distance performs even better than travel time, which is considered to be more appropriate for measuring the accessibility between cities. The explanation may be related to the fact that the travel time has changed along with the continuous improvement of transportation infrastructure in the study area. However, what we actually used is a constant travel time derived from Google Map service, which could not track such changes[12]. The following analysis only takes geographical distance into account.

The semi-log functional form using geographical distance performs best, according to the goodness-of-fit and AIC criteria. Together with two regional dummy variables, the geographic distance to the central city can explain 70% of the spatial variance of house prices in this model. The corresponding negative gradient is -0.0011, indicating that for one kilometre farther away a city lies from the central city, the average house price will decrease by about 0.11% when holding the regional effects constant. Moreover, house prices in Jiangsu province are significantly lower than those in Zhejiang, and Anhui is even cheaper. Finally, the estimation results of four time dummy variables show that overall house prices rose continuously during the study period, though we do not report the results [13].

§ 2.5.2 House price gradient of both central and subcentral cities

To investigate the distance penalties of both central and subcentral cities on interurban house prices, we extended the framework of Heikkila et al.(1989), who considered the role of subcentres in a polycentric city. For our interurban augmented model, we assumed a competitive relationship among three subcentral cities but a complementary relationship between each of them and the central city. Thus, access to subcentral cities is measured by the distance to the nearest subcentral city. We then assigned either the semi-log or log-log distance-decay form to both central and subcentral cities, resulting in four models with different functional combinations.

12 We thank one anonymous referee for noting this point.

13 The interaction terms of distance measure and time dummies are also included to test the hypothesis that the interurban house price gradient would flatten over time. However, we found no evidence supporting this proposition.

TABLE 2.2 Estimation of central-city house price gradient

	Linear/Dist	Linear/Time	Semi-log/Dist	Semi-log/Time	Log-log/Dist	Log-log/Time
	House price	House price	Ln(House price)	Ln(House price)	Ln(House price)	Ln(House price)
Constant	6699.1337***	6636.7919***	8.7021***	8.7009***	9.0524***	9.0272***
	(9.66)	(8.25)	(82.59)	(73.12)	(45.59)	(48.71)
Distance to central city	-3.6720	–	-0.0011***	–	–	–
	(-1.24)		(-2.84)			
Ln(Distance to central city)	–	–	–	–	-0.1198**	–
					(-2.58)	
Travel time to central city	–	-3.5313	–	-0.0012**	–	–
		(-0.89)		(-2.27)		
Ln(Travel time to central city)	–	–	–	–	–	-0.1156***
						(-2.62)
Dummy: Jiangsu	-2893.4657***	-3018.9746***	-0.4331***	-0.4702***	-0.4260***	-0.4428***
	(-2.79)	(-3.04)	(-3.08)	(-3.40)	(-2.93)	(-3.10)
Dummy: Anhui	-3670.9179***	-3917.2915***	-0.6136***	-0.6767***	-0.7090***	-0.7308***
	(-2.75)	(-3.04)	(-3.56)	(-4.03)	(-4.73)	(-5.04)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
F-stats	30.27	29.51	69.17	64.73	63.22	61.92
R^2	0.512	0.506	0.706	0.692	0.687	0.682
Adj.R^2	0.492	0.486	0.679	0.665	0.660	0.656
AIC	3832.490	3835.200	76.90	86.64	90.04	93.03
Sample size	210	210	210	210	210	210

Notes: ** and *** denote significance at the 5% and 1% level, respectively. The t values shown in parentheses are calculated based on clustered standard errors that are robust to serial correlation of the same unit and heteroskedasticity over time.

The results shown in Table 2.3 reveal that both central and subcentral cities have significantly negative distance effects on house prices when using the semi-log/log-log form and log-log/semi-log functional form. Those results also show that the semi-log/log-log form is the best one in terms of adjusted R^2 and AIC criteria. However, the distance penalties are no longer significant, even at 10% significance level, for the central city in the log-log/log-log model or for the subcentral city in the semi-log/semi-log model. Thus, it is certainly correct to infer that the log-log function is more appropriate than the semi-log function for subcentral cities. It should be kept in mind that the central city always has a macro-effect that influences a larger radius while the subcentral city only has a local micro-effect. In that light, it seems that the choice of functional form is sensitive to the influence sphere of the centre. The log-log function performs better when the area of influence is relatively small, while the semi-log function is more appropriate if the area is larger. These findings are in line with those of Osland et al. (2007), who found that the exponential (semi-log) function performs best when the estimation is based on a large area, while the power (log-log) function performs best if the data is restricted to a small area.

TABLE 2.3 Distance effects of both central and subcentral cities

	Semi-log/ Semi-log	Semi-log/ Log-log	Log-log/ Semi-log	Log-log/ Log-log
	Ln(House price)	Ln(House price)	Ln(House price)	Ln(House price)
Constant	8.7260*** (80.38)	8.9212*** (115.18)	8.9805*** (59.59)	9.0699*** (43.06)
Distance to central city	-0.0008*** (-1.98)	-0.0008*** (-2.03)	–	–
Ln(Distance to central city)	–	–	-0.0779** (-2.06)	-0.0607 (-1.27)
Distance to subcentral city	-0.0006 (-0.93)	–	-0.0010* (-1.69)	–
Ln(Distance to subcentral city)	–	-0.0701*** (-4.60)	–	-0.0746*** (-4.66)
Dummy: Jiangsu	-0.4339*** (-3.06)	-0.4142*** (-2.99)	-0.4269*** (-2.89)	-0.4196*** (-2.96)
Dummy: Anhui	-0.6733*** (-4.01)	-0.6530*** (-3.84)	-0.7598*** (-5.24)	-0.7454*** (-5.18)
Time effects	Yes	Yes	Yes	Yes
F-Stats	61.78	72.32	60.03	65.76
Adj.R^2	0.680	0.710	0.675	0.693
AIC	75.102	51.06	79.38	65.70
Sample size	210	210	210	210

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level,respectively. The *t* values shown in parentheses are calculated based on clustered standard errors that are robust to serial correlation of the same unit and heteroskedasticity over time.

Unlike the semi-log model that only includes the effect of the central city, adding the effects of subcentral cities raises the adjusted R^2 from 0.679 to 0.710. Their added effects explain 3% more variance in interurban house prices and decrease the magnitude of the distance penalties of the central city by about 27%. Since we use different functional forms for central and subcentral cities, we cannot compare the magnitudes of their distance penalties directly.

TABLE 2.4 House price and wage gradient after controlling for city amenities and characteristics

	House price model			Wage model
	Regional dummies	Amenities + Regional dummies	Amenities + Location variables	Amenities + Location variables
	Ln(House price)	Ln(House price)	Ln(House price)	Ln(Wage)
Constant	8.4666*** (77.39)	7.6520*** (17.77)	8.0267*** (22.26)	11.0058*** (45.97)
Dummy: Jiangsu	-0.4895*** (-3.46)	0.0946*** (0.66)	–	–
Dummy: Anhui	-0.8410*** (-6.90)	-0.1816 (-1.34)	–	–
Coastal city	–	0.3230** (2.32)	0.3523*** (3.98)	0.0227 (0.45)
Winter temperature	–	0.0791** (2.58)	0.0627* (1.72)	-0.0086 (-0.40)
Smoke and dust emissions	–	-0.0008 (-1.56)	-0.0013*** (-2.73)	0.0004 (0.60)
Doctor	–	0.0010** (2.58)	0.0008*** (2.72)	0.0004* (1.79)
Student/teacher ratio	–	-0.0050 (-0.38)	-0.0022 (-0.15)	-0.0354*** (-3.46)
Arable land	–	-0.0003* (-1.83)	-0.0003 (-1.33)	-0.0004** (-2.38)
Distance to central city	–	–	-0.0008*** (-2.80)	-0.0003* (-1.67)
Ln(Distance to subcentral city)	–	–	-0.0329* (-1.72)	-0.0131 (-1.20)
Time effects	Yes	Yes	Yes	Yes
F-Stats	61.04	85.58	89.47	63.22
Adj.R^2	0.622	0.788	0.794	0.746
Sample size	210	207	207	207

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level,respectively. The *t* values shown in parentheses are calculated based on clustered standard errors that are robust to serial correlation of the same unit and heteroskedasticity over time.

§ 2.5.3 House price gradient after controlling for city amenities and characteristics

As suggested by the theoretical model, the negative house price gradient with respect to higher-tier cities should persist after controlling for city amenities and characteristics. Before estimating this, we first investigate the compensating house price differentials for urban amenities and characteristics. As noted earlier, the average house prices in Zhejiang, Jiangsu and Anhui provinces differ significantly from each other. This observation is further supported by the estimation results of column 1 in Table 2.4, which only contains two regional dummy variables. We assume that the observed house price differentials across provinces are actually proxies for the differences in amenities. The result of testing this hypothesis is shown in column 2 of Table 2.4. After including the variables of city amenities and characteristics, the regional effects of Jiangsu and Anhui province fall dramatically and are no longer significant, which offers some support for our hypothesis.

The six variables of city amenities and characteristics, together with the two regional dummy variables, account for nearly 80% of the house price variance. As a group, the amenity and characteristic variables are statistically highly significant at the 1% significance level (the joint F-statistic is 39.98 where the 1% critical value is 2.90), and each has the anticipated sign. Among these variables, winter temperature, bordering an ocean and number of doctors have significantly positive effects, while arable land per capita has a negative effect at a significance level of 10% or better. The unpleasant effect of smoke and dust emissions is marginally insignificant.

The third column of Table 2.4 reports the estimation results of the model with both amenity variables and two distance measures. The two distance variables in which we are most interested still have significantly negative effects: distance to the central city is significant at the 1% level, while distance to the subcentral city at the 10% level. Compared to the semi-log / log-log model that only includes two distance variables and two regional dummies, the magnitude of the central-city house price gradient in this model does not change, but the distance penalties of the subcentral city decrease by about 50%. The point estimates of city amenities and characters are quite robust as they do not differ much from the results in column 2. Perhaps the most obvious change is that the negative effect of particulate matter becomes highly significant.

§ 2.6 Decomposition of interurban house price gradient

The previous section has provided estimates of the impact of urban hierarchy distances on house prices. According to the theoretical model, the decline in the interurban house price gradient could be attributed either to productivity disadvantages or amenity disadvantages. This section will empirically decompose the interurban house price gradient based on Equation (7) and reveal which component contributes more to the negative price effects of remoteness from higher-tier cities.

In doing so, we first estimated the wage gradient, which is reported in the fourth column of Table 2.4. Overall, the distance and amenity variables perform less successfully in the wage model than in the house price model, given the lower adjusted R^2 of wage regression. The central city still imposes a statistically significant distance penalty on wages, but its magnitude is less than the penalty on house prices. The negative coefficient of distance to subcentral cities, on the other hand, is no longer significant in the wage model. Among the significant wage determinants are two man-made amenities, namely the number of doctors and the ratio of students to teachers, as well as the area of arable land. Unlike the households, firms seem not to value the climate and environmental amenities as they are not significant in explaining wage differences. In contrast, firms strongly prefer man-made amenities, especially the human capital that is partially reflected in the student/teacher ratio. Of course, access to higher-tier cities is valued by both firms and households.

To decompose the house price gradient, we need to know the share of the household budget that is spent on housing (k). Since there is no official estimate of general housing expenditure in China, the parameter is simply calibrated from the owner-occupied housing market by the equation $k = (P \times R \times S + U)/I$[14]. In this equation, the numerator is the total annual housing expenditure in 2010, including imputed rent ($P \times R \times S$) and utility charges (U), and the nominator is the disposable income per capita (I) in 2010. For the imputed rent equation ($P \times R \times S$), P is the national average house sale price in 2010, R is the discount rate, which takes the value of 4.83% (average of five-year deposit rate during 2006-2010), and S is the average amount of living space per person, which equals 30 m^2. Finally, our estimated from this equation is 0.430.

TABLE 2.5 Decomposition of interurban house price gradient

	Average distance (km)	Distance penalties	Amenity component	Productivity component	Amenity share (%)	Productivity share (%)
Distance to central city	305.71	-2.4902E-01	-3.9167E-02	-2.0985E-01	15.73	84.27
Ln(Distance to subcentral city)	4.42	-1.4530E-01	-1.1058E-03	-1.3424E-01	7.61	92.39

Notes: Because of rounding in Table 2.4, the results shown here cannot be accurately calculated by readers.

With the parameter k, the estimated house price gradient and the wage gradient in hand, Equation (7) can be used to decompose the negative effects of urban hierarchy

14 We did not consider the private rental market when calculating the share of housing expenditure, since it is seriously underdeveloped in China.

distance into two components: productivity disadvantage and amenity disadvantage [15]. The second column of Table 2.5 reports the penalties of distance to the central city and the subcentral city at the mean values (shown in column 1). The remaining columns give the amount of the amenity component and the productivity component as well as their corresponding shares. For distance to either the central or the subcentral city, both components are negative. This empirical finding sheds light on why the cities that are proximal to higher-tier cities have higher house prices. It is more expensive there not only because these nearby cities can provide higher wages due to the firms' productivity advantage but also because households are willing to pay for access to consumer services that are only available in higher-tier cities. Yet, the productivity component explains the majority of urban hierarchy distance penalties on interurban house prices, namely about 85% of the distance penalties of the central city and 92% of the penalties of the subcentral city. In other words, households' valuation of access to higher-tier consumer services only plays a marginal role in determining the house prices.

According to Equation (7), our decomposition of the interurban house price gradient is sensitive to the parameter k – the ratio of housing expenditure to household budget. A small value, say less than 0.37, will lead to a counterintuitive finding: the amenity component would make no contribution to the negative house price gradient or even have a positive effect on house prices. In other words, households are found to be less willing to live near higher-tier cities, holding the city amenities and characteristics constant. But our estimated share, 0.430, seems pretty high from the perspective of housing affordability, given that the average ratio of the 31 OECD countries is 0.225 and the value 0.3 is often seen as the cut-off point for unaffordability [16]. The question then arises whether the Chinese housing market is unaffordable enough to consider our decomposition results robust. Chen et al. (2010) assessed housing affordability in Shanghai and estimated the ratio of monthly mortgage payments to monthly disposable income (MIR) over the period 2006-2008 at 0.62, 0.69 and 0.60, respectively. Not surprisingly, a nationwide study shows that even the households in the 60-80% income quintile usually face an MIR exceeding 0.40 (Yang and Chen 2014). Thus, it may be inferred that the Chinese housing market is indeed unaffordable and that our decomposition results are robust and reliable.

§ 2.7 Conclusion and discussion

15 Although the subcentral city takes the form of a log-log function, its decomposition is similar to a semi-log function: $\left(\frac{d\log p}{d\log D}\right)^V = \frac{d\log p}{d\log D} - \frac{1}{k}\frac{d\log w}{d\log D}$

16 The three missing OECD countries are Chile, Israel and Mexico.

While most studies have attributed the house price differences across cities to the differentials in city-specific amenities and characteristics, this paper focuses on the spatial dimension of the determinants of interurban house prices, i.e., the effect of urban hierarchy distance. We carried out our analysis under a general spatial equilibrium framework. Location decisions of firms and households jointly predict a declining pattern of house prices with distance from higher-tier cities. This negative house price gradient combines two aspects. First, firms in the higher-tier cities and their nearby areas are able to pay higher wages due to the productivity advantage, thereby driving up house prices. Second, households are willing to pay a premium on house prices for access to higher-order services. The theoretical findings are tested with the aggregate data of a specific hierarchical urban system in the Pan-Yangtze River Delta.

Both central and subcentral cities are found to impose statistically significant distance penalties on interurban house prices if we can correctly specify the distance-decay functions. The choice of forms for the functions is sensitive to the influential radius of the targeted higher-tier cities: the semi-log function is the best choice for the central city, while a log-log decay function is better for subcentral cities. The negative effects of urban hierarchy distance on house prices are robust, even after we control for city amenities and characteristics. We also find evidence of compensating house price differentials in terms of climate, environmental and healthcare amenities. The most counterintuitive finding embedded in the estimation of the central-city gradient − that the use of travel time does not improve the model's performance − is probably due to the fact that our time-point measure cannot truly reflect the cost of travel and changes therein during the study period.

To decompose the house price gradient, the wage gradient is also estimated. The results show that distances to the central and subcentral cities have negative impacts on wages, though the penalties of subcentral cities are not statistically significant. In particular, the slopes of house price gradients are much steeper than those of wage gradients, which may be taken as preliminary evidence of the existence of an amenity premium. Yet, the decomposition results reveal that the 'amenity component' contributes very little; the 'productivity component' contributes strongly to the negative house price gradients. This discrepancy is in line with the wage (growth) gradient decomposition studies by Beeson and Eberts (1989) and Partridge et al. (2010), who also found that the productivity component was much more important in determining the wage (growth) differences. Although the decomposition results obtained in this study are conditional on devoting a relatively large share of the household expenditure to housing ($k = 0.43$), we believe that our findings are robust and reliable given the highly unaffordable housing market in China.

Our empirical findings should be interpreted with caution because of a few methodological flaws. First, due to the general lack of data on housing markets, we chose to include only cities at the prefecture level (or above) of the PYRD hierarchical urban system. That choice limited the sample size and could thereby affect the robustness of the estimation results. Since some other city clusters have recently been growing rapidly in China, such as the Pearl River Delta and the Bohai Bay Economic Rim, future studies could be based on a large data set that combines all of these urban hierarchies. Second, studies on house price *dynamics* have suggested the existence of spatial interaction between intercity housing markets, which may result in spatial autocorrelation. Our failure to take this into account here may have led to inefficient estimators. In fact, the spatial autocorrelation of house prices has been extensively discussed in intracity studies (Yu et al. 2007; McMillen 2010; Osland 2010). Still, investigations of cross-sectional interurban housing markets are rather rare and warrant attention in the future. Third, we exclude land from the production of composite goods. That is, the benefits that accrue to households from having access to higher-tier cities will be completely capitalized in house prices and, in turn, in land prices. In the future, it would be interesting to investigate whether these benefits can also be capitalized in wages and whether urban hierarchy distance has a significant effect on land prices. China would provide a natural setting for testing the latter hypothesis because it has an explicit urban land market.

References

Ahlfeldt, G. (2011). If Alonso was right: Modeling accessibility and explaining the residential land gradient. Journal of Regional Science, 51(2), 318-338.

Alonso, W. (1964). Location and land use: toward a general theory of land rent. Cambridge MA: Harvard University Press.

Ashworth, J., & Park, S. C. (1997). Modeling regional house prices in the UK. Scottish Journal of Political Economy, 44(3), 225-246.

Beeson, P. E., & Eberts, R. W. (1989). Identifying productivity and amenity effects in interurban wage differentials. The Review of Economics and Statistics, 71(3), 443-452.

Chen, J., Hao, Q., & Stephens, M. (2010). Assessing housing affordability in post-reform China: A case study of Shanghai. Housing Studies, 25(6), 877-901.

de Bruyne, K., & van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. Applied Economics, 45(13), 1673-1689.

Ding, C. (2013). Transport development, regional concentration and economic growth. Urban Studies, 50(2), 312-328.

Drake, L. (1993). Modelling UK house prices using cointegration: An application of the Johansen techniques. Applied Economics, 25(9), 1225-1228.

Fujita, M., Krugman, P., & Mori, T. (1999). On the evolution of hierarchical urban systems. European Economic Review, 43(2), 209-251.

Gallin, J. (2006). The long-run relationship between house prices and income:

Evidence from local housing markets. Real Estate Economics, 34(3), 417-438.
Glaeser, E. L., Gyourko, J., & Saks, R. E. (2006). Urban growth and housing supply. Journal of Economic Geography, 6(1), 71-89.
Gyourko, J., Mayer, C., & Sinai, T. (2010). Dispersion in house price and income growth across markets. In E. Glaeser (Eds.), Agglomeration Economics. Chicago: The University of Chicago Press.
Gyourko, J., & Tracy, J. (1991). The structure of local public finance and the quality of life. Journal of Political Economy, 99(4), 774-806.
Heikkila, E., Gordon, P., Kim, J. I., Peiser, R. B., Richardson, H. W., & Dale-Johnson, D. (1989). What happened to the CBD-distance gradient? Land values in a policentric city. Environment and Planning A, 21(2), 221-232.
Huang, H., & Tang, Y. (2012). Residential land use regulation and the US housing price cycle between 2000 and 2009. Journal of Urban Economics, 71(1), 93-99.
Ihlanfeldt, K. R. (2007). The effect of land use regulation on housing and land prices. Journal of Urban Economics, 61(3), 420-435.
Li, Q., & Chand, S. (2013). House prices and market fundamentals in urban China. Habitat International, 40, 148-153.
Malpezzi, S. (1996). Housing prices, externalities, and regulation in U.S. metropolitan areas. Journal of Housing Research, 7(2), 209-241.
McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: a nonparametric approach. Journal of Urban Economics, 40(1), 100-124.
McMillen, D. P. (2010). Issues in spatial data analysis. Journal of Regional Science, 50(1), 119-141.
Meen, G. (2002). The time-series behavior of house prices: A transatlantic divide? Journal of Housing Economics, 11(1), 1-23.
Mills, E. S. (1967). An aggregation model of resource allocation in a metropolitan area. American Economic Review, 57(2), 197-210.
Muth, R. F. (1969). Cities and housing: the spatial pattern of urban residential land use. Chicago: University of Chicago Press.
Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. Journal of Real Estate Research, 32(3), 289-320.
Osland, L., Thorsen, I., & Gitlesen, J. P. (2007). Housing price gradients in a region with one dominating center. Journal of Real Estate Research, 29(3), 321-346.
Ottensmann, J. R., Payton, S., & Man, J. (2008). Urban location and housing prices within a hedonic model. The Journal of Regional Analysis and Policy, 38(1), 19-35.
Ozanne, L., & Thibodeau, T. (1983). Explaining metropolitan housing price differences. Journal of Urban Economics, 13(1), 51-66.
Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. Journal of International Economics, 78(1), 126-140.
Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2010). Recent spatial growth

dynamics in wages and housing costs: Proximity to urban production externalities and consumer amenities. Regional Science and Urban Economics, 40(6), 440-452.

Potepan, M. J. (1996). Explaining intermetropolitan variation in housing prices, rents and land prices. Real Estate Economics, 24(2), 219-245.

Qin, B., & Han, S. S. (2013). Emerging polycentricity in Beijing: Evidence from housing price variations,2001-05. Urban Studies, 50(15), 2006-2023.

Rappaport, J. (2007). Moving to nice weather. Regional Science and Urban Economics, 37(3), 375-398.

Roback, J. (1982). Wages, rents, and the quality of life. Journal of Political Economy, 90(6), 1257-1278.

Rosen, S. (1979). Wage-based indexes of urban quality of life. In P. Mieszkowski, & M. Straszheim (Eds.), Current Issues in Urban Economics. Baltimore: The Johns Hopkins University Press.

Söderberg, B., & Janssen, C. (2001). Estimating distance gradients for apartment properties. Urban Studies, 38(1), 61-79.

Waddell, P., Berry, B. J. L., & Hoch, I. (1993). Residential property values in a multinodal urban area: New evidence on the implicit price of location. Journal of Real Estate Finance and Economics, 7(2), 117-141.

Wang, Y., Shao, L., Murie, A., & Cheng, J. (2012). The maturation of the neo-liberal housing market in urban China. Housing Studies, 27(3), 343-359.

Yang, Z., & Chen, J. (2014). Housing affordability and housing policy in urban China. Berlin and New York: Springer.

Yu, D., Wei, Y. D., & Wu, C. (2007). Modeling spatial dimensions of housing prices in Milwaukee, WI. Environment and Planning B : Planning and Design, 34(6), 1085-1102.

Zheng, S., Cao, J., Kahn, M. E., & Sun, C. (2014). Real estate valuation and cross-boundary air pollution externalities: Evidence from Chinese cities. Journal of Real Estate Finance and Economics, 48(3), 398-414.

Zheng, S., & Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. Journal of Urban Economics, 63(2), 743-757.

Zheng, S., Kahn, M. E., & Liu, H. (2010). Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. Regional Science and Urban Economics, 40(1), 1-10.

3 Network externalities in Chinese housing markets: A spatial econometric approach

Submitted for review. Co-author: Jan de Haan.

Abstract: The spatial variation of interurban house prices and the spatial clustering pattern cannot be fully explained by local-specific characteristics; cross-city spillovers also play an important role in the formation of house prices. Existing studies that consider the spatial aspect usually include a spatial lag of house prices as an indicator of house price interaction. However, the underlying theoretical foundation of such spatial lag is rather weak. This paper investigates a special form of spatial interaction: city network externality. Such network spillovers can be properly modelled by the spatial lag of X model and spatial Durbin error model in spatial econometrics. Using panel data for the Pan-Yangtze River Delta (PYRD) in eastern China, we present evidence for positive network spillovers.

Keywords: House prices, City network externalities, Spatial econometrics, China

JEL: R12, R23, R30

§ 3.1 Introduction

In the spatial equilibrium framework of Rosen (1979) and Roback (1982), house prices of cities are determined by local productivities and amenities (Glaeser et al. 2014). Some local-specific indicators that reflect these two aspects, together with local housing supply conditions, form the mainstream specification of empirical house price models (e.g.,Ozanne and Thibodeau 1983; Malpezzi 1996; Potepan 1996; Zheng et al. 2010). Nevertheless, the fact that house prices are geographically clustered, which is still prevalent after reasonably controlling for local-specific characteristics, suggests that cross-city spillovers might be also important in the formation process of house prices.

The spillover of interurban house prices is well documented in the time series analysis of house price dynamics. In the UK housing market, for instance, the lagged changes of house prices in Greater London can be used to predict other regions' price dynamics in current period (Giussani and Hadjimatheou 1991; Holly et al. 2011). Further, such propaganda of house prices is not necessarily restricted to a hierarchical pattern – from a core city to periphery cities; it can also be present in a more general sense. Pollakowski and Ray (1997) revealed that house price shocks in one area can Granger cause subsequent shocks in other areas at the spatial level of both U.S. census divisions and primary metropolitan statistical areas.

This paper investigates the (static) house price spillovers from a cross-sectional perspective. We are particularly interested in the question whether cross-city spillovers are responsible for explaining the house price variation and hence for the spatial clustering patterns. Several attempts have been devoted to this issue by using recently developed spatial econometric models, such as the well-known spatial autoregressive model (SAR), spatial error model (SEM) and some of their variants (Fingleton 2008; Fingleton and Le Gallo 2008; Baltagi et al. 2014; Brady 2014). All of these studies found highly significant estimates for the spatial lag of house prices, confirming the existence of cross-city spillovers.

Existing studies using spatial econometrics attributed the house price spillovers either to displacement effects (e.g.,Fingleton 2008) or to yardstick competition (Brady 2014)[1]. However, whether cross-city spillovers are truly caused by such mechanisms is difficult to judge only from the significant spatial autoregressive parameter of the SAR model, because this model has inherent identification problems (Gibbons and Overman 2012). The present paper, instead, investigates the house price spillovers from a city network externalities perspective. In other words, we seek to examine whether the (average) house price in a city depends on the market size of neighbouring urban concentrations. In an urban hierarchy, it is well documented that the house prices of hinterland urban areas are much lower than that in higher-tier urban cores and that the house price differences are positively related to the distance between them, which bears the spillovers of higher-tier cities (Partridge et al. 2009; de Bruyne and van Hove 2013; Gong et al. 2016). Furthermore, Partridge et al. (2009) shows that local market potential, a measure of the aggregate personal income of surrounding regions, has a significantly positive effect on urban wages and house prices. The importance of market potential underlines the idea of city network externalities (Boix and Trullén 2007): each city interacts with other cities (not necessarily the higher-tier cities) in the network and benefits from such connectivity. Our analysis follows this tradition, and we assume that the effect of network externalities on house prices arises

1 The displacement mechanism assumes that a high house price signal in one market will force demand to be displaced to and attract supply from nearby markets. As such, the spatial lag of house prices will be present in the reduced form house price equation. Yardstick competition simply assumes that home buyers and developers take the actions of their counterparts in neighbouring markets into account when they make their buying and selling strategy, so that house prices are connected with each other.

not only from the productivity channel represented by market potential, but also from the amenity channel. The mechanism of amenity effect is closely related to the concept of 'borrowed size', whereby a city can perform better in terms of higher-order amenities without enlarging its own size through borrowing functions or performance from its neighbours (Alonso 1973; Meijers and Burger 2015).

Unlike the commonly used market potential measure, which represents the aggregated market demand weighted by inverse distance (Harris 1954), this paper uses the toolbox of spatial econometrics to investigate the effect of network spillovers on house prices, as the theoretical foundation of network externality can be perfectly fitted into the exogenous interaction assumption of spatial econometrics. Based on a panel data set of the Pan-Yangtze River Delta in eastern China, we find significant evidence for the presence of positive network spillovers. These results add to the literature on Chinese interurban housing markets by analysing its spatial aspects, which has been absent in most of the studies explaining house price variation across cities in China (e.g.,Zheng et al. 2010; Li and Chand 2013; Zheng et al. 2014).

The remainder of the paper is organised as follows. Section 3.2 briefly reviews the literature focusing on the spatial interaction of house prices. The theoretical foundation of city network externality on house prices is presented in section 3.3. Section 3.4 discusses the empirical spatial econometric models, followed by the data description in section 3.5. Section 3.6 reports the empirical results, and section 3.7 concludes.

§ 3.2 Literature on spatial spillovers of house prices

When assessing the value of a property, the sellers and buyers are very likely to take recent transaction prices of nearby properties as a reference. As such, the price of a property has direct influence on the prices of nearby properties, which is known as the adjacency effect or spillover effect. Can (1990, 1992) was the first to use spatial econometrics in order to incorporate the spillovers of house prices into the traditional hedonic model and found that the spatial models are superior to the conventional ones[2]. Since then, spatial econometric modeling based on three different interaction assumptions – endogenous interaction, exogenous interaction and correlated effects – has become a standard tool for hedonic house price analysis, for example in estimating the benefits of improvement of air quality and water supply (Kim et al. 2003; Anselin et

2 Another strategy, which relates to the field of geostatistics, directly specifies the covariance of residuals of hedonic models as a function of the distance between locations (Basu and Thibodeau 1998; Bourassa et al. 2007).

al. 2010)[3]. Among the family of spatial econometric specifications, the spatial autoregressive model (SAR) with endogenous interaction and the spatial error model (SEM) with correlated effects are the most popular approaches. Recently, Osland (2010) introduced the spatial Durbin model (SDM), with both endogenous and exogenous interaction, into the hedonic analysis of property prices.

House price spillovers also seem to be prevalent between cities' housing markets given the fact of geographical clustering of house prices. Such spillover effects have received increasing attention in regional house price studies. For example, Fingleton (2008) proposed a SAR-type cross-sectional house price model for local authority districts of England. Later on, this model was extended to incorporate spatially dependent disturbances (Fingleton and Le Gallo 2008). Baltagi et al. (2014) expanded the cross-sectional data set used by Fingleton (2008) to a panel data and estimated a house price model with spatial lag and random hierarchical error components. In markets outside the UK, Brady (2014) examined the spatial diffusion of house prices across continental U.S. states, using a spatial impulse response function derived from a single equation spatial autoregressive panel model. Holly et al. (2010) also proposed a spatio-temporal house price model for U.S. states, in which the spatial correlation is assumed to be attributed to common shocks.

Not surprisingly, endogenous interaction and correlated effects are still the main focus of these studies; the endogenous interaction is often difficult to justify, and SAR-type models cannot clearly tell us whether there is truly an endogenous interaction in the house price formation process (Gibbons and Overman 2012). On the other hand, the exogenous interaction of house prices, which is well established in economic theory, has been largely overlooked in the applied literature. The New Economic Geography (NEG) predicts that factor prices, such as wages, house prices and land rents, are higher in those areas with better access to major consumer and supplier markets (Head and Mayer 2004). This implies the interdependence between the house price of a city and the market size of neighbouring cities, which can also be interpreted as city network externality. Using the measure of market potential, which aggregates the market demand of other places through an inverse distance weighting scheme (Harris 1954), Hanson (2005) and Partridge et al. (2009) provided strong evidence of such network spillovers on U.S. county wages and/or house price. With regard to our focus on house prices, spatial econometric models based on the exogenous interaction assumption can properly deal with the network spillovers. Thus, spatial econometrics offers us an alternative to test for cross-city spillovers of house prices caused by network externality.

3 Endogenous interaction assumes that the house price of a city depends directly on the house prices of other cities, while exogenous interaction assumes that the house price of a city depends on other cities' house price determinants. The assumption of correlated effects is that the dependence of house prices stems from omitted house price determinants that are spatially correlated or from common shocks (Elhorst 2010a).

For a very long time, studies on Chinese regional house prices are largely absent in the literature because of the lack of housing transactions data. Only recent years have witnessed the emergence of studies on the role of fundamentals in explaining regional house prices (Li and Chand 2013), especially the influence of urban environmental and climate conditions (Zheng et al. 2009; Zheng et al. 2010; Zheng et al. 2014). In contrast, the spatial dimension of regional house prices is less investigated. Gong et al. (2016) explored the spillover effects of higher-tier cities on the house prices of small cities from the perspective of an urban hierarchy. This study, however, does not pay attention to the spillovers of neighbouring cities, which will be addressed in this paper. Hanink et al. (2012) considered the spatial dependence and spatial heterogeneity in Chinese county-level house prices using the SEM model and Geographically Weighted Regression (GWR), respectively. However, cross-city spillovers cannot be properly investigated by the SEM specification. Therefore, this paper also contributes to the literature by analysing the spatial aspects of interurban housing markets in the biggest developing economy, China.

§ 3.3 Network externalities on interurban house prices

Let us consider an economy that consists of a set of cities. These cities are linked by trade and migration, but workers are assumed not to commute between cities for working purpose. In spatial equilibrium where the marginal migrant is indifferent across cities, the urban house price of a city i (P_i) depends on the quality of life (A_i) and urban productivity (W_i) of that city (Glaeser et al. 2001):

$$P_i = p\left(A_i, W_i\right) \tag{1}$$

Quality of life refers to urban amenities, and has two components: common amenities (c_i) and higher-order amenities (a_i). The former ones are those natural and man-made amenities that are consumed locally and regularly by consumers so that their effects are largely confined to the city border, such as temperature, basic healthcare and education services. Higher-order amenities, on the other hand, are likely to be concentrated in a few big cities and have a broader influence on other areas because they require a sufficiently large market potential to be sustained. For instance, in the classical framework of Central Place Theory, the central urban core provides higher-order functions for the smaller urban areas in the hinterland. This market structure induces the effect of "borrowed size" whereby small cities can somewhat "borrow" the higher-order functions from their neighbouring large cities through easy access (Alonso 1973).

However, a modern urban system seems to show some network relationships that are beyond the hierarchical interaction suggested by Central Place Theory (Capello 2000). The city network paradigm, which nests the possibilities of both hierarchical and non-hierarchical structures, seems to be a more comprehensive theory to describe the spatial organisation of cities. 'Borrowing size' in a city network paradigm exhibits broader interaction patterns; it may occur between any two neighbouring cities, not only from large to small cities, but also between cities of the same rank or even from small to large cities (Boix and Trullén 2007). Indeed, large cities need small cities to help them maintain more higher-order amenities that cannot be supported by their own size. Meanwhile, small cities can share those surplus higher-order amenities through network accessibility, allowing them to perform better (Meijers and Burger 2015). Such 'borrowing size' effect in the context of city network is thus referred to as 'city network externality' and we will use this term throughout the paper. Empirical evidence for the effect of city network externality on presence of higher-order amenities has recently emerged. For instance, in an analysis of the distribution of metropolitan functions across Western European countries, Meijers et al. (2016) noted that network connectivity positively contributes to the presence of those higher functions. In this regard, the quality of urban amenities presented in city i is a function of its own urban size (s_i) and the urban sizes of its neighbouring cities (θs_{-i}), $A_i = A\,(c_i, s_i, \theta s_{-i})$, where c_i is a bundle of common amenities.

On the productivity side, network externalities also play an important role. Small cities that are readily accessible to large cities can borrow the technological externalities of those major urban cores, and hence improve the productivity without increasing their own size (Phelps et al. 2001). Beyond such vertical interaction, a more general form of network externalities on productivity side is the 'market access' effect stressed by New Economic Geography (NEG) – being access to large consumer and supplier markets contributes to the productivity of an area by saving on transportation costs (Fujita et al. 1999). That is, major urban cores in the urban system also benefit from the relatively large neighbouring markets. Many studies have revealed that market potential, a similar concept to population potential which has been suggested by Alonso (1973) as an index of 'borrowed size', positively contributes to the wage level of an area (Brakman et al. 2004; Hanson 2005). In line with these facts, a city's productivity level can be written as: $W_i = W\,(l_i, s_i, \theta s_{-i})$, where l_i indicates a set of locational characteristics.

After including the amenities and productivity components into equation (1), the reduced-form house price equation becomes $P_i = P\,(c_i, l_i, s_i, \theta s_{-i})$. This expression clearly shows that the house price of city i depends on an interaction term (θs_{-i}), representing the effect of city network externalities.

§ 3.4 Empirical models

§ 3.4.1 Spatial econometric models

There are several alternatives that can model network spillovers based on different interaction assumptions in spatial econometrics. One approach assumes that city network spillovers directly enter into the right-hand-side of the house price equation, which can be modelled by the spatial lag of X model (SLX) (LeSage and Pace 2009; Gibbons and Overman 2012; Vega and Elhorst 2015):

$$\mathbf{p} = \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{p}$ denotes a vector of observations of house prices, $\mathbf{X}$ is a matrix of observations on exogenous house price determinants, $\mathbf{WX}$ denotes the spatial lag of exogenous independent variables, and $\boldsymbol{\epsilon}$ represents the independently and identically distributed disturbances. The parameter vector $\boldsymbol{\theta}$ thus measures the magnitude of spillovers of independent variables. The SLX model, which has been largely overlooked, is actually an appealing tool in applied studies because of its superiority in avoiding identification issues and its flexibility in measuring spillover effects (Gibbons and Overman 2012; Vega and Elhorst 2015). In practice, the SLX model may suffer from a multicollinearity problem. However, our study is largely free of this problem because not all the variables have cross-city effects according to the theoretical setup.

Apart from network externalities, house price spillovers can also arise from other mechanisms, such as spatially correlated omitted variables and common shocks. The failure to properly model such spatial dependence will lead to inconsistent estimates of network spillovers. Conditional on the presence of spatial dependence in the residuals, the spatial Durbin error model (SDEM) is preferred, which takes the form (LeSage and Pace 2009):

$$\begin{aligned} \mathbf{p} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\theta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &= \lambda\mathbf{M}\boldsymbol{\epsilon} + \mathbf{u}, \end{aligned} \tag{3}$$

where the error terms $\boldsymbol{\epsilon}$ follow a spatial autoregressive process and $\mathbf{u}$ denotes the independently and identically distributed disturbances. The matrix $\mathbf{M}$, which captures the interaction of error terms, could be the same as $\mathbf{W}$ or not.

Pure house price spillovers can also occur, as suggested by yardstick competition whereby the house price formation process of a city takes into account the price signal of other cities (Brady 2014). In this case, the spatial Durbin model (SDM), which has attracted increasing attention recently, can be estimated:

$$\mathbf{p} = \rho\mathbf{Mp} + \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{4}$$

where the term $\mathbf{MP}$ captures the spillovers of house prices[4]. However, including

4 Conditional on the common factor restriction $\boldsymbol{\theta} + \rho\boldsymbol{\beta} = \mathbf{0}$, the SDM model collapses to the well-known spatial error model (SEM) which assumes that the error term follows a spatial autoregressive process. If the true model is SEM, the estimation of SDM model is preferred because it can produce unbiased estimates even if omitted variables are correlated with the explanatory variables and follow a spatial autoregressive process. However, Gibbons and Overman (2012) demonstrated that SDM can only solve a particular type of omitted variable problem, and it should not be seen as a general solution.

endogenous interactions in the model is somewhat risky; one can easily obtain significant spatial autoregressive parameter ρ in applied work, while it cannot be readily identified (e.g.,Gibbons and Overman 2012). This parameter might also pick up the information of omitted variables or even the nonlinearity in the **WX** variables if they are misspecified (Corrado and Fingleton 2012). Thus, the interpretation of the causal effect of pure spillovers is problematic.

If the parameter vector $\boldsymbol{\theta}$ in model (4) is insignificant, the SDM model collapses to the SAR model (Anselin 1988):

$$\mathbf{p} = \rho \mathbf{M}\mathbf{p} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{5}$$

Again, the interpretation of this model is difficult. The parameter ρ in this model can reflect pure spillovers of house prices, but it could also indicate that network externalities work indirectly through spillovers of house prices. For example, a positive population shock to city i will drive up house prices of this city. Afterwards, the house prices of neighbouring cities might also increase just because households change their expectations based on the price signal of city i. This is very likely to happen in housing markets where market participants are characterized by bounded rationality. However, this model cannot tell which mechanism the parameter ρ exactly points to.

Models (2) – (5) will be estimated accordingly in the following section. As our purpose is to examine the network spillovers on interurban house prices, we are particularly interested in models (2) and (3) because they can perfectly deal with the theoretical foundation of city network externalities. The most popular specifications, models (4) and (5), are mainly estimated for comparison purposes.

§ 3.4.2 Measuring cross-city spillovers

Due to the presence of spatial weight matrixes **W** (or **M**) in spatial models, the interpretation of the parameter estimates is a bit complicated, especially for the SAR and SDM models. In this paper, we use the partial derivative approach proposed by LeSage and Pace (2009) to calculate the direct effect – the effect of changes of the kth variable in a city on its own house prices – and the indirect effect – the effect of changes of the kth variable in a city on the house prices of other cities. By definition, the indirect effects represent the cross-city spillovers that we are interested in.

In the SAR model, the partial derivatives of the expectations of **p** with respect to the kth independent variable can be expressed as

$$\left[\frac{\partial E(\mathbf{p})}{\partial x_{1k}} \cdots \frac{\partial E(\mathbf{p})}{\partial x_{nk}}\right] = (\mathbf{I} - \rho \mathbf{M})^{-1} \beta_k = \mathbf{S}_k(\mathbf{M}). \tag{6}$$

Similarly, the partial derivative matrix for the SDM model can be expressed as $(\mathbf{I} - \rho \mathbf{M})^{-1} [\beta_k + \mathbf{W}\theta_k]$. The diagonal and non-diagonal elements of the partial derivative matrix $\mathbf{S}_k(\mathbf{W})$ in (6) measure the direct effects and indirect effects, respectively. Since these effects differ across the cities in the sample, LeSage and Pace (2009) suggests to report the direct effect as the average of the diagonal elements and the spillovers as the average of the row (column) sums of the non-diagonal elements. In the case of the SLX and SDEM models, the spillover effects are exactly equal to the parameter estimates θ_k. Note that, in the SAR model, the ratio of spillover effect to direct effect is constant across variables whereas there are no such restrictions in the SLX, SDEM and SDM models (Elhorst 2010a).

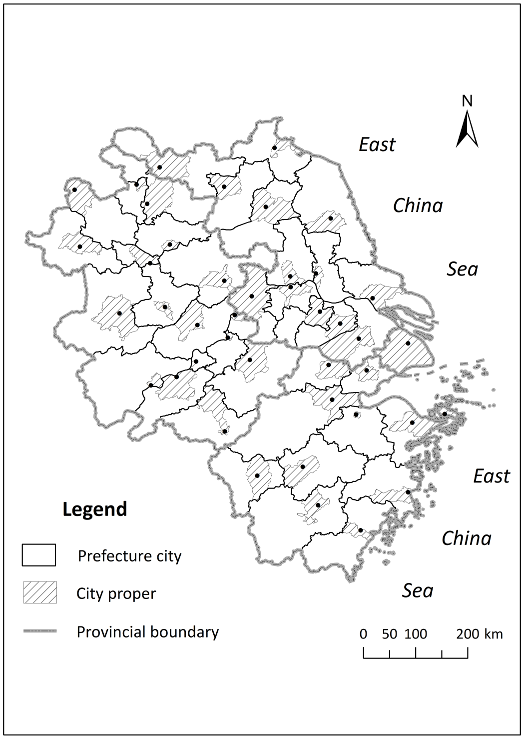


FIGURE 3.1 Cities in Pan-Yangtze River Delta

§ 3.5 Data

We empirically analyse the cross-city house price spillovers between 42 cities (prefecture cities or municipalities under the central government) of Pan-Yangtze River Delta (PYRD) in eastern China from 2006 to 2010[5] (Figure 3.1). The cities in PYRD

5 Prefecture cities form the second level of Chinese administrative system, under which are city districts and

form a 'city network' connected through railways, highways and telecommunication networks. Some formal planning with regard to this area is currently under discussion by scholars and policy makers, aiming to facilitate further economic integration. Therefore, we can expect the presence of significant interaction between the housing markets of the cities in this area.

TABLE 3.1 Description of variables

Variables	Description
House prices	Real average sale price of newly sold residential buildings in the city proper (Yuan/m2); deflated by CPI (base year of 2000); 2006-2010
Winter temperature	Average temperature of December, January and February (Centigrade); 2006-2010
Smoke and dust emission	Annual amount of industrial smoke and dust emissions per real GDP in the city territory (Tons per 100 million Yuan); 2006-2010
Student/Teacher ratio	The ratio of student to teacher in the city territory; 2006-2010
Doctor	Number of doctors per 10,000 inhabitants in the city territory; 2006-2010
Coast	=1 if the *city proper* borders an ocean; =0 otherwise
Arable land	Arable land per capita of the city territory in the year 2004 (m^2 per capita)
Population density	Urban population density of the city territory (person per km^2); 2006-2010
Land	Land area of the city territory (km^2)

The panel data set is compiled from various sources, such as the city- and province-level statistical yearbooks and the China City Statistical Yearbook. We have no access to property transaction data sets so that it is impossible for us to build a constant-quality house price measure. House price in this paper refers to the real average sale price of newly sold residential buildings in the *city proper* (see footnote 5). The city characteristics that have a local effect are captured by variables on natural and environmental conditions, human amenities, location and supply conditions. We use winter temperature and intensity of smoke and dust emission to measure the natural and environmental conditions of each city. The education and healthcare performance of a city, which reflect the level of human amenities, are approximated by the ratio of students to teachers and the number of doctors per thousand inhabitants, respectively. We also include a dummy variable 'coast' to indicate whether the city proper borders an ocean. The inclusion of arable land per capita aims to capture the construction land supply potential. To facilitate the efficient use of urban land and to ensure the grain

counties (or county-level cities); the city districts make up the city proper ('*shiqu*') of a prefecture city. The municipality under the central government is positioned in the first level, but has similar subdivisions with prefecture cities.

security, Land Use Planning is compulsory in each city and limits the conversion of arable land to construction land. We expect that the lower arable land per capita will reduce the construction land supply and hence drive up house prices. Urban size is of our main interest in this paper and we investigate two aspects of urban size: intensity and scale. The former one is measured by urban population density, while the latter one is approximated by land area of the city. The definition of each variable is reported in Table 3.1 and more details can be found in appendix. Note that house price and its determinants pertain to different spatial aggregation level, which can partly avoid the endogeneity between house prices and urban size.

The geographical distance between two cities used for constructing the spatial weight matrix refers to the straightforward distance between the city hall of the two cities. Among the 861 city pairs, the distance between the most separated cities reaches 803 km, while the closest two cities are only 21 km away. The average distance that separates a city pair is 305 km. Spatial weight matrixes are also constructed based on travel time, which means the shortest driving time between two cities without traffic. These figures are extracted from Google Maps in the year 2011. One has to drive 693 minutes for the two most distant cities, while only 45 minutes for the nearest two cities. In average, the city pair is separated by a 267 minutes journey.

§ 3.6 Results

§ 3.6.1 Nonspatial model

The house price models without cross-city spillovers are first estimated and serve as the benchmark. The results of the pooled model estimated by Ordinary Least Squares (OLS) and the random effect model estimated by Maximum Likelihood (ML) are reported in the first two columns of Table 3.2[6]. We prefer the random effects model to fixed effects model because of several reasons. First, in our model there are several time-constant variables including one of our focus variables, the effects of which cannot be estimated by fixed effects model. Second, some variables have little within-group variation, which affects the precision of fixed effects estimators. Third, the fixed effects model discards the cross-sectional information that we are most interested in.

All of the parameter estimates of the pooled model have expected signs and are statistically significant at 1% significance level except for the variable arable land per

6 The ML estimation of random effect model is performed by an iterative two-stage procedure suggested by Breusch (1987).

TABLE 3.2 Estimates of nonspatial model and SLX model

	Dependent variable = Ln(House prices)			
	Pooled model OLS	RE ML	SLX_G(RE) ML ($\mathbf{W} = \mathbf{W}_G^{0\text{-}160}$)	SLX_T(RE) ML ($\mathbf{W} = \mathbf{W}_T^{0\text{-}150}$)
Winter temperature	0.0617***	0.0262***	0.0258***	0.0260***
	(5.14)	(2.74)	(2.75)	(2.77)
Ln(Smoke and dust	-0.1231***	-0.1396***	-0.1297***	-0.1325***
emission)	(-5.24)	(-4.37)	(-4.12)	(-4.22)
Ln(Student/Teacher	-0.4662***	-1.3482***	-1.1673***	-1.1815***
ratio)	(-3.82)	(-7.24)	(-6.21)	(-6.39)
Doctor	0.0358***	0.0253***	0.0268***	0.0243***
	(6.01)	(2.69)	(2.93)	(2.66)
Coast	0.2537***	0.3009**	0.2230*	0.2104*
	(4.78)	(2.54)	(1.93)	(1.78)
Ln(Arable land)	-0.0082	-0.0160	-0.0497	-0.0856
	(-0.11)	(-0.13)	(-0.40)	(-0.71)
Ln(Population density)	0.1561***	0.1712***	0.1181*	0.1006*
	(5.29)	(2.91)	(1.87)	(1.63)
Ln(Land)	0.1721***	0.1817**	0.1897***	0.1630**
	(4.93)	(2.55)	(2.72)	(2.36)
$\mathbf{W} \times$ Ln(Population			0.2719***	0.2795***
density)			(3.04)	(3.17)
$\mathbf{W} \times$ Ln(Land)			0.2590**	0.1691
			(2.00)	(1.35)
Constant	6.4704***	9.2550***	5.3201**	6.6990***
	(8.23)	(6.48)	(2.38)	(3.21)
R-Squared	0.823	0.923	0.925	0.925
Corr-Squared		0.764	0.785	0.787
Log-likelihood	24.544	59.043	63.849	63.973
CD test	19.551***	7.6129***	7.955***	8.051***
Sample size	210	210	210	210

Notes: Corr-Squared is the squared correlation between fitted and actual value. t-values are reported in the parentheses. $\mathbf{W}_G^{0\text{-}160}$ and $\mathbf{W}_T^{0\text{-}150}$ denote the spatial interaction structure between a city and its neighbouring cities within the distance band 0 – 160 km and within the travel time band 0 – 150 min, respectively. The CD test, which detects the global cross-sectional dependence of residuals, tends to standard normal distribution under the null hypothesis. ***, ** and * indicate a 1%, 5%, 10% significance level, respectively.

capita which implies no significant influence of land supply constraint. After controlling for random city-specific effects, the results in the second column do not show any noticeable changes compared to the results of pooled model. In general, a warmer winter, less industrial smoke and dust emission, a better education and healthcare condition, and bordering to an ocean increases the house price of a city. Note that the estimated effect of education quality in the random effect model is much higher than that in the pooled model, while the influence of winter temperature is weakened drastically. As expected, the two variables measuring urban size have statistically and economically significant effects on house prices in both models. Interestingly, an increase in urban density has almost the same effect as an expansion in urban scale. A 1 percent increase of urban population of a city will drive up house prices by around 0.17%. The fixed effects estimation, including only the time-variant variables, also confirms the importance of climate, education quality and urban population density in determining the house prices.

Overall, the explanatory variables we have chosen perform satisfactorily as indicated by a relatively high *Corr-Squared* statistic (0.764) which represents the squared correlation between actual and fitted value. However, the CD test (Pesaran 2004) detects significant global cross-section dependence in residuals, suggesting the existence of cross-city spillovers[7].

§ 3.6.2 Results of spatial models

Estimation of SLX model

The spatial weights matrix **W** is vital to measuring the city network spillovers as **W** carries the underlying spatial interaction structure. In this paper, we expect that the network externalities are only noticeable within a certain radius; at some farther distance between the two cities, network spillovers vanish.

Such spatial interaction structure can be captured by different weight matrixes. Based on geographical distance, we first divide the cities surrounding city *i* into three distance bands, namely 0 – 160 km, 160 – 320 km and 320 – 480 km. A spatial weight matrix for each distance band is then constructed. For instance, for a city *j* within distance band 0 – 160 km, the spatial weight w_{ij} of $\mathbf{W}_G^{0\text{-}160}$ is defined as

$$w_{ij,i\neq j} = d_{ij}^{-2}, \quad \text{for} \quad 0 \leq d_{ij} < 160. \tag{7}$$

Geographical distance has some intrinsic pitfalls; it does not take into account the physical obstacles, such as mountains and bays. So we have also constructed spatial

7 The CD test is constructed based on the average of pair-wise correlations of the residuals of each cross-sectional unit. As $N \rightarrow \infty$, this test tends standard normal distribution under the null hypothesis of no cross-sectional correlation.

weight matrixes based on travel time, which represent the shortest driving time between two cities. Using the same strategy as for the distance-based matrixes, three time-based matrixes, $\mathbf{W}_T^{0\text{-}150}$, $\mathbf{W}_T^{150\text{-}300}$ and $\mathbf{W}_T^{300\text{-}450}$, are formed, corresponding to the time band 0 – 150 min, 150 – 300 min and 300 – 450 min. Following usual practice in spatial econometrics, all the spatial weight matrixes are row-standardized.

For the different distance/time bands, we calculated the correlation coefficient between the house price of a city and the spatial lag of population density of neighbouring cities. The correlation coefficients reported in Table 3.3 show that the house price of a city is indeed related to the population density of cities within the distance band 0 – 160 km (= 0.304) and within the time band 0 – 150 min (= 0.358). As the neighbouring cities are farther away, the correlation coefficients fall dramatically towards to zero or even become negative. The results confirm our hypothesis that network externalities have a local spillover effect; it only influences the nearby cities.

Given the nature of network externalities, we estimated the SLX model in equation (2) using the two matrixes, $\mathbf{W}_G^{0\text{-}160}$ and $\mathbf{W}_T^{0\text{-}150}$, and the ML estimators are shown in the third (SLX_G) and fourth column (SLX_T) of Table 3.2[8]. For the variables of local-specific characteristics (excluding population density and land area), both of the two SLX models produce similar estimates with respect to nonspatial models.

TABLE 3.3 Correlation coefficients between house prices and spatial lags of population density

	Ln(House prices)		Ln(House prices)
× Ln(Population density)		× Ln(House prices)	
$\mathbf{W}_G^{0\text{-}160}$	0.304	$\mathbf{W}_T^{0\text{-}150}$	0.358
$\mathbf{W}_G^{160\text{-}320}$	0.098	$\mathbf{W}_T^{150\text{-}300}$	0.036
$\mathbf{W}_G^{320\text{-}480}$	-0.190	$\mathbf{W}_T^{300\text{-}450}$	-0.057

Notes: For the definition of matrix $\mathbf{W}_G^{0\text{-}160}$ and $\mathbf{W}_T^{0\text{-}150}$, see notes of Table 3.2. All the other matrixes are defined in a similar way.

After including local network spillovers based on geographical distance neighbours ($\mathbf{W}_G^{0\text{-}160}$), the effect of population density on its own house prices decreases by about one third (from 0.17 to 0.12) and becomes less significant, while the direct effect of land area remains relatively stable. The network spillovers are much more important now, as shown by the large and statistically significant estimates of spatial lag of population density and land area. A similar finding occurs when we specify the neighbours based on travel time ($\mathbf{W}_T^{0\text{-}150}$), except that the expansion in urban scale has

8 We also estimate the model based on the remaining four matrixes. The parameter estimates are very unstable compared to the nonspatial model because they fail to properly measure the spatial interaction structure. The results are available upon request.

no spillovers on other cities. It is worth mentioning that our travel time measure is a post-measure that is collected after the study period so that it may not reflect the true interaction structure in our sample. Therefore, we insist on the findings of SLX_G model and our following analysis will be based on the geographical distance measure[9].

Although we have included network spillovers into our model, there is still significant global cross-sectional dependence in the residuals according to CD test. Such dependence might be caused by omitted spatially correlated variables, common shocks or pure spillovers of house prices. Thus it is necessary to estimate a SDEM or SDM model, which controls for the remaining dependence and hence produces more reliable estimates of network spillovers.

Estimation of SDEM model

Unlike the city network externalities, the presence of spatial dependence in residuals or pure house price spillovers is not necessarily confined to the scope of nearby neighbours, as pointed out by Pollakowski and Ray (1997). Indeed, when households coming from a large city form their decisions, they are more likely to refer to the price signal of a large, distant city rather than a small, nearby city. A similar argument was made by Fingleton and Le Gallo (2008) who stated that, in an economic sense, big cities may be less remote than their distance suggests, while very small cities may in fact be more isolated. Therefore, we believe that the spatial weights matrix based on economic distance measures will better capture the remaining spatial dependence structure.

We define a distance measure that combines geographical distance and economic similarities. To do so, we first measure the 'economic similarity' (*es*) of two cities, say city *i* and *j*, as the difference in their disposable income, that is $es_{ij} = |income_i - income_j|$. To avoid the potential endogeneity of this distance measure, income in the year 2000 is used. The economic-geographical distance (EG_{ij}) between city *i* and *j* is then calculated by

$$EG_{ij} = \sqrt{\left(\frac{es_{ij}}{std(es)}\right)^2 + \left(\frac{d_{ij}}{std(d)}\right)^2} \tag{8}$$

where $std(es)$ and $std(d)$ denote the standard deviation of economic similarities and geographical distance, respectively. The corresponding spatial weight matrix, $\mathbf{W}_{EG}$, is specified in the same way as in equation (8), with distance band being set to 0 – 1.5.

Table 3.4 reports the (robust) Lagrange Multiplier (LM) tests (Anselin et al. 2008; Elhorst 2010b) for the existence of spatially lagged dependent variable and spatial

9 We also conducted analyses based on time distance measure. The findings are similar to those based on the geographical distance measure.

error correlation in the SLX_G model based on different spatial weight matrixes. Assuming the remaining dependence structure is still confined to the neighbours in physical distance space, the SDM model is a better choice and the results will be discussed latter. On the other hand, if a city is assumed to interact with the cities that are nearby on the economic-geographical space, the LM tests are in favor of the SDEM specification.

TABLE 3.4 LM tests on residuals of SLX model

	Residuals of SLX_G model estimated in Table 3.2			
	LM spatial lag	Robust LM spatial lag	LM spatial error	Robust LM spatial error
$\mathbf{W}_G^{0\text{-}160}$	57.784*	59.056*	18.162*	9.434*
$\mathbf{W}_{EG}^{0\text{-}1.5}$	0.091	1.760	17.504*	19.173*

Notes: For the definition of matrix $\mathbf{W}_G^{0\text{-}160}$, see notes of Table 3.2. $\mathbf{W}_{EG}^{0\text{-}1.5}$ has the similar definition but are constructed based on economic-geographic distance. The LM and robust LM tests, developed by Anselin et al. (2008) and Elhorst (2010b) for the spatial panel data, are based on the residuals of SLX_G model estimated in Table 3.2 and follow the $\chi^2(1)$ distribution under null hypothesis. * denotes the 1% significance level.

The SDEM model is estimated by a ML procedure suggested by Elhorst (2014); the results are reported in second column of Table 3.5[10]. For the sake of comparison, the first column replicates the estimates of SLX_G model. Based on the economic-geographical distance matrix, we find a highly significant spatial autoregressive process in residuals of the SLX_G model. After controlling for the spatial error correlation, the influence of population density on its own house prices becomes highly significant at 1% significance level. The point estimates of population density and land area as well as their spillovers effects are almost in line with the estimates of SLX_G model, showing the robustness of SLX_G model in measuring the agglomeration spillovers. In contrast, the estimates for the local-specific characteristics show a noticeable discrepancy between the two models. For example, smoke and dust emission and the ratio of students to teacher no longer significantly affect the house prices, whereas land supply constraint becomes an important house price determinant in SDEM model. This discrepancy might be due to the fact that the spatial pattern of some local-specific variables is closely related to the spatial pattern of the residuals of SLX_G model.

Estimation of SDM and SAR model

As previously discussed, if the spatial interaction of house prices after controlling for

10 The following random effects SDM and SAR model are also estimated by ML procedure. The matlab routine can be found at http://www.regroningen.nl/elhorst/.

TABLE 3.5 Estimates of SDEM, SDM and SAR models

	Dependent variable = Ln(House prices)			
	SLX_G (RE) ML ($\mathbf{W} = \mathbf{W}_G^{0\text{-}160}$)	SDEM_G (RE) ML ($\mathbf{W} = \mathbf{W}_G^{0\text{-}160}$) ($\mathbf{M} = \mathbf{W}_{EG}^{0\text{-}1.5}$)	SDM_G(RE) ML ($\mathbf{W} = \mathbf{W}_G^{0\text{-}160}$) ($\mathbf{M} = \mathbf{W}_G^{0\text{-}160}$)	SAR_G ML ($\mathbf{M} = \mathbf{W}_G^{0\text{-}160}$)
Winter temperature	0.0258*** (2.75)	0.0601*** (3.14)	0.0122 (1.52)	0.0126 (1.57)
Ln(Smoke and dust emission)	-0.1297*** (-4.12)	-0.0219 (-0.74)	-0.0409 (-1.54)	-0.0420 (-1.59)
Ln(Student/Teacher ratio)	-1.1673*** (-6.21)	-0.6218*** (-3.07)	-0.7090*** (-4.33)	-0.7080*** (-4.29)
Doctor	0.0268*** (2.93)	0.0026 (0.31)	0.0065 (0.85)	0.0069 (0.90)
Coast	0.2230* (1.93)	0.1560* (1.68)	0.1793* (1.91)	0.1671* (1.77)
Ln(Arable land)	-0.0497 (-0.40)	-0.2906** (-2.40)	-0.1091 (-1.09)	-0.1178 (-1.24)
Ln(Population density)	0.1181* (1.87)	0.1278*** (2.65)	0.2042*** (3.95)	0.1909*** (4.10)
Ln(Land)	0.1897*** (2.72)	0.1702*** (2.80)	0.1737*** (3.07)	0.1757*** (3.12)
$\mathbf{W} \times$ Ln(Population density)	0.2719*** (3.04)	0.2493*** (2.91)	-0.0686 (-0.85)	
$\mathbf{W} \times$ Ln(Land)	0.2590** (2.00)	0.2490** (2.02)	-0.0564 (-0.52)	
$\mathbf{M} \times$ Error		0.7364*** (14.25)		
$\mathbf{M} \times$ Ln(House prices)			0.5510*** (9.48)	0.5280*** (9.96)
Constant	5.3201** (2.38)	5.5096** (2.67)	4.4485** (2.43)	3.8541*** (3.05)
R-Squared	0.925	0.940	0.948	0.948
Corr-Squared	0.785	0.788	0.819	0.812
Log-likelihood	63.849	82.073	95.108	94.721
Sample size	210	210	210	210

Notes: Corr-Squared is the squared correlation between fitted and actual value. *t*-values are reported in the parentheses. For the definition of the various spatial matrixes, see notes of Table 3.2. The SDEM, SDM and SAR models are estimated by the ML procedure introduced in Elhorst (2014). ***, ** and * indicate 1%, 5%, 10% significance level, respectively.

network spillovers occurs based on geographical proximity, the SDM model is a better specification. The third column of Table 3.5 shows the ML estimates of SDM model. Again, the point estimate of spatial lag of house prices is statistically significant, suggesting the existence of remaining dependence arising from other channels. Compared to the SDEM model, which models the spatial interaction in the residuals, the magnitude of the direct effect of population density in the SDM model increases from 0.13% to 0.21%. Most importantly, the spatial lags of population density and land area have negative signs, which finding contradicts network spillovers, but these effects are not statistically significant. In this case, the SDM model collapses to the SAR model which only includes the spatial lag of house prices. The SAR model estimates presented in the fourth column of Table 3.5 are largely in line with the results of SDM model. The results of the SDM and SAR models suggest that the spatial lag of house prices also contains the information of network spillovers, which cannot be distinguished from other mechanisms that can result in spatial dependence.

§ 3.6.3 Network spillovers

All of the four models, SLX, SDEM, SDM and SAR, can be used to model the network spillovers in the housing markets. The first two models directly reflect our theoretical foundation and generate local network spillovers whereas the last two models generate global spillovers which are not easy to justify[11]. Table 3.6 summarizes the direct and spillovers effects of the four models. For the SLX and SDEM model, the direct and spillover effects are the corresponding point estimates. On the other hand, the partial derivative approach is needed to calculate the direct and spillover effects for the SDM and SAR models.

The direct effect of land area is almost the same among the four spatial models, while the direct effect of population density estimated by the SDM and SAR models is much more pronounced than that in the SLX and SDEM models. In contrast, the SDM and SAR models estimate much lower network spillovers of both population density and land area than the SLX and SDEM models do. In the SDM specification, there is no significant network spillover at all. Since the spatial lag of house prices mixes various sources of spatial interaction, and because the global spillovers assumption is not consistent with our theoretical foundation, our interpretation is based on the SLX and SDEM models, and in particular on the latter model which considers the remaining spatial dependence in residuals.

11 Local spillovers are those spillovers occuring only between a city and its neighbouring cities connected by a spatial weight matrix. In contrast, global spillovers are those spillovers that originate from a city and transmit to all other cities.

TABLE 3.6 The estimated direct effects and spillovers

	RE	SLX	SDEM	SDM	SAR
Direct effects					
Ln(Population density)	0.1712***	0.1181*	0.1278***	0.2148***	0.2059***
	(2.91)	(1.87)	(2.65)	(3.90)	(3.84)
Ln(Land)	0.1817**	0.1897***	0.1702***	0.1832***	0.1914***
	(2.55)	(2.72)	(2.80)	(2.68)	(3.01)
Spillovers					
Ln(Population density)		0.2719***	0.2493***	0.0941	0.1989***
		(3.04)	(2.91)	(0.66)	(3.07)
Ln(Land)		0.2590**	0.2490**	0.0895	0.1850***
		(2.00)	(2.02)	(0.38)	(2.56)

Notes: For SLX and SDEM model, the direct effect and network spillovers are exactly the point estimates, whereas the partial derivative approach is used for SDM and SAR model. The inferences of direct effects and spillovers in SDM and SAR model are based on 1000 simulations using the variance-covariance matrix implied by the ML estimates (LeSage and Pace 2009; Elhorst 2010a). *t*-values are reported in the parentheses. ***, ** and * indicate 1%, 5%, 10% significance level, respectively.

The direct effect of the SDEM model shows that the influence of land area is bigger than the effect of population density, suggesting that, in current China, city growth is likely to be characterized by an expansion of urban scale rather than an increase in intensity. The network spillover is even more noticeable. If a city becomes 1% denser and larger, the total house price increases of neighbouring cities are about 0.25%, whereas its own house price only rises by 0.13% and 0.17%, respectively. Considering that each city on average has 8 neighbours within the radius of 160 km, the network spillover on each neighbouring city is by average around 0.03%, which is much lower than the magnitude of the direct effect.

§ 3.6.4 Discussion

As previously discussed, the estimation of network spillovers depends on the choice of spatial weight matrix because this matric reflects the underlying interaction structure. In this paper, the distance band used for constructing the spatial weight matrix is somewhat arbitrarily chosen. To check the robustness of our findings, we replicated our analysis based on 4 distance bands: 120, 200, 240 and 280 km. Table B1 in the appendix reports the SDEM estimation of direct effect and network spillovers based on different matrixes. Clearly, the estimation of the direct effect is very robust to the choice of spatial matrix. The estimation of network spillovers, on the other hand, shows some variation, though small. When the definition of neighbours is restricted to the radius of 120 km, the network spillovers decrease substantially with the spillover of land area becoming insignificant. As we include more cities as neighbours, the magnitude of network spillover becomes a bit larger (see the results of distance band 240 and 280 km), which is in line with our expectation. Nevertheless, given that the spatial weight matrix is based on a squared inverse distance function, the majority of the network spillovers still falls into the nearby neighbours despite the total number of neighbours being increased. In this sense, we believe that the evidence on network spillovers presented in this paper is reliable.

The existence of network spillovers means that, all else being equal, house price in a city surrounded by large cities are much higher than those in a city that has small neighbours. If we see high house prices as a sign of the 'triumph of the city', our findings would indicate a core-periphery structure in our study area. We checked this implication using Moran's I plot, which is based on a spatial weight matrix which assumes that every city within the 160 km radius of a specific city has the same influence on that city (a bit different from the matrix defined by equation (7))[12]. The global Moran's I plot (Figure C1 in the Appendix) clearly shows a positive correlation between a city's house price and the neighbouring cities' urban population[13]. In particular, almost all of the cities that are surrounded by small cities have relatively lower house prices. The local Moran's I map, also known as local indicators of spatial association (LISA) (Anselin 1995), depicts a detailed clustering pattern. In general, we find a 'successful' group (high prices – high population) in the east of the study area and a 'lost' group (low price – low population) in the western part. There is one exception: a coastal city (Nantong) in the east with large neighbours is characterized by relatively low prices. Yet, the pattern that a city surrounded by small cities has high house prices is not supported. A lesson learned is that the vast number of 'lost' cities in the western part will not flourish in the near future, since they are unable to benefit from network spillovers of big cities.

§ 3.7 Concluding remarks

Most studies attributed the spatial variation of interurban house prices to local-specific characteristics. However, the spatial clustering pattern of house prices cannot be fully explained by these local-specific variables, pointing to the importance of spillovers. To account for spatial interaction, spatial econometrics is becoming the standard toolbox in the analysis of house prices. In particular, the spatial model with spatial lag of house prices has been widely used by researchers. Nevertheless, SAR-type models have been heavily criticized because the endogenous interaction is difficult to justify.

12 Both the global Moran's I plot and the LISA map are calculated by the software 'Geoda' which is available at https://geodacenter.asu.edu.

13 The Moran's I statistic is 0.193. Based on the distribution of 999 simulations of spatially random distributed data, it is significant at 5% significance level.

This paper differs from conventional spatial analysis of house prices in that we investigate spillovers caused by city network externalities. In a city network system, the house price of a city is influenced by the urban size of accessible neighbouring cities, because the performance of amenities and productivity advantage of that city, which are the two basic components of house prices, can be somewhat 'borrowed' from its neighbours. The network spillovers justify the assumption of exogenous interactions in spatial econometrics which has been overlooked in applied studies. Hence, we argue that, when analyzing house price spillovers, the SLX and SDEM models are attractive alternatives to SAR-type models.

Using a panel data set of Pan-Yangtze River Delta (PYRD) in eastern China, the SLX model, which incorporates exogenous interaction, strongly supports the presence of network spillovers. Even after controlling for the spatial interaction in residuals, the effect of network externality is still significant. On the other hand, the SAR-type models, SAR and SDM, cannot properly measure the local network spillovers. They estimated a less amount of the network spillovers than the SLX and SDEM model did. Our findings are in line with studies based on the measure of market potential, such as Partridge et al. (2009). The evidence underlines the importance of cross-city spillovers in the formation of house prices. Especially cities that are proximal to super big cities are likely to have higher house prices than their own local-specific characteristics suggest. This point should be remembered when assessing 'house price bubbles': taking cross-city spillovers into account may lead to opposite conclusions.

This paper is also relevant to the increasing studies that focus on 'borrowed size', which is currently used to explain the faster growth of small and medium-sized cities in Europe (Meijers et al. 2016). While most studies investigate the 'borrowing size' concept from a functional view by examining the presence of metropolitan functions, such as science, sport, political-administrative functions and cultural amenities (e.g.,Burger et al. 2015; Meijers et al. 2016), this paper provides new evidence from the perspective of house prices. Furthermore, our results suggest that 'borrowed size' might also make sense in explaining the city growth in China, though China has significantly different social-economic conditions from Western Europe. This calls for the making of more regional policies that involve more collaboration and integration between cities. However, it should be noted that, despite the existence of network externalities, it is still not easy for small cities in more peripheral areas to achieve fast development.

References

Alonso, W. (1973). Urban zero population growth. Daedalus, 102(4), 191-206.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1995). Local indicators of spatial association – LISA. Geographic Analysis, 27, 93-115.

Anselin, L., Le Gallo, J., & Jayet, H. (2008). Spatial panel econometrics. In L. Mátyás, & P. Sevestre (Eds.), The econometrics of panel data (Vol. 46, pp. 625-660). Berlin: Springer.

Anselin, L., Lozano-Gracia, N., Deichmann, U., & Lall, S. (2010). Valuing access to water - A spatial hedonic approach, with an application to Bangalore, India. Spatial Economic Analysis, 5(2), 161-179.

Baltagi, B. H., Fingleton, B., & Pirotte, A. (2014). Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices. Journal of Urban Economics, 80, 76-86.

Basu, S., & Thibodeau, T. G. (1998). Analysis of Spatial Autocorrelation in House Prices. Journal of Real Estate Finance and Economics, 17(1), 61-85.

Boix, R., & Trullén, J. (2007). Knowledge, networks of cities and growth in regional urban systems. Papers in Regional Science, 86(4), 551-574.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. The Journal of Real Estate Finance and Economics, 35(2), 143-160.

Brady, R. R. (2014). The spatial diffusion of regional housing prices across U.S. states. Regional Science and Urban Economics, 46, 150-166.

Brakman, S., Garretsen, H., & Schramm, M. (2004). The spatial distribution of wages: Estimating the Helpman-Hanson model for germany. Journal of Regional Science, 44(3), 437-466.

Breusch, T. S. (1987). Maximum likelihood estimation of random effects models. Journal of Econometrics, 36(3), 383-389.

Burger, M. J., Meijers, E. J., Hoogerbrugge, M. M., & Tresserra, J. M. (2015). Borrowed size, agglomeration shadows and cultural amenities in North-West Europe. European Planning Studies, 23(6), 1090-1109.

Can, A. (1990). The Measurement of Neighborhood Danymics in Urban House Prices. Economic Geography, 66(3), 254-272.

Can, A. (1992). Specification and estimation of hedonic housing price mdoels. Regional Science and Urban Economics, 22(3), 453-474.

Capello, R. (2000). The city network paradigm: Measuring urban network externalities. Urban Studies, 37(11), 1925-1945.

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? Journal of Regional Science, 52(2), 210-239.

de Bruyne, K., & van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. Applied Economics, 45(13), 1673-1689.

Elhorst, J. P. (2010a). Applied spatial econometrics: Raising the bar. Spatial Economic Analysis, 5(1), 9-28.

Elhorst, J. P. (2010b). Spatial panel data models. In M. M. Fischer, & A. Getis (Eds.), Handbook of applied spatial analysis (pp. 377-407). Berlin: Springer.

Elhorst, J. P. 2014. Spatial Econometrics: From cross-sectional data to spatial panels. Heidelberg New York Dordrecht London: Springer.

Fingleton, B. (2008). Housing supply, housing demand, and affordability. Urban Studies, 45(8), 1545-1563.

Fingleton, B., & Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: finite sample properties. Papers in Regional Science, 87(3), 319-339.

Fujita, M., Krugman, P., & Venables, A. J. 1999. "The spatial economy: Cities, regions, and international trade". Cambriage MA: The MIT Press.

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? Journal of Regional Science, 52(2), 172-191.

Giussani, B., & Hadjimatheou, G. (1991). Modeling regional house prices in the United Kingdom. Papers in Regional Science, 70(2), 201-219.

Glaeser, E. L., Gyourko, J., Morales, E., & Nathanson, C. G. (2014). Housing dynamics: An urban approach. Journal of Urban Economics, 81, 45-56.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. Journal of Economic Geography, 1(1), 27-50.

Gong, Y., Boelhouwer, P., & de Haan, J. (2016). Interurban house price gradient: Effect of urban hierarchy distance on house prices. Urban Studies, 53(15), 3317-3335.

Hanink, D. M., Cromley, R. G., & Ebenstein, A. Y. (2012). Spatial variation in the determinants of house prices and apartment rents in China. Journal of Real Estate Finance and Economics, 45(2), 347-363.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. Journal of International Economics, 67(1), 1-24.

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. Annals of the Association of American Geographers, 44(4), 315-348.

Head, K., & Mayer, T. 2004. "The empirics of agglomeration and trade". In Handbook of Regional and Urban Economics, edited by J. V. Henderson, & J. F. Thisse (Ed.), (Vol. 4, pp. 2609-2669). Amsterdam: Elsevier.

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

Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of Urban Economics, 69(1), 2-23.

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. Journal of Environmental Economics and Management, 45(1), 24-39.

LeSage, J. P., & Pace, R. K. 2009. Introduction to spatial econometrics. Boca Raton: Taylor & Francis CRC Press.

Li, Q., & Chand, S. (2013). House prices and market fundamentals in urban China. Habitat International, 40, 148-153.

Malpezzi, S. (1996). Housing prices, externalities, and regulation in U.S. metropolitan areas. Journal of Housing Research, 7(2), 209-241.

Meijers, E. J., & Burger, M. J. (2015). Stretching the concept of 'borrowed size'. Urban Studies, Forthcoming.

Meijers, E. J., Burger, M. J., & Hoogerbrugge, M. M. (2016). Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe. Papers in Regional Science, 95(1), 181-199.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. Journal of Real Estate Research, 32(3), 289-320.

Ozanne, L., & Thibodeau, T. (1983). Explaining metropolitan housing price differences. Journal of Urban Economics, 13(1), 51-66.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. Journal of International Economics, 78(1), 126-140.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo Working Paper, No. 1229, 1-40.

Phelps, N. A., Fallon, R. J., & Williams, C. L. (2001). Small firms, borrowed size and the urban-rural shift. Regional Studies, 35(7), 613-624.

Pollakowski, H. O., & Ray, T. S. (1997). Housing price diffusion patterns at different aggregation levels: An examination of housing market efficiency. Journal of Housing Research, 8(1), 107-124.

Potepan, M. J. (1996). Explaining intermetropolitan variation in housing prices, rents and land prices. Real Estate Economics, 24(2), 219-245.

Roback, J. (1982). Wages, rents, and the quality of life. Journal of Political Economy, 90(6), 1257-1278.

Rosen, S. (1979). Wage-based indexes of urban quality of life. In P. Mieszkowski, & M. Straszheim (Eds.), Current Issues in Urban Economics. Baltimore: The Johns Hopkins University Press.

Vega, S. H., & Elhorst, J. P. (2015). The SLX model. Journal of Regional Science, 55(3), 339-363.

Zheng, S., Cao, J., Kahn, M. E., & Sun, C. (2014). Real estate valuation and cross-boundary air pollution externalities: Evidence from Chinese cities. Journal of Real Estate Finance and Economics, 48(3), 398-414.

Zheng, S., Fu, Y., & Liu, H. (2009). Demand for urban quality of living in China: Evolution in compensating land-rent and wage-rate differentials. Journal of Real Estate Finance and Economics, 38(3), 194-213.

Zheng, S., Kahn, M. E., & Liu, H. (2010). Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. Regional Science and Urban Economics, 40(1), 1-10.

Appendices

Appendix A.Variable compilation

House prices. The only data set available to us is the total transaction price of all the newly sold residential buildings, from which we derived the average unit price. A few cities only have combined sales data for all the buildings (commercial, residential and mixed used), but, according to the data in other cities, residential buildings account for the great majority of total transactions. The average unit price for residential buildings in these few cities is estimated by correcting the average unit price of all buildings; the correction coefficient is the average ratios of residential price to mixed price in neighbouring cities.

Student/Teacher ratio. This ratio is calculated based on the aggregate data on primary and regular secondary schools. The teachers and students in regular institutions of higher education (universities or colleges) are excluded from the calculation.

Population density. We do not have consistent data on urban population and urbanisation rate in each year for each city in our sample, but we do have the data for total permanent population (including urban population and rural population). In 2000 and 2010 population census year, the urbanisation rate can be accurately calculated. We assume a linear growth pattern for urbanisation during the decades, and so we can estimate the corresponding urbanisation rate during our sample year. With the urbanisation rate and total population in hand, we can estimate the urban population in each year.

Appendix B.Tables

TABLE B1 Direct effect and spillovers based on different spatial weight matrixes, SDEM

	$\mathbf{W}_G^{0\text{-}120}$	$\mathbf{W}_G^{0\text{-}160}$	$\mathbf{W}_G^{0\text{-}200}$	$\mathbf{W}_G^{0\text{-}240}$	$\mathbf{W}_G^{0\text{-}280}$
Direct effects					
Ln(Population density)	0.1228***	0.1278***	0.1348***	0.1319***	0.1361***
	(2.60)	(2.65)	(2.80)	(2.65)	(2.74)
Ln(Land)	0.1412**	0.1702***	0.1753***	0.1716***	0.1770***
	(2.32)	(2.80)	(2.85)	(2.77)	(2.78)
Spillovers					
Ln(Population density)	0.1198*	0.2493***	0.2484***	0.2765***	0.2894***
	(1.78)	(2.91)	(2.70)	(2.76)	(2.69)
Ln(Land)	0.0589	0.2490**	0.2504*	0.2747*	0.3127*
	(0.63)	(2.02)	(1.93)	(1.94)	(1.93)

Notes: The direct effects and network spillovers are estimated by the spatial Durbin error model specified in equation (3), with $\mathbf{M} = \mathbf{W}_{EG}^{0\text{-}1.5}$. The spatial weight matrixes are defined in the same way as those defined in Table 3.2. t-values are reported in the parentheses. ***, ** and * indicate 1%, 5%, 10% significance level, respectively.

TABLE B2 Fixed effects estimation of time-variant variables

Dependent variable = Ln (House prices)					
Winter temperature	Ln (Smoke and dust emission)	Ln (Student /Teacher ratio)	Doctor	Ln (Population density)	R-squared
0.0226***	0.0200	-1.0950***	0.0088	2.0576***	0.730
(0.0065)	(0.0337)	(0.3369)	(0.0142)	(0.3739)	

Notes: The robust standard errors are reported in the parentheses. ***, ** and * indicate 1%, 5%, 10% significance level, respectively.

Appendix C.Figures

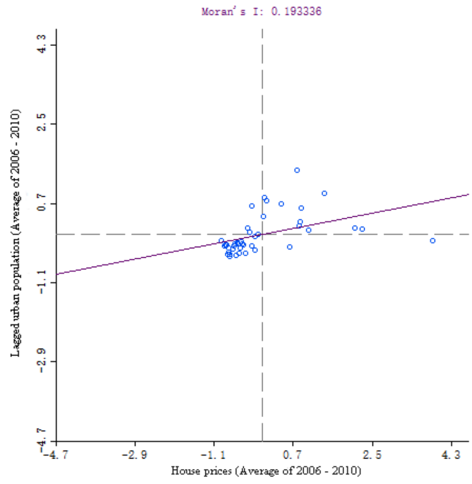


FIGURE C1 Global Moran's I plot

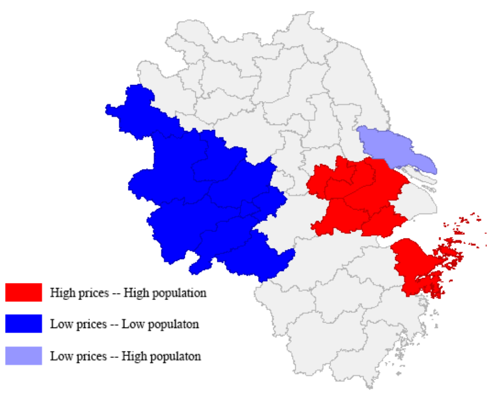


FIGURE C2 LISA cluster map

4 The clustering pattern of Chinese house price dynamics

Submitted for review

Abstract: This paper investigates the clustering pattern of house price dynamics in 34 major cities in China over the period 2005–2016. Hierarchical agglomerative clustering is implemented based on a distribution-based dissimilarity measure, the Kullback-Leibler divergence, which measures the similarity between house price appreciation trajectories. The clustering procedure reveals a broad two-cluster structure: one relatively homogeneous slow-growing market cluster and one red-hot market cluster which, however, has a higher degree of within-cluster heterogeneity. The two-cluster partition also indicates a geographical pattern that separates out Eastern China. However, this clustering pattern is mainly shaped by the market structure in the recent period after 2014. Prior to 2014, and especially before 2010, the interurban housing market in China could be considered a homogenous market in terms of house price changes.

Keywords: House price dynamics, housing market divergence, hierarchical clustering, Kullback-Leibler divergence, China

§ 4.1 Introduction

After decades of rapid growth in house prices, the Chinese housing market has begun to cool down since 2014, when Chinese economic growth also began to slow. This has caused widespread worries about the prospects of the Chinese economy given the important economic role of the real estate sector. To stabilize the housing market and achieve the economic growth targets, the central and local governments chose to actively engage in the housing market through policy interventions. Then, after the second half of 2015, the housing markets in some cities heated up again while other cities' housing markets remained stagnant. For example, as of June 2016, house prices in Beijing had increased 20 percent compared to June 2015, whereas house prices in Kunming, the capital of a Western province, are nearly stable.

Facing the great divergence in house price dynamics between cities, the government regulation has to resort to diverging, local-oriented policy tools, the design of which heavily depends on our clear understanding about the segmentation of interurban housing market. What is the segregation pattern of the Chinese housing markets? Is the housing market divergence a new phenomenon or a long-established pattern? Does geography play a role in fragmenting the housing markets? This paper attempts to shed light on these questions by conducting classification analysis on city-level housing markets in China. By means of cluster analysis, the divergent housing market structure can be well described by a few homogeneous clusters, within which the markets are very similar to each other but the differences between clusters are significant.

A key element in classifying real estate markets is the similarity criterion. The delineation of intra-city housing submarkets, for example, can be based on the similarity in housing attributes and/or the similarity in shadow prices of those attributes (e.g., Goodman and Thibodeau, 1998; Watkins, 2001; Bhattacharjee et al., 2016). In the case of classifying the interurban real estate market, a large amount of studies have been based on the similarity in market performance, like the dynamics of property rent or price (e.g.,Hamelink et al., 2000; Jackson, 2002). This paper follows the paradigm of market performance approach. However, unlike the previous literature that use distance measures to represent similarities, I introduce a distribution-based dissimilarity measure, the Kullback-Leibler (*KL*) divergence (Kullback, 1968; Kullback and Leibler, 1951), which reflects the structural difference between Data Generating Processes (DGP) that generate the house price dynamics of different cities.

This paper then applies the hierarchical clustering method to 34 major cities' housing markets in China over the period 2005–2016, aiming to investigate the cross-market divergence pattern. The temporal stability of divergence pattern is also examined by performing the cluster analysis on sub-periods. In general, these cities can be broadly grouped into two clusters, one cluster containing relatively homogeneous slow-growth markets and the other containing red-hot markets in Eastern China, which have a much higher degree of heterogeneity. That is, the latter cluster can be further partitioned into sub-clusters. Such a clustering pattern is mainly shaped by the market structure in the recent period after 2014. Throughout the sample period, the Chinese interurban housing market has experienced significant structural changes, particularly in the later years; it has shifted from a homogenous market structure to a divergent one. Besides, this paper also examines whether the geographical demarcation and city-tier division schemes, which are frequently referred to when defining homogenous housing market groups in practice, are consistent with the divergence pattern of housing markets.

While the literature on homogeneous grouping of commercial property markets is extensive, very few studies focus on the cluster analysis of housing markets. Some exceptions are Abraham et al. (1994) on grouping U.S. metropolitan housing markets and Hepşen and Vatansever (2012) on clustering Turkish housing markets. Dong et al. (2015) and Guo et al. (2012) also made attempts to partition the Chinese city-level housing markets into few homogeneous clusters. However, both of the studies are subject to a relatively short period with no more than five years and thus fail to examine the temporal evolution of the segmentation structure. In this regard, the current paper greatly contributes to the understanding of the evolutionary divergence pattern of Chinese housing markets.

The remainder of this paper is organised as follows. Section 4.2 briefly reviews the literature on the clustering of housing markets. The dissimilarity measure and clustering method are described in Section 4.3, followed by an introduction to the data and some stylized facts in Section 4.4. Section 4.5 reports the clustering results, tests the structural changes and discusses the findings. Finally, a short summary is provided in Section 4.6.

§ 4.2 Previous literature

The intra-city and inter-urban real estate market is likely fragmented due to market imperfections. Defining and identifying intra-city housing submarkets thus has various advantages. It can significantly improve the prediction power of house price models, help lenders and investors to better price the risk associated with financing homeownership, and reduce the search cost for housing consumers (Goodman and Thibodeau, 2007). Similarly, the cluster analysis of interurban real estate market also brings considerable benefits. This section mainly reviews the studies on classifying the interurban real estate market.

One benefit of homogenous grouping of real estate markets across cities is aiding in real estate portfolio diversification. The grouping strategy has initially been to conform to the geographical regions created for administrative purposes, such as the U.S. eight-region system used by the Bureau of Economic Analysis (BEA)[1]. However, Malizia and Simons (1991), using the standard deviation of demand-side indicators (employment, for example) as the criterion of homogeneity within categories, found that this eight-region system does not perform well.This calls for a classification scheme based on the characteristics of property markets rather than solely on regional proximity.

Using the time series data of real estate market characteristics, many studies, mostly on commercial real estate markets, employ clustering methods to perform the

1 The eight regions are New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountains and Far West.

classification analysis, based on some similarity measures, such as Euclidean distance and correlation coefficient. Goetzmann and Wachter (1995) looked into the segmentation structure of 21 metropolitan U.S. office markets based on effective rents and the structure of 22 markets based on vacancy data. In line with the suggestion of Malizia and Simons (1991), the K-means clustering revealed bicoastal relationships among cities; that is, some east and west coast cities tend to be clustered together regardless of the great distances between them. The resulting clustering pattern in the paper is then tested by a bootstrap procedure. Outside the U.S., Jackson (2002) applied the hierarchical clustering method to the retail property markets of 60 towns and cities in Great Britain and identified seven homogeneous groups based on average retail rental value growth.

Hoesli et al. (1997) applied various clustering techniques to 156 retail, office and industrial markets in the UK, attempting to reveal the extent to which property markets are grouped by property type or by area. Property type is found to be the dominant factor in determining different market behaviours; it is overlaid by the geographical factor, which emphasises the role of London. A later study (Hamelink et al. 2000) extends the work of Hoesli et al. (1997) by testing more property type/ region combinations, such as the 3 property types $\times$ 3 super-regions combination and the 3 property types $\times$ 13 standard regions combination[2]. The results confirmed the findings of Hoesli et al. (1997), revealing a strong property-type dimension and a weak broad geographical dimension.

Compared to the large body of literature on the homogeneous clustering of commercial real estate markets, clustering analyses of housing markets are relatively limited, with a few notable exceptions. Abraham et al. (1994) identified three meaningful homogeneous clusters: an East Coast group, a West Coast group and a central U.S. group. More recently, Hepşen and Vatansever (2012) applied hierarchical clustering method to 71 Turkish metropolitan housing markets and revealed three clusters with different rental return levels. Using a combination of wavelet analysis and expert experience, Guo et al. (2012) first divided the time series of house prices indexes of 70 Chinses cities over the period 2005 - 2010 into a few distinct sub-periods. The DBSCAN clustering algorithm was then applied and partitioned these markets into 6 clusters and 5 un-clustered markets based on the characteristics of each sub-period. With a two-stage clustering procedure, Dong et al. (2015) divided the housing markets of 283 cities in China into three clusters and thirteen sub-clusters. The first stage of classification is based on the similarity in demand and supply fundamentals and the second state further divides the clusters formed in stage one according to the similarity

2 The three super-regions are London, the South (the rest of the South East, East Anglia, and the South West) and the North (East Midlands, West Midlands, Wales, North West, Yorkshire and Humberside, the North, and Scotland). In the 13 cases, London is further divided.

of market performance (housing sale value and house prices).

Besides, van Dijk et al. (2011), using a latent-class panel time series model, divided the Dutch regional housing markets into two clusters according to criterion whether the markets can be modelled by a common house price model. This classification logic is based on the similarity in structural parameters of the regional house price model, which is different from the previous inter-market classification studies but in line with most studies on identifying the intra-city submarkets. Owing to the lack of continuous time series data on housing market fundamentals in a long period, this paper follows the traditional wisdom and performs the cluster analysis based on the house price appreciation trajectories.

§ 4.3 The Kullback-Leibler discrepancy measure and clustering method

§ 4.3.1 The Kullback-Leibler divergence between two housing markets

To assign a local housing market into a corresponding cluster based on its house price growth pattern, a measure that reflects the difference between two house price appreciation series is needed, such as the Euclidean distance. This paper introduces the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951; Kullback 1968), which measures the 'distance' between two probability distributions. Suppose that two house price growth series, say $y_{p,t}$ and $y_{q,t}$ $(t = 1, 2, \cdots, T)$, are generated by probability density functions (pdf) $P(t)$ and $Q(t)$, respectively. Then, the structural dissimilarity between housing markets p and q can be reflected by the discrepancy between the house price distributions of $P(t)$ and $Q(t)$. The Kullback-Leibler divergence is a measure that calculates the divergence of distribution $Q(t)$ from the distribution $P(t)$ and follows the form

$$KL\,(P;Q) = \int \ln \frac{P(t)}{Q(t)} P(t)dt. \tag{1}$$

Note that KL divergence is not symmetric, that is $KL(P;Q) \neq KL(Q;P)$. For ease of classification, a symmetric measure, known as J divergence (Kullback 1968), is defined as

$$JKL(P;Q) = KL(P;Q) + KL(Q;P). \tag{2}$$

This measure has all the properties of a distance measure except triangular inequality and has been widely used in cluster analysis (e.g.,Kakizawa et al. 1998; Bengtsson and Cavanaugh 2008)[3].

Assume that the house price appreciation of a city i, $y_{i,t}$, is generated from an $AR(P)$

3 The triangular inequality of JKL divergence means that $JKL\,(P,Q) \leq JKL\,(P,R) + JKL\,(R,Q)$

process

$$y_{i,t} = c_i + \phi_{1,i} y_{i,t-1} + \phi_{2,i} y_{i,t-2} + \cdots + \phi_{p,i} y_{i,t-p} + \epsilon_{i,t} \tag{3}$$

where c_i is the average growth rate of city i and reflects the long-run growth trend driven by city-specific characteristics such as population, income growth and so forth, $\epsilon_{i,t} \sim N\left(0, \sigma_i^2\right)$ is the independent and identically distributed Gaussian error. In total, $\theta_i = \left(c_i, \phi_{1,i}, \phi_{2,i}, \cdots, \phi_{p,i}, \sigma_i^2\right)'$ is the parameter vector to be estimated in the model. Conditional on the first p observations, the joint probability density function for the house price appreciation in city i becomes

$$\begin{aligned} f(y_{i,T}, \cdots, y_{i,p+1} | y_{i,p}, \cdots, y_{i,1}; \theta_i) = & \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_i^2}} \\ & \exp\left(-\frac{\left(y_{i,t} - c_i - \phi_{1,i} y_{i,t-1} - \cdots - \phi_{p,i} y_{i,t-p}\right)^2}{2\sigma_i^2}\right). \end{aligned} \tag{4}$$

By maximizing the natural logarithm of likelihood function (4), one obtains the conditional maximum likelihood estimators (MLE) for θ, which are identical to the ordinary least squares (OLS) estimators of Equation (3). In compact fashion, Equation (4) can be expressed as

$$f(y_{i,T}, \cdots, y_{i,p+1} | y_{i,p}, \cdots, y_{i,1}; \theta_i) = \left(2\pi\sigma_i^2\right)^{-\frac{T-p}{2}} \exp\left(-\frac{\left(y_i - X_i\beta_i\right)'\left(y_i - X_i\beta_i\right)}{2\sigma_i^2}\right) \tag{5}$$

where $y_i = (y_{i,p+1}, \cdots, y_{i,T})'$, $X_i = \left((1, 1, \cdots, 1)', (y_{i,p}, \cdots, y_{i,T-1})', (y_{i,p-1}, \cdots, y_{i,T-2})', \cdots, (y_{i,1}, \cdots, y_{i,T-p})'\right)$, and $\beta_i = (c_i, \phi_{1,i}, \phi_{2,i}, \cdots \phi_{,p,i})'$.

Now one has another city's house price appreciation series, $y_{j,t}$, generated by the parameter vector θ_j conditional on the observations $\{y_{j,1}, \cdots, y_{j,p}\}$. According to Equation (1), the KL divergence between the two cities' housing markets will be:

$$KL(y_i; y_j) = \int_{y_{p+1}, \cdots y_T} \log \frac{f(y_T, \cdots, y_{p+1} | y_{i,p}, \cdots, y_{i,1}; \theta_i)}{f(y_T, \cdots, y_{p+1} | y_{j,p}, \cdots, y_{j,1}; \theta_j)} f(y_T, \cdots, y_{p+1} | y_{i,p}, \cdots, y_{i,1}; \theta_i)\, dy. \tag{6}$$

Substituting Equation (5) into Equation (6), one obtains the computational form of the KL divergence:

$$\begin{aligned} KL(y_i; y_j) &= \frac{T-p}{2}\left(\log\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2} - 1\right) + \frac{\left(X_i\beta_i - X_j\beta_j\right)'\left(X_i\beta_i - X_j\beta_j\right)}{2\sigma_j^2} \\ KL(y_j; y_i) &= \frac{T-p}{2}\left(\log\frac{\sigma_i^2}{\sigma_j^2} + \frac{\sigma_j^2}{\sigma_i^2} - 1\right) + \frac{\left(X_j\beta_j - X_i\beta_i\right)'\left(X_j\beta_j - X_i\beta_i\right)}{2\sigma_i^2}. \end{aligned} \tag{7}$$

The symmetric measure $JKL(y_i; y_j)$ can be easily obtained by summing up the term $KL(y_i; y_j)$ and $KL(y_j; y_i)$, and this symmetric measure will be used for the cluster analysis in the next step.

Unlike the Euclidean distance which measures the straight-line distance between two sample house price appreciation series, the *JKL* divergence measures the structural difference between two distributions that can generate the sample series. Thus, the *JKL* divergence is considered to be more consistent with the 'true' difference between housing markets.

§ 4.3.2 The clustering method

After obtaining the *KL* divergence matrix across cities, the hierarchical clustering method, particularly the bottom-up agglomerative method, is employed to assign the cities into relatively homogeneous sub-groups. The procedure begins by treating each city as an individual cluster and merging the two cities (say y_i and y_j) that have the lowest dissimilarity, measured by $JKL(y_i; y_j)$, into one cluster C_i. The next step involves updating the dissimilarity between a formed cluster and other clusters (or individual cities) according to linkage criteria. There are several linkage criteria available, and in this paper the widely used *average-linkage* method is employed; that is, the dissimilarity between the two clusters is equal to the average dissimilarity between a city in cluster C_i and a city in cluster C_j. Let N_i and N_j be the number of cities belonging to clusters C_i and C_j, respectively. The dissimilarity between clusters C_i and C_j is defined as $JKL(C_i; C_j) = \sum_{y_i \in C_i} \sum_{y_j \in C_j} JKL(y_i; y_j)/N_i N_j$. By repeating this process, a hierarchical tree linking the nearest neighbours is generated, which is known as a 'dendrogram'. Finally, one can cut the tree at the desired level and obtain the corresponding clusters.

Hierarchical clustering is silent on determining the correct number of clusters. This can be achieved by optimizing some cohesion and separation measures. One widely used example of such measures is the Silhouette statistic (Rousseeuw 1987); the number that can maximize the average Silhouette values is chosen as the correct number of clusters. One drawback of the Silhouette statistic is that it is not well defined for the individual clusters that have only one member, which, according to Figure 4.3, is very likely to happen in this study. This paper uses a heuristic approach to determine the number of clusters: the "elbow" approach.

The elbow approach attempts to find a balance between the increase in within-cluster cohesion and the decrease in data compactness. Within-cluster cohesion is measured by the sum of within-cluster distances $S_w(k) = \sum_{i,j \in C_k}^{k} d_{ij}$ where d_{ij} is the distance measure that can be either Euclidean distance or *JKL* divergence. The smaller the $S_w(k)$ is, the higher the cohesion of a cluster. Compactness is measured by the number of between-cluster city pairs $N_b(k) = \sum_{i=1}^{k-1} \sum_{j=k+1}^{k} n_i n_j$ where n_i is the number of objects in cluster C_i. By this measure, the uneven partition is considered to be more compact than the even partition. When one more cluster is added, the S_w is always decreasing while the N_b is always increasing. The process should stop at cluster k, where continuing to increase N_b cannot offer much of a decrease in S_w. Now, I define statistic $A_k = [S_w(1) - S_w(k)] / N_b(k)$, which means the average cohesion gain of $k$clusters. Note that because the numerator of A_k is exactly the sum of between-cluster relationships, A_k can also be interpreted as the average between-cluster distance. If one plots the A_k on the Y-axis and the number of clusters k on the X-axis, it can be found that from some k onward, the remarkably flattens (see Figure A1, for example). The "elbow" point is deemed to be the appropriate number of clusters. I test the effectiveness of the elbow approach on two data sets exhibited in Charrad et al. (2014), which comprehensively uses 27 indicators presented in the literature to determine the number of natural clusters. The elbow approach turns out to correctly identify the number of clusters as recovered by Charrad et al. (2014).

§ 4.4 Data and stylized facts

§ 4.4.1 House price index and appreciation

This paper analysed the monthly house price dynamics of 34 major cities in China from July 2005 to June 2016 ($T = 132$). These cities cover municipalities directly under the central government, provincial (autonomous regions) capitals and vital economic centres, and hence their price changes attract the majority of public attention. For all of the sample cities, the system of "Price Indices of Newly Constructed Residential Buildings in 35/70 Large- and Medium-sized Cities" (70 Cities Index), which is compiled by the National Bureau of Statistics of China (NBSC), publishes month-over-month house price changes[4]. The series of monthly house price changes in 34 cities will be the main input in the classification analysis. The "35/70 Cities Index" was launched in 1997 and reports, on a quarterly basis, the year-over-year index for 35 major cities. In July 2005, the system was expanded to cover 70 Large- and Medium-sized cities and began to report monthly. Also since 2005, house price changes have been calculated through a so-called "matching approach" (Wu et al. 2014), which can better control for quality changes[5]. The price index compilation strategy was slightly adjusted in January 2011, but this change would not significantly affect the consistency of the house price index. Finally, to provide an intuitive perception of house price dynamics during the sample period, I convert the month-over-month price changes into a fixed-base house price index through the

4 Haikou, the capital of Hainan province, is excluded from our analysis. As a popular tourist resort, Haikou's housing market has some distinct characteristics, and its house price dynamics clearly deviate from the other cities during the sample period.

5 The "matching" model used for the NBSC index is analogous to the repeat sales model. In each month, local statistical authorities collect housing transaction information from different housing complexes. The houses within the same housing complex have similar structural and locational characteristics. Thus, for each housing complex, comparing the average transaction prices of different periods roughly produces a quality-adjusted house price index. The city-level index is the weighted average of all complex-level indexes.

chaining algorithm (reference base = June 2015).

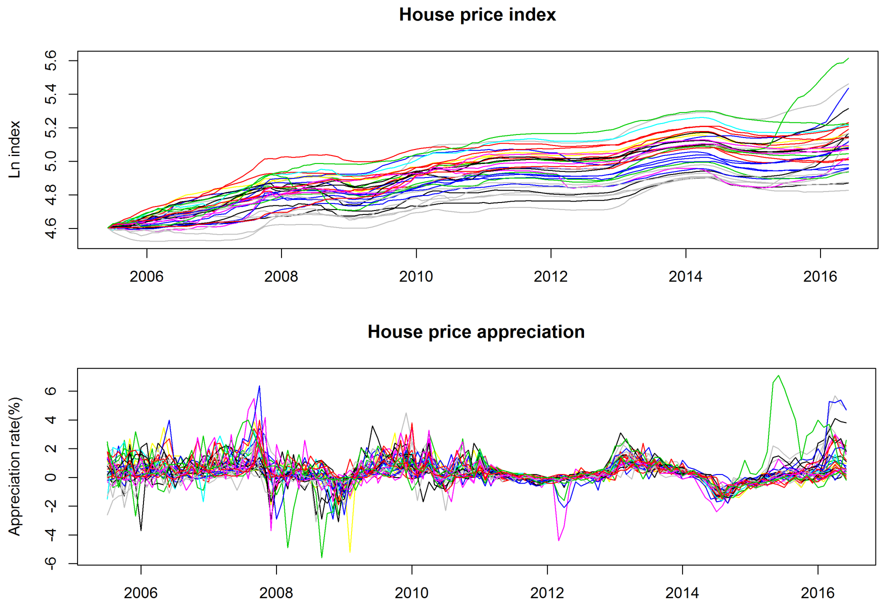


FIGURE 4.1 House price index (ln transformation) and monthly house price appreciation of 34 cities

§ 4.4.2 Stylized facts

The upper panel of Figure 4.1 plots the house price indexes of 34 cities. While a common upward trend can be easily noticed throughout the whole sample period, most of the cities experienced two or three episodes of rising and falling prices. The first common episode of price decreases occurs in approximately 2008-2009, right after the global financial crisis. However, house prices bounced back very quickly and then entered a relatively stable period until 2013. After a national upward trend started in 2013, house prices dropped again in 2014. Recently, particularly after the second half of 2015, house prices in some cities recovered with tremendous price increases. Although a national trend in house prices is noticeable, cities differ from each other in terms of their house price trajectories. For example, some cities obviously have higher growth rates than others.

The lower panel of Figure 4.1 depicts the house price appreciation rates. The house price growth rates are quite volatile in the first half of the sample period and in the most recent period after 2014, while during the period 2011-2013, house price dynamics are relatively stable. Using the difference of the logarithmic house price index (e.g., $\log(Index_{2016}) - \log(Index_{2016})$), I calculate the total house price appreciation for the sample period, as well as for two subsample periods: July 2005-December 2010 and January 2011-June 2016; the results are shown in Figure 4.2. Shenzhen enjoys the largest price appreciation at 100.80%, followed by Beijing, Xiamen and Guangzhou (all of which are eastern cities). Kunming, Taiyuan and Hohhot (either central or western cities) have the lowest price appreciation (below 30%) during the last decade[6]. It seems that geographical proximity is a meaningful way to divide the national housing market. For the vast majority of cities, especially those that have lower house price appreciation during the sample period, housing returns are mainly accumulated during the first 5 to 6 years. A few of the most developed cities, such as Shenzhen, Guangzhou, Shanghai and Nanjing, are the exceptions; their price appreciation in the second half of the period overwhelmingly surpasses their price growth in the first period. The two distinct patterns of house price dynamics also indicate the divergence of interurban housing markets to some extent.

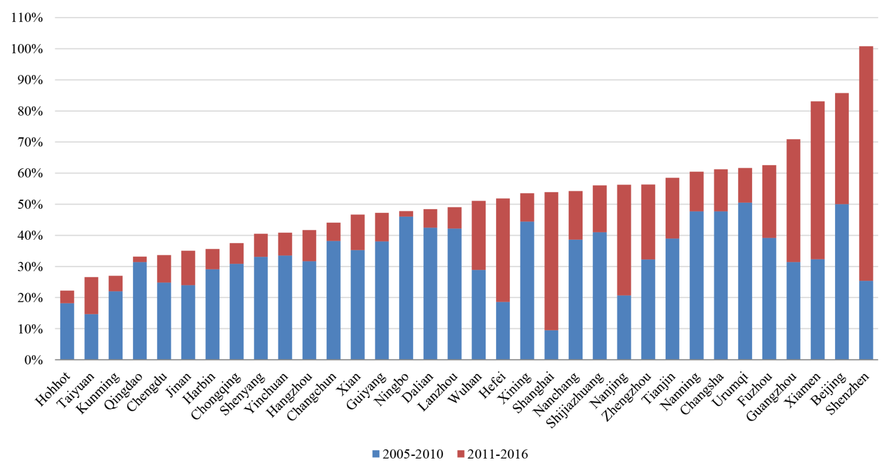


FIGURE 4.2 House price appreciation for 34 cities

§ 4.5 Results

§ 4.5.1 *JKL* divergence vs Euclidean distance

6 The "70 Cities Index" has been largely criticized for underestimating price growth. See, for example, Wu et al. (2014). However, this is the only accessible public house price index that can provide consistent information for a large number of cities over a relatively long time period.

To ensure that the Kullback-Leibler divergence accurately measures the difference between housing markets, one has to first choose the appropriate order p for the $AR(p)$ process in Equation (3). The selection of orders is based on both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). With setting the maximum lag order to 12. The BIC prefers no more than 3 lags for all of the house price appreciation series. The AIC criterion presents a similar pattern, but in quite a few cities more than 3 lags have been chosen. While both AIC and BIC suggest several choices of optimal order, one should choose a sufficiently large order to remove the serial correlation in residuals. In this regard, $AR(3)$ specification performs quite well for the vast majority of the house price appreciation series. Thus, the $AR(3)$ process is chosen as the data generating process[7].

TABLE 4.1 Descriptive statistics of *JKL* divergence and Euclidean distance

	Min	1st quartile	Median	3rd quartile	Max	Mean	Standard deviation	Coefficient of variance
JKL	25.755	111.310	167.108	276.226	1502.533	221.542	173.214	0.783
Euclidean	4.798	8.773	10.301	12.147	21.714	10.792	3.049	0.283

Notes: There are 561 city pairs in total.

Aside from the *JKL* divergence matrix, the Euclidean distance matrix is also calculated and will serve as the benchmark in the following analysis. The descriptive statistics of *JKL* divergence and Euclidean distances are presented in Table 4.1. The average *JKL* divergence of 561 city pairs is 221.54, with a standard deviation of 173.21; for Euclidean distance, the two statistics are 10.79 and 3.05. As indicated by the coefficient of variation (CV), the *JKL* divergence (CV = 0.78) is distributed in a much more dispersed manner than the distribution of the Euclidean distance, which has a CV of only 0.28. The housing markets difference measured by *JKL* divergence is in line with the difference measured by Euclidean distance to some extent, but the two are not very close; the Pearson's correlation between the two measures is 0.64, while the Spearman's rank correlation is only 0.55. Furthermore, as seen from the quartile statistics, both the distribution of *JKL* divergence and Euclidean distance are right-skewed, but the former is skewed much more severely. Thus, in the majority of the 34 cities, house price dynamics may not be so different from each other. This can be confirmed by the multidimensional scaling (MDS) shown in Figure 4.3.

Figure 4.3 presents some similarities and disparities of the two measures. For example, both separate Shenzhen (indexed by 24) as an "outlier" and identify a few

7 I randomly selected 15 cities out of the sample and reported their *AR(3)* estimation results in the appendix. I also calculated the Kullback-Leibler divergence based on *AR(2)* and *AR(1)* specifications. The results do not differ much from the results based on *AR(3)*. The Pearson correlation between *AR(3) JKL* divergence and *AR(2) JKL* divergence is 0.998, and it is 0.992 with respect to the *AR(1) JKL* divergence.

distinctive cities relatively far away from the main city group that includes the majority of the cities, such as Nanjing(11), Hefei(14) and Xiamen(16). There are certain differences between the two measures, however. The *JKL* divergence suggests that Shanghai(10) is another "outlier", while this is not prominent in the map of Euclidean distance.

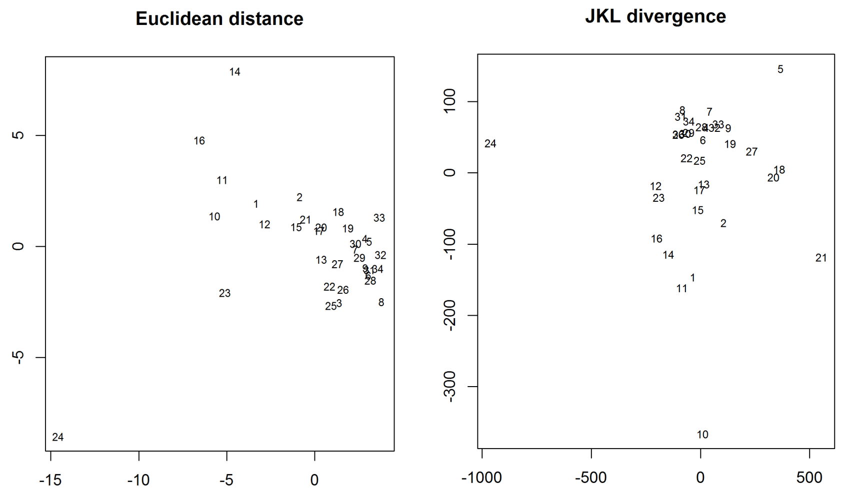


FIGURE 4.3 Multidimensional scaling of Euclidean distance and *JKL* divergence

§ 4.5.2 Classification results

Figure 4.4 plots the dendrograms generated by hierarchical agglomerative clustering method, based on both Euclidean distance and *JKL* divergence. The information hidden in the dendrogram is largely consistent with what one can learn from the Multidimensional scaling (Figure 4.3). A visual check of the dendrograms suggests that there should be one cluster containing most of the cities, a group including only a few cities, and a few individual groups. Furthermore, note that if one wants more clusters ($k > 5$ for instance), the tree of *JKL* divergence might yield better and more meaningful clusters than the Euclidean distance tree, which will produce too many individual groups composed of only one entity[8].

The elbow plot depicted in Figure A1 suggests a four-cluster solution for classifying the 34 housing markets. In addition to the elbow approach, I also report the average

8 Note that the dendrogram structure definitely relies on the *linkage* method used to calculate the dissimilarity between clusters. Thus, this inference may not hold for the dendrogram generated by other *linkage* methods, such as *complete linkage*.

Silhouette index across all clusters that have more than one member. The Silhouette index ranges from -1 to 1, and a positive larger value toward 1 indicates a better demarcation. In this sense, a two-cluster solution seems to be the best solution for both the Euclidean distance and *JKL* divergence, according to the results of Table A1. However, one should keep in mind that the decision based on the Silhouette index might be not optimal due to the appearance of individual clusters, for which the Silhouette index is not defined. Therefore, I only use the Silhouette index as a robustness check. Given that the average Silhouette value at $k = 4$ is also much larger than the values for any $k > 4$, it is believed that the four-cluster solution is a reasonable choice.

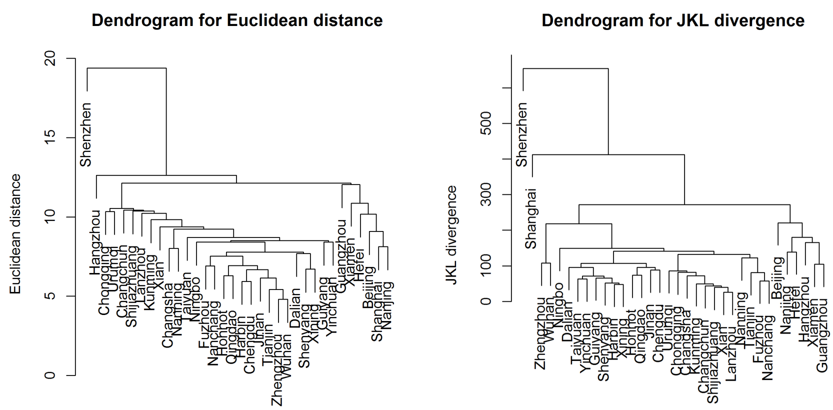


FIGURE 4.4 Dendrogram based on Euclidean distance and *JKL* divergence (average linkage)

The four clusters generated by Euclidean distance and *JKL* divergence, which are shown in Table 4.2, are almost identical with each other. The only difference is that the former separates Hangzhou(12) from Cluster 2 as an independent cluster while the latter separates Shanghai(10). In the following analysis, I mainly focus upon the classification results based on *JKL* divergence. Among the four clusters (see also Figure 4.5), Cluster 1 merges the majority (76%) of the sample cities and is composed mainly of less developed Central, Western and Northeast cities. Cluster 2 is relatively small and includes 6 cities, which are the main centres of the three most developed economic regions in Eastern China: the Pan-Yangtze River Delta (PYRD) containing Nanjing(11), Hangzhou(12) and Hefei(14); the Pan-Pearl River Delta (PPRD) including Guangzhou(23) and Xiamen(16); and the Jing-Jin-Ji Economic Region containing Beijing(1). Moreover, two distinct cities, Shenzhen(24) and Shanghai(10) belonging to the PPRD and PYRD respectively, stand out and form two individual clusters; their house price dynamics are substantially different from each other and from the remaining cities. The four-cluster solution explains the cross-city housing market structure reasonably well. The average distances between the housing markets within Cluster 1 and Cluster 2 are 136.53 and 184.02, respectively, much smaller than the sample average distance (221.54). The extremely low standard deviations of Cluster 1 and Cluster 2 also confirm such findings.

TABLE 4.2 Cluster membership and statistics based on Euclidean distance and *JKL* divergence

	Euclidean distance		*JKL* divergence		
	Membership	Average within distance	Membership	Average within distance	Average monthly growth rate (%)
Cluster 1	2,3,4,5,6,7,8,9,13,15, 17,18,19,20,21,22, 25,26,27,28,29,30, 31,32,33,34	9.10 (1.49)	2,3,4,5,6,7,8,9,13,15, 17,18,19,20,21,22, 25,26,27,28,29,30, 31,32,33,34	136.53 (68.24)	0.35 (0.09)
Cluster 2	1,10,11,14,16,23	10.70 (1.57)	1,11,12,14,16,23	184.02 (45.72)	0.50 (0.13)
Cluster 3	24		24		0.78
Cluster 4	12		10		0.41
Sample		10.79 (3.05)		221.54 (173.21)	0.39 (0.13)

Notes: For the cluster membership, only the ID of the city is presented. Readers can refer to Table B1 for the names of the cities. The numbers in parentheses are the standard deviations. The average within distance is the mean of all city-pair distances within the same cluster. After obtaining the mean monthly growth rate for each city throughout the sample period, the average monthly growth rate reported in the table is the average of mean growth rate across cities in the same cluster.

In a broad sense, the classification results can reduce to a two-cluster solution. One is Cluster 1, which contains mainly slow-growing markets with an average monthly growth rate of 0.35%. The other combines Clusters 2, 3 and 4, which contain markets with relatively higher appreciation rates (see Table 4.2) and are considered to be "red-hot" markets. There is a much higher degree of heterogeneity within these red-hot markets, however. Compared to the difference in house price appreciation rates, the most striking divergence between the two broad groups is that, referring back to Figure 4.2, the red-hot markets experienced tremendous house price appreciation in the second half of the period, whereas house price growth during the first half of the period makes the most important contribution for the cities in Cluster 1. One can also easily identify a spatial pattern in the house price dynamics of cities with red-hot markets located in Eastern China, as well as others that are mainly in the remaining "peripheral" regions.

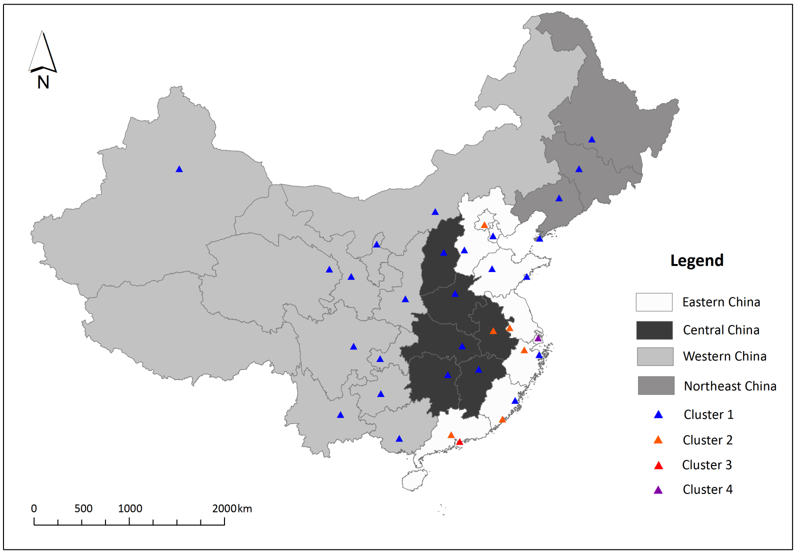


FIGURE 4.5 Spatial distribution of sample cities and their membership based on *JKL* divergence

§ 4.5.3 Do structural changes matter?

The classification analysis of the whole period relies on the premise that the housing market structure of a city does not change. In other words, the relationship between any two cities' housing markets is stable throughout the sample period. This assumption may not hold given the evolving conditions of the nascent Chinese housing market. To explore whether structural changes affect previous classification results, I split the sample into subsamples and perform the classification analysis on each subsample. If there is no structural change, the classification results will be highly consistent among the subsamples, as well as with the clusters presented in Table 4.2. Otherwise, some quite different cluster demarcations will be observed.

The house price appreciation trajectories depicted in Figure 4.1 indicate that house price changes among sample cities are quite volatile during the first half of the period and the recent period. Thus, I divide the sample period into three sub-periods: 2005–2010 (66 observations), 2011–2013 (36 observations) and 2014–2016 (30 observations). For each sub-period, the *JKL* divergence is calculated based on an *AR(3)* specification. The average *JKL* distances between cities as well as the associated standard deviations are reported in the last row of Table 4.3. The average distance between housing markets in the period 2014–2016 (196.79) is much larger than the average distance in the first (92.14) and second periods (71.23), indicating a remarkable divergence of housing markets in the recent period. While the overall housing market difference throughout the whole sample is approximately 220, as measured by *JKL* divergence, it can be concluded that the highly divergent market conditions in the third period account for the most important component.

The hierarchical clustering method is then applied to the three subsamples, providing more details about the structural evolution of housing markets. To obtain comparable clusters across the different periods, the dendrograms of these three periods should be cut at a common "height" (divergence threshold). To do this, I first determine the appropriate number of clusters (k) for the 2014–2016 sample, which should be 5 according to Figure A2. This demarcation requires a minimum height (h) approximately 94.87. I then cut the dendrograms of the first and second period at the same height and obtain 3 clusters for both of the periods. According to the elbow plots in Figure A2, the 3-cluster demarcation is not the optimal solution for the first two periods. But only in this way can the demarcation solutions of the three sub-periods be directly compared with each other.

TABLE 4.3 Cluster membership and statistics for different sample periods

	2005-2010		2011-2013		2014-2016		
	$k = 3, h = 94.87$		$k = 3, h = 94.87$		$k = 5, h = 94.87$		
	Membership	Average within distance	Membership	Average within distance	Membership	Average within distance	Average monthly growth rate (%)
Cluster 1	1,2,3,4,5,6, 7,8,9,11,12, 13,14,15,16, 17,18,19,22, 23,24,25,26, 27,28,29,30, 31,32,33,34	72.16 (45.72)	2,3,4,5,6,7, 8,9,11,14, 15,16,17,18, 20,21,22,25, 26,27,28,29, 30,31,32,33, 34	41.66 (24.42)	3,4,5,6,7,8, 9,19,22,25, 26,27,28,29, 30,31,32,33, 34	29.86 (11.96)	-0.12 (0.09)
Cluster 2	20,21	36.37 (20.21)	12,13,19	52.86 (30.65)	2,13,15,17, 18,20,21	52.71 (26.38)	0.21 (0.12)
Cluster 3	10		1,10,23,24	34.67 (11.24)	1,10,12,23	46.41 (24.12)	0.56 (0.27)
Cluster 4					11,14,16	47.65 (16.20)	0.89 (0.16)
Cluster 5					24		1.82
Sample		92.14 (73.90)		71.23 (66.48)		196.79 (275.54)	0.17 (0.46)

Notes: The same notes as Table 4.2.

In both the 2005–2010 period and the 2011–2013 period, one observes highly integrated, homogeneous cross-city housing markets. Most of the cities are classified into the same cluster, with a few exceptions, Shanghai(10) for instance. The most remarkable structural change between these two periods could be that a few developed eastern cities, such as Beijing(1), Shenzhen(24), Guangzhou(23) and Hangzhou(12), began to deviate from the main group during the 2011–2013 period.

When turning to the 2014–2016 period, the interurban housing market diverges much more. House price growth in Shenzhen(24), where the average monthly growth rate is 1.82%, obviously stands out from the other cities. Aside from the red-hot markets that are already separated in the 2011–2013 period (1, 12 and 23 for example), a few more rapid-growth markets with monthly growth rates of 0.89% (Cluster 4), such as Hefei(11), Fuzhou(14) and Xiamen(16), are also isolated from the main group. Furthermore, there is another cluster (Cluster 2) that deviates from the main group but not by much; this group has a relatively low average monthly growth rate (0.21%) and mainly contains some lower-tier centres in Eastern and Central China, such as Ningbo(13), Fuzhou(15), Nanchang(16) and Wuhan(21). Note that a parsimonious three-cluster solution for the 2014–2016 housing markets, which merges Cluster 1 with Cluster 2, and fuses Cluster 3 with Cluster 4, is highly consistent with the classification results based on the whole period (Table 4.2). By grouping Cluster 1 and Cluster 2 in Table 4.3, one obtains precisely Cluster 1 in Table 4.2. Similarly, Cluster 2 in Table 4.2 is almost identical to the combination of Clusters 3 and 4 in Table 4.3.

To summarize the findings, the Chinese interurban housing market in the early periods (2005–2013) can be considered a homogeneous market; only a few markets have distinctive dynamic patterns of house prices. In the recent period (after 2014), the interurban housing market begins to diverge; not only do the markets of the most important centres of the three main economic regions stand out remarkably, the markets of some lower-tier centres are also isolated. Given this sudden structural change, it is not surprising to see that the clustering pattern of the 2014–2016 period almost determines the classification results based on the whole period.

§ 4.5.4 The effectiveness of geographical and economic divisions

The classification results suggest that geography seems to play a role in the homogeneous clustering of the city-level housing markets. In this section, I will test the effectiveness of the traditional geographical demarcation in describing the interurban housing market structure. This classification scheme was introduced by NBSC in 2011 and divides all of China into four regions: Eastern, Central, Western and Northeast China (see Figure 4.5). In addition, I also examine whether the city-tier system of demarcation based on economic factors can produce meaningful groups of housing markets. There are presently several versions of the city-tier division system available, published by different research institutes or real estate agencies. In this paper I refer to the "*China60*" city-tier system compiled by Jones Lang LaSalle (JLL) (JLL 2015), which rates 60 Chinese cities according to a range of economic indicators and classifies these cities into seven different tiers. I make a slight adjustment to the original demarcation of "*China60*" and reduce the seven-tier system to a four city-tier system. A detailed introduction to the geographical and economic demarcation system can be found in Appendix B.

The McClain and Rao (1975) index (MR index for short), which is defined as the ratio of average within-cluster distance to average between-cluster distance, is employed to assess the power of the two demarcation systems in explaining the interurban housing market structure. The smaller the value, the better the classification scheme performs. To evaluate whether the classification systems really make sense and are significantly different from some random allocations, I simulate the distribution of the MR index under the null hypothesis of random division. Specifically, I generate B random partition samples by randomly assigning the sample cities into clusters with the same sizes as those of the "real" partition and calculate the MR index of each sample. Together with the MR index calculated from the "real" partition, one obtains $B + 1$ values in total, and the p-value is the fraction of the measures that are smaller than and equal to the real MR index. Note that the analysis of this section is based on the JKL divergence matrix of the whole sample.

TABLE 4.4 MR indexes of different demarcation schemes

	NBSC 4-region classification	Adjusted NBSC 2-region classification	JLL 4-tier classification	4-cluster classification in Table 4.2
MR index	0.9760	0.7321	0.7021	0.3971
p-value	0.3807	0.0014	0.0006	0.0000

Notes: The p-value is calculated based on 9999 random permutation samples.

The MR index of the four-region NBSC geographical demarcation, shown in Table 4.4, is 0.9760 with the significance p-value of 0.3807, suggesting that this widely used demarcation scheme is not significantly different from random demarcation schemes and thus cannot explain the interurban housing market structure. However, according to the clustering results of Figure 4.5, the cities in Eastern China indeed behave in a distinct fashion. I therefore test a broader geographical classification scheme with Eastern China in one cluster and the remaining regions in the other. As expected, the adjusted NBSC two-region classification system, which has a lower MR value, is more powerful in explaining the interurban housing market structure and rejects the null hypothesis of random partition at the 1% significance level.

Compared to the larger MR values of geographical demarcation schemes, the JLL 4-tier classification scheme based on socio-economic conditions clearly outperforms the division system that is purely based on geography, though there is some correlation between economic development and localities. Although the economic division and the broad East – Remainder geographical division can make sense in classifying the interurban housing markets, they still produce a large amount of "noise" compared to the clustering solutions reported in Table 4.2. This is mainly because the developed cities' housing markets, such as the markets within Eastern China and Tier 1 cities, diverge more among each other than the undeveloped cities do.

§ 4.5.5 Discussion

House price dynamics are driven by shifts in demand factors, such as income, population and credit market conditions, and by changes in supply factors, such as construction costs and regulation constraints; they are sometimes even driven by behavioural factors, such as spillovers. One weakness of the cluster analysis in this paper is that the markets are clustered solely based on the time-series behaviour of house price changes, but the underlying factors that drive the house price behaviour and clustering process are not clear. However, from the existing literature, one can still make some inferences about the driving mechanisms behind these trajectories. A recent study (Fang et al. 2016) of Chinese house price appreciation reveals that the house price appreciation of first-tier cities (Beijing, Shanghai, Guangzhou and Shenzhen) nearly doubles the increases in disposable income, whereas the price growth of second-tier (i.e., most of the remaining cities in this paper) and third-tier cities is strongly in accordance with income growth. Meanwhile, they also report that the urban population living in the four first-tier cities has increased by 46%, while the population of the second-tier cities has increased by 18% and that of third-tier cities has almost remained stable. On the supply side, Li and Chand (2013) state that supply factors, including construction costs and land prices, play a role in determining the house prices of developed provinces in Eastern China. Thus, it can be tentatively concluded that the slow-growing markets in Cluster 1 (Table 4.2) are mainly driven by income growth, while the red-hot markets in Eastern China (Clusters 2, 3 and 4 in Table 4.2) are more influenced by population growth and housing supply. Moreover, another potential weakness of this paper lies in the fact that the cities' house price dynamics are assumed to be independent from each other after accounting for the time-series structure. This is probably not true given the spillover effect between housing markets that is evidenced in a large amount of literature (e.g.,Holly et al. 2011).

The clustering method in this paper attempts to group the markets that have similar growth trajectories, no matter what the underlying structure is. This logic is different from the clustering logic of van Dijk et al. (2011), who tried to group cities that can be described by a common house price model ($\beta_i = \beta_j$). Furthermore, although the markets within the same cluster tend to move synchronously, this cannot guarantee the property of market cointegration as studied in, for example, MacDonald and Taylor (1993).

The classification results show that diversifying the housing portfolio across space and cities indeed brings some benefits, especially in the recent period when the interurban market has been more fragmented. Beyond this, the results can also benefit policy makers in both the central and local governments. Because the Chinese interurban housing market has undergone significant structural changes in the recent period after 2014, national policy guidance – monetary policy, for instance – would no longer be appropriate for the overall divergent market. The central government has already called on the local governments of those red-hot markets, such as Shanghai, Shenzhen and Nanjing, to play a more active role in tailoring local-oriented policy. Aside from these red-hot markets, the markets of some lower-tier centres (Cluster 2 in Table 4.3), such as Ningbo and Wuhan, also need special attention. Of course, considering the changing circumstances of Chinese housing markets, structural changes might be expected in the future, and policy guidance needs to be updated accordingly. However, the housing market clusters presented in this paper can still provide useful policy guidance in the near future. For housing researchers who focus on aggregation levels above the city, grouping markets based on the city-tier system is a better choice than the geographical four-region division. However, a broad geographical demarcation, which emphasises the role of Eastern China, can still make sense to some extent.

§ 4.6 Conclusion

This paper is a response to the lively debate about interurban housing market divergence in China. I investigated the clustering pattern of 34 cities' housing markets according to their house price appreciation trajectories over the period of July 2005 to June 2016. The hierarchical agglomerative clustering method was employed to perform the partition. In particular, I adopted a distribution-based statistic – Kullback-Leibler (*KL*) divergence – to measure the dissimilarity between markets, which can enable inferences about future market discrepancy. Specifically, the KL divergence was calculated based on the assumption that the house price appreciation series is generated by an *AR(3)* process.

As a response to the debate, I found that the interurban housing market is indeed fragmented and can be broadly partitioned into two clusters. One cluster, which is very large, is mainly composed of markets with low house price growth that are mostly located in Central, Western and Northeast China. The other cluster is a combination of the most important centres in Eastern China that have flourishing housing markets. This cluster has a higher degree of heterogeneity and can be further partitioned into smaller groups. However, I noted that the market divergence seems to be a new phenomenon appearing after 2014; before that year, the interurban housing market in China was relatively homogeneous. The classification results of the recent period (2014–2016) also suggest that not only the red-hot markets of major regional centres but also the markets of some lower-tier centres require special attention.

This paper also tested the usefulness of the widely used geographical demarcation and city-tier system in describing the interurban housing market structure. The four-region geographical demarcation scheme created by the National Bureau of Statistics of China plays no role in terms of explaining housing market structure, but a super-region demarcation scheme, namely 'Eastern China – Remaining periphery', makes sense to some extent. The city-tier system based on socio-economic conditions is a superior system in applied housing market analysis. However, it still produces considerable noise due to the large discrepancies among the cities within higher tiers.

Although this analysis offers no conclusions regarding the driving mechanism underlying the clustering pattern, it is inferred from the literature that the slow-growing cluster is likely to be driven by income increases, while the red-hot markets are probably driven by supply factors and population growth. The classification results also aid housing portfolio managers in diversifying their investments, policy makers in tailoring proper policies for specific markets, and regional researchers in aggregating city markets properly. However, one should bear in mind that structural changes in the future may influence the robustness of the clusters.

References

Abraham, J. M., Goetzmann, W. N., & Wachter, S. M. (1994). Homogeneous groupings of metropolitan housing markets. Journal of Housing Economics, 3(3), 186-206.

Bengtsson, T., & Cavanaugh, J. E. (2008). State-space discrimination and clustering of atmospheric time series data based on Kullback information measures. Environmetrics, 19(2), 103-121.

Bhattacharjee, A. Castro, E., Maiti, T., & Marques, J. (2016). Endogenous spatial regression and delineation of submarkets: A new framework with application to housing markets. Journal of Applied Econometrics, 31(1), 32-57.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. Journal of Statistical Software, 61(6), 1-36.

Dong, J., Li, X., Li, W., & Dong, Z. (2015). Segmentation of Chinese urban real estate market: A demand-supply distribution perspective. Annals of Data Science, 2(4), 453-469.

Fang, H., Gu, Q., Xiong, W., & Zhou, L. (2016). Demystifying the Chinese housing boom. In M. Eichenbaum, & J. Parker (Eds.), NBER Macroeconomics Annual

2015 (Vol. 30, pp. 105-166). Chicago: University of Chicago Press.
Goetzmann, W. N., & Wachter, S. M. (1995). Clustering methods for real estate portfolios. Real Estate Economics, 23(3), 271-310.
Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. Journal of Housing Economics, 7(2), 121-143.
Goodman, A. C., & Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets. Real Estate Economics, 35(2), 209-232.
Guo, K., Wang, J., Shi, G., & Cao, X. (2012). Cluster analysis on city real estate market of China: Based on a new integrated method for time series clustering. Procedia Computer Science, 9, 1299-1305.
Hamelink, F., Hoesli, M., Lizieri, C., & MacGregor, B. D. (2000). Homogeneous commercial property market groupings and portfolio construction in the United Kingdom. Urban Studies, 32(2), 323-344.
Hepșen, A., & Vatansever, M. (2012). Using hierarchical clustering algorithms for Turkish residential market. International Journal of Economics and Finance, 4(1), 138-150.
Hoesli, M., Lizieri, C., & MacGregor, B. (1997). The spatial dimensions of the investment performance of UK commercial property. Urban Studies, 34(9), 1475-1494.
Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of Urban Economics, 69(1), 2-23.
Jackson, C. (2002). Classifying local retail property markets on the basis of rental growth rates. Urban Studies, 39(8), 1417-1438.
JLL (2015). China60: From fast growth to smart growth. Report.
Kakizawa, Y., Shumway, R. H., & Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. Journal of American Statistical Association, 93(441), 328-340.
Kullback, S. (1968). Information theory and statistics. New York: Dover Publications.
Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22(1), 79-86.
Li, Q., & Chand, S. (2013). House prices and market fundamentals in urban China. Habitat International, 40, 148-153.
MacDonald, R., & Taylor, M. P. (1993). Regional house prices in Britain: long-run relationships and short-run dynamics. Scottish Journal of Political Economy, 40(1), 43-55.
Malizia, E. E., & Simons, R. A. (1991). Comparing regional classifications for real estate portfolio diversification. Journal of Real Estate Research, 6(1), 53-77.
McClain, J. O., & Rao, V. R. (1975). CLUSTISZ: A program to test for the quality of clustering of a set of objects. Journal of Marketing Research, 12(4), 456-460.
Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65.

van Dijk, B., Franses, P. H., Paap, R., & van Dijk, D. (2011). Modelling regional house prices. Applied Economics, 43(17), 2097-2110.
Watkins, C. A. (2001). The definition and identification of housing submarkets. Environment and Planning A, 33(12), 2235-2253.
Wu, J., Deng, Y., & Liu, H. (2014). House price index construction in the nascent housing market: The case of China. The Journal of Real Estate Finance and Economics, 48(3), 522-545.

Appendices

Appendix A.Hierarchical agglomerative clustering

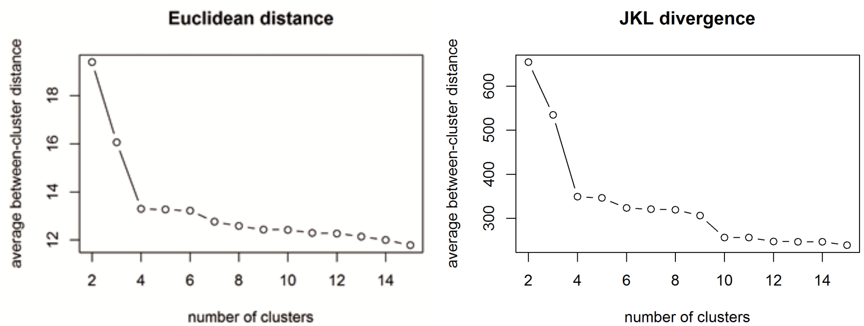


FIGURE A1 The elbow plot for Euclidean distance and *JKL* divergence

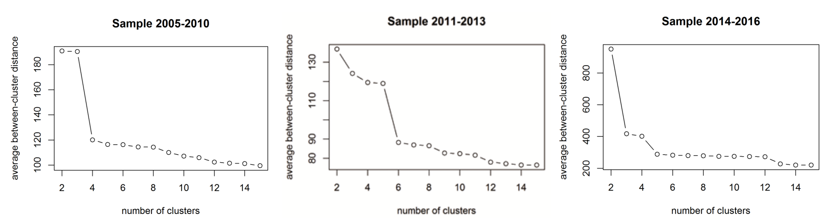


FIGURE A2 The elbow plot for first and second half of the sample period (*JKL* divergence)

TABLE A1 Average Silhouette index for different clusters

	2 clusters	3 clusters	4 clusters	5 clusters	6 clusters	7 clusters	8 clusters	9 clusters
Euclidean	0.4670	0.1966	0.2063	0.2033	0.1875	0.1259	0.1149	0.1144
JKL	0.6287	0.4963	0.4239	0.3469	0.2709	0.2424	0.2093	0.1528

Notes: Since Silhouette statistic is not well defined for the individual groups that have only one member, the average Silhouette indexes are calculated based on the groups that have more than one member.

Appendix B.The geographical and economic division

To better reflect the regional structure of socio-economic conditions, the National Bureau Statistics of China (NBSC) officially divided 32 municipalities, provinces and autonomous regions (excluding Hong Kong and Macao) into four economic regions according to geographical proximity, namely Eastern, Central, Western and Northeast China (see also Figure 4.5) . The Panel A of Table B1 lists the four regions and the cities belonging to them.

The Jones Lang LaSalle (JLL) launched their China Cities Research programme in 2006 and has sequentially published "China30", "China40" and "China50" before the release of "China60" in 2015. The "China60" assesses the relative position of each of 60 cities based on the analysis of a range of economic, business and property indicators and finally allocate the cities to one of the seven tiers: Alpha cities, Tier 1, Tier 1.5, Tier 2, Tier 3 Growth, Tier 3 Emerging and Tier 3 Early Adopter. In this paper, we reduce the seven tiers into four tiers by merging Alpha cities into Tier 1 and by combining Tier 3 Growth, Emerging and Early Adopter into Tier 3 cities. Then the 34 cities, which are all included in the "China60", are assigned into the relative tiers accordingly (see Table B1).

TABLE B1 The list of cities grouped by geographical divisions and by city-tiers

Panel A: Cities grouped by geographical divisions	
Eastern	Beijing (1), Tianjin (2), Shijiazhuang (3), Shanghai (10), Nanjing (11), Hangzhou (12), Ningbo (13), Fuzhou (15), Xiamen (16), Jinan (18), Qingdao (19), Guangzhou (23), Shenzhen (24)
Central	Taiyuan (4), Hefei (14), Nanchang (17), Zhengzhou (20), Wuhan (21), Changsha (22)
Western	Hohhot (5), Nanning (25), Chongqing (26), Chengdu (27), Guiyang (28), Kunming (29), Xian (30), Lanzhou (31), Xining (32), Yinchuan (33), Urumqi (34)
Northeast	Shenyang (6), Dalian (7), Changchun (8), Harbin (9)
Panel B: Cities grouped by tiers	
Tier 1	Beijing (1), Shanghai (10), Guangzhou (23), Shenzhen (24)
Tier 1.5	Tianjin (2), Shenyang (6), Nanjing (11), Hangzhou (12), Wuhan (21), Chongqing (26), Chengdu (27), Xian (30)
Tier 2	Dalian (7), Ningbo (13), Xiamen (16), Jinan (18), Qingdao (19), Zhengzhou (20), Changsha (22)
Tier 3	Shijiazhuang (3), Taiyuan (4), Hohhot (5), Changchun (8), Harbin (9), Hefei (11), Fuzhou (15), Nanchang (17), Nanning (25), Guiyang (28), Kunming (29), Lanzhou (31), Xining (32), Yinchuan (33), Urumqi (34)

Appendix C.The AR(3) estimation

TABLE C1 The AR(3) estimation of selected cities

	Lag 1	Lag 2	Lag 3	Intercept	R-squared	BG test
Tianjin	0.6423***	-0.0355	0.2180**	0.0922*	0.504	8.30**
	(0.088)	(0.105)	(0.093)	(0.052)		
Taiyuan	0.2663***	-0.0502	0.3104***	0.0847	0.179	27.64***
	(0.084)	(0.087)	(0.081)	(0.055)		
Shenyang	0.4988***	0.0611	0.1395	0.0922	0.380	3.33
	(0.089)	(0.098)	(0.087)	(0.057)		
Changchun	0.4762***	0.0576	-0.0414	0.1743**	0.249	26.12***
	(0.089)	(0.097)	(0.085)	(0.075)		
Shanghai	1.0636***	-0.2565*	0.0715	0.0772*	0.801	1.65
	(0.090)	(0.130)	(0.090)	(0.040)		
Ningbo	0.7572***	-0.0572	0.0777	0.0768	0.580	3.00
	(0.089)	(0.111)	(0.089)	(0.051)		
Hefei	0.7352***	0.2850***	-0.1992**	0.0982	0.640	13.24***
	(0.087)	(0.102)	(0.097)	(0.065)		
Nanchang	0.5306***	0.0566	0.1114	0.1336**	0.389	1.01
	(0.089)	(0.100)	(0.089)	(0.061)		
Zhengzhou	0.5888***	0.0147	0.1432	0.1086**	0.447	4.30
	(0.089)	(0.103)	(0.090)	(0.047)		
Changsha	0.5207***	0.2283**	-0.1019	0.1517**	0.418	6.32
	(0.088)	(0.095)	(0.082)	(0.070)		
Guangzhou	0.5805***	0.2104**	-0.0348	0.1422*	0.512	3.81
	(0.089)	(0.102)	(0.091)	(0.080)		
Chengdu	0.5154***	0.2242**	0.0024	0.0686*	0.471	2.07
	(0.089)	(0.096)	(0.088)	(0.041)		
Guiyang	0.5804***	-0.0476	0.0299	0.1588***	0.318	5.89
	(0.089)	(0.103)	(0.089)	(0.061)		
Xian	0.0792	0.2755***	0.2213**	0.1492*	0.179	2.79
	(0.087)	(0.084)	(0.086)	(0.078)		
Yinchuan	0.1882**	0.4658***	0.0623	0.0922*	0.369	1.38
	(0.090)	(0.081)	(0.090)	(0.054)		

Notes: The standard errors are reported in the parentheses. The BG test represents the Breusch-Godfrey test, following the χ_3^2 distribution under the null hypothesis of no serial correlation of order up to 3. ***, ** and * indicate 1%, 5%, 10% significance level, respectively.

5 Spatial interrelations of Chinese housing markets: Spatial causality, convergence and diffusion

Published as: Gong, Y., Hu, J., Boelhouwer, P., 2016. Spatial interrelations of Chinese housing markets: Spatial causality, convergence and diffusion. Regional Science and Urban Economics, 59:103-117.

Abstract: This paper comprehensively tests the spatial interrelationships of 10 housing markets in the Pan-Pearl River Delta (Pan-PRD) in China, including the properties of spatial causality, convergence and diffusion patterns. The pairwise Toda-Yamamoto Granger causality tests suggest widely existing leading-lag relationships between housing markets; a unidirectional causal flow from the eastern-central area to western China can be tentatively confirmed. However, there is a lack of sufficient evidence supporting pairwise long-run *cointegration* and *convergence*, indicating a diverged interurban housing market in the Pan-PRD. In the short run, the spatial-temporal diffusion model manifests the importance of the spillover effect from neighbouring cities in predicting one city's house price changes. Furthermore, the generalized impulse response functions (GIRFs) clearly depict the transmission pattern of shocks to one chosen city. The diffusion pattern is characterized by the fact that the shocks first spread to nearby cities with cities further away taking a longer time to respond.

Keywords: Spatial causality, long-run convergence, ripple effect, diffusion, house prices, China

JEL: C31, C33, R21, R31

§ 5.1 Introduction

After the financial crisis in 2008, many governments have been attempting to stimulate depressed housing markets through policy interventions. However, whether the interventions can work as expected relies heavily on our understanding of housing markets. To provide deeper insights into house price behaviour, many scholars advocate an investigation into a series of interrelated regional markets rather than a single national market (Meen 1996; Yunus and Swanson 2013). Indeed, the structure of regional housing markets is likely to vary significantly across space given the huge differences in local amenities, economic conditions, and regulation constraints, among other considerations. Simply aggregating a bundle of local housing markets into a national unit could lead to severe misunderstanding, particularly in a country in which the regional housing markets are highly differentiated.

Regional housing markets are neither identical nor independent. A large volume of literature has provided evidence supporting the interrelations of regional housing markets (e.g.,Giussani and Hadjimatheou 1991; Pollakowski and Ray 1997; Holly et al. 2011). Specifically, researchers find that house price changes in an area depend significantly on what occurred in other areas' housing markets. Among the various interrelations of local housing markets, long-run *integration*, which describes a situation in which local house prices maintain an equilibrium relation in the long-run, has long been a concern because of its policy implications. If local housing markets are highly integrated, a unified nation-wide housing policy will be sufficient; otherwise a basket of diversified, locally-oriented policies are necessary. Another parallel research agenda has concentrated on the so-called ripple effect whereby house price shocks to an area will gradually diffuse to other areas, with areas further away being slower to respond to the shocks. Statistical evidence for long-run integration and a ripple effect of regional house prices has been found, for example, in the UK markets by Alexander and Barrow (1994), Meen (1996), Cook (2003) and Holmes and Grimes (2008), although certain studies cast doubt on it (Drake 1995; Ashworth and Park 1997; Abbott and Vita 2013) [1].

While a large amount of empirical evidence for long-run and short-run patterns of house prices is already available, the underlying behavioural mechanisms are not yet clear. Meen (1999) proposed five possible explanations for the patterns in the UK market: migration, equity transfer, spatial patterns in the determinants of house prices, spatial arbitrage, and coefficient heterogeneity. Although the transitional economy of China makes its housing market significantly different from the UK market, we have observed the presence of such factors which can cause a certain pattern of house prices. For example, the loosening of *Hukou* restrictions has largely accelerated labour mobility between areas and consequently induced the equity transfer among regions [2]. The information transmission pattern, namely, that housing market information usually flows from "superstar" cities to "normal" cities (Wu and Deng 2015), raises the chance of spatial arbitrage. Hanink et al.(2012) showed significant

---

1 It should be noted that the mixture of the evidence is partly due to the different understanding of the term '*integration*' ('*convergence*') and '*ripple effect*'. We will discuss this in the literature review.

2 The "*Hukou*" (household registration) system in China was initially designed as a mechanism of monitoring population movements in early 1950s. Afterwards, it became a strong tool to restrain the rural-urban migration and the labour mobility between cities. Since 1980s, the power of "*Hukou*" system has been weakened through a series of reforms, but it remains in place to this day.

coefficient heterogeneity among Chinese county-level housing markets. However, whether these factors can result in long-run integration of regional housing markets remains unclear. From the national perspective, the current migration pattern, flowing from less developed Western China to developed Eastern China, is likely to induce the divergence of housing markets between the East and West rather than market integration. However, local housing markets within the East or the West have a larger chance to be integrated. Spatial patterns of house price determinants also provide us with a confusing hint regarding the long-run integration of local house prices. Province-level real GDP per capita, used as a proxy for income, is found to be convergent in Eastern and Western China, but not in Central and North-eastern China (Su and Chang 2013).

Given such arguments, the spatial interrelations of Chinese local housing markets appear to be an interesting question to answer. Indeed, much effort has been dedicated to this issue in recent years. For example, Wang et al. (2008) examined the long-run and short-run properties of house prices based on cities within 5 sub-national areas during the period 1997Q4 – 2007Q1. Huang et al. (2010b) conducted research on nine major Chinese cities during a similar time span (1999Q1 – 2008Q3), and Li and Li (2011) on nine cities in Pearl River Delta for the period 2001Q1 – 2010Q4. In general, these studies confirmed the spatial interrelations of housing markets among different cities and they found long-run equilibrium relationships between these markets.

Using a new data set of house price indexes for 10 cities within the Pan-Pearl River Delta (Pan-PRD) spanning from June 2005 to May 2015, this paper comprehensively investigates the spatial-temporal interrelations between city-level housing markets. Specifically, we are particularly interested in the following three questions. First, is there any 'spatial causality' in the interurban housing markets so that the historical house price information in one market can be used to predict the current house prices in other markets? Second, are the house price indexes of ten cities converged (integrated) or segmented in the long-run? Third, is there a distinct house price diffusion pattern so that shocks to one particular market can propagate to other markets gradually?

To our knowledge, this is the first paper that focuses on the spatial interrelations of housing markets in the Pan-PRD area in South China. This area is of great interest given its economic importance and policy implications. Since the reform and opening-up started at 1978, the cities of Pearl River Delta (PRD) in Guangdong province, such as Shenzhen and Guangzhou, have been rapidly developing due to their advantageous location and access to Hong Kong and Macao. Meanwhile, most Central and Western provinces, which provide a large amount of cheap labour for Guangdong and thus can be seen as the hinterland, still struggled with low economic growth. To narrow the gap of development between these areas, "Pan-Pearl River Delta Regional

Co-operation Framework Agreement" was signed by 11 relevant governments in 2004. This initiative aims to remove the trade barriers between cites, promote the economic linkages and interaction between eastern, central and western China, and finally achieve the economic integration of this area. The results of this paper shed light on the extent to which the cities in this cooperation framework are linked with each other and the degree to which their markets have been integrated. Thus, this paper might have great implications for policy makers.

Our results suggest widely existing pairwise leading-lag relations among the housing markets under investigation. That is, a city's housing market is generally interrelated with the markets of other cities. However, in contrast to most of the previous studies that support the long-run integration of interurban housing markets, we find rare evidence for pairwise cointegration relationships between cities in the Pan-PRD, and even less evidence for convergence. This discrepancy is probably due to the fact that we focus on a large and heterogeneous area, while previous studies are confined to a relatively small and homogeneous area or to the Chinese cities that have similar socio-economic conditions. Furthermore, a distinct house price diffusion pattern is confirmed; the generalized impulse response function (GIRF) shows that shocks to a city first spread to the nearby cities and then gradually to the distant cities.

The remainder of this paper is organised as follows. Section 2 briefly reviews the related literature, followed by the data description in section 3. The empirical examination of the leading-lag relationships, long-run integration and house price dynamic pattern are shown in section 4, 5 and 6, respectively. Finally, section 7 concludes the findings and derives certain implications.

§ 5.2 Previous literature

The focus on regional housing markets interaction dates to the observation of UK housing markets: house price disparities between South and North tended to increase in the 1980s, but tended to narrow again in the 1990s (Giussani and Hadjimatheou 1991). This behaviour inspires the discussion on regional market integration and the 'ripple effect' hypothesis.

The long-run properties of regional house prices are usually examined under the cointegration framework. MacDonald and Taylor (1993) and Alexander and Barrow (1994) found general evidence for cointegration relationships between regions within either the South or the North of the UK, although the South/North segmentation still appears to exist. In the U.S. housing markets, Yunus and Swanson (2013) documented systematic cointegration among 9 census regions, the degree of which has further intensified after the subprime crisis.

Certain researchers take the idea of cointegration one step further and are interested in the long-run *convergence*, which describes a situation in which local house prices converge towards a constant equilibrium relationship in the long-run (Meen 1996) [3]; that is a more stringent concept than *cointegration*. Cointegration is necessary but not sufficient for long-run convergence of regional markets. House price convergence necessitates that two house price series are cointegrated with a cointegrating vector following the form (1,-1), as well as that they are co-trending, which means no deterministic trend in the cointegrating vector (Holly et al. 2011; Abbott and Vita 2013). In accordance with this tradition, Meen (1996) tentatively suggested three groups (namely the South, the North and the Midlands) in the UK within which house prices may be converged. However, a later study by Abbott and Vita (2013), using the pairwise approach, offered no evidence in support of overall convergence or 'club convergence' [4]. Controversially, Holmes et al. (2011) applied the pairwise approach to the US market and found significant supportive evidence of long-run convergence between state house prices, as well as between MSA house prices.

Since Meen (1999) noted that long-run convergence is equivalent to the long-run stationarity of deviations of regional house prices from the national average, another strand of studies uses the unit root test of the ratio of regional/national house prices to investigate long-run convergence properties. Although Meen (1999) failed to prove the stationarity of region/national ratios using an augmented Dickey-Fuller (ADF) test, Cook (2003) successfully reversed the negative findings by using threshold autoregressive (TAR) and momentum threshold autoregressive (MTAR) tests, which can allow for asymmetric adjustments. The researcher contended that the failure of previous studies is due to the neglect of significant asymmetry in the convergence process. More recently, the stationarity of regional/ national house price ratios in the UK market was confirmed by Holmes and Grimes (2008) who conducted the unit root test on the first principal component (FPC). By applying non-linear unit root tests and linear unit root tests with structural breaks, Canarella et al. (2012) documented conflicting evidence in favour of the stationarity of U.S. metropolitan house price indices to a national house price index [5].

---

3 The convergence here is commonly referred to as stochastic convergence from the time-series point of view. It does not imply that all the local house prices are equalized across regions. However, another notion of convergence that house prices will ultimately converge to the common level in the long run is also investigated by, for example, Kim and Rous (2012) and Blanco et al. (2016).

4 The pairwise approach is essentially similar with Engle-Granger two-step cointegration procedure (Engle and Granger 1987), but pre-specifies the cointegrating coefficients to be (1,-1) in the first step. In this case, the normal unit root statistic can be used to test the unit root of cointegrating residuals in the second step. If the null hypothesis of unit root and linear trend are rejected, the house prices are considered to be converged.

5 In addition to the methods we noted, the battery of approaches that have been dedicated to the long-run convergence of house prices also includes the Kalman filter/time varying parameter (TVP) estimation (Drake 1995) and the so-called synchronicity approach (Miles 2015).

In the UK market, house price changes are usually first observed in London, and then spread to other regions, with the distant regions echoed last. This behaviour is often referred to as a 'ripple effect'. This phenomenon is proved by Meen (1996) who revealed that the speed of adjusting to an equilibrium relation with the South East for each region clearly falls as moving to the North. However, Ashworth and Park (1997) suspected this assumption because they found that the other regions have the common timing of responses to shocks from the South East. Generally, a ripple effect could mean the propagation of shocks emanating from a 'dominant' market such as London to the remaining markets, ignoring the response time of each market. Thus, MacDonald and Taylor (1993) intuitively exhibited a ripple effect from Great London to other regions by using impulse response functions. Alexander and Barrow (1994), using the Granger causality test, detected a causal flow from south to the north with the South East being the most likely exogenous region.

Compared with the term 'ripple effect', U.S. researchers appear to prefer the term 'spatial diffusion', which emphasises the influence on one specific housing market originating from neighbouring markets, not solely from a certain 'dominant' market. An example is Pollakowski and Ray (1997) who revealed the importance of an area's historic price change information in predicting other areas' price change at both the primary metropolitan statistical area (PMSA) level and the subnational census division level.

A recent development in modelling house price diffusion is to incorporate the spatial correlations of housing markets into the conventional time-series models. For example, Holly et al. (2011) proposed a spatial diffusion model in which the house price changes of a region are affected by the short-term and the long-run house price changes both in London and in neighbouring regions. Additionally, the spillover effect of neighbouring regions is evident in the estimation results. Similarly, when modelling U.S. state house prices, Brady (2014) adopted a so-called single-equation spatial autoregressive panel model, which incorporates a "spatial regressor" that is common to spatial autoregressive models, into the panel model. The spatial impulse response functions (SIRFs) support the existence of spatial diffusion.

The relevant studies on long-run and short-run properties of local housing markets in the literature are not exhausted. Certain other examples include Stevenson (2004) on the market of the Republic of Ireland, Berg (2002) on the Swedish second-hand market, Balcilar et al. (2013) on the 5 major metropolitan area markets of South Africa, and Luo et al. (2007) on state capital cities in Australia. More recently, a few studies focusing on Asian housing markets have emerged. For example, Lean and Smyth (2013) documented a ripple effect from the most developed states to the less developed states of Malaysia. A ripple effect from the central city to the suburbs is also demonstrated in Singapore (Liao et al. 2015). In Taiwan, both Lee and Chien (2011) and Chen et al. (2011) offered partial evidence for long-run stable relationships across an inter-city housing market; however, Taipei, the economic centre and capital city, appears not to be the Granger causality of regions in Southern Taiwan.

For historical reasons, urban private housing markets in China were not established until the late 1990s. The lack of continuous house price records makes it difficult to investigate the interaction of local housing markets; however, few attempts have been made by scholars in mainland China. Wang et al. (2008) examined spatial interrelations of house prices among the cities within 5 sub-national areas during the period 1997Q4 – 2007Q1 [6]. Within each of the five sub-national markets, Johansen cointegration test suggests the existence of at least two cointegration relationships; moreover, they found heterogeneous diffusion patterns. Meanwhile, Huang et al. (2010b), focusing on the pair-wise relationships among nine major Chinese cities during a similar time span (1999Q1 – 2008Q3), also presented evidence for generally existing long-run equilibrium relationships. Li and Li (2011) found cointegration relationships for the 9 cities in Pearl River Delta for the period 2001Q1 – 2010Q4; furthermore three submarkets are identified based on Granger Causality test. In addition, Huang et al. (2010a) used a two-stage procedure to test the ripple effect hypothesis in 19 cities based on a period from January 2008 to April 2010. The ripple effect is supported by the evidence that popular and vibrant cities that have greater price fluctuations, such as Guangzhou and Shenzhen, are likely to achieve a turning point earlier than other less active cities.

This paper attempts to offer a comprehensive investigation into the interrelations of Chinese housing markets in an economic co-operation framework – the Pan-PRD, which has not yet been considered by previous studies. First, we examine whether housing markets depend on each other through a Granger causality test. Second, the pairwise long-run *cointegration* and *convergence* properties are tested, respectively. Finally, when modelling the house price diffusion pattern, we consider spatial dependence in accordance with the treatment by Holly et al. (2011).

§ 5.3 Data

§ 5.3.1 The "Pan-Pearl River Delta"

This paper focuses on the housing market interrelations among prefecture-cities [7]. The

6 The five sub-national areas are Northern Coast Area, Central Coast Area, Southern Coast Area, Central Area and Western Area.

7 A prefecture city is an administrative division of China, ranking in the second level of administrative structure. Under its administration are the counties (county-level cities) and districts, of which the districts constitute the city proper. The housing market we noted in this paper pertains to the *city proper*.

whole area under investigation is the so-called "Pan-Pearl River Delta" (Pan-PRD) in South China. The Pan-PRD is a regional co-operation framework launched in June 2004, which has the objective to remove the barriers to the flow of production factors and finally establish a common market. This framework is composed of 11 geographically contiguous spatial units, including 9 provinces: Guangdong, Fujian, Jiangxi, Hunan, Guangxi, Guizhou, Yunnan, Sichuan and Hainan; plus 2 special administrative regions: Hong Kong and Macao (known as "9+2") (refer to Figure 5.1).

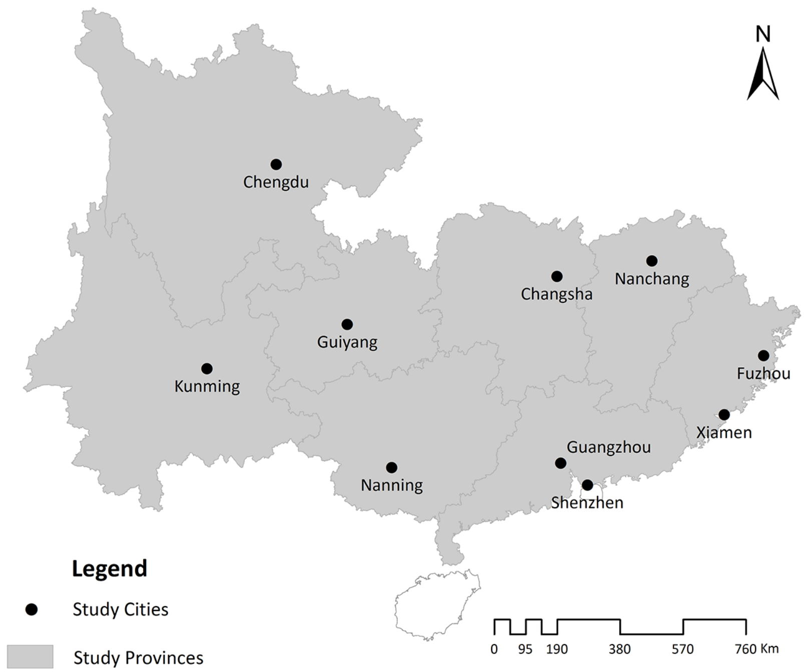


FIGURE 5.1 "Pan-Pearl River Delta" and study cities

The Pan-PRD is already the largest economic bloc in China, representing 20% of China's total land area, 36% of its population and 40% of its GDP (2004 figure). The Pan-PRD spans across several geographic and economic zones that are formulated by the central government. Guangdong, Fujian, Hong Kong and Macao, bordering the South China Sea, are categorised as part of Eastern China; Hunan and Jiangxi, connected with the Yangtze River, belong to Central China, and the remaining four inland provinces are divided into Western China. Obviously, economic development in the Pan-PRD area is far from integration, and regional disparities remain notable. In general, aside from Hong Kong and Macao, the eastern provinces, Guangdong and Fujian, are much more developed than the remaining inland provinces. In particular, Guangdong is undoubtedly the leading province due to its production capabilities and its economic integration with Hong Kong.

The 10 cities of interest in the Pan-PRD are the capitals of 8 provinces, Guangzhou (Guangdong), Fuzhou (Fujian), Nanchang (Jiangxi), Changsha (Hunan), Nanning (Guangxi), Guiyang (Guizhou), Kunming (Yunnan) and Chengdu (Sichuan), and 2 special economic zones (SEZs), Shenzhen (Guangdong) and Xiamen (Fujian). Haikou, the capital of Hainan province, is excluded because its house price development path is clearly different from the others [8]. With increasing economic cooperation in the Pan-PRD, we believe that the dependence between the 10 housing markets is also strengthened.

§ 5.3.2 House price index

The availability of the house price data has been the largest obstacle to examining house price behaviour on temporal and spatial dimensions. Because a truly private housing market was not developed in most Chinese cities until 1998, house price indices used for measuring the movement of house prices are rare. "Price Indices for Real Estate in 35/70 Large- and Medium-sized Cities" (70 Cities Index), compiled and published by the National Bureau of Statistics of China (NBSC), is the only public accessible index system that can provide consistent information on house prices over a long period.

The 70 Cities Index system was first established in 1997 and originally covered 35 major cities. The system published year-over-year quarterly price changes for land transactions, housing sales and housing rentals. However, with the rapid growth of housing transactions and house prices after 2000, the index system could no longer accurately reflect house price movement and was widely criticised by the public. Therefore, NBSC updated the survey and calculation method, which can be called the "matching approach" (Wu et al. 2014), to obtain the quality-adjusted price index, and the scope was expanded to 70 large- and medium-sized cities. Since July 2005, the 70 Cities Index system has been formally reported on a monthly basis, including year-over-year and month-over-month Laspayres indices reported monthly. In January 2011, the survey and calculation strategy for the 70 Cities Index was refined again. Since then, the chained Laspayres index was also made available (base year of 2010).

We apply the "Price Indices of Newly Constructed Residential Buildings" (NCRB Index), drawn from the 70 Cities Index system, for the 10 cities in our empirical analysis,

8 As the capital of Hainan, the sole tropical island in China, Haikou is a famous tourist city. Home buyers from outside constitute a very large share of housing demand so that is no surprise that the housing market of Haikou has certain distinctive characteristics and differs from others.

covering the period from June 2005 to May 2015, for a total of 120 observations. Quarterly year-over-year indexes prior to 2005 are discarded from the analysis. First, the index in this period was likely unreliable due to the rough survey and the calculation methodology. Second, during the period after 2005, house prices are very volatile due to active housing market transactions. Thus, monthly data are more appropriate to model this volatility in house prices.

For our analysis, we joined together the June 2005 – December 2010 index and the January 2011 – May 2015 index. When compiling the house price index by chaining the month-over-month house price changes, at first, January 2011 was set as a reference base because the NCRB Index did not report consistently before and after 2011. Since the month-over-month index before 2010 is not transitive, there is a drift in the early years of house price index [9]. Thereafter, the index series was re-based to June 2005.

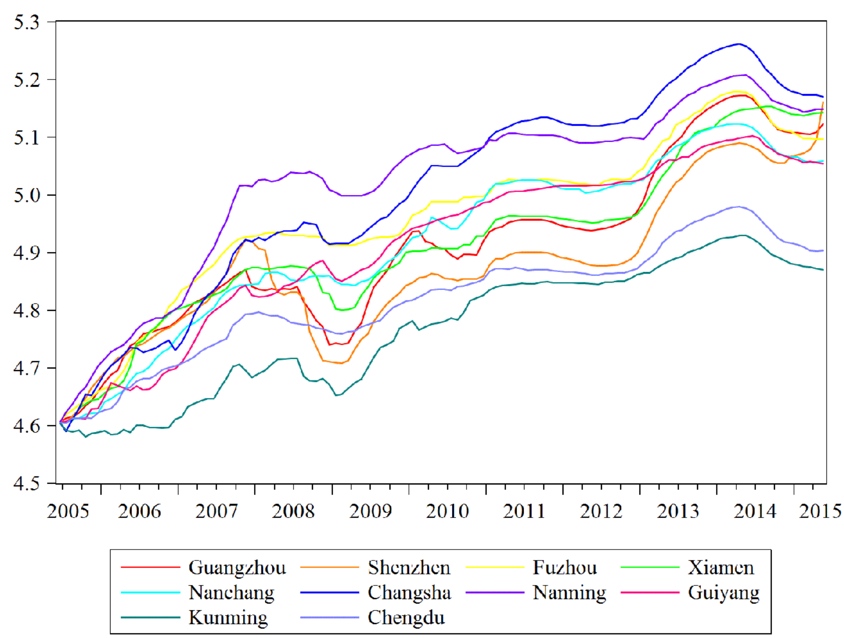


FIGURE 5.2 "Pan-Pearl River Delta" and study cities

Figure 5.2 shows the natural logarithm of house price indices for 10 cities. A casual examination of the figure suggests that house price developments of the 10 cities

9 We note that the chained month-over-month index is not identical to the chained year-over-year index. However, the house price series obtained from chaining these two indexes are highly correlated (the correlation coefficient is greater than 0.97). There are only small differences for the period before January 2007.

follow a similar upward trend throughout the study period. The house prices of Guangzhou and Shenzhen may be much more volatile than those of other cities, particularly during the period after the 2008 global financial crisis, which experienced a sharp decrease of house prices.

§ 5.4 Spatial leading-lag relationships

§ 5.4.1 Toda-Yamamoto Granger causality test

The first question of whether there is a spatial leading-lag relationship or a causal flow between intercity housing markets can be examined using Granger causality tests. Suppose two house price series, p_{1t} and p_{2t}, take the form of VAR(k)

$$\begin{aligned} p_{1t} = c_{10} &+ a_{11}^{1} p_{1t-1} + a_{11}^{2} p_{1t-2} + \cdots + a_{11}^{k} p_{1t-k} \\ &+ a_{12}^{1} p_{2t-1} + a_{12}^{2} p_{2t-2} + \cdots + a_{12}^{k} p_{1t-k} + \epsilon_{1t} \end{aligned} \tag{1}$$

$$\begin{aligned} p_{2t} = c_{20} &+ a_{21}^{1} p_{1t-1} + a_{21}^{2} p_{1t-2} + \cdots + a_{21}^{k} p_{1t-k} \\ &+ a_{22}^{1} p_{2t-1} + a_{22}^{2} p_{2t-2} + \cdots + a_{22}^{k} p_{2t-k} + \epsilon_{2t} \end{aligned} \tag{2}$$

The series p_{2t} is said to Granger cause p_{1t} if the historical values of p_{2t} can contribute to predicting p_{1t} in equation (1). This is equivalent to a test of the null hypothesis $H_0 : a_{12}^{1} = a_{12}^{2} = \cdots = a_{12}^{k} = 0$. The standard causality test procedure requires that the series p_{1t} and p_{2t} are stationary. If they are both $I(1)$, a pre-test of cointegration is needed. In this paper, we act in accordance with the Toda and Yamamoto (1995) (TY) procedure, which overcomes the limitations of standard methodology in a manner that can allow the series to be integrated or cointegrated of an arbitrary order. In other words, the TY procedure estimates an augmented VAR(k+d) system in which d is the maximum order of integration of the time series in the system. The Granger causality tests are then performed on the first k coefficient matrices by using a standard Wald test, ignoring the coefficients matrices of the last d lagged vectors.

§ 5.4.2 Results

We begin by determining the integration orders of 10 (log) house price series. The commonly used Augmented Dickey-Fuller test (Dickey and Fuller 1979) is performed, and the results are reported in the first and fourth columns of Table 5.1. The null hypothesis of unit root for *level* variables of Guangzhou and Xiamen is rejected at the approximately 5% significance level (the p-value of Xiamen is 5.76%), indicating a trend stationary process for these two cities. Conversely, all the *first difference* series are stationary. That is, house price indexes of 8 of 10 cities are unit root processes and are integrated of order one, i.e., $I(1)$. The ADF tests provide preliminary evidence that house prices of different cities are integrated of different orders.

The ADF test has very low power in distinguishing highly persistent stationary processes from non-stationary processes; the power is lower when a deterministic trend is included in the test. Thus, an efficient unit root test, the Dickey-Fuller generalized least square (DF-GLS) test proposed by Elliott et al. (1996), is also conducted. The results in the second and fifth column of Table 5.1 are largely consistent with the ADF test, except that the level house price determination of Xiamen is a unit root process.

TABLE 5.1 Unit root test

	level			*1st difference*		
	ADF	DF-GLS	KPSS	ADF	DF-GLS	KPSS
Guangzhou	-3.886(2)**	-3.361(2)**	0.080(9)	-3.467(1)**	-3.494(1)**	0.056(8)
Shenzhen	-2.179(1)	-2.209(1)	0.121(9)*	-3.477(0)**	-3.510(0)**	0.078(8)
Fuzhou	-2.538(2)	-1.399(2)	0.157(9)**	-3.642(1)**	-3.574(1)***	0.085(8)
Xiamen	-3.391(2)*	-2.404(2)	0.103(9)	-3.784(1)**	-3.642(1)***	0.086(8)
Nanchang	-1.411(1)	-0.957(1)	0.199(9)**	-6.037(0)***	-5.233(0)***	0.051(8)
Changsha	-1.852(1)	-0.962(1)	0.226(9)***	-6.596(0)***	-3.899(0)***	0.046(8)
Nanning	-2.188(1)	-0.862(1)	0.248(9)***	-4.579(0)***	-4.382(0)***	0.083(8)
Guiyang	-1.044(1)	-0.745(1)	0.286(9)***	-6.289(0)***	-5.908(0)***	0.042(8)
Kunming	-1.365(1)	-1.764(1)	0.148(8)**	-7.070(0)***	-5.161(0)***	0.092(7)
Chengdu	-2.623(2)	-2.144(2)	0.127(9)*	-3.628(1)**	-3.347(1)**	0.055(8)

Notes: All the models include an intercept and a deterministic trend. Numbers shown in parentheses are the lag length or bandwidth. For the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller 1979), the lag length for *level* variables is selected using the Bayesian Information Criterion (BIC) with the maximum lag length being set to 12. DF-GLS unit root tests (Elliott et al. 1996) use the same lag length chosen for ADF test. For KPSS test (Kwiatkowski et al. 1992), the bandwidth is selected by Newey-West method. In all the cases, the lag length for *first difference* variables equals to the lag length for *level* variables minus one. The null hypothesis for ADF and DF-GLS tests is having a unit root, but stationary for the KPSS test. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 5.1(column 3 and 6) also presents the results of the KPSS test (Kwiatkowski et al. 1992), which has the null hypothesis of stationary. The KPSS results support that Xiamen's house price series is $I(0)$, in accordance with the ADF test but in contradiction to the DF-GLS test. The evidence for the integration order of the Xiamen price series appears to be mildly confusing. Given that the KPSS statistic is very close to the 10% critical value (0.119), it is reasonable to assume Xiamen's price process to be I(1) for the following analysis [10].

10 One reviewer pointed out that house prices are very likely to display an asymmetric adjustment, reducing the

Because not all the house price series are integrated of the same order, the pairwise Toda-Yamamoto Granger causality procedure is preferable in this analysis. According to information criteria, such as AIC and BIC, the optimal lag lengths for most of the city-pair VARs are 2 and 3. To largely ensure that the residuals are close to white noise, we proceed using the VAR($3 + d$) system in which d is the maximum integration order of the city pair.

The results of Toda-Yamamoto Granger causality tests are reported in Table 5.2, where the null hypothesis is that column cities do not Granger cause row cities. Overall, in most of the city pairs, we find significant evidence of bilateral or unidirectional leading-lag relations; however, the causality pattern is very complicated [11]. Certain literature focusing on housing market interaction, such as Clapp et al. (1995) and Chen et al. (2011), has claimed that the house price interrelation (or causality) only occurs between neighbouring markets. Our results cast doubt on this conclusion because leading-lag relationships are found in many city pairs where the two cities are separated by very long distances. Conversely, there is no spatial causality in certain short-distance city pairs, such as Xiamen and Fuzhou. Similarly, Pollakowski and Ray (1997) and Luo et al. (2007) also found significant causality between non-contiguous regions. Such complicated causality patterns may be largely due to economic relations rather than behavioural reasons (Pollakowski and Ray 1997).

To further examine the results, the cities in the system are divided into two groups: the eastern-central group including the first six cities in Table 5.2 and the western group containing the remaining cities. We tentatively find a general unidirectional causal flow from the eastern-central area to western China. The historical house price information of all eastern-central cities, except Xiamen, can significantly contribute to predicting the house prices of western cities. The opposite, conversely, can hardly be true given the largely insignificant Granger causality test results in the lower-left panel. However, among the four western cities, the housing markets of Kunming and Chengdu appear to play a role in predicting house price behaviours in eastern-central cities. The above findings are closely related to the socio-economic disparities between eastern, central and western China. Considering that eastern-central cities are generally more developed than western ones, it can be expected that their market dynamics can lead housing market behaviours in the remaining cities. Chengdu is an exception in western cities, given its status as the financial and economic centre of western China. Consequently, the mutual leading-lag relationships between Chengdu and most

power of traditional unit root tests. Thus, we employed the momentum threshold autoregressive (MTAR) asymmetric unit root test proposed by Enders and Granger (1998) to test the integration order of 10 cities' house price indexes. However, the results, which are available upon request, do not turn over the finding that all the cities are $I(1)$ process except for Guangzhou.

11 It should be noted that, throughout the paper, when we say "a market leads or causes another market in the Granger sense", we cannot exclude the possibility that such correlation is caused by common shocks.

TABLE 5.2 Toda-Yamamoto Granger causality test

	Guangzhou	Shenzhen	Fuzhou	Xiamen	Nanchang	Changsha	Nanning	Guiyang	Kunming	Chengdu
Guangzhou		25.63***	6.47*	50.54***	11.13***	14.10***	11.25***	9.47**	4.41	9.16**
		(0.000)	(0.091)	(0.000)	(0.011)	(0.003)	(0.010)	(0.024)	(0.220)	(0.027)
Shenzhen	13.82		9.14**	18.93***	6.05*	18.76***	8.25**	14.14***	6.22*	8.23**
	(0.003)		(0.027)	(0.000)	(0.109)	(0.000)	(0.041)	(0.003)	(0.101)	(0.041)
Fuzhou	2.00	1.52		7.52*	5.04	7.97**	8.52**	6.49*	8.26**	6.78*
	(0.572)	(0.678)		(0.057)	(0.169)	(0.047)	(0.036)	(0.090)	(0.041)	(0.079)
Xiamen	18.73***	2.32	3.38		4.10	2.82	5.28	2.01	0.28	0.94
	(0.000)	(0.510)	(0.336)		(0.251)	(0.420)	(0.153)	(0.570)	(0.963)	(0.815)
Nanchang	4.47	7.90**	8.25**	6.42*		14.61***	12.22***	17.52***	11.44***	11.55***
	(0.215)	(0.048)	(0.041)	(0.093)		(0.002)	(0.007)	(0.001)	(0.010)	(0.009)
Changsha	1.03	5.31	7.48*	6.11*	6.74*		13.36***	34.22***	13.29***	7.99**
	(0.794)	(0.150)	(0.058)	(0.106)	(0.081)		(0.004)	(0.000)	(0.004)	(0.046)
Nanning	8.91**	14.22***	4.92	3.52	5.08	9.51**		19.42***	8.64**	8.70**
	(0.031)	(0.003)	(0.178)	(0.318)	(0.166)	(0.023)		(0.000)	(0.035)	(0.034)
Guiyang	0.41	2.37	2.32	0.67	5.31	5.07	5.29		1.38	7.46*
	(0.938)	(0.500)	(0.509)	(0.880)	(0.150)	(0.167)	(0.152)		(0.711)	(0.058)
Kunming	7.27*	11.59***	1.79	5.96	7.30*	27.03***	8.31**	11.68***		10.32**
	(0.064)	(0.009)	(0.616)	(0.113)	(0.063)	(0.000)	(0.040)	(0.009)		(0.016)
Chengdu	5.57	5.24	16.81***	17.69***	15.25***	8.21**	6.20*	7.24*	7.19*	
	(0.135)	(0.155)	(0.001)	(0.001)	(0.002)	(0.042)	(0.102)	(0.064)	(0.066)	

Notes: Toda-Yamamoto Granger causality test is performed under VAR(3+d) system where d is the maximum integration order of house price processes in the city pair. Both intercept and deterministic time trend are included in the VAR system. The null hypotheses are that house prices in column cities do not Granger cause house prices in row cities. The asymptotic Wald statistics are reported, and the corresponding p-values are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

eastern-central cities are no surprise. Furthermore, within either the eastern-central or western group, the significant leading-lag relationship (at least for one direction) can be found in every city pair.

Given the leading position of Guangdong in the Pan-PRD, we expect that the housing markets of the two cities under its territory, namely Guangzhou and Shenzhen, will lead the markets in other cities. Indeed, Guangzhou and Shenzhen impose a significant leading influence on nearly all the other housing markets, and they are less predictable on the basis of previous information from other markets. However, these cities' dominant role is not unique. Certain other cities appear to also have similar 'exogenous' properties, such as Nanchang.

§ 5.5 Long-run properties

§ 5.5.1 Cointegration and convergence test

The previous section reveals that the 10 cities' housing markets in the Pan-PRD are interrelated with each other. In this section, we go further to ask the question of whether these markets are tied together in the long-run, i.e., if they hold a long-run equilibrium relationship. To answer this question, the long-run cointegration and convergence properties of house prices are investigated.Two $I(1)$ house price series p_{1t} and p_{2t} are said to be cointegrated if a linear combination of p_{1t} and p_{2t} is stationary. Since we are interested in the pairwise cointegration of house price indexes of two cities, the Engle-Granger (EG) two-step procedure (Engle and Granger 1987) is employed in this paper, which has been applied by, for example, MacDonald and Taylor (1993) to a similar question.

The first step is to estimate the long-run equilibrium relationship by the following equation

$$p_{it} = D + \beta p_{2t} + u_t \tag{3}$$

where D is deterministic terms that may contain a constant, a deterministic trend or both. The cointegration test involved in the second step is then based on testing the unit root of residual series u_t. If u_t is stationary, we say that p_{1t} and p_{2t} are cointegrated with $(1, -\beta)$. Because of the spurious regression under the null hypothesis of non-cointegration in the first stage, the residual-based ADF test in the second stage does not have the standard Dickey-Fuller distribution. Therefore, critical values for cointegration test simulated by MacKinnon (1996) are used in this paper.

As noted by Holly et al.(2011) and Abbott and Vita (2013), conditional on cointegration, the long-run convergence of house prices necessitates two additional conditions: (1) the cointegrating vector corresponding to house price series being (1,-1), and (2) the lack of a deterministic trend being presented in cointegration space. The long-run convergence property is tested by the so-called pair-wise approach, which has been used by Holmes et al. (2011) and Abbott and Vita (2013). Specifically, the cointegrating vector is pre-specified with form (1,-1), and then any standard unit root test can be directly used to test the stationarity of cointegrating residuals.

TABLE 5.3 Engle-Granger pairwise cointegration results

	Shenzhen	Fuzhou	Xiamen	Nanchang	Changsha	Nanning	Guiyang	Kunming	Chengdu
Shenzhen		-0.633	-2.053	-0.837	-0.532	-0.520	-0.919	-0.723	-0.067
		-3.094	-13.706	-4.413	-2.092	-1.557	-3.967	-2.718	-0.280
Fuzhou	-0.221		-0.855	-2.094	-1.716	-1.981	-2.204	-2.239	-2.711
	-1.289		-3.683	-5.553	-5.963	-7.077	-9.371	-7.201	-10.044
Xiamen	-1.875	-1.129		-0.791	-1.362	-1.248	-1.482	-1.339	-0.831
	-13.728	-4.692		-2.390	-5.642	-3.800	-5.885	-4.200	-3.591
Nanchang	-0.501	-2.423	-0.604		-3.792*	-2.136	-3.086*	-3.172	-2.615
	-3.266	-6.421	-1.918		-27.395**	-6.507	-19.217*	-14.187	-14.058
Changsha	-0.025	-1.774	-1.279	-3.722*		-1.657	-3.331*	-4.069**	-2.381
	-0.129	-6.234	-5.748	-26.892**		-3.287	-20.505*	-20.546*	-11.213
Nanning	0.309	-1.762	-0.693	-1.693	-1.162		-1.988	-1.938	-1.593
	1.225	-6.234	-2.155	-5.007	-2.189		-6.342	-5.058	-4.286
Guiyang	-0.580	-2.387	-1.498	-3.086*	-3.435*	-2.453		-2.907	-2.680
	-3.231	-10.249	-6.588	-19.214*	-21.148*	-8.041		-14.028	-15.741
Kunming	-0.511	-2.793	-1.438	-3.470*	-4.479**	-2.856	-3.118		-3.169
	-2.369	-8.815	-4.608	-15.207	-22.088*	-7.390	-14.779		-13.768
Chengdu	0.330	-2.919	-0.661	-2.569	-2.420	-2.017	-2.601	-2.843	
	1.649	-10.753	-2.936	-13.714	-11.287	-5.444	-14.893	-12.465	

Notes: The two statistics in each cell are, respectively the τ statistic and z statistic of ADF test with the null hypothesis of no cointegration. The MacKinnon (1996) critical values are used. In each cointegration equation, the row cities are defined as dependent variables. The constant is included in the cointegration space in all city pairs except the Nanchang-Guiyang pair, in which the constant is not significant. * and ** indicate the 5% and 1% significance level, respectively.

§ 5.5.2 Empirical results

Table 5.3 demonstrates the results of the pairwise Engle-Grange cointegration test. Guangzhou is excluded from cointegration analysis because it is $I(0)$ according to our unit root test results. A brief view of the results indicates that cointegration relationships rarely exist between cities, in contrast to the widely existing leading-lag relationships. Among the city pairs that are tied together in the long-run, three cities, Nanchang, Changsha and Guiyang, form a ‘cointegration club’ within which every city cointegrates with each other. In addition, the significant long-run equilibrium relationship can also be observed in the city pair of Changsha-Kunming. We also note that none of the three eastern cities, namely Shenzhen, Fuzhou and Xiamen, is cointegrated with each other or with the remaining central and western cities. This might indicate that in the long-run, the housing market conditions in eastern cities still significantly differ from the markets of the remaining cities, although much effort has been made to promote the integration process of Pan-PRD's economy.

TABLE 5.4 Pairwise convergence results with pre-specified coefficients (1,-1)

	Shenzhen	Fuzhou	Xiamen	Nanchang	Changsha	Nanning	Guiyang	Kunming	Chengdu
Shenzhen		-0.919	-1.838	-1.071	-0.875	-0.667	-1.077	0.563	1.040
Fuzhou	-0.642		-1.060	-0.796	-0.940	-1.009	-1.448	0.146	1.044
Xiamen	-2.109	-0.927		-0.840	-1.047	-0.880	-1.510	0.803	1.109
Nanchang	-0.853	-1.843	-0.733		-0.546	-0.612	-2.994**	0.206	0.588
Changsha	-1.144	-1.607	-1.529	-2.729		-0.648	-0.465	0.348	0.614
Nanning	-0.480	-2.144	-1.132	-2.090	-0.594		-0.577	0.174	0.392
Guiyang	-0.882	-2.170	-1.468	-3.075*	-2.900*	-2.451		-0.009	-0.056
Kunming	-0.480	-3.383*	-1.464	-3.550**	-4.012**	-3.770**	-2.949*		-1.154
Chengdu	0.509	-4.018**	-0.491	-1.810	-2.973*	-3.443*	-1.542	-2.124	

Notes: The null hypothesis of no convergence is tested based on the residual from $p_{it} - p_{jt}$. The results of ADF test (Dickey and Fuller 1979) are reported in the lower triangle, whereas the upper triangle shows the results of DF-GLS tests (Elliott et al. 1996). In the unit root test process, the constant is included, and the lag length (not reported) is automatically selected by the Bayesian Information Criterion (BIC). * and ** indicate the 5% and 1% significance level, respectively.

After examining the cointegration, we proceed to test the more restricted long-run convergence properties. The pairwise cointegration results with pre-specified coefficients (1,-1) are shown in Table 5.4. The ADF test in the lower triangle suggests that among the four cointegration city pairs, there are three convergent pairs: Nanchang-Guiyang, Changsha-Guiyang and Changsha-Kunming. In addition, another seven city pairs, which are not cointegrated in the Engle-Granger cointegration test are found to be significantly converged. Considering that cointegration is a necessary condition of convergence, the convergent results of these seven pairs are a surprise. These contradictory results might be due to the low power of the ADF test in detecting the unit root. To verify the ADF test, we also perform the more efficient DF-GLS test, the results of which are reported in the upper triangle of Table 5.4. This time we find only one significantly convergent pair, the Nanchang-Guiyang pair, which is also cointegrated. Note that regardless which unit root test we used, city pairs that are convergent are rare. Therefore, it is reasonable to conclude that cointegration or convergence is unlikely to widely exist among the nine cities, which indicates a diverged interurban housing market in the Pan-PRD.

To check the robustness of the cointegration and convergence test results based on the two-step procedure, we also conduct two additional tests: the pairwise Johansen cointegration test and the two-step convergence test based on the momentum threshold autoregressive (MTAR) unit root test (Enders and Granger 1998) which can allow an asymmetric adjustment. The trace statistics of the Johansen procedure, computed based on a VAR(3) specification with unrestricted intercept and no trend in VAR, are reported in the lower triangle of Table A1. The null hypothesis is that the column city is not cointegrated with the row city. For those cointegrated city pairs, the upper triangle of Table A1 reports the results of the log-likelihood ratio (LR) test, which is used to test the cointegrating vector restriction (1,-1). Table A2 displays the pairwise cointegration results with pre-specified coefficients (1,-1) based on the MTAR unit root test. Both of these two powerful methods identify a similar cointegration and convergence pattern between cities with the two-step procedure does; the results are largely in line with the lower triangle results of Table 5.4. Thus, we are confident of the previous finding that only very few city pairs are found to be cointegrated or convergent. For the following analysis, we mainly rely on the results of the two-step procedure.

§ 5.6 House price diffusion pattern

§ 5.6.1 Spatial-temporal house price diffusion model

Previous analysis suggests that most of the city pairs do not hold a long-run equilibrium relationship; however, a few pairs do. When modelling the house price dynamics of a city, we should consider the interrelation with both the cointegrated cities and the non-cointegrated cities. In other words, we should consider the influence from the cities that can impose a long-term effect and the cities that only have a transitory effect. In addition, the spatial dimension should also be considered because it is likely that the effect imposed by nearby cities is stronger than the influence of distant cities. The spatial-temporal house price diffusion model adopted in this paper can fully capture the characteristics along both spatial and temporal dimensions. This model is a variant of spatial-temporal diffusion model proposed by Holly et al. (2011) (the Holly model), which has been applied to investigating the effects of language border on the diffusion of house prices in Belgian markets by Helgers and Buyst (2016). However, unlike the Holly model, we do not designate a 'dominant' city, which is assumed to have contemporaneous effects on non-dominant cities. The reason for abandoning the 'dominant' city from our model specification is that the general lack of pairwise long-run cointegration relationships between cities found previously suggests that there is no city that can be seen as the long-run forcing for other cities. In our model specification, the house price series in the system excluding p_{it} is split into two groups because of the existence of a 'cointegration club': one group (denoted by C) being cointegrated with p_{it} and the other (denoted by O) not. A first order error correction specification for p_{it} is given by

$$\Delta p_{it} = \phi_{i0}\left(p_{i,t-1} - \beta_i \bar{p}^C_{i,t-1}\right) + a_i + a_{i1}\Delta p_{i,t-1} + b_{i1}\Delta \bar{p}^C_{i,t-1} + c_{i1}\Delta \bar{p}^O_{i,t-1} + \epsilon_{it}, \quad (4)$$

where $\bar{p}^C_{i,t-1}$ and $\bar{p}^O_{i,t-1}$ are the spatially lagged variables, defined by

$$\begin{cases} \bar{p}^C_{i,t-1} = \sum w_{ij} p_{jt}, & \text{if } p_{jt} \text{ belongs to cointegrating group} \\ \bar{p}^O_{i,t-1} = \sum w_{ij} p_{jt}, & \text{if } p_{jt} \text{ belongs to non-cointegrating group} \end{cases}$$

The weight, $w_{ij} \geq 0$, which describes the spatial interaction between city i and j, can be constructed either based on a contiguity measure or certain distance measures. Here, the weight is simply calculated by a simple inverse distance function

$$w_{ij} = 1/d_{ij} \tag{5}$$

where d_{ij} is the straightforward distance between the CBDs of city i and city j. In accordance with tradition, the weights are arranged in a row-standardized spatial weight matrix W.

Because p_{it} is cointegrated with each member in the cointegrating group C, it is expected to be cointegrated with $\bar{p}^C_{i,t-1}$ as well. The cointegrating parameter β_i can be estimated in advance and treated as known in estimating the equation (4). Even if city i has no cointegrated counterpart, the model can also be conducted by simply setting the error correction coefficient (ϕ_{i0}) to zero.

§ 5.6.2 Generalized impulse response function (GIRF)

After obtaining the parameter estimates of model (4) by ordinary least squares (OLS), we can construct the spatial-temporal impulse response functions for simulating and forecasting purposes. We begin by writing the system of equations (4) in matrix form

$$\Delta \mathbf{p}_t = \mathbf{a} + \mathbf{\Pi} \mathbf{p}_{t-1} + \mathbf{\Gamma} \Delta \mathbf{p}_{t-1} + \epsilon_t \tag{6}$$

where $\mathbf{\Gamma} = \mathbf{A}_1 + \mathbf{B}_1 + \mathbf{C}_1$, $\mathbf{p}_t = (p_{1t}, p_{2t}, \cdots, p_{nt})'$, $\mathbf{a} = (a_1, a_2, \cdots, a_n)'$, $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \cdots, \epsilon_{nt})'$,

$$\mathbf{\Pi} = \begin{bmatrix} \phi_{10} & 0 & \cdots & 0 & 0 \\ 0 & \phi_{20} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \phi_{n-1,0} & 0 \\ 0 & 0 & \cdots & 0 & \phi_{n0} \end{bmatrix} - \begin{bmatrix} \phi_{10}\beta_1 \mathbf{w}'_{1,C} \\ \phi_{20}\beta_2 \mathbf{w}'_{2,C} \\ \vdots \\ \phi_{n-1,0}\beta_{n-1} \mathbf{w}'_{n-1,C} \\ \phi_{n0}\beta_n \mathbf{w}'_{n,C} \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ 0 & a_{21} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,1} & 0 \\ 0 & 0 & \cdots & 0 & a_{n1} \end{bmatrix}, \mathbf{B}_1 = \begin{bmatrix} b_{11} \mathbf{w}'_{1,C} \\ b_{21} \mathbf{w}'_{2,C} \\ \vdots \\ b_{n-1,1} \mathbf{w}'_{n-1,C} \\ b_{n1} \mathbf{w}'_{n,C} \end{bmatrix}, \quad \text{and} \quad \mathbf{C}_1 = \begin{bmatrix} c_{11} \mathbf{w}'_{1,O} \\ c_{21} \mathbf{w}'_{2,O} \\ \vdots \\ c_{n-1,1} \mathbf{w}'_{n-1,O} \\ c_{n1} \mathbf{w}'_{n,O} \end{bmatrix}$$

where $\mathbf{w}'_{i,C}$ and $\mathbf{w}'_{i,O}$ represent the i_{th} row of spatial weight matrixes connecting with the cointegration group and the non-cointegration group, respectively. Matrix $\mathbf{A}_1$ indicates their own short-run influence, and matrixes $\mathbf{B}_1$ and $\mathbf{C}_1$ represent the short-run impacts of the cities from the cointegration group and the non-cointegration group, respectively.

The equation (6) can be rewritten as a form of vector autoregression (VAR)

$$\mathbf{p}_t = \mathbf{a} + \mathbf{\Phi}_1 \mathbf{p}_{t-1} + \mathbf{\Phi}_2 \mathbf{p}_{t-2} + \epsilon_t \tag{7}$$

where $\mathbf{\Phi}_1 = \mathbf{I}_n + \mathbf{\Pi} + \mathbf{\Gamma}$ and $\mathbf{\Phi}_2 = -\mathbf{\Gamma}$. The VAR model (7) can then be used for impulse response analysis. Suppose that the shock, ϵ_{it}, which will propagate to other cities, is characterised by the variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1,n-1} & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2,n-1} & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{n-1,1} & \sigma_{n-1,2} & \cdots & \sigma_{n-1,n-1} & \sigma_{n-1,n} \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{n,n-1} & \sigma_{nn} \end{bmatrix}$$

where $\sigma_{ij} = E(\epsilon_{it}\epsilon_{jt})$, which can be consistently estimated from the OLS residuals $\hat{\epsilon}_{it}$ of the individual regressions, namely by $\hat{\sigma}_{ij} = T^{-1}\sum_{t=1}^{T} \hat{\epsilon}_{it}\hat{\epsilon}_{jt}$ and $\hat{\sigma}_{ii} = T^{-1}\sum_{t=1}^{T} \hat{\epsilon}_{it}^2$. To allow for possible contemporaneous correlation across cities, we consider the generalized impulse response function (GIRF) advanced in Pesaran and Shin (1998). The impulse response of a unit (one standard error) shock to house price in a city on the remaining cities at a horizon h periods ahead will be provided by

$$\begin{aligned} &\mathbf{g}_i(h) = E\left(\mathbf{p}_{t+h} | \epsilon_{it} = \sqrt{\sigma_{ii}}, \widetilde{S}_{t-1}\right) - E\left(\mathbf{p}_{t+h} | \widetilde{S}_{t-1}\right) = \frac{\mathbf{\Psi}_h \mathbf{\Sigma} \mathbf{e}_i}{\sqrt{\sigma_{ii}}} \\ &\text{for} \quad i = 1, \cdots, n; h = 0, 1, \cdots, H \end{aligned} \tag{8}$$

where $\widetilde{S}_{t-1}$ is the information set at time $t-1$ and $\mathbf{e}_i$ is an $n \times 1$ vector of zeros with the exceptions of its i_{th} element, which is unity, and

$$\mathbf{\Psi}_h = \mathbf{\Phi}_1 \mathbf{\Psi}_{h-1} + \mathbf{\Phi}_2 \mathbf{\Psi}_{h-2}, \tag{9}$$

with $\mathbf{\Psi}_0 = \mathbf{I}_n$ and $\mathbf{\Psi}_h = \mathbf{0}$ for $h < 0$.

§ 5.6.3 Empirical results

According to the previous cointegration test, four cities, Nanchang, Changsha, Guiyang and Kunming, are cointegrated with at least one of the other cities. For these four cities, we should include the error correction term $\left(p_{i,t-1} - \beta_i \bar{p}_{i,t-1}^C\right)$ in their house price dynamic specifications. In contrast, for the remaining cities, the error correction term and the term $\Delta\bar{p}_{i,t-1}^C$ in equation (4) can be eliminated.

The cointegration results between the four cities and the spatial lag of their cointegrated counterparts are shown in Table 5.5. As expected, the τ and z statistics provide significant evidence of cointegration for Nanchang, Changsha and Kunming. In the case of Guiyang, we cannot reject the null hypothesis of no cointegration at the 5% significance level; however, the statistics are marginally significant at the 10% level. The third column of Table 5.5 reports the estimated long-run relationships (β_i). The estimated β_i are approximately distributed around the value of unity, except for Kunming.

TABLE 5.5 Cointegration between the four cointegrating cities and spatial lag of their cointegrated counterparts

	τ statistic	z statistic	β
Nanchang	-3.561*	-25.128*	0.843441
Changsha	-4.541**	-34.150**	1.287945
Guiyang	-3.089	-18.209	0.871291
Kunming	-4.069**	-20.546*	0.602230

Notes: The test is based on one equation regression in which we take column cities as dependent variables and the spatial lags of their cointegrated counterparts as independent variables. The first two columns report the τ statistic and z statistic of ADF test with the null hypothesis of no cointegration. The third column reports the estimation of β. Note that the t-ratio for β is invalid in this case. * and ** indicate the 5% and 1% significance level, respectively.

With β_i being determined, the spatial-temporal house price dynamic model for each city can be estimated using ordinary least squares (OLS). The estimation results are summarized in Table 5.6 where the lag-orders are set to 2. These models perform reasonably well because the Breusch-Godfery test suggests no serial correlation in each regression's residuals at least at the 5% significance level.

The error correction terms, which appear in the model specification of four cities, are all significant at the 10% significance level, three of which are significant at the 5% level or above. That is, the four cities' short-run dynamics are influenced by the deviation from the long-run equilibrium relationship. The coefficient ϕ_{i0} indicates that the house price of Changsha responds to the disequilibrium much more rapidly than that of the other three cities.

We now turn to the influence of short-term dynamics. Not surprisingly, the first-lag price changes are significantly positive in all equations. The second-order lagged price dynamics also play a role in the price equation for Fuzhou and Xiamen. Similarly, the lagged price changes from neighbouring cities (either from the cointegration group or the non-cointegration group) are also found to be statistically significant in all equations, except for Changsha and Kunming. This confirms the existence of cross-city spillover effects from the neighbouring cities, which is in accordance with the findings of Holly et al. (2011) for the UK market and Helgers and Buyst (2016) for Belgian housing markets.

TABLE 5.6 Estimation results of the spatial-temporal house price diffusion model

City	Constant	Error correction	Own lag effects		Lag effects of cointegration group		Lag effects of non-cointegrated group		Adjusted R^2	Serial correlation
		ϕ_{i0}	a_{i1}	a_{i2}	b_{i1}	b_{i2}	c_{i1}	c_{i2}		
Guangzhou	0.0007 (0.001)	–	0.4326*** (0.098)	0.0765 (0.108)	–	–	0.6540*** (0.182)	-0.3067* (0.161)	0.514	5.452
Shenzhen	0.0011 (0.001)	–	0.6464*** (0.106)	-0.0038 (0.106)	–	–	0.9324** (0.197)	-0.7141*** (0.200)	0.527	6.035
Fuzhou	0.0005 (0.001)	–	0.3074*** (0.100)	0.2321* (0.097)	–	–	0.5384*** (0.140)	-0.2231 (0.141)	0.548	0.588
Xiamen	0.0008 (0.001)	–	0.1673* (0.096)	0.1624* (0.093)	–	–	0.7223*** (0.167)	-0.1577 (0.180)	0.477	2.344
Nanchang	0.0548* (0.028)	-0.0755* (0.040)	0.3530*** (0.104)	0.0723 (0.108)	0.0737 (0.098)	-0.0436 (0.096)	0.3720** (0.171)	-0.0430 (0.174)	0.425	4.708
Changsha	-0.2594*** (0.062)	-0.2006*** (0.048)	0.3654*** (0.093)	0.0726 (0.092)	0.1437 (0.158)	-0.0934 (0.159)	0.1763 (0.192)	0.1628 (0.195)	0.541	1.271
Nanning	0.0003 (0.001)	–	0.4869*** (0.106)	0.0244 (0.101)	–	–	0.3624** (0.171)	0.0928 (0.170)	0.615	3.111
Guiyang	0.0414** (0.020)	-0.0696** (0.034)	0.4129*** (0.095)	-0.0258 (0.092)	0.4800*** (0.109)	-0.2064* (0.111)	0.0976 (0.155)	-0.0413 (0.165)	0.467	7.778*
Kunming	0.1915*** (0.059)	-0.1092*** (0.034)	0.3423*** (0.102)	-0.0546 (0.098)	0.1323 (0.109)	-0.1476 (0.106)	0.3847 (0.263)	0.0062 (0.232)	0.258	0.855
Chengdu	-0.0002 (0.000)	–	0.2419** (0.099)	0.0420 (0.098)	–	–	0.3315*** (0.120)	0.1753 (0.124)	0.550	1.934

Notes: The estimation results are based on equation (4) with lag-orders being set to 2. For the cities except Nanchang, Changsha, Guiyang and Kunming, we eliminate the error correction term and the term 'lag effects of cointegration group' because these cities do not have cointegrated counterparts. The "serial correlation" column reports the Breusch-Godfery serial correlation statistic test that distributes approximately as χ_3^2 under the null hypothesis of no residual serial correlation. *, ** and *** indicate the 10%, 5% and 1% significance level, respectively.

The regression results of spatial-temporal diffusion models presented in Table 5.6 depict a complicated dynamic system in which the historical house price changes of a city not only affect its own price changes but also influence the price changes in other cities directly or indirectly through their neighbouring cities, or through the long-run equilibrium. To intuitively illustrate the diffuse nature of house prices in a complicated system, we provide the generalized impulse response functions, which can trace the time profile of shocks both over time and space.

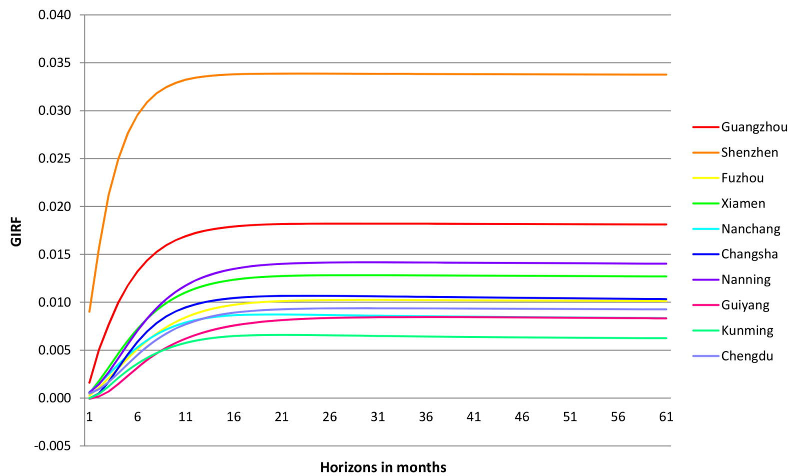


FIGURE 5.3 Generalized impulse responses of a positive unit shock (one standard error) to Shenzhen house prices

Figure 5.3 plots the generalized impulse responses of all the cities to a positive unit shock (one standard error) to the house prices of Shenzhen, one of the most developed cities in the Pan-PRD. The positive shock gradually diffuses to the remaining cities, significantly raising the house prices in the whole area (being confirmed by the bootstrap confidence interval in Figure B1 in the Appendix). However, the magnitude of the spillover effect differs across the region. Given the one standard error shock to Shenzhen, its own house prices soar approximately 3.5%, followed by Guangzhou, which rises by approximately 2%. Conversely, the increases of the other cities' house prices are approximately 1%. This indicates a diverged interurban housing market between developed and less-developed cities. For the sake of comparison, Figure 5.4 portrays the responses to a positive stand error shock to Changsha, a city in Central China and cointegrating with the other three cities. It is clear that the unit shock to Changsha generates relatively homogenous effects on all other cities' house prices (house prices increases are approximately between 0.8% and 1%), except for Shenzhen and Kunming. The effect on Shenzhen house prices is not significant, as indicated by the bootstrap error bounds shown in Figure B2 (refer to the Appendix). This information further supports our conclusion regarding the divergence of a few cities' housing markets, such as that of Shenzhen[12].

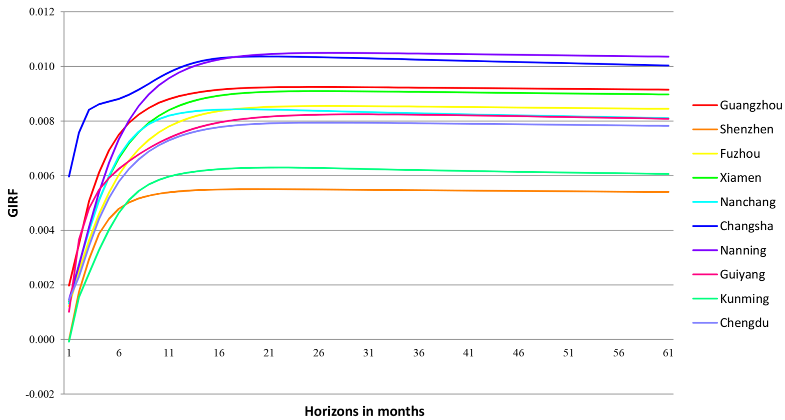


FIGURE 5.4 Generalized impulse responses of a positive unit shock (one standard error) to Changsha house prices

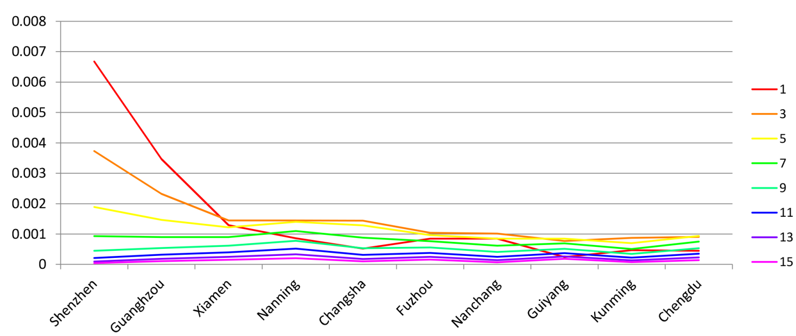


FIGURE 5.5 Generalized impulse responses of house price changes to one stand error shock to Shenzhen house prices

Figure 5.3 and Figure 5.4 also display a certain diffusion pattern in a manner that certain cities' response to shocks is more rapid than the others. To further examine the spatial-temporal diffusion pattern, Figure 5.5 depicts the impulse responses of house price changes to one standard error shock to Shenzhen house prices (the cities in the horizontal axis are ordered by distance). The first month after shock witnesses much higher house price increases in Shenzhen and its neighbouring cities than in cities far away. In the following few months, the house price changes of distant cities begin to

12 The impulse responses of the shock to other cities, which are not reported for space consideration, can lead to the similar finding that the overall interurban housing market is diverged. The responses are available upon request.

catch up, but remain slightly behind the neighbouring cities. Finally, house price changes in each city are nearly identical to each other after the seventh month. This behaviour clearly describes a diffusion pattern with the cities that are close to Shenzhen responding to shocks more rapidly and drastically.

§ 5.7 Conclusion and implications

Three aspects of the spatial interrelations of the 10 cities' housing markets in the Pan-PRD, namely spatial causality, convergence and diffusion, are carefully examined in this paper, based on the monthly house price indexes covering the period from June 2005 to May 2015. Among the 10 cities' housing markets, the Toda-Yamamoto Granger causality test reveals a complicated inter-market correlation pattern. It can be tentatively concluded that there is a causal flow from eastern-central China to the West considering that house prices of eastern-central cities are helpful in predicting house prices of western cities, but not vice versa.

In spite of the widely existing leading-lag interrelations, the Engle-Granger cointegration test provides very limited evidence for long-run cointegration among the cities. We find 4 cointegrated pairs of 36 city combinations. The evidence for convergence is rare too. Overall, the housing markets in the Pan-PRD are diverged, particularly between developed eastern and less developed western cities. The finding of divergence in the housing markets in Pan-PRD area contradicts most of the previous studies (e.g.,Wang et al. 2008; Huang et al. 2010b; Li and Li 2011) that support long-run cointegration of housing markets within a relatively homogeneous area. This suggests the possibility of 'club integration' and we indeed find a 'cointegration club' among the three cities in Central China.

In the short-run, the estimation results of the spatial-temporal diffusion model show that the house price change of a city can be influenced by its own lagged price changes, the spillovers from neighbouring cities, or the long-run forces from the cointegrated counterparts. Furthermore, the generalized impulse response functions (GIRF) confirm the divergence between developed and less developed housing markets because the shocks to Shenzhen can notably raise its own house prices but have limited influence on other cities' house price. However, a house price diffusion pattern can be conformed because the propagation of the shocks is approximately in accordance with the distance decay.

Similar to most of the studies on the Asian market, this paper is also limited by the short time-period, which is a notable issue when our analysis is concerned with long-run properties. This short time period of observation warns us that the results should be treated with caution. However, these results should have relevance to investors, policy makers and regulators. First, the leading-lag relationship among regional housing markets and the house price diffusion pattern could be useful for investors and portfolio managers to adjust their real estate portfolio accordingly. Second, a few implications can be drawn for policy makers and regulators. The lack of market convergence in the long-run could suggest that a local market-oriented housing policy will be more appropriate than a unified national policy. Indeed, this supposition has attracted the attention of policy makers. Recently, the central government announced a new policy to stimulate the housing market by reducing the down payment for second homes from 30% to 20%. An innovation of this policy is that it allows the local governments of four first-tier cities, Beijing, Shanghai, Shenzhen and Guangzhou, to make their own decisions according to the local market conditions[13]. Moreover, the results offer us a perspective on the degree of regional economic integration in the Pan-PRD. We observed that, during the following decade after the launch of Pan-PRD which aims to promote the integration of regional development, the developed eastern cities such as Shenzhen and Guangzhou, still appear to be deviated from the remaining cities, at least from the perspective of housing market integration. This suggests a need for regional policies that can facilitate the further decentralisation of economic activities, such as industrial policies.

References

Abbott, A., & Vita, G. D. (2013). Testing for long-run convergence across regional house prices in the UK: A pairwise approach. Applied Economics, 45(10), 1227-1238.

Alexander, C., & Barrow, M. (1994). Seasonality and cointegration of regional house prices in the UK. Urban Studies, 31(10), 1667-1689.

Ashworth, J., & Park, S. C. (1997). Modeling regional house prices in the UK. Scottish Journal of Political Economy, 44(3), 225-246.

Balcilar, M., Beyene, A., Gupta, R., & Seleteng, M. (2013). 'Ripple' effects in South African house prices. Urban Studies, 50(5), 876-894.

Berg, L. (2002). Prices on the second-hand market for Swedish family houses: Correlation, causation and determinants. International Journal of Housing Policy, 2(1), 1-24.

Blanco, F., Martín, V., & Vazquez, G. (2016). Regional house pice convergence in Spanish during the housing boom. Urban Studies, 53(4), 775-798.

Brady, R. R. (2014). The spatial diffusion of regional housing prices across U.S. states. Regional Science and Urban Economics, 46, 150-166.

Canarella, G., Miller, S., & Pollard, S. (2012). Unit roots and structural change: An

13 Document No [2015] 128 issued by Ministry of Finance of Peoples' Republic of China.

application to US house price indices. Urban Studies, 49(4), 757-776.
Chen, P.-F., Chien, M.-S., & Lee, C.-C. (2011). Dynamic modeling of regional house price diffusion in Taiwan. Journal of Housing Economics, 20(4), 315-332.
Clapp, J. M., Dolde, W., & Tirtiroglu, D. (1995). Imperfect information and investor inferences from housing price dynamics. Real Estate Economics, 23(3), 239-269.
Cook, S. (2003). The convergence of regional house prices in the UK. Urban Studies, 40(11), 2285-2294.
Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427-431.
Drake, L. (1995). Testing for convergence between UK regional house prices. Regional Studies, 29(4), 357-366.
Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. Econometrica, 64(4), 813-836.
Enders, W., & Granger, C. W. J. (1998). Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. Journal of Business & Economic Statistics, 16(3), 304-311.
Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. Econometrica, 55(2), 251-276.
Giussani, B., & Hadjimatheou, G. (1991). Modeling regional house prices in the United Kingdom. Papers in Regional Science, 70(2), 201-219.
Hanink, D. M., Cromley, R. G., & Ebenstein, A. Y. (2012). Spatial variation in the determinants of house prices and apartment rents in China. Journal of Real Estate Finance and Economics, 45(2), 347-363.
Helgers, R., & Buyst, E. (2016). Spatial and temporal diffusion of housing prices in the presence of a linguistic border: Evidence from Belgium. Spatial Economic Analysis, 11(1), 92-122.
Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of Urban Economics, 69(1), 2-23.
Holmes, M. J., & Grimes, A. (2008). Is there long-run convergence among regional house prices in the UK. Urban Studies, 45(8), 1531-1544.
Holmes, M. J., Otero, J., & Panagiotidis, T. (2011). Investigating regional house price convergence in the United States: Evidence from a pair-wise approach. Economic Modelling, 28(6), 2369-2376.
Huang, F., Li, C., & Li, Y. (2010a). Ripple Effect of Housing Prices among Chinese Deputy Provincial Cities Based on an Alternative Approach. International Journal of Business Administration, 1(1), 19-24.
Huang, F., Zhou, Y., & Li, C. (2010b). Ripple effect of housing prices fluctuations among nine cities of China. Management Science and Engineering, 4(3), 41-54.
Kim, Y. S., & Rous, J. J. (2012). House price convergence: Evidence from US state and metropolitan area panels. Journal of Housing Economics, 21(2), 169-186.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root: How sure are we that economic series have a unit root. Journal of Econometrics, 54(1-3), 159-178.

Lean, H. H., & Smyth, R. (2013). Regional house prices and the ripple effect in Malaysia. Urban Studies, 50(5), 895-922.

Lee, C.-C., & Chien, M.-S. (2011). Empirical modeling of regional house prices and the ripple effect. Urban Studies, 48(10), 2029-2047.

Li, J., & Li, H. (2011). The spatial diffusion and convergence of urban housing price in the Pearl River Delta. Journal of Chongqing University (Social Science Edition), 17(6), 17-23.

Liao, W.-C., Zhao, D., Lim, L. P., & Wong, G. K. M. (2015). Foreign liquidity to real estate market: Ripple effect and housing price dynamics. Urban Studies, 52(1), 138-158.

Luo, Z. Q., Liu, C., & Picken, D. (2007). Housing price diffusion pattern of Australia's state capital cities. International Journal of Strategic Property Management, 11(4), 227-242.

MacDonald, R., & Taylor, M. P. (1993). Regional house prices in Britain: long-run relationships and short-run dynamics. Scottish Journal of Political Economy, 40(1), 43-55.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. Journal of Applied Econometrics, 11(6), 601-618.

Meen, G. (1996). Spatial aggregation, spatial dependence and predictability in the UK housing market. Housing Studies, 11(3), 345-372.

Meen, G. (1999). Regional house prices and the ripple effect: A new interpretation. Housing Studies, 14(6), 733-753.

Miles, W. (2015). Regional house price segmentation and convergence in the US: A new approach. Journal of Real Estate Finance and Economics, 50(1), 113-128.

Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. Economics Letters, 58(1), 17-29.

Pollakowski, H. O., & Ray, T. S. (1997). Housing price diffusion patterns at different aggregation levels: An examination of housing market efficiency. Journal of Housing Research, 8(1), 107-124.

Stevenson, S. (2004). House price diffusion and inter-regional and cross-border house price dynamics. Journal of Property Research, 21(4), 301-320.

Su, C. W., & Chang, H. L. (2013). Is income converging in China? A flexible Fourier stationary test. Economics of Transition, 21(2), 341-356.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66, 225-250.

Wang, S., Yang, Z., & Liu, H. (2008). Empirical study on urban housing price interactions among regional marekts in China. Research on Financial and Economic Issues, 6, 122-129.

Wu, J., & Deng, Y. (2015). Intercity information diffusion and price discovery in housing markets: Evidence from Google Searches. The Journal of Real Estate Finance and Economics, 50(3), 289-306.
Wu, J., Deng, Y., & Liu, H. (2014). House price index construction in the nascent housing market: The case of China. The Journal of Real Estate Finance and Economics, 48(3), 522-545.
Yunus, N., & Swanson, P. E. (2013). A closer look at the U.S. housing market: Modelling relationships among regions. Real Estate Economics, 41(3), 542-568.

Appendices

Appendix A

TABLE A1 The results of pairwise Johansen cointegration test

	Shenzhen	Fuzhou	Xiamen	Nanchang	Changsha	Nanning	Guiyang	Kunming	Chengdu
Shenzhen									
Fuzhou	7.60								1.95
Xiamen	8.69	7.44							
Nanchang	7.25	11.11	9.05		7.19**		0.16	1.57	
Changsha	10.02	9.83	7.27	20.27**			4.45*	8.80**	
Nanning	10.74	12.66	9.01	12.77	10.26			2.05	
Guiyang	6.81	14.52	10.30	18.03*	18.22*	13.74		0.01	
Kunming	2.40	13.20	6.46	16.89*	25.04**	18.55*	17.07*		
Chengdu	6.75	16.61*	4.12	15.46	13.93	12.94	15.03	13.22	

Notes: The lower triangle cells report the trace statistic of the pairwise Johansen cointegration test under the null hypothesis $H_0 : r = 0$. The test is based on a VAR(3) specification, with unrestricted intercept and no trend in VAR. For the co-integrated city pairs, the upper triangle cells report the log-likelihood ratio (LR) test for the cointegrating vector restriction (1,-1). * and ** denote 5% and 1% significance level, respectively.

TABLE A2 Pairwise convergence with pre-specified coefficients (1,-1) base on the MTAR unit root test

	Shenzhen	Fuzhou	Xiamen	Nanchang	Changsha	Nanning	Guiyang	Kunming	Chengdu
Shenzhen									
Fuzhou	2.08(1)								
Xiamen	3.72(1)	0.70(2)							
Nanchang	1.54(1)	1.43(1)	0.48(1)						
Changsha	3.19(1)	3.93(2)	1.40(2)	4.87(1)					
Nanning	0.18(1)	2.76(1)	1.96(2)	2.15(1)	0.32(1)				
Guiyang	1.90(1)	2.38(1)	1.55(2)	4.73(1)	4.66(1)	3.03(1)			
Kunming	0.91(1)	5.17*(1)	2.17(1)	7.81**(1)	8.29**(1)	6.01*(1)	4.10(1)		
Chengdu	0.24(1)	4.34(1)	4.60(2)	1.67(1)	3.72(1)	5.96*(1)	1.20(1)	2.71(1)	

Notes: The null hypothesis of no convergence is tested based on the residual from $p_{it} - p_{jt}$. In all the models a constant is included, and the lag length is shown in the parentheses. The 5% and 1% critical values are 5.02 and 7.10, respectively. * and ** denote 5% and 1% significance level, respectively.

Appendix B: Bootstrap GIRF confidence intervals

The methods for computing the bootstrap confidence intervals of the generalized spatial-temporal impulse response functions are borrowed from Holly et al.(2011). The estimated model (7) is first used to generate B bootstrap samples. The bth bootstrap sample can be obtained by the following Data Generation Process (DGP)

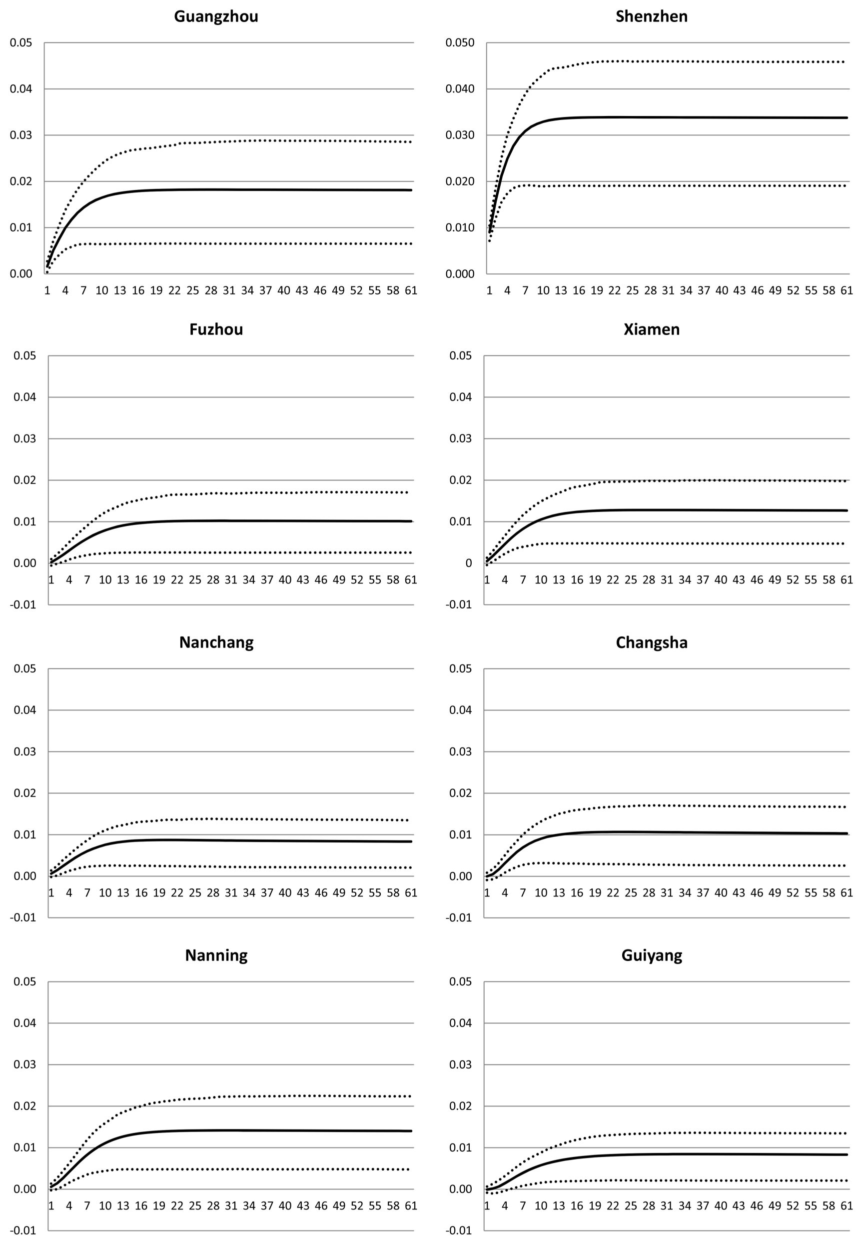
$$\mathbf{p}_t^{(b)} = \hat{\mathbf{a}} + \hat{\mathbf{\Phi}}_1 \mathbf{p}_{t-1}^{(b)} + \hat{\mathbf{\Phi}}_2 \mathbf{p}_{t-2}^{(b)} + \hat{\boldsymbol{\epsilon}}_t^{(b)}, \tag{B.1}$$

where $\hat{\boldsymbol{\epsilon}}_t^{(b)} = \hat{\mathbf{\Sigma}}^{1/2} \mathbf{v}_t^{*(b)}$. The elements of $\mathbf{v}_t^{*(b)}$ are recursively replaced by the values that are randomly drawn from the transformed residual matrix $\hat{\mathbf{\Sigma}}^{-1/2} (\hat{\boldsymbol{\epsilon}}_1, \hat{\boldsymbol{\epsilon}}_2, \ldots, \hat{\boldsymbol{\epsilon}}_t)$. Note that in equation (B.1), the first 2 observations are replaced by the original data.

When obtaining the bootstrap sample $\mathbf{p}_t^{(b)}$, we estimate the model (7) again and produce the b_{th} bootstrap GIRF

$$\mathbf{g}_i^{(b)}(h) = \frac{\hat{\mathbf{\Psi}}_h^{(b)} \hat{\mathbf{\Sigma}}^{(b)} \mathbf{e}_i}{\sqrt{\hat{\sigma}_{ii}^{(b)}}}, \quad \text{for} \quad i = 1, \cdots, n; h = 0, 1, \cdots, H. \tag{B.2}$$

The lower and upper bands of $100(1-\alpha)$% confidence interval are equivalent to the $\alpha/2$ and $1-\alpha/2$ quantiles of B $\mathbf{g}_i^{(b)}(h)$ for each i and h.


Guangzhou
0.05
0.04
0.03
0.02
0.01
0.00
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Shenzhen
0.050
0.040
0.030
0.020
0.010
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Fuzhou
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Xiamen
0.05
0.04
0.03
0.02
0.01
0
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Nanchang
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Changsha
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Nanning
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Guiyang
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61

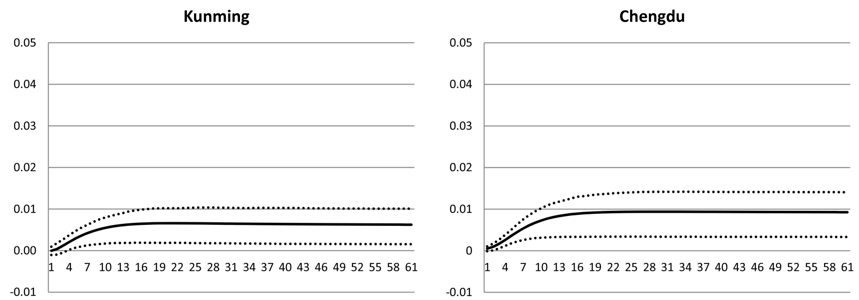
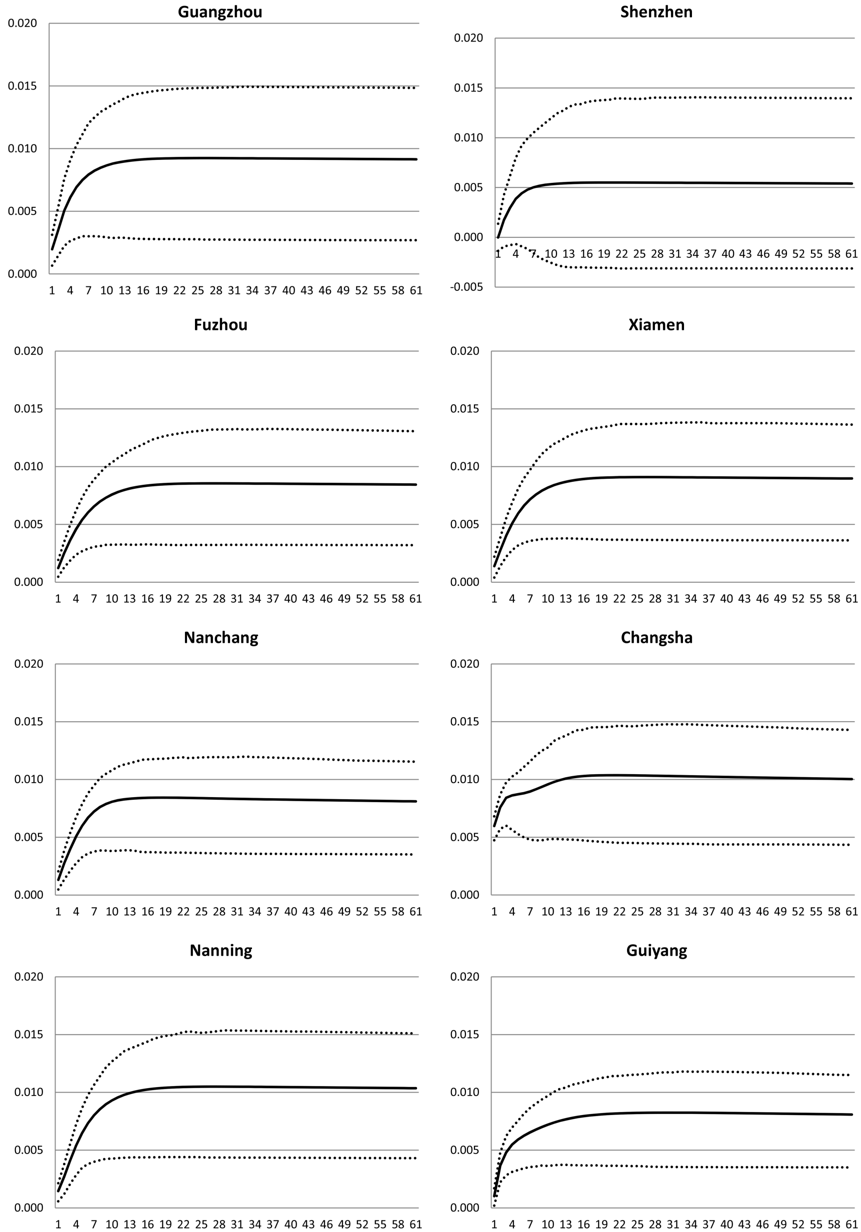


FIGURE B1 90% bootstrap error bounds for GIRF of a positive unit shock (one s.e.) to Shenzhen house prices (based on 1000 bootstrap samples)


Guangzhou
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Shenzhen
0.020
0.015
0.010
0.005
0.000
-0.005
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Fuzhou
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Xiamen
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Nanchang
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Changsha
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Nanning
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61
Guiyang
0.020
0.015
0.010
0.005
0.000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61

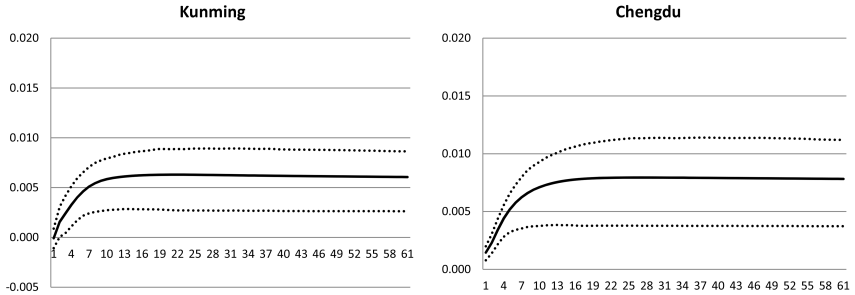


FIGURE B2 90% bootstrap error bounds for GIRF of a positive unit shock (one s.e.) to Changsha house prices (based on 1000 bootstrap samples)

6 Accounting for spatial variation of land prices in hedonic imputation house price indexes: A semi-parametric approach

Submitted for review. Co-author: Jan de Haan.

Abstract: Location is capitalized into the price of the land the structure of a property is built on, and land prices can be expected to vary significantly across space. We account for spatial variation of land prices in hedonic house price models using geospatial data and a semi-parametric method known as mixed geographically weighted regression. To measure the impact on aggregate price change, quality-adjusted (hedonic imputation) house price indexes are constructed for a small city in the Netherlands and compared to price indexes based on more restrictive models, using postcode dummy variables or no location information at all. We find that, although taking spatial variation of land prices into account improves the model performance, the Fisher house price indexes based on the different hedonic models are almost identical. The land and structures price indexes, on the other hand, are sensitive to the treatment of location.

Keywords: Geospatial information, hedonic modeling, land and structure prices, mixed geographically weighted regression, residential property.

JEL: C14, C33, C43, E31, R31.

§ 6.1 Introduction

Housing markets have two distinct features: every house is unique and houses are sold infrequently. This is problematic for the construction of house price indexes because the usual matched-model method, where the prices of goods are tracked over time, breaks down. Hedonic regression methods and repeat sales methods deal with these problems. The uniqueness of properties is mainly due to location. Within a single neighborhood, the value of two properties with similar structures can differ significantly, depending on the exact location.

Repeat sales indexes fully control for location since they track the prices of the 'same' properties over time (in a regression framework). The problem with repeat sales methods is threefold. First, because they only use matched pairs of houses during the sample period, these methods ignore single sales and are therefore inefficient and prone to sample selection bias. Second, standard repeat sales methods do not adjust for quality changes of the individual houses. Third, these methods cannot provide information on the shadow prices of the property characteristics and thus do not allow the estimation of, for example, price indexes of the land the structure sits on. Given these problems with repeat sales methods, we focus on hedonic regression methods.

Traditional hedonic price indexes also have a number of disadvantages. First, data on housing characteristics must be available. Second, location is typically included in hedonic models at some aggregated level, such as postcode areas, rather than at the individual property level, potentially leading to 'location bias', which is a form of omitted variable bias. Third, land is often not included as an independent variable, again potentially giving rise to bias and making it impossible to estimate price indexes for land. Geospatial data, i.e. information on the exact location of the dwellings in terms of geographic coordinates such as longitude and latitude, can help attenuate the latter disadvantages. Our aim is to show how this can be done and how hedonic house price indexes can be constructed accordingly.

A general problem with the estimation of hedonic models for housing is omitted variables bias. Not properly accounting for location can be a major cause of bias and often leads to spatial autocorrelation of the error terms. As mentioned above, the easiest way to deal with the problem is to include dummy variables for postcode areas. Another straightforward approach, which has also been frequently investigated empirically, is to include explanatory variables for all kinds of amenities. While being of interest because it provides information on the shadow prices of the amenities, this method is rather data intensive and, just like the inclusion of dummy variables, cannot fully capture location effects. As a result, some omitted variables bias and spatial autocorrelation will likely remain.

In recent years, more sophisticated methods have been put forward to handle the problem of spatial autocorrelation. Spatial error models attempt to explicitly model the spatial autocorrelation while spatial lag models include the value of neighbor properties in the model. Both methods can be used in a time dummy hedonic framework, where the model is estimated on pooled data for the whole sample period and price indexes are computed from the time dummy coefficients (Hill et al. 2009; Dorsey et al. 2010). Also within the time dummy hedonic framework, Thanos et al. (2016) comprehensively control for both spatial and temporal effects in computing house price index. It is also possible to apply these spatial (and temporal) methods in a hedonic imputation framework (Rambaldi and Rao, 2011; 2013). Another method uses a spatio-temporal filter which eliminates spatial autocorrelation in order to estimate an index for a dwelling with specific characteristics (Pace et al., 1998; Tu et al., 2004; Sun et al., 2005).

A disadvantage of the above parametric methods is that a spatial weight matrix has to be specified a priori but that its precise structure is unknown. Nonparametric or semi-parametric methods are more suitable to account for spatial dependence. Semi-parametric methods have become increasingly popular. The effect of variables relating to location, for example, can be estimated nonparametrically in 'characteristics space', whereas the effect of variables relating to the structure of the property can be estimated parametrically, as in traditional hedonic models.

In this paper, we assume that location affects the price of land but not the price of structures. That is, we postulate that land prices vary across space whereas the price of structures is 'fixed'. We deal with this type of spatial nonstationarity using a semi-parametric approach known as Mixed Geographically Weighted Regression (MGWR) in which the land prices are estimated by Geographically Weighted Regression (GWR), a nonparametric method proposed by Brunsdon et al. (1996) and Fotheringham et al. (1998b). An additional advantage is that we will be able to plot a continuous surface of land prices.

Apart from the fact that it deals with spatial nonstationarity in a straightforward way, GWR enables us to model the local form of autocorrelation. Moreover, it allows land prices to vary not only across space but also across time by estimating the model for each period separately. The latter is a prerequisite for the construction of hedonic imputation price indexes. In conclusion, (M)GWR is a flexible approach, which can be seen as a generalization of traditional hedonic methods.

We are specifically targeting statistical agencies engaged in the compilation of house price indexes. This has several consequences. The agencies should have access to geocoded data, but this is hardly a problem these days. The methods applied should be relatively easy to explain. Most importantly, the price indexes should be non-revisable. This means that the use of the time dummy method, where previously published index numbers change when the sample period is extended and new data is added, is ruled out. This strengthens the case for constructing hedonic imputation indexes.

Furthermore, our paper tries to fill a gap in the recent *Handbook on Residential Property Price Indices* (Eurostat et al., 2013) in which the use of geospatial data in the estimation of hedonic house price models is not very well covered. The Handbook uses data for detached dwellings sold in the Dutch city of “A” from the first quarter of 2005 to the second quarter of 2008 to illustrate the various methods. We exploit sales data for the city of “A” also but extend the data set in three dimensions. We have data from the first quarter of 1998 to the fourth quarter of 2007, so our data set covers a period of 10 years. Note that we will use annual rather than quarterly data in our empirical work. The range of characteristics for the structures is broader than that in the

Handbook. Finally, we include houses other than detached dwellings.

The paper proceeds as follows. Section 6.2 outlines some basic ideas. Our hedonic model is linear, with non-transformed property price as the dependent variable and size of land and size of structures as explanatory variables. A normalized version, with price per square meter of living space as the dependent variable, is discussed as well. We also address the inclusion of additional characteristics to describe the quality of structures, including age of the structure to adjust for depreciation. Section 6.3 describes how we treat location. As mentioned before, location is capitalized into the price of land, and we would expect land prices to differ at the property level. The GWR and MGWR models and the way in which they are estimated are explained in detail. Section 6.4 shows how we calculate hedonic imputation indexes. Section 6.5 presents empirical evidence for the Dutch city of "A" and discusses the results. Section 6.6 concludes and identifies potential improvements.

§ 6.2 A simplification of the 'builder's model'

§ 6.2.1 Some basic ideas

Our starting point is the 'builder's model' proposed by Diewert et al. (2011; 2015). It is assumed that the value of a property i in period t, p_i^t, can be split into the value v_{iL}^t of the land the structure sits on and the value v_{iS}^t of the structure:

$$p_i^t = v_{iL}^t + v_{iS}^t. \tag{1}$$

The value of land for property i is equal to the plot size in square meters, z_{iL}^t, times the price of land per square meter, α^t, and the value of the structure equals the size of the structure in square meters of living space, z_{iS}^t, times the price of structures per square meter, β^t. After adding an error term u_i^t with zero mean, model (1) becomes

$$p_i^t = \alpha^t z_{iL}^t + \beta^t z_{iS}^t + u_i^t. \tag{2}$$

The (shadow) prices of both land and structures in (2) are the same for all properties, irrespective of their location. In section 6.3 we relax this assumption and allow for spatial variation of, in particular, the price of land. The 'builder's model' takes depreciation of the structures into account, a topic we address in section 6.2.2.

Equation (2) can be estimated on data of a sample S^t of properties sold in period t. This approach, however, suffers from at least three problems. First, the model has no intercept term, which hampers the interpretation of R^2 and the use of standard tests in Ordinary Least Squares (OLS) regression. Second, a high degree of collinearity between land size and structure size can be expected, so that α^t and β^t will be estimated with low precision. Finally, heteroskedasticity is likely to occur since the absolute value of the errors tends to grow with increasing property prices.

Our next step is to divide the left hand side and right hand side of equation (2) by structure size z^t_{iS}, giving

$$p^{t*}_i = \alpha^t r^t_i + \beta^t + \epsilon^t_i, \tag{3}$$

where $p^{t*}_i = p^t_i / z^t_{iS}$ is the *normalized property price*, i.e. the value of the property per square meter of living space, $r^t_i = z^t_{iL} / z^t_{iS}$ denotes the ratio of plot size to structure size, and $\epsilon^t_i = u^t_i / z^t_{iS}$. This resolves the first two problems as the model now has an intercept term and a single explanatory variable.

However, the normalization is unlikely to resolve the issue of unstable parameter estimates. Estimating (3) by OLS regression is equivalent to estimating (2) by Weighted Least Squares (WLS) using weights equal to $1/z^t_{iS}$. That is, dividing by z^t_{iS} adjusts for heteroskedasticity when the error variance in (2) would be proportional to the square of structure size. This kind of error variance seems quite extreme, so this weighting system may not help reduce the heteroskedasticity problem. Also, the ratios r^t_i (as well as the normalized values p^{t*}_i) may exhibit relatively little dispersion.

Some statistical agencies publish changes in normalized rather than unadjusted property prices, often prices per square meter of structures, to adjust for compositional change of the properties sold. We do not recommend this approach because it is changes in unadjusted property prices and price changes most people will be interested in. Yet, given that (3) is a straightforward regression model, including an intercept term, we do favor specification (3) over (2).

§ 6.2.2 Adding structures characteristics

A potential weakness of hedonic modeling for housing is omitted variables, leading to biased (OLS) parameter estimates and predicted prices. Omitted variables in the models (2) and (3) can relate to land or structures. Improving the treatment of land is the topic of section 6.3. In the present section, we discuss the inclusion of additional characteristics for structures. There are two main issues: depreciation and renovation of structures have been ignored so far, and the use of size as the only price-determining feature seems too simplistic.

Following Diewert et al.(2015), we initially assume a straight-line depreciation model. The adjusted value of the structure is $\beta^t \left(1 - \delta^t \alpha^t_i\right) z^t_{iS}$, where δ^t is the depreciation rate and α^t_i is age of the structure. Information on renovations at the level of individual dwellings is unavailable so that $-\delta^t \alpha^t_i$ measures the effect of *net* depreciation, i.e. the combined effect of 'true' depreciation and renovation. Written in linear form, the adjusted structures value is $\beta^t z^t_{iS} - \beta^t \delta^t \alpha^t_i z^t_{iS}$. Adding the second term to the right-hand side of equation (2) yields

$$p^t_i = \alpha^t z^t_{iL} + \beta^t z^t_{iS} - \beta^t \delta^t \alpha^t_i z^t_{iS} + u^t_i. \tag{4}$$

We do not know the exact age of the structures, but we do know the building period in decades, from which we can calculate approximate age in decades. Thus, age in our data set is an ordinal (categorical) variable. The net depreciation rate is of course ordinal as well. Using multiplicative dummy variables D_{ia}^t that take on the value 1 if in period t property i belongs to age category a ($a = 1, \cdots, A$) and the value 0 otherwise, and after reparameterizing such that $\beta^t z_{iS}^t$ is no longer a separate term, model (4) becomes $p_i^t = \alpha^t z_{iL}^t + \sum_{a=1}^{A} \gamma^t D_{ia}^t z_{iS}^t + u_i^t$. To be able to use standard techniques, we modify this model as follows:

$$p_i^t = \alpha^t z_{iL}^t + \sum_{a=1}^{A} \gamma_a^t D_{ia}^t z_{iS}^t + u_i^t. \tag{5}$$

No restrictions are placed on the parameters γ_a^t, and the new functional form is neither continuous nor smooth. This is somewhat problematic from a theoretical point of view, because it is at odds with the initial straight-line depreciation model. On the other hand, our approach introduces some flexibility. Age of the structures is not only important for modeling depreciation, it can also be seen as an attribute of the dwelling itself in that houses built in a particular decade are more in demand than other houses, perhaps for their architectural style or other age-related attributes.

Diewert et al. (2015) also show how to incorporate the number of rooms. The new value of the structures becomes $\beta^t \left(1 - \delta^t a_i^t\right) \left(1 + \mu^t z_{iR}^t\right) z_{iS}^t$, where μ^t is the parameter for the number of rooms z_{iR}^t. The linear form for this expression is $\beta^t z_{iS}^t + \beta^t \mu^t z_{iR}^t z_{iS}^t - \beta^t \delta^t a_i^t z_{iS}^t - \beta^t \delta^t \mu^t a_i^t z_{iR}^t z_{iS}^t$. Using dummies D_{ir}^t for the number of rooms with the value 1 if in period t the property belongs to category r ($r = 1, \cdots, R$) and the value 0 otherwise, and reparameterizing again, the extended version of (5)becomes

$$p_i^t = \alpha^t z_{iL}^t + \sum_{a=1}^{A} \gamma_a^t D_{ia}^t z_{iS}^t + \sum_{r=1}^{R} \lambda_r^t D_{ir}^t z_{iS}^t + \sum_{a=1}^{A} \sum_{r=1}^{R} \eta_{ar}^t D_{ia}^t D_{ir}^t z_{iS}^t + u_i^t. \tag{6}$$

Next, in order to save degrees of freedom, we ignore the 'second-order' effects due to the interaction terms $D_{ia}^t D_{ir}^t$, yielding

$$p_i^t = \alpha^t z_{iL}^t + \sum_{a=1}^{A} \gamma_a^t D_{ia}^t z_{iS}^t + \sum_{r=1}^{R} \lambda_r^t D_{ir}^t z_{iS}^t + u_i^t = \alpha^t z_{iL}^t + \left[\sum_{a=1}^{A} \gamma_a^t D_{ia}^t + \sum_{r=1}^{R} \lambda_r^t D_{ir}^t\right] z_{iS}^t + u_i^t \tag{7}$$

The second expression shows that the price of structures, i.e. the price per square meter of living space, equals $\gamma_a^t + \lambda_r^t$ for properties in age class a ($a = 1, \cdots, A$) and category r ($r = 1, \cdots, R$) for number of rooms. A high degree of multicollinearity can occur among the various structures components, but we do not worry about this because we are only interested in the combined effect. Multicollinearity between these components and plot size might still be a problem though. Dividing the first expression in (7) by z_{iS}^t gives

$$p_i^{t*} = \theta^t + \alpha^t r_i^t + \sum_{a=1}^{A-1} \gamma_a^t D_{ia}^t + \sum_{r=1}^{R-1} \lambda_r^t D_{ir}^t + \epsilon_i^t \tag{8}$$

We included an intercept term θ^t and then excluded dummy variables for age class A and category R for the number of rooms to identify the model.

Model (8) is a straightforward estimating equation for the overall property price per square meter of living space. Additional categorical variables for the structures can be incorporated in a similar way as the number of rooms. As a matter of fact, we will use type of house instead of the number of rooms in our empirical work.

§ 6.3 Land and spatial nonstationarity

§ 6.3.1 Location and the price of land

Location is the most important omitted variable in the hedonic models presented so far. In many empirical studies, location is treated as a 'separate characteristic' by including additive locational dummy variables in models for the *overall* property price. This is not the solution we prefer. Location is definitely capitalized into property prices. However, at least within relatively small regions or cities, the price of structures is most likely to be more or less constant across space. It is the price of the land the structure is built on that can vary significantly across different locations, even within a single neighborhood. The question then arises as to how this spatial variation, or spatial nonstationarity as it is sometimes referred to, in the price of land should be modeled.

We could make the simplifying assumption that the price of land varies across postcode areas but is the same within each postcode area k ($k = 1, \cdots, K$) and denoted by α_k^t. This idea is widely used in empirical studies, such as Diewert and Shimizu (2013) who estimated the 'builder's model' for Tokyo. Using *multiplicative* postcode dummy variables D_{ik}, which take on the value of 1 if property i belongs to k and the value 0 otherwise, an improved version of model (7) for the unadjusted property price is

$$p_i^t = \sum_{k=1}^{K} \alpha_k^t D_{ik} z_{iL}^t + \sum_{a=1}^{A} \gamma_a^t D_{ia}^t z_{iS}^t + \sum_{r=1}^{R} \lambda_r^t D_{ir}^t z_{iS}^t + u_i^t, \tag{9}$$

and an improved version of model (8) for the normalized property price is

$$p_i^{t*} = \theta^t + \sum_{k=1}^{K} \alpha_k^t D_{ik} r_i^t + \sum_{a=1}^{A-1} \gamma_a^t D_{ia}^t + \sum_{r=1}^{R-1} \lambda_r^t D_{ir}^t + \epsilon_i^t \tag{10}$$

The assumption of equal land prices within postcode areas could be too crude, depending of course on the level of detail of the postcode system. Generalized versions of the models (9) and (10) are obtained by assuming that the price of land can differ at the individual property level, i.e. at the micro location. We denote the property-specific land price by α_i^t, yielding

$$p_i^t = \alpha_i^t z_{iL}^t + \sum_{a=1}^{A} \gamma_a^t D_{ia}^t z_{iS}^t + \sum_{r=1}^{R} \lambda_r^t D_{ir}^t z_{iS}^t + u_i^t \tag{11}$$

and

$$p_i^{t*} = \theta^t + \alpha_i^t r_i^t + \sum_{a=1}^{A-1} \gamma_a^t D_{ia}^t + \sum_{r=1}^{R-1} \lambda_r^t D_{ir}^t + \epsilon_i^t. \tag{12}$$

Models (11) and (12) obviously cannot be estimated by standard regression techniques. In section 6.3.2 we will discuss a semi-parametric approach that does allow us to estimate these models. Because the method utilizes data on the prices of neighboring properties (in addition to the price of property *i* itself) to estimate α_i^t, it is not necessarily true that the use of models (11) or (12) will lead to aggregate price indexes that are very different from those found by using models (9) or (10).

§ 6.3.2 Mixed Geographically Weighted Regression

One method that deals with spatial nonstationarity of property prices is the 'expansion method' (Casetti, 1972; Jones and Casetti, 1992). The property price, or in our case the price of land, can be viewed as an unknown function of the property's location in terms of latitude x_i and longitude y_i or a similar geographic coordinate system. This function can be approximated using a Taylor-series expansion of some order; typically, second-order approximations are applied. The expansion method makes use of geospatial data but is basically parametric as it calibrates a prespecified parametric model for the trend of land prices across space (Fotheringham et al. 1998a).

The method we will apply, referred to as *Geographically Weighted Regression* (GWR), deals with spatial nonstationarity in a truly nonparametric fashion (Brunsdon et al. 1996; Fotheringham et al. 1998b). Let us remove the structures characteristics from model (11) for a moment and thus consider land as the only independent variable. Using $\alpha_i = \alpha\,(x_i, y_i)$, the model becomes

$$p_i = \alpha(x_i, y_i) z_{iL} + u_i. \tag{13}$$

Note that we have dropped the superscript *t* for convenience; it should be clear that we estimate all models for each time period separately. Note further that land prices can be estimated for each location in the area under study, not just for the sample observations, enabling us to plot a continuous surface of land prices.

Model (13) can be estimated using a moving kernel window approach, which is essentially a form of WLS regression. In order to obtain an estimate for the price of land $\alpha(x_i, y_i)$ for property *i*, a weighted regression is run where each related observation *j*, i.e. each neighboring property, is given a weight $w_{ij} (i \neq j)$. The weights should follow a monotonic decreasing function of distance d_{ij} between (x_i, y_i) and (x_j, y_j). There is a range of possible functional forms from which we have chosen the frequently-used *bi-square function*

$$w_{ij} = \begin{cases} \left(1 - d_{ij}^2/h^2\right)^2 & \text{if} \quad d_{ij} < h \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where h denotes the bandwidth. The choice of bandwidth involves a trade-off between bias and variance. A larger bandwidth generates an estimate with larger bias but smaller variance whereas a smaller bandwidth produces an estimate with smaller bias but larger variance. The usual solution is to select the optimal bandwidth by minimizing the *cross-validation* (CV) statistic

$$CV(h) = \sum_{i=1}^{n} \left[p_i - \hat{p}_{\neq i}(h)\right]^2 \tag{15}$$

where $\hat{p}_{\neq i}(h)$ is the predicted price of property i where the observations for i have been omitted from the calibration process.

The nonparametric GWR approach to dealing with spatial nonstationarity of the price of land has to be adjusted for the fact that models (11) and (12) include structures characteristics with spatially fixed parameters. This leads to a specific instance of the semi-parametric Mixed GWR (MGWR) approach discussed by Brunsdon et al. (1999), where some parameters are spatially fixed and the remaining parameters are allowed to vary across space. To outline the estimation procedure, it will be useful to change over to matrix notation. Denoting the number of observations by n, model (11) can be written in matrix form as

$$\mathbf{P} = \mathbf{Z}_L \otimes \boldsymbol{\alpha} + \mathbf{Z}_S \boldsymbol{\beta} + \mathbf{u}, \tag{16}$$

where $\boldsymbol{\alpha} = (\alpha(x_1, y_1), \alpha(x_2, y_2), \cdots, \alpha(x_n, y_n))^T$ is a vector of land prices to be estimated, $\otimes$ is an operator that multiplies each element of $\boldsymbol{\alpha}$ by the corresponding element of $\mathbf{Z}_L$, $\mathbf{Z}_S$ is the matrix of structures characteristics included in model (11), given by

$$\mathbf{Z}_S = \begin{pmatrix} D_{11}z_{1S} & D_{12}z_{1S} & \cdots & D_{1j}z_{1S} \\ D_{21}z_{2S} & D_{22}z_{2S} & \cdots & D_{2j}z_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}z_{nS} & D_{n2}z_{nS} & \cdots & D_{nj}z_{nS} \end{pmatrix},$$

and $\beta = (\beta_1, \beta_2, \cdots, \beta_n)^T$ is the vector of coefficients relating to $\mathbf{Z}_S$ to be estimated.

We follow Fotheringham et al. (2002), who proposed an estimation method that is less computationally intensive than the method described by Brunsdon et al. (1999). We will broadly describe the actual estimation procedure and present the estimators for the parameters, but we do not provide the exact MGWR algorithm. For details, the readers can refer to Fotheringham et al. (2002), Mei et al. (2006), and Geniaux and Napoléone (2008). To economize on notation, we write the GWR projection or hat matrix as

$$\mathbf{S} = \begin{pmatrix} z_{1L} \left[\mathbf{Z}_L^T \mathbf{W}(x_1, y_1) \mathbf{Z}_L\right]^{-1} \mathbf{Z}_L^T \mathbf{W}(x_1, y_1) \\ z_{2L} \left[\mathbf{Z}_L^T \mathbf{W}(x_2, y_2) \mathbf{Z}_L\right]^{-1} \mathbf{Z}_L^T \mathbf{W}(x_2, y_2) \\ \vdots \\ z_{nL} \left[\mathbf{Z}_L^T \mathbf{W}(x_n, y_n) \mathbf{Z}_L\right]^{-1} \mathbf{Z}_L^T \mathbf{W}(x_n, y_n) \end{pmatrix}$$

where $\mathbf{W}(x_i, y_i) = \mathrm{diag}\left[w_1(x_i, y_i), w_2(x_i, y_i), \cdots, w_n(x_i, y_i)\right]$. The calibration of the model consists of four steps:

1. regressing each column of $\mathbf{Z}_S$ against $\mathbf{Z}_L$ using the GWR calibration method and computing the residuals $\mathbf{Q} = (\mathbf{I} - \mathbf{S})\mathbf{Z}_S$;

2. regressing the dependent variable $\mathbf{P}$ against $\mathbf{Z}_L$ using the GWR approach and then computing the residuals $\mathbf{R} = (\mathbf{I} - \mathbf{S})\mathbf{P}$;

3. regressing the residuals $\mathbf{R}$ against the residuals $\mathbf{Q}$ using OLS in order to obtain the estimates $\hat{\boldsymbol{\beta}} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}$;

4. subtracting $\mathbf{Z}_S\hat{\boldsymbol{\beta}}$ from $\mathbf{P}$ and regressing this part against $\mathbf{Z}_L$ using GWR to obtain estimates $\hat{\alpha}(x_i, y_i) = \left[\mathbf{Z}_L^T \mathbf{W}(x_i, y_i) \mathbf{Z}_L\right]^{-1} \mathbf{Z}_L^T \mathbf{W}(x_i, y_i)(\mathbf{P} - \mathbf{Z}_S\hat{\boldsymbol{\beta}})$.

The predicted values for the property prices can be expressed as

$$\hat{\mathbf{P}} = \mathbf{S}(\mathbf{P} - \mathbf{Z}_S\hat{\boldsymbol{\beta}}) + \mathbf{Z}_S\hat{\boldsymbol{\beta}} = \mathbf{L}\mathbf{P}, \tag{17}$$

with $\mathbf{L} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\mathbf{Z}_S \left[\mathbf{Z}_S^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{Z}_S\right]^{-1} \mathbf{Z}_S^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})$.

The parameter estimates and the predicted property prices obviously depend on the choice of weights, hence on the choice of bandwidth h. The optimal value for h is determined by minimizing the CV statistic given by (15). In the case of MGWR, the CV statistic is equivalent to (Mei et al., 2006)

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{p_i - \hat{p}_i(h)}{1 - l_{ii}(h)}\right]^2 \tag{18}$$

where $\hat{p}_i(h)$ is the predicted price for property i and $l_{ii}(h)$ is the i_{th} diagonal element of matrix $\mathbf{L}$ in equation (17).

§ 6.4 Hedonic imputation price indexes

This section addresses the issue of estimating quality-adjusted property price indexes. Suppose that sample data is available for periods $t = 0, \cdots, T$, where 0 is the base period (the starting period of the time series we want to construct), and suppose model (12) has been estimated separately for each period. The predicted property prices, obtained using MGWR, are given by $\hat{p}_i^t = \hat{\alpha}_i^t z_{iL}^t + \left[\hat{\theta}^t + \sum_{a=1}^{A-1} \hat{\gamma}_a^t D_{ia}^t + \sum_{r=1}^{R-1} \hat{\lambda}_r^t D_{ir}^t\right] z_{iS}^t$. For short, we write the predicted price of structures, $\hat{\theta}^t + \sum_{a=1}^{A-1} \hat{\gamma}_a^t D_{ia}^t + \sum_{r=1}^{R-1} \hat{\lambda}_r^t D_{ir}^t$, as $\hat{\beta}_i^t$ and the predicted overall property price as $\hat{p}_i^t = \hat{\alpha}_i^t z_{iL}^t + \hat{\beta}_i^t z_{iS}^t$ $(t = 0, \cdots, T)$.

We denote the sample of properties sold in the base period by S^0. The hedonic imputation Laspeyres property price index going from period 0 to period t is defined by

$$P_{Laspeyres}^{0t} = \frac{\sum_{i \in S^0} \hat{p}_i^{t(0)}}{\sum_{i \in S^0} \hat{p}_i^0} \tag{19}$$

Equation (19) may need some explanation. All quantities are equal to 1, reflecting the fact that each property is considered unique. The index is not affected by compositional change because it is based on a single sample. Most, if not all, of the properties sold in period 0 are not re-sold in period t, and the 'missing prices' have to be imputed by $\hat{p}_i^{t(0)}$. We have also replaced the observed base period prices p_i^0 by the predicted values $\hat{p}_i^0$, a method known as *double imputation*[1].

The $\hat{p}_i^{t(0)}$ are estimated period t constant-quality property prices, i.e. estimates of the prices that would prevail in period t for properties sold in period 0 if the properties' price-determining characteristics were equal to those of the base period, which serves to adjust for quality changes of the individual properties. These constant-quality prices are estimated by $\hat{p}_i^{t(0)} = \hat{\alpha}_i^t z_{iL}^0 + \hat{\beta}_i^{t(0)} z_{iS}^0$, where $\hat{\beta}_i^{t(0)} = \hat{\theta}^t + \sum_{a=1}^{A-1} \hat{\gamma}_a^t D_{ia}^0 + \sum_{r=1}^{R-1} \hat{\lambda}_r^t D_{ir}^0$ denotes the estimated constant-quality price of structures.

Substitution of $\hat{p}_i^0 = \hat{\alpha}_i^0 z_{iL}^0 + \hat{\beta}_i^0 z_{iS}^0$ and $\hat{p}_i^{t(0)} = \hat{\alpha}_i^t z_{iL}^0 + \hat{\beta}_i^{t(0)} z_{iS}^0$ into (19) yields

$$P_{Laspeyres}^{0t} = \frac{\sum_{i \in S^0} \left[\hat{\alpha}_i^t z_{iL}^0 + \hat{\beta}_i^{t(0)} z_{iS}^0\right]}{\sum_{i \in S^0} \left[\hat{\alpha}_i^0 z_{iL}^0 + \hat{\beta}_i^0 z_{iS}^0\right]} = \hat{s}_L^0 \frac{\sum_{i \in S^0} \hat{\alpha}_i^t z_{iL}^0}{\sum_{i \in S^0} \hat{\alpha}_i^0 z_{iL}^0} + \hat{s}_S^0 \frac{\sum_{i \in S^0} \hat{\beta}_i^{t(0)} z_{iS}^0}{\sum_{i \in S^0} \hat{\beta}_i^0 z_{iS}^0} \tag{20}$$

where $\sum_{i \in S^0} \hat{\alpha}_i^t z_{iL}^0 / \sum_{i \in S^0} \hat{\alpha}_i^0 z_{iL}^0$ is a price index of land and $\sum_{i \in S^0} \hat{\beta}_i^{t(0)} z_{iS}^0 / \sum_{i \in S^0} \hat{\beta}_i^0 z_{iS}^0$ is a price index of structures. Equation (20) decomposes the overall house price index into structures and land components; the weights $\hat{s}_L^0 = \sum_{i \in S^0} \hat{\alpha}_i^0 z_{iL}^0 / \sum_{i \in S^0} \left[\hat{\alpha}_i^0 z_{iL}^0 + \hat{\beta}_i^0 z_{iS}^0\right]$ and $\hat{s}_S^0 = \sum_{i \in S^0} \hat{\beta}_i^0 z_{iS}^0 / \sum_{i \in S^0} \left[\hat{\alpha}_i^0 z_{iL}^0 + \hat{\beta}_i^0 z_{iS}^0\right]$ are estimated shares of land and structures in the total value of property sales in period 0. The double imputation method ensures that the weights sum to unity.

The price indexes of land and structures in (20) are Laspeyres-type indexes and can be written as weighted averages of price relatives for the individual properties. For example, the Laspeyres price index of land can be written as $\sum_{i \in S^0} \hat{s}_{iL}^0 \left(\hat{\alpha}_{iL}^t / \hat{\alpha}_{iL}^0\right)$, where the weights $\hat{s}_{iL}^0 = \hat{\alpha}_i^0 z_{iL}^0 / \sum_{i \in S^0} \hat{\alpha}_i^0 z_{iL}^0$ for the price relatives $\hat{\alpha}_{iL}^t / \hat{\alpha}_{iL}^0$ reflect the shares of

1 Hill and Melser (2008) discuss different types of hedonic imputation methods in the context of housing. For a general discussion of the difference between hedonic imputation indexes and time dummy indexes, see Diewert et al. (2009) and de Haan (2010).

the properties in the estimated value of land (implicitly) sold in period 0. Properties with relatively large value shares, like properties in wealthy and sought-after neighborhoods with large plot sizes and high land prices, therefore have a big influence on the index.

An alternative to the Laspeyres index is the hedonic double imputation Paasche price index, defined on the sample S^t of properties sold in period $t (t = 1, \cdots, T)$:

$$P^{0t}_{Paasche} = \frac{\sum_{i \in S^t} \hat{p}^t_i}{\sum_{i \in S^t} \hat{p}^{0(t)}_i}. \tag{21}$$

The imputed constant-quality prices $\hat{p}^{0(t)}_i$ are estimates of the prices that would prevail in period 0 if the property characteristics were those of period t, which are estimated as $\hat{p}^{0(t)}_i = \hat{\alpha}^0_i z^t_{iL} + \hat{\beta}^{0(t)}_i z^t_{iS}$, where $\hat{\beta}^{0(t)}_i = \hat{\theta}^0 + \sum_{a=1}^{A-1} \hat{\gamma}^0_a D^t_{ia} + \sum_{r=1}^{R-1} \hat{\lambda}^0_r D^t_{ir}$ denotes the period 0 constant-quality price of structures. By substituting the constant-quality prices and the predicted prices $\hat{p}^t_i = \hat{\alpha}^t_i z^t_{iL} + \hat{\beta}^t_i z^t_{iS}$ into (21), the hedonic imputation Paasche index can be written as

$$P^{0t}_{Paasche} = \frac{\sum_{i \in S^t} \left[\hat{\alpha}^t_i z^t_{iL} + \hat{\beta}^t_i z^t_{iS}\right]}{\sum_{i \in S^t} \left[\hat{\alpha}^0_i z^t_{iL} + \hat{\beta}^{0(t)}_i z^t_{iS}\right]} = \hat{s}^{t(0)}_L \frac{\sum_{i \in S^t} \hat{\alpha}^t_i z^t_{iL}}{\sum_{i \in S^t} \hat{\alpha}^0_i z^t_{iL}} + \hat{s}^{t(0)}_S \frac{\sum_{i \in S^t} \hat{\beta}^t_i z^t_{iS}}{\sum_{i \in S^t} \hat{\beta}^{0(t)}_i z^t_{iS}} \tag{22}$$

where $\sum_{i \in S^t} \hat{\alpha}^t_i z^t_{iL} / \sum_{i \in S^t} \hat{\alpha}^0_i z^t_{iL}$ and $\sum_{i \in S^t} \hat{\beta}^t_i z^t_{iS} / \sum_{i \in S^t} \hat{\beta}^{0(t)}_i z^t_{iS}$ are Paasche price indexes of land and structures, which are weighted by $\hat{s}^{t(0)}_L = \sum_{i \in S^t} \hat{\alpha}^0_i z^t_{iL} / \sum_{i \in S^t} \left[\hat{\alpha}^0_i z^t_{iL} + \hat{\beta}^{0(t)}_i z^t_{iS}\right]$ and $\hat{s}^{t(0)}_S = \sum_{i \in S^t} \hat{\beta}^0_i z^t_{iS} / \sum_{i \in S^t} \left[\hat{\alpha}^0_i z^t_{iL} + \hat{\beta}^{0(t)}_i z^t_{iS}\right]$. The weights are now of a hybrid nature and reflect the shares of land and structures in the estimated total value of property sales in period t, evaluated at base period prices.

A drawback of the above indexes is that they are based on the sample of either the base period or the comparison period t, but not on both samples. When constructing an index going from 0 to t, the sales in both periods should ideally be taken into account in a symmetric fashion. The double imputation Fisher price index

$$P^{0t}_{Fisher} = \left[P^{0t}_{Laspeyres} \times P^{0t}_{Paasche}\right]^{\frac{1}{2}} \tag{23}$$

does so by taking the geometric mean of the Laspeyres and Paasche price indexes. Note that, because the Fisher index number formula is not consistent in aggregation, it is not possible to provide an exact decomposition of the Fisher property index into structures and land components.

Double imputation Laspeyres, Paasche and Fisher property price indexes and the land price indexes based on the more restrictive hedonic models (10) or (8) are found by replacing $\hat{\alpha}^0_i$ and $\hat{\alpha}^t_i$ in (20) and (22) by the corresponding postcode-specific estimates

$\hat{\alpha}_k^0$ and $\hat{\alpha}_k^t$ or the city-wide estimates $\hat{\alpha}^0$ and $\hat{\alpha}^t$. In the latter case, the estimated land price index of course equals $\hat{\alpha}^t/\hat{\alpha}^0$, irrespective of the index number formula used.

§ 6.5 Empirical evidence

§ 6.5.1 The data set

The data set we utilize was provided by the Dutch Association of Real Estate Agents. It contains residential property sales for a small city (population is around 60,000) in the northeastern part of the Netherlands, the city of "A", and covers the first quarter of 1998 to the fourth quarter of 2007. Statistics Netherlands has geocoded the data. We decided to exclude sales on condominiums and apartments since the treatment of land deserves special attention in this case. The resulting total number of sales in the data set during the ten-year period is 6,058, representing approximately 75% of all residential property transactions in "A".

The data set contains information on the time of sale, transaction price, a range of structures characteristics, and land characteristics. We included only three structures characteristics in our models, i.e., usable floor space, building period and type of house. For land, we used plot size and postcode or latitude/longitude. Initially, we deleted 43 observations with missing values or prices below €10,000, properties with more than 10 rooms and those with ratios of plot size to structure size (usable floor space) larger than 10 as well as transactions in rural areas. Finally, we removed 32 outliers or influential observations detected by Cook's distance and were left with 5,983 observations during the sample period.

Table A1 in the Appendix reports summary statistics by year for the numerical variables. Both the average transaction price and the price per square meter significantly increased from 1998 to 2007. Average land size and usable floor space were quite stable over time. The urban area of the city of "A" seems to have expanded along the east-west axis; the standard deviation of the *x* coordinate in later years is generally much larger than that in earlier years.

§ 6.5.2 Estimation results for hedonic models

Given the small size of the city of "A" and the resulting low number of observations, we decided to use annual rather than quarterly data. We estimated three normalized hedonic models: model (8), which does not include location (denoted by OLS), model (10) with 9 postcode dummy variables (OLSD), and model (12) with property-specific land prices (MGWR).

When estimating the MGWR model, we used the adaptive bi-square function to construct the weighting scheme, given that the transactions were not evenly distributed across space. In this case, the bandwidth is generally referred to as the window size, and the choice of window size is equivalent to the choice of the number of nearest neighbors. To find the optimal value, we varied the window size from 10% to 95% using a 5% interval and selected the size that yielded the lowest CV score as given by equation (18). Each annual sample has a unique optimal window size. The CV scores indicated that a 10% window size was preferred for most of the years, except for 1999, 2000 and 2002, with an optimal size of 15%, and 2003, with an optimal size of 30%. However, for the construction of price indexes, we would prefer using the same window size for all years, especially since the number of sales is almost evenly spread across the whole period. So we chose a window size of 10% for each year, leading to 60 nearest neighbors that were used in the estimation of the annual MGWR models.

TABLE 6.1 Parameter estimates for structures characteristics, 2007

	OLS	OLSD	MGWR
Intercept	1480.70***	1405.41***	1395.76***
	(46.93)	(53.71)	(57.51)
Building period:1960-1970	-370.48***	-389.50***	-398.40***
	(25.94)	(36.67)	(41.75)
Building period:1971-1980	-311.17***	-261.50***	-323.50***
	(23.36)	(33.96)	(41.75)
Building period:1981-1990	-232.93***	-173.08***	-226.14***
	(23.37)	(32.59)	(42.87)
Building period:1991-2000	-58.64***	-49.34*	-115.13***
	(21.64)	(26.55)	(37.26)
Terrace	-285.65***	-264.34***	-187.28***
	(35.17)	(35.24)	(37.32)
Corner	-281.36***	-274.54***	-192.85***
	(31.77)	(31.18)	(34.07)
Semidetached	-122.89**	-149.50***	-96.93**
	(47.96)	(47.57)	(48.73)
Duplex	-151.08***	-147.24***	-104.56***
	(30.60)	(30.17)	(31.03)

*Notes:*Standard errors are reported in parentheses; ***, ** and * denote significance at the 1% , 5% and 10% level, respectively.

As an illustration, Table 6.1 shows the 2007 parameter estimates for the structures characteristics. Almost all of the estimates differ significantly from zero at the 1% level. To some extent they vary across the different models. For example, the OLS intercept term is relatively high compared to the OLSD and MGWR intercepts. Note that, since dummy variables for houses built after 2000 and for detached houses were not included, the intercept measures the price in euros of structures per square meter of living space for detached houses built after 2000. In accordance with a priori expectations, detached dwellings are more expensive than other types of houses. For all models, there is a clear tendency for the structures to become less expensive as they are getting older.

TABLE 6.2 Summary statistics for estimated land prices

	OLS	OLSD		MGWR				
		Mean	S.D.	Min	Max	Median	Mean	S.D.
1998	116.80	131.50	31.14	72.30	231.03	122.66	125.49	28.66
1999	154.64	178.50	34.85	105.95	223.66	174.07	167.77	30.39
2000	239.77	239.41	36.24	138.53	319.32	251.34	241.83	44.27
2001	214.54	235.58	47.59	110.41	295.01	229.52	226.70	48.77
2002	234.77	245.11	38.41	156.15	323.63	255.05	242.23	40.89
2003	166.07	185.11	44.23	82.12	248.23	179.93	172.26	44.55
2004	186.40	197.19	29.75	104.95	254.20	197.70	195.41	33.78
2005	226.13	224.11	36.55	127.53	299.74	214.19	205.89	35.17
2006	202.84	195.77	30.85	125.90	274.24	207.43	201.27	32.05
2007	214.87	236.73	27.96	141.46	286.91	235.07	229.25	30.99

Notes: For OLS, the land price estimates are reported. For OLSD, the columns show the weighted mean and standard deviation of the estimated land prices for 9 postcode areas where the weights are equal to the share of transactions within each postcode area. For MGWR, the columns provide summary statistics for the land price estimates of all transacted properties.

Table 6.2 contains summary statistics for the estimated price per square meter of land from the three models. The three average land price series exhibit a similar pattern over time, which differs substantially from the changes in the average transaction price of the properties (see Table A1 in the Appendix). After a sharp increase in 1999, the estimated average land price fluctuated during a couple of years, experienced a dramatic drop in 2003, and then increased again.

As mentioned earlier, a virtue of MGWR is that it allows us to plot a continuous map with estimated prices of land per square meter. For the year 2007, such a map is depicted in Figure 6.1 for the city of "A", where the land prices have been rescaled to the range [0,1]. The postcode areas are indicated as well. While the spatial pattern in Figure 6.1 is largely consistent with the pattern found using the OLSD model (shown in Figure A1 in the appendix), the MGWR land prices estimates do vary within some of the areas. This suggests that the use of postcode dummies, as in the OLSD model, is a rather crude strategy to incorporate spatial variation of land prices.

To formally compare the performance of the three hedonic models, two statistics were calculated, the Corrected Akaike Information Criterion (AICc) and the Root Mean Square Error (RMSE). The AICc takes into account the trade-off between goodness of fit and degrees of freedom. The AICc expressions for the OLS and OLSD models can be found in Hurvich and Tsai (1989). And for MGWR models, it is defined by

$$AICc = 2nln(\hat{\sigma}) + nln(2\pi) + n\left(\frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})}\right)$$

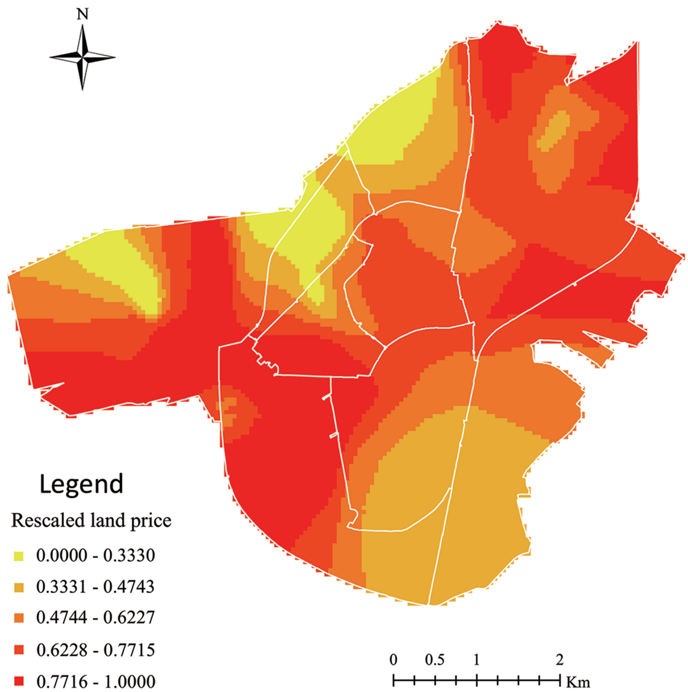


FIGURE 6.1 Price of land per square meter, 2007

where $\hat{\sigma}$ is the estimated standard deviation of the error term and $tr(\mathbf{S})$ the trace of the hat matrix described in section 6.3.2. The RMSE measures the variability of the absolute prediction errors of the models and is given by

$$RMSE = \frac{1}{n}\sqrt{\sum_i (p_i - \hat{p}_i)^2}.$$

Table 6.3 shows the AICc and RMSE and their differences for the three models. A rule of thumb states that if the difference in the AICc for two models is larger than 3, a significant difference exists in their performance (Fotheringham et al. 2002). It can be seen that the OLSD model performs much better than the OLS model in all years, as we would expect, and in turn that the MGWR model outperforms the OLSD model (except for 2003, when the difference is insignificant). The same ranking is found if the RMSE is used to assess the various models. These results confirm the earlier finding that land prices vary across space, both across and within postcode areas.

Although MGWR is obviously better suited to model the variation of land prices and to predict property prices, the OLSD model does a surprisingly good job. In several years,

TABLE 6.3 Model comparison

	Dependent variable = House price per square metre of living space											
	OLS		OLSD				MGWR					
	AICc	RMSE	AICc	ΔAIC_{DO}	RMSE	$\Delta RMSE_{DO}$	AICc	ΔAIC_{MD}	ΔAIC_{MO}	RMSE	$\Delta RMSE_{MD}$	$\Delta RMSE_{MO}$
1998	6487.71	91.32	6372.45	-115.26	80.89	-10.43	6366.21	-6.24	-121.50	77.85	-3.04	-13.47
1999	7056.56	146.82	6990.52	-66.04	136.14	-10.68	6982.93	-7.59	-73.63	131.28	-4.86	-15.54
2000	7216.89	151.11	7164.26	-52.63	142.00	-9.11	7127.51	-36.75	-89.38	133.30	-8.70	-17.81
2001	7380.41	147.00	7294.10	-86.31	134.36	-12.64	7279.66	-14.44	-100.75	128.87	-5.49	-18.13
2002	7718.63	152.44	7643.97	-74.66	141.19	-11.25	7632.34	-11.63	-86.29	135.62	-5.57	-16.82
2003	7769.06	159.02	7702.07	-66.99	148.23	-10.79	7701.91	-0.16	-67.15	143.93	-4.30	-15.09
2004	7968.62	159.66	7947.61	-21.01	154.80	-4.86	7927.92	-19.69	-40.70	147.91	-6.89	-11.75
2005	8060.84	161.52	7993.88	-66.96	150.93	-10.59	7984.11	-9.77	-76.73	145.10	-5.83	-16.42
2006	8597.81	175.46	8565.36	-32.45	168.94	-6.52	8517.73	-47.63	-80.08	157.67	-11.27	-17.79
2007	9006.25	177.18	8960.58	-45.67	169.24	-7.94	8929.11	-31.47	-77.14	159.98	-9.26	-17.20

Notes: ΔAIC_{DO} is equal to AICc for OLSD minus AICc for OLS; ΔAIC_{MD} and ΔAIC_{MO} are equal to AICc for MGWR minus AICc for OLSD and OLS, respectively; $\Delta RMSE_{DO}$, $\Delta RMSE_{MD}$ and $\Delta RMSE_{MO}$ have a similar meaning.

for example in 1998, 1999 and 2003, the inclusion of postcode dummy variables accounts for the major part of the variance in overall property prices, almost as much as the MGWR model does.

§ 6.5.3 Hedonic imputation price indexes

Changes in average property prices and their land and structure components are affected by compositional change and quality change of the traded properties. The hedonic house price indexes and the land and structures components that we estimated control for these effects. We estimated chained rather than direct indexes because imputing the 'missing prices' over a very long period of time may not be useful and because the value shares of land and structures will then be updated annually. A drawback of chaining is that the resulting price indexes cannot be exactly decomposed because they are not consistent in aggregation.

In Figures 6.2-6.4, the estimated double imputation hedonic Laspeyres, Paasche and Fisher price indexes for the overall property are plotted, based on the three models (OLS, OLSD, and MGWR). A comparison of Figures 6.2 and 6.3 shows that, for each model, the chained Laspeyres index sits above the Paasche index, as expected. The Laspeyres and Paasche indexes based on OLSD and MGWR are very similar; for the Laspeyres index, the difference can even hardly be noticed. This result is in accordance with our finding that the OLSD model captures the spatial nonstationarity of land prices reasonably well.

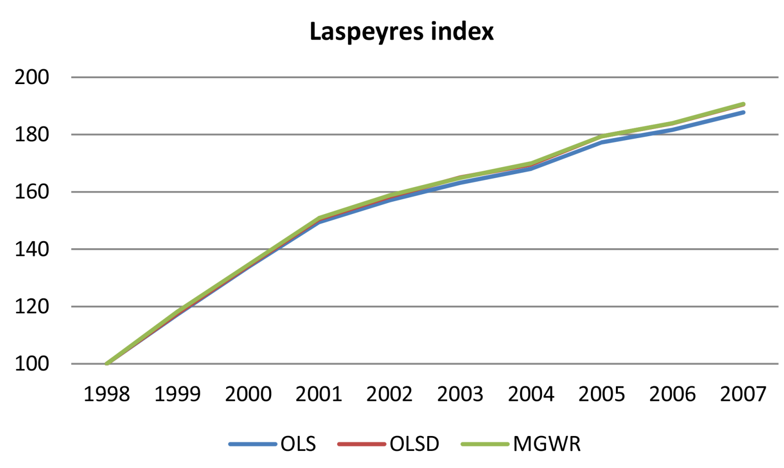


FIGURE 6.2 Hedonic imputation Laspeyres house price index

Not using location information at all does make a difference though, at least for the

Laspeyres and Paasche house price indexes. The OLS-based Laspeyres and Paasche indexes seem to be biased downwards and upwards, respectively. However, the biases almost cancel out in the Fisher index: the OLS-based Fisher index is very similar to the OLSD-based and MGWR-based Fisher indexes. In other words, the hedonic imputation Fisher house price index is insensitive to the treatment of location in the hedonic model, which is a surprising result.

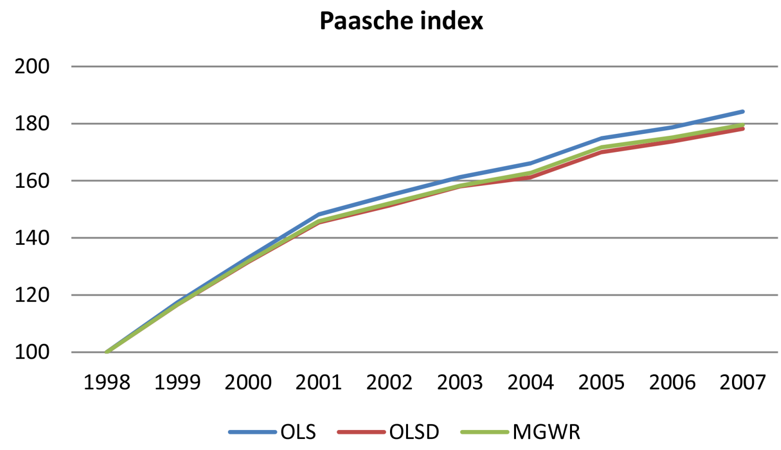


FIGURE 6.3 Hedonic imputation Paasche house price index

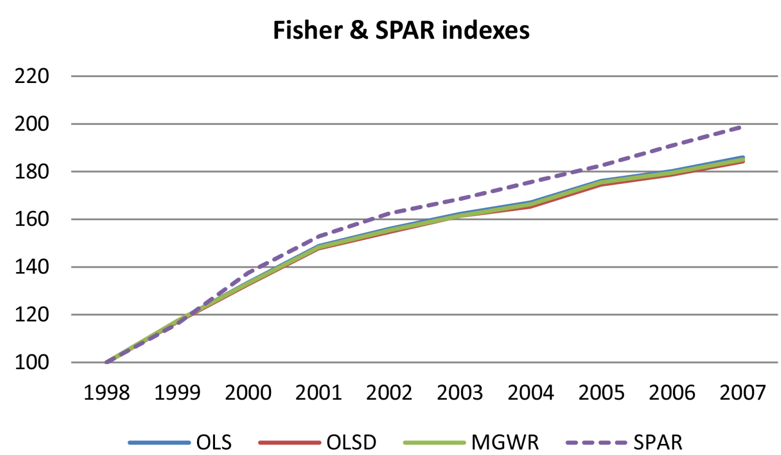


FIGURE 6.4 Hedonic imputation Fisher house price index and SPAR index

The house price index for the Netherlands published by Statistics Netherlands is also plotted in Figure 6.4. This official index is based on the Sale Price Appraisal Ratio (SPAR) method (de Haan et al. 2009; de Vries et al. 2009). Our hedonic indexes show a much more modest price increase. There may be two reasons for this. First, house prices in the city of "A" appreciated less compared to the rest of the country. Second, our indexes better adjust for quality changes while the SPAR method only adjusts for compositional change of the properties sold. We think that the second reason is more important.

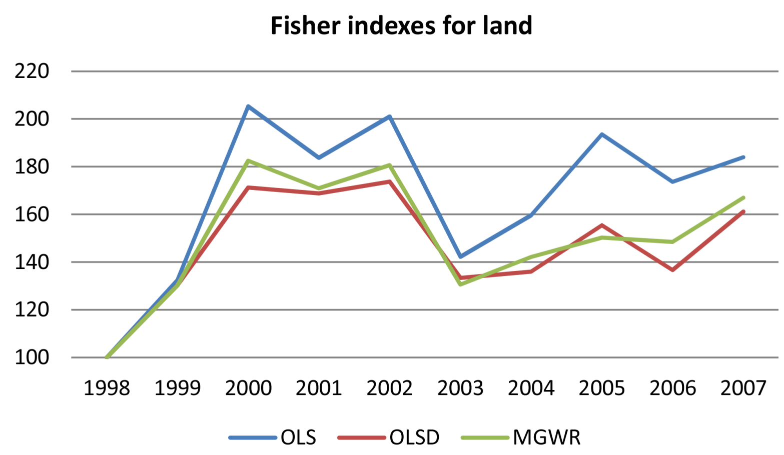


FIGURE 6.5 Hedonic imputation Fisher price indexes for land

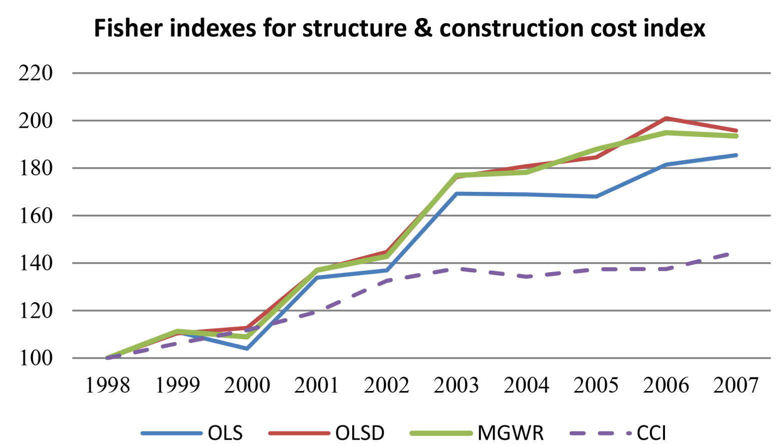


FIGURE 6.6 Hedonic imputation Fisher price indexes for structures and official construction cost index

The picture changes when we look at the Fisher indexes for the price of land in Figure 6.5. The OLSD- and MGWR-based indexes, which explicitly account for location, are similar, although the MGWR-based index is less volatile, at least during 2003-2007. However, the OLS-based index seems to be significantly upward biased. For example, between 1999 and 2000 as well as between 2003 and 2005, the OLS-based index rises much faster than the other two indexes.

Figure 6.6 shows the Fisher price indexes for structures based on the three models. Again, the OLSD-based and MGWR-based indexes are similar. The OLS-based index sits below the other indexes, but the difference is less pronounced than for land. This is in line with our expectations: location should affect the price of land and is modeled accordingly, but it should leave the price of structures unaffected.

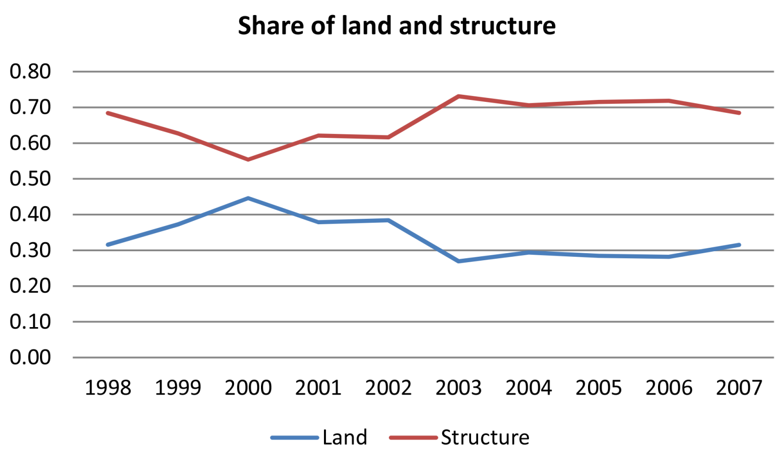


FIGURE 6.7 Estimated value shares of land and structures, MGWR-based

Figure 6.7 shows the MGWR-based value share estimates for both structures and land. Prior to 2003, these shares are quite volatile, but from 2003 on they remain fairly constant. The average estimated shares for structures and land across the entire sample period are 0.67 and 0.33. The OLS- and OLSD-based shares show similar patterns and levels; the average shares for structures are 0.68 and 0.66, respectively, hence for land 0.32 and 0.34. Given that the estimated value share of structures is twice as large as that of land, the overall house price indexes are affected most by changes in structures prices. Yet, combining Figures 6.4, 6.5, 6.6 and 6.6 suggests that the increase in house prices between 1998 and 2001 has been driven mainly by the increase of land prices: both the (average) price of land and its value share show a sharp increase.

§ 6.5.4 Discussion

Figures 6.5, 6.6 and 6.7 raise a number of issues. The first issue is the volatility of the land and structures price indexes. Volatile series can of course be expected with sparse data (and without smoothing). Another cause might be multicollinearity. Diewert et al. (2015) found that multicollinearity (between land and structure size) led to price changes for land and structures which consistently had opposite signs. To deal with this form of multicollinearity, some studies (e.g.,Diewert et al. 2009; Diewert and Shimizu 2013; Francke and van de Minne 2016) included exogenous information in the hedonic models; they all used the officially published construction cost index as the measure of price change for structures. Put differently, their models do not provide an endogenously determined price index of structures. We do not follow their approach because, as we discuss in the next paragraph, multicollinearity does not seem to be the most important issue and because the trend of the endogenous price index of structures might be more consistent with the evolution of the market values of structures.

In Figure 6.8, the MGWR-based Fisher price indexes for land and structures from Figures 6.5 and 6.6 are copied. In some years, for example in 2003 when the land price index suddenly falls and starts to sit below the structures price index, the price changes for land and structures have opposite signs, but in other years the price changes are in the same direction. The variance inflation factor (VIF) for the ratio of plot size to structure size did not point to significant multicollinearity either. Further, there is a considerable amount of variation in these ratios in our data set; see Table A1. We therefore suspect that multicollinearity is not the main issue.

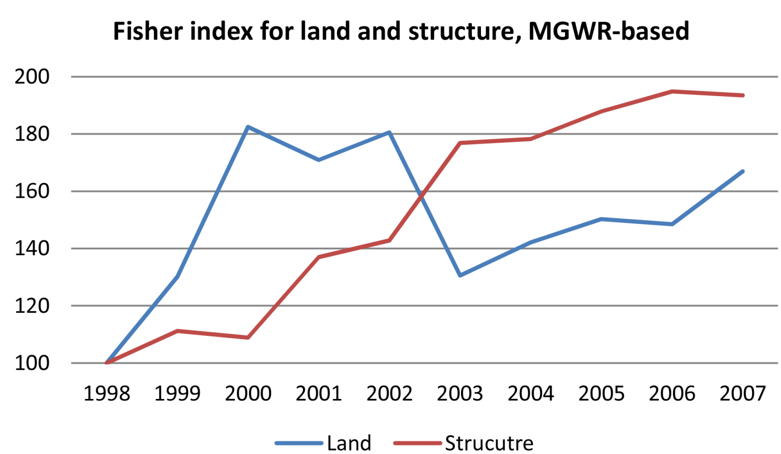


FIGURE 6.8 Chained Fisher price indexes for land and structures, MGWR-based

The second issue is whether the trends of the (Fisher) price indexes for land and structures are plausible. For land, this cannot be checked since information on the price change of land is not available for the Netherlands [2]. For structures we use the official nationwide construction cost index (CCI) for new dwellings as a benchmark. This price index, rebased to 1998=100, is shown in Figure 6.6 as well. Our structures price indexes rise much faster than the construction cost index, especially during the second half of the sample period when the construction cost index flattens.

At first, a construction cost index does not necessarily have to coincide with an implicit price index for structures derived from a hedonic model. Since structure is producible, it is believed that, in a completely competitive market, the construction cost is equal to the market value of structure (Davis and Heathcote 2007; Davis and Palumbo 2008). This equality might be held in a stable market where the developers can take a sufficiently long time to meet the demand. However, the market of structures in reality tends to be less competitive due to the restriction of new constructions and the high cost of replacing old structures with new ones. In this case, it is very likely that a persistent markup is imposed on structure prices and hence it is no surprising to see the structure price index sitting above the construct cost index. This disparity can be even more striking during a housing boom, which is exactly the case of this study. Kuminoff and Pope (2013), who estimated the land values for US metropolitan areas using a refined hedonic approach that mitigates the omitted variable bias, presented a similar finding that the increase of market value of structures exceeds the growth of replacement cost in the booming period in some places. On the other hand, the flattening of the construction cost index between 2003 and 2007 has been subject of debate in the Netherlands. The discussion arose because the construction cost index increased by only 4.9%, which was even lower than the increase in the CPI of 5.8%, while house prices were still rapidly rising.

Nevertheless, a divergence that large is still a bit worrying. One of the reasons for the strong increase of our structures price indexes could be omitted variables bias – resulting in quality-change bias – because we included only a few structures characteristics in the hedonic models. Unless they are highly collinear with included variables, adding characteristics will lead to better quality adjustment for structures and lower the price indexes for structures if, as can be expected during this period of booming house prices, the quality of structures has improved over time. One obvious omitted variable that is related to depreciation of the structures is the level of maintenance.

The third issue concerns the share of land in the value of properties sold, which was estimated at roughly one third across the sample period. van de Minne and Francke (2012) estimated the share of land for properties (excluding

2 Municipalities do have information on what are sometimes referred to as realizations of the value of land sold to developers of residential construction. These realizations are determined residually, but it is doubtful whether they accurately measure the 'true' value of land.

apartments/condominiums) sold during 2003-2010 in the city of 's Hertogenbosch at 0.39 on average. In a follow-up study (Francke and van de Minne 2016), where they made a distinction between the part of the land plot that the structure sits on and the part used as gardens, the estimate was almost 0.50. It is not unreasonable to find that the value share of land for the city of "A" is lower than that for 's Hertogenbosch. The city of "A" lies in a less prosperous part of the Netherlands with fewer amenities that households appreciate, and we expect this to have a downward effect on the price of land rather than the price of structures, hence on the value share of land.

de Groot et al. (2015), also using hedonic models to decompose property values into land and structures components, estimated the price of land for most Dutch cities, though unfortunately not for "A". They found substantial cross-city differences. For example, the price per square meter of land in 2005 was estimated at 717 euros for the capital city of Amsterdam, 308 euros for 's-Hertogenbosch, and 184 euros for Leeuwarden. Like "A", Leeuwarden is a city in the northeastern part of the Netherlands but bigger. In light of their findings, our MGWR estimates of the average price of land for the city of "A", 206 euros in 2005 (Table 6.2), and the value share of land are not surprisingly low after all.

§ 6.6 Summary and conclusions

Land is often not explicitly included in hedonic models for house prices, which can bias the results. Ignoring spatial nonstationarity of land prices can also generate bias. As far as we know, the present paper is the first attempt to account for nonstationarity of land prices in the construction of hedonic imputation house price indexes. We linearized the 'builder's model' proposed by Diewert et al. (2015), allowed the price of land to vary across individual properties, and estimated the model for the normalized property price (the price of the property per square meter of living space) by MGWR, a semi-parametric method, on annual data for the Dutch city of "A". We then constructed chained imputation Laspeyres, Paasche and Fisher indexes, and compared these indexes with price indexes based on more restrictive models, i.e. a model where land prices vary across postcode areas and a model with no variation in land prices and, both estimated by OLS.

The Fisher house price indexes were quite insensitive to the choice of model, but the Laspeyres and Paasche indexes for the 'fixed' land price model differed from those for the models where location was explicitly included. The use of postcode area dummy variables produced price indexes very similar to indexes obtained by MGWR. Hill and Scholz (2014), who treated location as a 'separate characteristic' in their hedonic models in that they estimated property-specific shift terms for the overall property price, also concluded that the use of geocoded information did not significantly improve hedonic imputation house price indexes compared to indexes based on models with postcode dummy variables. This result is reassuring for statistical agencies that do not have the expertise or resources to apply more sophisticated methods. It should be noted that the similarity between OLSD-based and MGWR-based house price indexes could also be due to the small size and homogeneity of the city "A" where relatively little variation of land prices can be expected.

Apart from being able to capture spatial variation of land prices at the property level, the MGWR model has two additional advantages. A potential problem with the OLSD model is that if a large number of postcode areas are distinguished, observations in some areas may not be available, leading to difficulties in the construction of hedonic imputation price indexes. The MGWR method deals with this problem by using data of the nearest neighbors which are not necessarily confined to a particular postcode area. Most importantly, the use of nearest-neighbor information in the (semi-parametric) MGWR method makes it possible to properly account for spatial effects in the absence of detailed information on amenities, such as the availability of, and distance to, public transport, green space, schools, shopping centers, and so on.

For some purposes, separate price indexes for land and structures are needed. As was demonstrated already by Diewert et al. (2015), the decomposition into land and structures using hedonic modeling is not straightforward and raises several statistical and functional form issues. First, our MGWR-based price indexes of land and structures for the city of "A" are quite volatile, in spite of the use of annual data, which can be attributed to the sparse data in combination with possibly multicollinearity (though we believe this is less important). Second, the structures price index increases much faster than expected, perhaps due to omitted variables or quality-change bias, i.e. a failure to fully control for changes in structures characteristics. Third, the estimated value share of land seems rather low. The above-mentioned problems may have played a role here, but the low land share could also be a real phenomenon: households do not value a square meter of land in the city of "A" as much as they would do in more prosperous cities with more and better amenities. Anyhow, in future work it would be useful to re-examine our models and compare the results for the city of "A" with those for bigger and more densely populated cities in the western part of the country, like Amsterdam, Rotterdam or The Hague. Having more observations might also enable us to estimate biannual or even quarterly price indexes.

Functional form problems might be even more important. The original 'builder's model' is nonlinear, in particular due to the treatment of net depreciation. We linearized the model, which basically means we ignored interaction terms. Another potential type of misspecification arises from the linear relationship between land price and plot size in our models. As Diewert et al. (2015), Francke and van de Minne (2016) and others have argued that the marginal price of land tends to decrease with plot size.

Diewert et al. (2015) accounted for this form of nonlinearity by using linear splines. In future work we may modify our 'normalized' models by using linear splines as well and estimating different parameters for the plot size to structure size ratio for different categories of lot size or by explicitly specifying some nonlinear function of this ratio.

References

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis, 28(4), 281-298.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1999). Some notes on parametric significance tests for geographically weighted regression. Journal of Regional Science, 39(3), 497-524.

Casetti, E. (1972). Generating models by the expansion method: Applications to geographical research. Geographical Analysis, 4(1), 81-91.

Davis, M. A., & Heathcote, J. (2007). The price and quantity of residential land in the United States. Journal of Monetary Economics, 54(8), 2595-2620.

Davis, M. A., & Palumbo, M. G. (2008). The price of residential land in large US cities. Journal of Urban Economics, 63(1), 352-384.

de Groot, H. L. F., Marlet, G., Teulings, C., & Vermeulen, W. (2015). Cities and the Urban Land Premium. Cheltenham: Eaward Elgar.

de Haan, J. (2010). Hedonic price indexes: A comparison of imputation, time dummy and 're-pricing' methods. Jahrbücher für Nationalökonomie und Statistik, 230(6), 772-791.

de Haan, J., van de Wal, E., & de Vries, P. (2009). The measurement of house prices: A review of the sale price appraisal ratio method. Journal of Economic and Social Measurement, 34(2,3), 51-86.

de Vries, P., de Haan, J., van de Wal, E., & Mariën, G. (2009). A house price index based on the SPAR method. Journal of Housing Economics, 18(3), 214-223.

Diewert, W. E., de Haan, J., & Hendriks, R. (2011). The decomposition of a house price index into land and structures components: A hedonic regression approach. The Valuation Journal, 6(1), 58-105.

Diewert, W. E., de Haan, J., & Hendriks, R. (2015). Hedonic regressions and the decomposition of a house price index into land and structure components. Econometric Reviews, 34(1-2), 106-126.

Diewert, W. E., Heravi, S., & Silver, M. (2009). Hedonic imputation versus time dummy hedonic indexes. In W. E. Diewert, J. S. Greenlees, & C. R. Hulten (Eds.), Price Index Concepts and Measurement (pp. 161-196). Chicago: University of Chicago Press.

Diewert, W. E., & Shimizu, C. (2013). Residential property price indexes for Tokyo. UBC Vancouver School of Economics Discussion Papers 2013-16, The University of

British Columbia.
Dorsey, R. E., Hu, H., Mayer, W. J., & Wang, H. (2010). Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle. Journal of Housing Economics, 19(2), 75-93.
Eurostat, ILO, IMF, OECD, UNECE, & World Bank (2013). Handbook on Residential Property Price Indices. Luxemburg: Publications Office of the European Union.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (1998a). Scale issues and geographically weighted regression. In N. J. Tate, & P. M. Atkinson (Eds.), Modelling Scale in Geographical Information Science (pp. 123-140). Chichester: Wiley.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester: John Wiley & Sons.
Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998b). Geographically weighted regression: a natrual evolution of the expansion method for spatial data analysis. Environment and Planning A, 30(11), 1905-1927.
Francke, M. K., & van de Minne, A. M. (2016). Land, structure and depreciation. Real Estate Economics, Forthcoming.
Geniaux, G., & Napoléone, C. (2008). Semi-parametric tools for spatial hedonic models: An introduction to mixed geographically weighted regression and geoadditve models. In A. Baranzini, J. Ramirez, C. Schaerer, & P. Thalmann (Eds.), Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation (pp. 101-127). New York: Springer.
Hill, R. J., & Melser, D. (2008). Hedonic imputation and the price index problem: An application to housing. Economic Inquiry, 46(4), 593-609.
Hill, R. J., Melser, D., & Syed, I. (2009). Measuring a boom and bust: The Sydney housing market 2001 - 2006. Journal of Housing Economics, 18(3), 193-205.
Hill, R. J., & Scholz, M. (2014). Incorporating geospatial data in house price indexes: A hedonic imputation approach with splines. Graz Economics Papers 2014-05, University of Graz.
Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.
Jones, J. P., & Casetti, E. (1992). Applications of the Expansion Method. London: Routledge.
Kuminoff, N. V., & Pope, J. C. (2013). The value of residential land and structures during the great housig boom and bust. Land Economics, 89(1), 1-29.
Mei, C., Wang, N., & Zhang, W. (2006). Testing the importance of the explanatory variables in a mixed geographically weighted regression model. Environment and Planning A, 38(3), 587-598.
Pace, R. K., Barry, R., Clapp, J. M., & Rodriquez, M. (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. Journal of Real Estate Finance and Economics, 17(1), 15-33.

Rambaldi, A. N., & Rao, D. S. P. (2011). Hedonic predicted house price indices using time-varying hedonic models with spatial autocorrelation. School of Economics Discussion Paper Series 432, University of Queensland.

Rambaldi, A. N., & Rao, D. S. P. (2013). Econometric modeling and estimation of theoretically consistent housing price indexes. Centre for Efficiency and Productivity Analysis Working Paper Series WP04/2013, University of Queesland.

Sun, H., Tu, Y., & Yu, S.-M. (2005). A spatio-temporal autoregressive model for multi-unit residential market analysis. The Journal of Real Estate Finance and Economics, 31(2), 155-187.

Thanos, S., Dubé, J., & Legros, D. (2016). Putting time into space: the temporal coherence of spatial applications in the housing market. Regional Science and Urban Economics, 58, 78-88.

Tu, Y., Yu, S.-M., & Sun, H. (2004). Transaction-Based Office Price Indexes: A Spatiotemporal Modeling Approach. Real Estate Economics, 32(2), 297-328.

van de Minne, A. M., & Francke, M. K. (2012). De waardebepaling van grond en opstal [The determination of the value of land and structures]. Real Estate Research Quarterly, 11, 14-24.

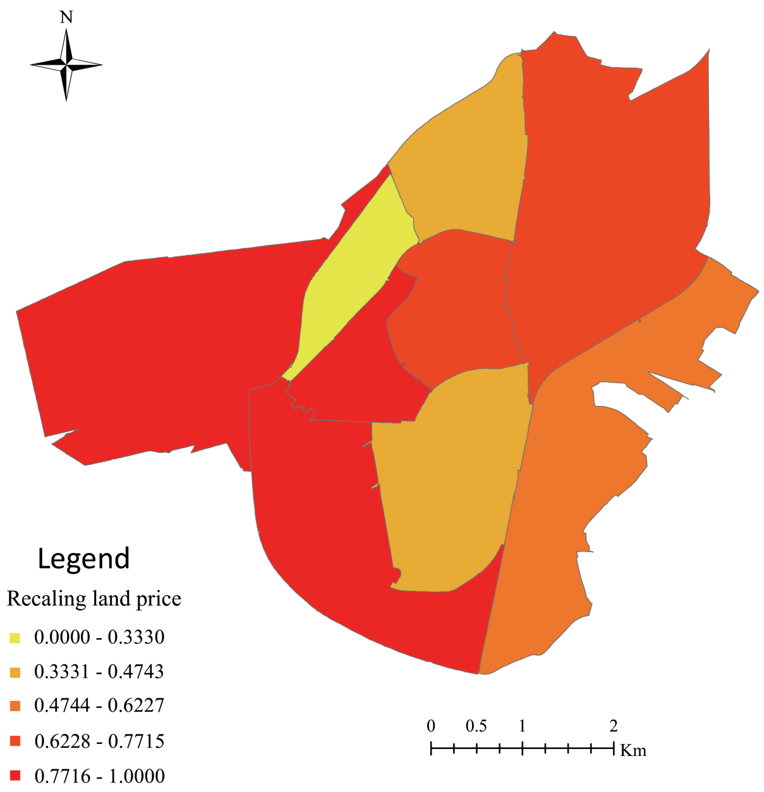


FIGURE A1 Price of land per square meter, 2007, OLSD model

TABLE A1 Summary statistics by year

	Total	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
# of obs.	5983	545	549	559	574	597	597	612	618	651	681
Transaction price (Euro)											
Mean	157073.87	95124.15	117936.77	131907.96	144672.16	151363.75	162956.98	174998.71	180882.00	191491.09	198546.51
S.D.	72782.29	40240.34	53569.32	54793.53	58064.72	53220.31	63278.10	82975.61	68777.60	76120.61	83639.92
Standardized price (Euro)											
Mean	1232.38	742.30	930.70	1039.71	1168.13	1240.63	1287.24	1353.89	1420.07	1469.62	1518.50
S.D.	374.83	206.31	273.33	279.98	293.14	285.56	285.87	296.73	294.31	321.20	348.89
Lot size (m^2)											
Mean	251.57	234.08	259.73	242.23	239.68	239.20	250.46	261.38	248.93	263.15	270.98
S.D.	148.16	135.05	169.59	132.98	120.00	115.39	145.76	163.19	136.00	149.26	187.52
Floor space (m^2)											
Mean	125.87	126.00	125.42	126.48	123.34	122.05	125.29	126.57	125.89	128.52	128.39
S.D.	30.61	23.59	31.99	31.97	29.59	28.16	29.87	36.90	30.29	31.14	30.09
Ratio of lot size to floor space											
Mean	1.96	1.81	2.04	1.89	1.93	1.97	1.96	2.01	1.93	2.01	2.04
S.D.	0.82	0.77	0.99	0.72	0.72	0.80	0.84	0.80	0.72	0.78	0.95
x-coordinate											
Mean	233733.81	233972.85	234200.97	234180.34	233948.97	234007.39	233624.00	233480.63	233519.69	233222.34	233385.19
S.D.	1796.35	1453.72	1427.35	1551.87	1716.67	1713.60	1794.99	1984.82	1927.09	1918.80	1948.29
y-coordinate											
Mean	558597.10	558739.46	558805.54	558830.14	558660.23	558721.99	558522.02	558397.61	558549.11	558429.21	558410.25
S.D.	1414.88	1436.14	1463.14	1428.62	1424.92	1410.80	1451.63	1413.94	1354.34	1322.63	1381.24

7 Conclusion

This book is a collection of five chapters dedicated to discovering and understanding the spatial dimension of house prices, especially the spatial aspects of the Chinese interurban housing market after the introduction of market forces. The free mobility of labour and capital between cities in a market-oriented economy in contemporary China following the economic reform launched in the late 1970s shapes the spatial distribution of economic activities across cities, which in turn has profound effects on nascent urban private housing markets. Chapters 2 through 5 seek to illustrate the spatial pattern of city house prices and their dynamics and reveal the role that location plays in the formation of such a spatial pattern. Chapter 6, unlike the previous chapters, focuses on an intra-urban housing market in the Netherlands and investigates the spatial variation in land prices caused by the spatial fixity in local amenities, as well as its influence on the house price index. The present chapter summarises the main findings of this book, discusses its strengths and weaknesses, draws a few policy implications and advances several ideas for future research.

§ 7.1 Main findings

§ 7.1.1 Explaining the spatial distribution of interurban house prices

Chapters 2 and 3 are concerned with the spatial distribution of interurban house prices in the urban system of the Pan-Yangtze River Delta (Pan-YRD) in Eastern China, which contains 1 municipality directly under the central government, 3 capital cities of provinces, and 38 prefecture cities. Each of those cities comprises a city proper (*shiqu*) made up of districts and several hinterland jurisdictions (counties or county-level cities)[1]. An urban housing market is defined as the market within the boundary of the city proper. The city-level house prices are calculated as the average per unit sale price of newly sold residential buildings without controlling for housing characteristics. The

1 The Chinese administrative division consists of five levels: province (municipality, autonomous region, special administrative region); prefecture city (prefecture, autonomous prefecture, league); county (city district, county-level city, autonomous county); town (sub-district); and village (neighbourhood).

questions asked about the spatial distribution of house prices are:

What is the spatial distribution of house prices across cities? How can that pattern be explained? What role does location play in shaping the interurban house price pattern?

To conduct the analysis, a panel data set covering the period 2006-2010 is compiled from various sources. The location of a city is captured by a set of distance measures, such as distance to central cities and distance to the nearest sub-central city. Several city-specific characteristics are also collected, such as winter temperature, pollution levels, healthcare services and quality of education system.

Chapter 2 treats the Pan-YRD urban system as a three-tier hierarchical system in which Shanghai, the municipality directly under the central government, is designated the central city of this system, with three provincial capitals, Nanjing, Hangzhou and Hefei, as the sub-central cities. This hierarchy division is in accordance with the city functions as outlined in the Outline of National Urban System Planning (2005-2020). The Pan-YRD urban hierarchy is assumed to follow the central place theory, which states that higher-tier cities have larger local markets and provide higher-order services and products for lower-tier cities. Thus, from the Rosen-Roback spatial general equilibrium framework (Rosen 1979; Roback 1982), it is inferred that the farther away a city is located from higher-tier cities, the lower the house price is in that city. Such penalties imposed by distance from higher-tier cities are attributed to two channels. First, firms in peripheral cities are less productive because they cannot benefit from the agglomeration spillovers of higher-tier cities; thus, the wages there are lower. Second, households in peripheral cities need the house price to be lower as a compensation for their difficulties in access to the unique consumer services that are only available in higher-tier cities. The theoretical framework presented in Chapter 2 further allows one to decompose the house price gradient and assess which component contributes more to the distance penalties.

Under the assumption that sub-central cities substitute for each other but complement the central city (i.e., a third-tier city is affected by both the central city and the nearest sub-central city), Chapter 2 identifies significant distance penalties of both the central city and sub-central cities if the distance-decay function forms are properly specified. It seems that the distance-decay function forms should be chosen in light of the influential radius of targeted higher-tier cities: the semi-log function is the best choice for the central city which has broad influences, while a log-log decay function is better for sub-central cities that only affect a relatively small radius of area. This is in line with the findings of Osland et al. (2007), who concluded that the exponential (semi-log) function performs best when the estimation is based on a large area, while the power (log-log) function performs best if the data is restricted to a small area.

The results in Chapter 2 also show that a warmer winter, less industrial smoke and dust emission, better healthcare condition and bordering an ocean tend to increase the house price of a city. What happens to the house price gradient after controlling for these city characteristics and amenities? The previously revealed distance penalties of higher-tier cities are still statistically significant, at least at the 10% significance level. However, the magnitude of distance penalties of sub-central cities decreases a lot.

To decompose the interurban house price gradient, the wage gradient is also estimated. The slopes of wage gradients towards higher-tier cities are flatter than those of house price gradients, which may be taken as preliminary evidence of the existence of amenity premiums for the higher-tier cities and their neighbours. Formal decomposition confirms the contribution of an amenity component in explaining the negative house price gradient, yet it only accounts for a very small share, less than 20% in general, implying that lower house prices in more peripheral cities are mainly due to the differences in urban productivities.

Chapter 3 also explains the spatial structure of interurban house prices in a general spatial equilibrium framework. Chapter 3 differs from Chapter 2 in its treatment of the Pan-YRD urban system as a city network in which each city interacts with all the other cities in a parallel manner, not necessarily the vertical interaction with higher-tier cities, with each city benefitting from such connectivity (Boix and Trullén 2007). As such, the productivity and amenity performance of a city relies not only on its own urban size but also on the size of neighbouring urban concentrations. The latter is referred to as city network externalities. The city network externalities on the productivity side lie in the 'market access' effect stressed by New Economic Geography – having a larger aggregate and undifferentiated market potential, which is measured by the population or income within a broader region, contributes to the productivity advantage of a city by saving on transportation costs (Fujita et al. 1999; Head and Mayer 2004). On the amenity side, the city network externalities are reflected in the so-called 'borrowing size' effect (Alonso 1973). On one hand, a city can maintain more higher-order amenities or functions than its own size supports through borrowing size from neighbouring cities. Meanwhile, cities that offer such support can share those surplus higher-order amenities through network accessibility, thus improving their performance (Meijers and Burger 2015). Note that the city network spillovers are not fully independent from the spillovers of higher-tier cities discussed in Chapter 2, as higher-tier cities usually have a large urban size.

The empirical part of Chapter 3 presents urban size in two ways: urban scale measured by land area and urban intensity measured by urban population density. Note that the variables on urban size, as well as on city-specific characteristics, are measured on the whole territory of the city, which can partly avoid the endogeneity between house prices (of the market of city proper) and urban size. The city network externalities in the Chinese interurban housing markets are then modelled by various models of spatial econometrics, in which the spatial weight matrix carries the interaction structure between cities. A simple correlation test shows that the house price of a city is mostly correlated to the urban population density of immediate neighbours, say, for example, the neighbouring cities within a radius of 160 km.

Among the several model specifications of spatial econometrics aiming for capturing different kinds of spillovers, the theoretical foundation of city network externalities reasonably justify the spatial lag of X model (SLX) (Gibbons and Overman 2012; Vega and Elhorst 2015). The estimation results of the SLX model strongly support the presence of network spillovers of both land area and urban population density; the amount of spillovers is even larger than the direct effect imposed by urban size. Such findings are quite robust, even after controlling for other forms of spillovers, such as those arising from yardstick competition, in a spatial Durbin error model (SDEM) (LeSage and Pace 2009). Two other types of spatial models are also estimated. The first is the spatial autoregressive model (SAR), which compresses all the forms of spillovers into the single parameter of the spatial lag of dependent variables. The other is the spatial Durbin model (SDM), which includes the spatial lags of both dependent and independent variables. The results of the SAR model suggest that the network spillovers of urban size are almost equivalent with the direct effect, while the SDM reports no significant network spillovers of urban size. However, the SAR and SDM suffer from the inherent identification problem that one cannot tell apart the network externalities from other forms of spillovers (Gibbons and Overman 2012). Thus, the results of the SLX and SDEM should be more reliable.

In summary, the house prices of the Pan-YRD urban system exhibit a 'core – periphery structure', with large urban cores having high house prices and small peripheral cities lower prices. Both the agglomeration spillovers from higher-tier cities from the urban hierarchy point of view and the spillovers from neighbouring cities in the city network paradigm contribute to shape such an agglomeration pattern.

The spatial regularity of house prices in China also applies to the interurban housing markets of western countries. Partridge et al. (2009) examined whether the urban hierarchy spillovers and the spillovers of neighbouring counties affect the spatial distribution of median housing rents of over 3000 U.S. counties in the year 2000. The results indicate that both types of spillovers play a role and that the influence of urban hierarchy spillovers tends to be larger. Since Chapters 2 and 3 of this research separately test the urban hierarchy externalities and city network externalities in the Chinese housing market, their relative importance cannot be directly distinguished.

Both China and America are large countries, so it is not surprising to find that geographical location affects the formation of house prices. In small countries, though, geographical location matters too. de Bruyne and van Hove (2013) investigated the role that location plays in shaping the house price pattern of 589 municipalities in Belgium, a small, densely populated country. They revealed that the distance and travel time to capitals (on both the national and provincial level) have a significant and substantial effect on municipality house prices. This finding stresses the importance of urban hierarchy spillovers, as the capitals are usually the most important economic centres and offer more job opportunities and a large range of services. However, de Bruyne and van Hove (2013) do not consider city network externalities.

§ 7.1.2 Discovering the spatial pattern and interrelationships of interurban house price dynamics

Chapters 4 and 5 consider the spatial patterns and interrelationships of interurban house price dynamics. The data source used in these two chapters is the "Price Indices of Newly Constructed Residential Buildings in 35/70 Large- and Medium-sized Cities", published monthly by the National Bureau of Statistics of China (NBSC). This price index system is compiled by a so-called "match" model that aims to control for quality changes (Wu et al. 2014)[2]. The questions addressed in these two chapters are:

Are house price dynamics across cities different from each other or are they homogeneous? What are the long-run and short-run relationships between them?

Chapter 4 answers the first question and focuses on the broad house price developments of cities across China. Specifically, Chapter 4 investigates the similarities between the house price appreciation trajectories between July 2005 and June 2016 of 34 major cities in China, including municipalities directly under the central government, provincial capitals and some vital economic centres. It explores the possibility of grouping these trajectories into a few homogeneous clusters. The literature usually measures (dis)similarity between house price development paths using Euclidean distance. Chapter 4, however, adopts a distribution-based dissimilarity measure, the Kullback-Leibler (*KL*) divergence (Kullback 1968), which has been applied in geological and environmental studies (Kakizawa et al. 1998; Bengtsson and Cavanaugh 2008). Under the assumption that the house price changes of a city follow an $AR(p)$ process, the *KL* divergence measures the dissimilarities of two aspects: the difference between the predictions and the difference between the prediction errors. In this manner, the *KL* divergence calculated from historical values is also an estimate of the divergence of future dynamics so that one can make statistical inference. Euclidean distance does not hold this virtue. Chapter 4 calculates the *KL* divergence between any city pair based on $AR(3)$ specification. When measuring the dissimilarity between cities, the *KL* divergence is consistent with Euclidean distance to

2 The "match" model used for the NBSC index is analogous to the repeat sales model. In each month, local statistical authorities collect the housing transaction information from different housing complexes. The houses within the same housing complex have similar structural and locational characteristics. Thus, for each housing complex, comparing the average transaction prices of different periods roughly produces the quality-adjusted house price index. The city-level index is the weighted average of all complex-level indexes.

some extent; the Pearson's correlation between them is 0.64.

With the dissimilarity matrix in hand, the hierarchical agglomerative clustering method, along with an *average-linkage* to calculate the dissimilarity between two clusters, is employed to perform the cluster analysis. It seems that the cluster results of *KL* divergence do not make much difference from the results of Euclidean distance. In general, the 34 house price growth trajectories can be broadly partitioned into two clusters. The first cluster includes mainly the Central, Western and Northeast Chinese cities with a relatively low average growth rate. The second cluster comprises the most important city centres in Eastern China with a flourishing housing market. The latter, however, has a higher degree of heterogeneity within the cluster and hence can be further divided into sub-clusters. For example, if four clusters are specified, Shenzhen and Shanghai stand out from the second cluster and form their own clusters.

Given the changing housing market condition in China, the temporal stability of cluster membership is also tested in Chapter 4. To do so, the whole sample period is divided into three intervals: July 2005-December 2010, January 2011-December 2013 and January 2014-June 2016. The inter-period comparison of cluster memberships shows that only in the most recent period after 2014 does the interurban housing market in China become highly fragmented. Before that, homogeneity characterises the house price growth of 34 cities, especially in the first period. It therefore comes as no surprise that the clustering pattern of the last period determines the clustering pattern based on the whole sample. Aside from the red-hot markets already identified across the entire timeframe, a new cluster, comprising mainly lower-tier centres in Eastern and Central China, also emerges in the period after 2014.

Obviously, the clustering pattern is highly associated with the cities' geography and economic position. Chapter 4 formally tests the usefulness of two widely used classification schemes in describing the housing market structure: the four-region geographical scheme (Eastern, Central, Western and Northeast) used by NBSC and the four-tier city system based on socio-economic conditions published by Jones Lang LaSalle. The four-region geographical scheme fails to explain any housing market structure, while a broad two-region geographical scheme (Eastern - Others) makes a certain amount of sense. The city-tier system is, of course, a superior solution to geographical demarcation. Therefore, regional housing market researchers should proceed with caution when aggregating the city-level housing markets based on geographical proximity.

The geography-based clustering pattern of Chinese housing markets is analogous to the pattern found in the housing markets of other countries. Using k-means clustering techniques, Abraham et al. (1994) partitioned 30 U.S. metropolitan housing markets into three homogeneous groups: a West Coast group, an East Coast group and a central U.S. group. However, as discussed previously, geography as a key determinant of housing market structure only makes sense at a very broad scale. For example, if more clusters are specified for the U.S. market, the West Coast group can be divided into two clusters, but not north and south. The same conclusion is also derived from UK commercial real estate markets. When performing the classification, both Hoesli et al. (1997) and Hamelink et al. (2000) found a strong property-type dimension and a weak broad geographical dimension, which identifies London as the core and the remaining cities as peripheral markets.

Compared to Chapter 4, which focuses on broad market classifications, Chapter 5 looks into detailed interrelationships between cities' house price developments and concentrates on a relatively small spatial scale: the housing markets of 10 vital cities in the Pan-Pearl River Delta (Pan-PRD). The Pan-PRD, located in South China, includes both developed Eastern regions and less developed Central and Western regions and is a regional cooperation framework established in 2004 that aims to remove trade barriers between regions and finally achieve the economic integration of this area. Specifically, Chapter 5 investigates leading-lag relationships, long-run convergence properties and diffusion patterns between markets based on the house price indexes from June 2005 to May 2015.

Leading-lag relationships, which mean that the historical house price information of leading markets can be used to predict the current house prices of lagging cities, are examined by the Toda-Yamamoto (TY) Granger causality test (Toda and Yamamoto 1995). Compared to the standard procedure that requires the tested series to be stationary, the TY procedure is more powerful and allows the series to be integrated or cointegrated of an arbitrary order. Given that the house price series in this analysis comprises both *I(0)* and *I(1)* process, the TY procedure is particularly preferable. The results suggest widely existing leading-lag relationships between housing markets. In contrast to Clapp et al. (1995) and Chen et al. (2011), who found house price interrelationships only among neighbouring markets, the leading-lag relationships in this research are beyond geographical proximity, emphasising the role that economic linkages play in shaping the spatial interaction of housing markets (Pollakowski and Ray 1997). Most importantly, the results tentatively reveal a unidirectional causal flow from the developed eastern-central areas to less-developed western China. This pattern is largely in line with the findings in the UK regional housing markets, where the house price changes are first observed in London or the southeast and then transmitted to the remaining areas (Alexander and Barrow 1994).

The concept of housing market convergence has several meanings. Some researchers think of it as long-run cointegration relationship, in that house price developments are tied together over the long-run (e.g.,MacDonald and Taylor 1993). In other words, there is a long-run equilibrium relationship between the house price developments of different markets and the markets do not move apart from each other. Others think of convergence as a tendency for a diminishing gap between different cities' house price levels over time (e.g., Kim and Rous 2012). Chapter 5 adopts the former understanding and examines the pairwise cointegration relationship between markets using the Engle-Granger (EG) two-step procedure. The fact that only a few city pairs are cointegrated indicates a generally diverged interurban housing market in Pan-PRD; the cities in eastern China in particular are found to be significantly different from the remaining markets. However, the housing markets of three cities in central China form a 'cointegration club'. Furthermore, Abbott and Vita (2013) test a more stringent concept of convergence in which the relative price ratio of two markets remains stable over the long-run. Conditional on the cointegration relationship, this concept necessitates two more conditions: (1) the cointegration vector being (1, -1) and (2) no deterministic trend in the cointegrating vector. The results suggest that the evidence for this type of convergence is even less. The finding of divergence in the urban Chinese housing market contradicts previous studies (e.g.,Wang et al. 2008; Li and Li 2011) that support the long-run cointegration of housing markets. The discrepancy might be because previous studies are confined to a relatively small and homogeneous area, while this study covers a larger and more heterogeneous region.

Chapter 5 then builds a spatial-temporal house price diffusion model to fully capture the house price development characteristics along both spatial and temporal dimensions. This model is a variant of the house price model proposed by Holly et al. (2011). In the model, the house price change of a city at time depends on three components: (1) the historical house price changes itself, (2) the historical house price changes of all other cities, (3) the long-run cointegration relationship with the other cities, if there is any. Note that other cities' influences over the long- and short-runs are compressed into a single variable through a spatial weight matrix, with nearby cities contributing more to the variable. The estimation of this model confirms previous findings and reveals a cross-city spillover effect from neighbouring cities in the short-run. Further, the generalised impulse response function of the model depicts a clear diffusion pattern that price shocks first spread to nearby cities, with those further away taking a longer time to respond.

While almost all of the studies support the short-run spillover effect between housing markets, there is no consensus about convergence over the long run. In UK regional housing markets, for instance, Alexander and Barrow (1994), Meen (1996) and Cook (2003), among others, present some positive evidence in favour of the housing markets cointegration or convergence, whereas Drake (1995) and Abbott and Vita (2013) cast doubt on that hypothesis. With regard to the U.S. metropolitan housing markets, Canarella et al. (2012) documented conflicting evidence about the long-run convergence. This research also provides mixed evidence for the Chinese interurban housing markets. Several reasons can be responsible for the contradictory findings, such as different interpretations of market convergence, the spatial scale for defining the housing markets and the time series used for analysis.

In summary, the house price developments in Chinese cities have not been homogeneous in the past decade. However, the market divergence seems to be a new phenomenon emerging after 2014, with the markets of a few developed eastern cities standing out. Indeed, an in-depth investigation into the Pan-PRD markets shows that the housing markets of eastern cities lead the market changes of the remaining cities; over the long run, the former markets clearly deviate away from the latter, although cross-city spillovers between cities can occur in the short-run. These features are coincident with the house price dynamic behaviour of many other countries.

§ 7.1.3 Exploring the effect of spatial factors on the construction of price index

Chapter 6 is concerned with the construction of the house price index for a Dutch city. Particular attention has been paid to the influence of the different treatment of location on the performance of house, land and structure price indexes. The related questions are:

How can the house price index be decomposed into a land price index and a structure price index? Does better treatment of location benefit the construction of a house price index?

To decompose the house prices, Chapter 6 follows the framework of 'builder's model', which states that the house value can be split into the value of the land and the value of the structure (Diewert et al. 2015). This common belief is at the root of house price decomposition models, such as the residual model and hedonic model (e.g., Davis and Heathcote 2007; Kuminoff and Pope 2013). Chapter 6 is conducted under the hedonic framework, which simultaneously estimates the shadow price of the structure and land. To do so, the price of a structure, which is producible, is assumed to be constant across the city, whereas the price of land, which is non-producible, is assumed to vary across the space. Three models are specified with different treatment of land prices. In the first model, land price remains constant across the city. The second model assumes that land price varies across postcode areas but is the same within each postcode area. In the third model, land price is assumed to be property-specific and thus can be different even within a postcode area. The first two models are easily estimated by ordinary least squares (OLS). The third, however, is not tractable by OLS; thus, a mixed geographically weighted regression (MGWR) model is introduced. Additionally, the MGWR model enables one to plot a continuous surface of land prices across the city.

The three models are estimated on the single-family housing market of a small Dutch city from 1998 to 2007. As expected, the second model, with land price varying across postcode areas, has more prediction power than the first model, which excludes the spatial variability of land prices; in turn, the more flexible MGWR model outperforms the relatively restricted second one. Thus, it might be not appropriate to assume the same land price within a postcode area.

The hedonic imputation Laspeyres, Paasche and Fisher indexes for house, land and structure prices are then computed based on the estimation results of the three aforementioned models. For the Laspeyres and Paasche index, the house price index is actually the weighted average of the land and structure price index. The Fisher index is the geometric mean of the Laspeyres and Paasche indexes. The results show that the Fisher house price indexes based on the different models are almost identical. However, not taking spatial variation of land prices into account indeed influences the Fisher land and structure price indexes; they are biased upward and downward, respectively. There is not much difference between the indexes based on MGWR and those based on the model considering a postcode-level variation of land prices, though. In short, a better treatment of location in the house price model does improve predicting power, but it does not greatly benefit the construction of house price index. This might be good news for statistical agencies, as a simple model can do a good job in terms of house price index construction.

According to the Fisher structure price index derived from the MGWR estimation, the structure prices have increased by more than 90% from 1998 to 2007. Meanwhile, the national construction cost index (CCI) published by Statistics Netherlands (CBS) only experienced a 30% increase, which challenges the structure price index produced in this chapter. The explanations for the significant disparity between these two indexes are twofold. First, construction cost can reflect the market price of structures accurately only if the market clears perfectly. However, in reality, this is hardly true, given that one cannot easily tear down the old structure and rebuild a new one. Thus, a construction cost index does not necessarily coincide with the structure price index derived from the hedonic framework. Second, the MGWR model in this chapter only includes very limited structural variables. If the omitted structural characteristics are improving constantly over time, the estimated structure price index is biased upwards.

Throughout the sample period, the value of land accounts for about one third of the total value of a house in the sample city, which is low compared to the share of 0.52 of national account during 2001-2007. Francke and van de Minne (2016) also estimated a land share between 40% and 50% for the Dutch city of 's-Hertogenbosch. The relatively low land share in this research might be attributed to the omitted variable bias; the omission of structural variables would bias the structure value upwards and hence the land value downwards. It is more likely, though, that the low share of land is because the city in this research lies in a less prosperous area with fewer amenities, which reduces the households' appreciation of land in that city. Davis and Palumbo (2008), using a residual approach, estimated the components of home values for 46 large U.S. metropolitan areas in the year 2004. For the Midwest metropolitan areas, which are less developed, they reported an average land share of 36%, which is very similar to the figure in this research.

§ 7.2 Reflections on the research

This dissertation mainly contributes to an understanding of spatial interactions of housing markets in China, where market forces were introduced in the late 1990s. Nevertheless, the chapters about the spatial distribution of house prices are relevant to any housing markets that operate under the market principle. Chapter 6 is slightly different and is based on the housing market of a Dutch city, but it is enlightening about the construction of land price index in China. Currently, the Hang Lung Centre for Real Estate of Tsinghua University publishes a quarterly quality-adjusted house price index for eight cities since 2006, as well as the land price index for 35 cities since 2004. The land price index is constructed using vacant land transactions. However, an increasing number of vacant land transactions are observed in the outer urban area, which will consequently cause systematic bias in the construction of a city-level land price index. The method proposed in Chapter 6 will serve as a good alternative to the estimation of land price index in Chinese housing markets.

However, as with other scientific studies, the chapters in this dissertation are subject to various flaws, and some findings need to be interpreted with caution. This section will discuss the weakness of this research and give some directions for future research. Potential policy implications of the findings are also discussed.

§ 7.2.1 Limitations

The data problem

Since the urban private housing market in China is still a young market, obtaining a high-quality data set of housing transactions is not easy. The housing authority registers housing transactions without much detailed information. Those authorities are also very cautious with their information; thus, public access to the data is extremely limited. The measure of house prices and their developments might therefore contain some "noise" that affects the reliability of the results.

Both Chapters 2 and 3 compare the house prices between different cities. An ideal house price measure in such analysis would be the price of a "standard" house in each city. However, such a quality-adjusted house price measure cannot be obtained without detailed housing characteristics. The only available information is the total value and areas (represented as square metres) of all the transacted properties of each city, which can allow one to calculate the average sale price. In this manner, the differences of average house prices between cities come from two sources: the disparities in city characteristics and the differences of housing market composition, among which only the former is of interest in this analysis. Failing to control for market composition differences in house price measure might influence the estimation results. In the Chinese context, however, this seems not to be a big problem. First, the standard measurement of house price per square metre is used, meaning that area, the most important housing characteristic, is controlled. Second, the most popular dwelling type in China is the multi-family apartment; there is only a very small market share for luxury apartments and houses. Third, average house prices are measured mainly based on newly residential buildings, which mitigates the influence of depreciation. Therefore, it is believed that house price disparities between cities are mainly caused by differences in city characteristics.

Even these crude housing market statistics need intensive negotiation with government agencies; moreover, for many of the cities, the data is not usable. The empirical foundations presented in Chapters 2 and 3 are thus constrained to the Pan-Yangtze River Delta in eastern China, and the housing markets of only 42 prefecture-level cities (municipalities) are used. This small sample might undermine the reliability of the estimation results, so the findings might not be generalised to the whole Chinese housing markets. Therefore, readers should interpret the results with caution.

Chapters 4 and 5, which investigate the spatial pattern of house price dynamics, utilise the house price indexes published monthly by the National Bureau of Statistics of China (NBSC). The NBSC indexes are compiled by a so-called matching approach so that the quality changes can be somewhat controlled for (Wu et al. 2014). However, the NBSC index is widely criticised for its underestimation of house price growth (e.g.,Wu and Deng 2015; Fang et al. 2016). If the house price developments of each city are systematically biased downwards to the same degree, it will not significantly affect the clustering pattern and the spatial interrelation pattern. It may, however, be possible that the bias of a higher house price growth series is much more severe than that of lower house price growth series, although there is no solid evidence supporting this idea. As such, the findings in Chapter 4 and 5 might be influenced to some extent. However, the NBSC indexes are the only accessible indexes that cover all the cities in the study for a relatively long period.

Chapter 5 examines the long-run relative relationships between housing markets. How long of a time series is enough for this long-term analysis? For the UK studies, MacDonald and Taylor (1993) utilise a time series of 19 years, Cook (2003) uses 29 years, and Holly et al. (2011) use 34 years. The length of time series for the U.S. studies is 36 years for Yunus and Swanson (2013) and 30 years for Gupta and Miller (2012). Consequently, the 10-year time series used in Chapter 5 seems to be rather limited for a long-term behaviour analysis, especially given that the transitory condition of the housing markets in that period. Thus, it is no surprise to find no evidence of long-run convergence or cointegration. In addition, it seems that all the time series analyses of

Asian housing markets suffer from the same data problem. For example, both of the studies on the Malaysian and Taiwanese housing markets use a time series of approximately 10 years (Chen et al. 2011; Lean and Smyth 2013).

Methodological weakness

Aside from the data problem, methodological issues also limit a deeper understanding of the spatial dimension. In Chapter 2, the results are obtained by running an OLS regression on a pooled data set. Although a few city-specific characteristics are included, these variables might not be sufficient to fully control for the city heterogeneities. A panel data specification with fixed or random effects might be a better choice. Additionally, the empirical model fails to consider the spatial dependence between house prices, even though it has been proven that spatial interdependence is prevailing among the markets (e.g.,Fingleton 2008). The spatial interdependence would probably lead to inefficient estimators and thus affect the statistical inference. Chapter 3 comprehensively investigates the spatial dependence of interurban house prices using spatial econometrics, paying particular attention to the spatial spillovers caused by city network externalities. To do so, urban size, measured by land area and urban population density, is included in the explanatory variables. However, according to the spatial general equilibrium framework, the urban population and house prices of a city are jointly determined, indicating the potential endogeneity of the empirical model. In Chapter 3, the endogeneity problem is partially mitigated by measuring house prices and urban sizes at different spatial scales. More sophisticated methods, such as instrumental variable approach, might better solve this problem.

Chapter 4 assigns the housing markets of 34 cities to a few homogeneous groups according to house price growth paths using a hierarchical cluster method. While the hierarchical cluster method produces a dendrogram that depicts how the cities are grouped into clusters step by step, it is difficult to determine the appropriate cut point. Chapter 4 uses an “elbow” approach to choose the number of clusters. However, the choice is still somewhat arbitrary and should be based on more objective criteria. Another weak point of the clustering method is that the analysis is solely based on the time-series behaviour of house price changes but pays no attention to the underlying market structure. In this sense, two totally different housing markets, one driven by demand factors and the other driven by supply factors, can be fused simply because they have similar growth rates.

Chapter 5 examines the spatial interrelationships between housing markets. The Granger causality test is employed to explore whether house price changes in the leading markets cause similar shifts in the lagging markets. In performing this test, however, one cannot exclude the possibility that such correlation is caused by common shocks, meaning that the leading-lag relationships are just the results of different responses of different markets to common factors and are not due to causal relationships. Chapter 5 then tests whether the relative house price ratio between markets remains stable in the long run under the co-integration framework. The results refuse the long-run cointegration or convergence relationships, indicating that the house price differences between markets are either narrowed or widened. However, the cointegration method delivers no answer on the tendency of relative house price ratios over time.

In Chapter 6, the hedonic framework is used to decompose the house price into land price and structure price. As criticised by many other researchers, however, this approach is prone to omitted variable bias. Consumers who buy an expensive land plot in a good neighbourhood also tend to spend more on structural materials. If these superior structural characteristics cannot be appropriately controlled for, their effect on house prices will be confounded with the value of land. Chapter 6 is not exceptional, either. To ascertain an accurate estimate of land and structure values, one has to include as many housing characteristics as possible.

§ 7.2.2 Future directions

Despite the data problem and methodological flaws, this research is still a good attempt to understand the spatial dimension of Chinese housing markets. However, much work needs to be done to develop a full picture about the spatial behaviour of housing markets.

A simple extension of this research is to test whether the agglomeration spillovers, of both higher-tier cities and neighbouring urban concentrations, shape the house price pattern across all of China and which spillovers play a more important role. In the U.S. context, Partridge et al. (2009) found that effects generated by urban hierarchy are generally larger than those of undifferentiated market potential. With regard to the Chinese housing markets, nothing has been determined about the relative importance of these two spillovers.

Given the importance of location in determining house prices, the question remains as to whether location also contributes to house price developments. Together with agglomeration spillovers, one can test whether the large cities and their neighbours experience more house price growth than remote, small cities. On the other hand, in agglomeration economies, house prices serve as an important channel of "centrifugal" forces that drive the decentralisation of population and economic activities to peripheral areas. Thus, it is interesting to learn how the sky-high house prices in China's super cities like Beijing, Shanghai and Shenzhen affect households' location decisions, especially the relocation decisions of young people.

Owing to the spatial fixity of houses, the housing market is no doubt a local market and is largely influenced by city-specific characteristics. In this sense, the effect of natural amenities, such as climate and environmental conditions, on house prices desires more attention. Given the rapid growth of household wealth, households are more willing to pay for the quality of life in the city (Zheng et al. 2009). Indeed, as indicated in Figure 1 of the Introduction, cities in northeast China, which have very cold winters, are among the cities with the slowest house price growth during the last decade. It is also reported that an increasing number of households from northeast China buys their second home in the southern islands of China and spends their winters there, where it is warmer. Thus, one can speculate that the flourishing housing markets of coastal cities in eastern China might be attributed to their friendly climate. Meanwhile, the productivity advantages in eastern areas are also greater. It is still unclear which component is more important in driving the growth of house prices.

House price developments are driven not only by local factors but also by national factors, such as monetary policy and macro business cycles. Chapter 4 tentatively suggests that, before 2014, house price growth in China might have been driven by a national component, but since then, regional and local components have played a larger role. Future research can formalise this idea and use factor models to disentangle the relative importance of national and local components in house price dynamics.

Long-term stable house price ratio between markets originates from the empirical observation of the UK housing market that house price disparities between North and South widened in the 1980s but tended to come back together in the 1990s (Giussani and Hadjimatheou 1991). Since then, substantial effort has been made to test such long-run equilibrium relationship empirically by using various time-series techniques on the UK housing markets and those around the world. However, a theoretical foundation is still lacking. A possible theory that can investigate the relative house price behaviour between markets is the New Economic Geography (NEG) model with a housing sector (Helpman 1998; Fujita et al. 1999). The NEG theory allows for the existence of multiple equilibria, with each having its own attractive basin in terms of determinant conditions, such as population share. If the shocks to determinant conditions are not beyond the threshold of the attractive basin, the relative house price relationship will return to the original equilibrium, namely the stable long-run equilibrium relationship that has been widely discussed. Otherwise, the original equilibrium will be broken, and a new equilibrium will be formed. In that case, the relationship of relative house prices between markets is not stable, but shows some tendency to change over time until a new equilibrium is established. The NEG framework seems to explain the evolution of Chinese city-level housing markets well. During the past decade, one can observe continuous migration from the less developed western area to the developed eastern area, which has possibly driven the transition from an old equilibrium to a new one. In this transition process, it is no surprise to find divergence of relative house price ratios between markets. However, there is still a long way to go before such a sophisticated model is built; moreover, a lot of empirical work needs to be done regarding the transitional path of relative house price relationships.

§ 7.2.3 Policy implications

The findings of this research have policy implications not only for housing policies but also regional development policies. Given the high home ownership rate in China and households' strong desire to own a home, there are significant public concerns about house prices and their dynamics. Government intervention is a standard tool to stabilise the housing markets and ensure affordability. After the establishment of private housing markets in 1998, central government agencies such as the State Council and the People's Bank of China played a major role in creating housing-related policies. These centralised policies applied to all local housing markets with differing conditions and worked as expected because, as shown in Chapter 4, local house price growth trajectories were very homogenous across the country at that time. However, the market divergence since 2010 have been increasingly prominent, with some developed cities standing out and developing along their own paths. Such market divergence in the recent period indicates that the centralised national policy will be helpless and calls for the government to resort to some regional- or local-based policies. For example, according to the results of Chapter 4, for most of the cities in Central and Western China, a unified policy framework will be enough. But for some import economic centres in Eastern China, the local governments have to tailor local policies based on their own market conditions. The need for diversified housing policies has been recognised by the policy makers in the practice. For example, in the recent intervention in housing markets after 2015, the central authorities did not introduce any monetary policies as they had done before; the policy instruments mainly came from local governments. However, when tailoring the local-based policies, one should also consider the interaction between local markets given the fact that the price changes in some markets can spread out to other lagged markets.

For almost all of the cities that have been exposed to tremendous house price growth, policy instruments have primarily sought to constrain demand. For example, as a response to the recent house price boom, many local governments have increased down payment requirements for mortgages, with an even higher requirement set for a second home. They also prohibit potential home-buyers who have worked in the city less than a period from gaining access to the market. In doing so, the housing demand is indeed suppressed in the short run by squeezing out marginal home-buyers from the market, and house price growth will temporarily slow down. However, from the spatial equilibrium point of view, the long-run housing demand will not diminish as long as these cities retain their productivity and amenity advantages arising from agglomeration economies; once the demand restrictions are loosened to some extent, the house prices will bounce back at an unexpected magnitude. In this regard, the local governments have to reassess their policy tools and focus on more about the supply side. For example, local governments, as the owners of urban land, can increase the supply of residential land. Of course, the housing supply is not unlimited but relies on topographical and planning restrictions. An alternative approach is to encourage the efficient use of the current housing stock. According to the China Household Financial Survey, the average housing vacancy rate of 6 large cities in 2013 stands at 22.38%[3]. Letting these vacant homes accommodate families can have a great impact on the market. Policy instruments, such as property taxes, should be introduced to motivate multi-home owners to place their extra dwellings on the market.

From a national perspective, however, the high house prices in developed cities are not entirely negative. Housing costs serve as an important spatial adjustment mechanism in balancing the distribution of economic activities across regions. High house prices in big cities force workers to relocate to small cities, which is good for the development of peripheral regions. The central government should play a role in this adjustment process. For example, a national or regional cooperation framework is necessary to guide the relocation of physical and human capitals to the peripheries. Higher-level planning can also help small and rural settlements functionally integrate with large urban concentrations. Furthermore, policymakers in China should consider issues such as whether to develop smaller cities linked by a fast transit network or to continue to build mega-cities.

Local governments of peripheral regions also need to rethink their policies of stimulating local economies. For a long time, small cities in peripheral areas have relied on the assumption that their economies can thrive through mass investment in construction. Thus, they build multilane roads in the city and new residential buildings in the urban fringe. To some extent, mass construction works by introducing job opportunities. It turns out, however, that the wide roads of many cities are utilised by only a few cars and that the newly developed areas become ghost neighbourhoods. People escape from peripheral cities because of both the productivity and amenity disadvantages. Would it not be better for these governments to invest more in providing high-quality public goods and services, such as education and healthcare services?

References

Abbott, A., & Vita, G. D. (2013). Testing for long-run convergence across regional house prices in the UK: A pairwise approach. Applied Economics, 45(10), 1227-1238.

Abraham, J. M., Goetzmann, W. N., & Wachter, S. M. (1994). Homogeneous groupings of metropolitan housing markets. Journal of Housing Economics, 3(3), 186-206.

3 The six cities are Beijing, Shanghai, Tianjin, Chongqing, Chengdu and Wuhan.

Alexander, C., & Barrow, M. (1994). Seasonality and cointegration of regional house prices in the UK. Urban Studies, 31(10), 1667-1689.

Alonso, W. (1973). Urban zero population growth. Daedalus, 102(4), 191-206.

Bengtsson, T., & Cavanaugh, J. E. (2008). State-space discrimination and clustering of atmospheric time series data based on Kullback information measures. Environmetrics, 19(2), 103-121.

Boix, R., & Trullén, J. (2007). Knowledge, networks of cities and growth in regional urban systems. Papers in Regional Science, 86(4), 551-574.

Canarella, G., Miller, S., & Pollard, S. (2012). Unit roots and structural change: An application to US house price indices. Urban Studies, 49(4), 757-776.

Chen, P.-F., Chien, M.-S., & Lee, C.-C. (2011). Dynamic modeling of regional house price diffusion in Taiwan. Journal of Housing Economics, 20(4), 315-332.

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

Cook, S. (2003). The convergence of regional house prices in the UK. Urban Studies, 40(11), 2285-2294.

Davis, M. A., & Heathcote, J. (2007). The price and quantity of residential land in the United States. Journal of Monetary Economics, 54(8), 2595-2620.

Davis, M. A., & Palumbo, M. G. (2008). The price of residential land in large US cities. Journal of Urban Economics, 63(1), 352-384.

de Bruyne, K., & van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. Applied Economics, 45(13), 1673-1689.

Diewert, W. E., de Haan, J., & Hendriks, R. (2015). Hedonic regressions and the decomposition of a house price index into land and structure components. Econometric Reviews, 34(1-2), 106-126.

Drake, L. (1995). Testing for convergence between UK regional house prices. Regional Studies, 29(4), 357-366.

Fang, H., Gu, Q., Xiong, W., & Zhou, L. (2016). Demystifying the Chinese housing boom. In M. Eichenbaum, & J. Parker (Eds.), NBER Macroeconomics Annual 2015 (Vol. 30, pp. 105-166). Chicago: University of Chicago Press.

Fingleton, B. (2008). Housing supply, housing demand, and affordability. Urban Studies, 45(8), 1545-1563.

Francke, M. K., & van de Minne, A. M. (2016). Land, structure and depreciation. Real Estate Economics, Forthcoming.

Fujita, M., Krugman, P., & Venables, A. J. (1999). The spatial economy: Cities, regions, and international trade. In. Cambriage MA: The MIT Press.

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? Journal of Regional Science, 52(2), 172-191.

Giussani, B., & Hadjimatheou, G. (1991). Modeling regional house prices in the United Kingdom. Papers in Regional Science, 70(2), 201-219.

Gupta, R., & Miller, S. M. (2012). "Ripple effects" and forecasting home prices in Los Angeles, Las Vegas, and Phoenix. The Annals of Regional Science, 48(3), 763-782.

Hamelink, F., Hoesli, M., Lizieri, C., & MacGregor, B. D. (2000). Homogeneous commercial property market groupings and portfolio construction in the United Kingdom. Urban Studies, 32(2), 323-344.

Head, K., & Mayer, T. (2004). The empirics of agglomeration and trade. In J. V. Henderson, & J. F. Thisse (Eds.), Handbook of Regional and Urban Economics (Vol. 4, pp. 2609-2669). Amsterdam: Elsevier.

Helpman, E. (1998). The size of regions. In D. Pines, E. Sadka, & I. Zilcha (Eds.), Topics in public Economics (pp. 33-54). New York: Cambridge University Press.

Hoesli, M., Lizieri, C., & MacGregor, B. (1997). The spatial dimensions of the investment performance of UK commercial property. Urban Studies, 34(9), 1475 -1494.

Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of Urban Economics, 69(1), 2-23.

Kakizawa, Y., Shumway, R. H., & Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. Journal of American Statistical Association, 93(441), 328-340.

Kim, Y. S., & Rous, J. J. (2012). House price convergence: Evidence from US state and metropolitan area panels. Journal of Housing Economics, 21(2), 169-186.

Kullback, S. (1968). Information theory and statistics. New York: Dover Publications.

Kuminoff, N. V., & Pope, J. C. (2013). The value of residential land and structures during the great housig boom and bust. Land Economics, 89(1), 1-29.

Lean, H. H.,& Smyth, R. (2013). Regional house prices and the ripple effect in Malaysia. Urban Studies, 50(5), 895-922.

LeSage, J. P., & Pace, R. K. (2009). Introduction to spatial econometrics. Boca Raton: Taylor & Francis CRC Press.

Li, J., & Li, H. (2011). The spatial diffusion and convergence of urban housing price in the Pearl River Delta. Journal of Chongqing University (Social Science Edition), 17(6), 17-23.

MacDonald, R., & Taylor, M. P. (1993). Regional house prices in Britain: long-run relationships and short-run dynamics. Scottish Journal of Political Economy, 40(1), 43-55.

Meen, G. (1996). Spatial aggregation, spatial dependence and predictability in the UK housing market. Housing Studies, 11(3), 345-372.

Meijers, E. J., & Burger, M. J. (2015). Stretching the concept of 'borrowed size'. Urban Studies, Forthcoming.

Osland, L., Thorsen, I., & Gitlesen, J. P. (2007). Housing price gradients in a region with one dominating center. Journal of Real Estate Research, 29(3), 321-346.

Partridge, M. D., Rickman, D. S., Ali, K.,& Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. Journal of

International Economics, 78(1), 126-140.

Pollakowski, H. O., & Ray, T. S. (1997). Housing price diffusion patterns at different aggregation levels: An examination of housing market efficiency. Journal of Housing Research, 8(1), 107-124.

Roback, J. (1982). Wages, rents, and the quality of life. Journal of Political Economy, 90(6), 1257-1278.

Rosen, S. (1979). Wage-based indexes of urban quality of life. In P. Mieszkowski, & M. Straszheim (Eds.), Current Issues in Urban Economics. Baltimore: The Johns Hopkins University Press.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66, 225-250.

Vega, S. H., & Elhorst, J. P. (2015). The SLX model. Journal of Regional Science, 55(3), 339-363.

Wang, S., Yang, Z., & Liu, H. (2008). Empirical study on urban housing price interactions among regional marekts in China. Research on Financial and Economic Issues, 6, 122-129.

Wu, J., & Deng, Y. (2015). Intercity information diffusion and price discovery in housing markets: Evidence from Google Searches. The Journal of Real Estate Finance and Economics, 50(3), 289-306.

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

Yunus, N., & Swanson, P. E. (2013). A closer look at the U.S. housing market: Modelling relationships among regions. Real Estate Economics, 41(3), 542-568.

Zheng, S., Fu, Y., & Liu, H. (2009). Demand for urban quality of living in China: Evolution in compensating land-rent and wage-rate differentials. Journal of Real Estate Finance and Economics, 38(3), 194-213.

Curriculum Vitae

Yunlong Gong was born on August 16th, 1987 in Shandong province, China. He obtained his bachelor degree in land resource management in 2009 at China University of Mining and Technology. He then started his master study at the same university and involved in a research project aiming to model the spatial diffusion of land prices, which was funded by National Natural Science Foundation of China. As a research master, he has (co)-authored several academic articles on the measurement and development of land prices. In 2012, he received a scholarship from China Scholarship Council to support his PhD research at OTB – Research for the Built Environment, Delft University of Technology. With this PhD project, he further developed his interest in the spatial interaction of land and housing markets.