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}), $A_i = A\left(c_i, s_i, \theta s_{-i}\right)$, 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: $W_i = W\left(l_i, s_i, \theta s_{-i}\right)$, 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 $P_i = P\left(c_i, l_i, s_i, \theta s_{-i}\right)$. 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 **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).

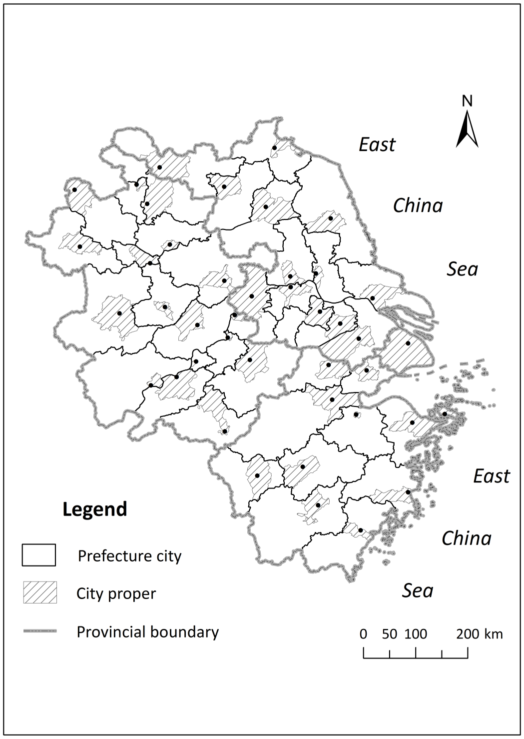

FIGURE 3.1 Cities in Pan-Yangtze River Delta

§ 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 emission)	-0.1231***	-0.1396***	-0.1297***	-0.1325***
	(-5.24)	(-4.37)	(-4.12)	(-4.22)
Ln(Student/Teacher ratio)	-0.4662***	-1.3482***	-1.1673***	-1.1815***
	(-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 density)			0.2719***	0.2795***
			(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

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

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***	0.0601***	0.0122	0.0126
	(2.75)	(3.14)	(1.52)	(1.57)
Ln(Smoke and dust	-0.1297***	-0.0219	-0.0409	-0.0420
emission)	(-4.12)	(-0.74)	(-1.54)	(-1.59)
Ln(Student/Teacher	-1.1673***	-0.6218***	-0.7090***	-0.7080***
ratio)	(-6.21)	(-3.07)	(-4.33)	(-4.29)
Doctor	0.0268***	0.0026	0.0065	0.0069
	(2.93)	(0.31)	(0.85)	(0.90)
Coast	0.2230*	0.1560*	0.1793*	0.1671*
	(1.93)	(1.68)	(1.91)	(1.77)
Ln(Arable land)	-0.0497	-0.2906**	-0.1091	-0.1178
	(-0.40)	(-2.40)	(-1.09)	(-1.24)
Ln(Population density)	0.1181*	0.1278***	0.2042***	0.1909***
	(1.87)	(2.65)	(3.95)	(4.10)
Ln(Land)	0.1897***	0.1702***	0.1737***	0.1757***
	(2.72)	(2.80)	(3.07)	(3.12)
$\mathbf{W} \times$ Ln(Population	0.2719***	0.2493***	-0.0686	
density)	(3.04)	(2.91)	(-0.85)	
$\mathbf{W} \times$ Ln(Land)	0.2590**	0.2490**	-0.0564	
	(2.00)	(2.02)	(-0.52)	
$\mathbf{M} \times$ Error		0.7364***		
		(14.25)		
$\mathbf{M} \times$ Ln(House prices)			0.5510***	0.5280***
			(9.48)	(9.96)
Constant	5.3201**	5.5096**	4.4485**	3.8541***
	(2.38)	(2.67)	(2.43)	(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

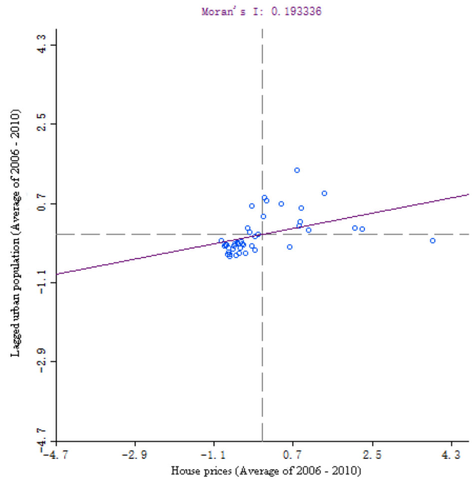

FIGURE C1 Global Moran's I plot

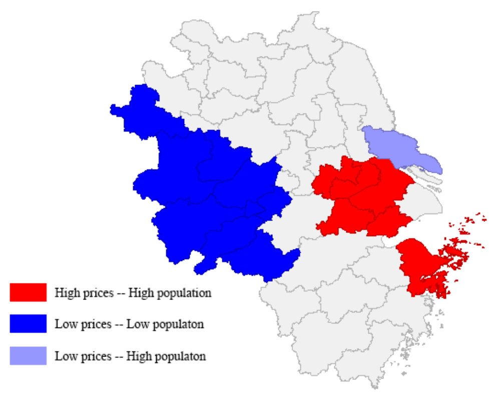

FIGURE C2 LISA cluster map