Schools, Land Markets and Spatial Effects

**Abstract:**

This paper uses a spatial modeling approach to explore the capitalisation effect of proximity to schools on land markets. The results suggest that adjacency to primary schools leads to considerable price premiums but there are no significant effects on middle schools and universities. The results also provide some reassurance that spatial simulations offer a useful representation of localised variations in values attached to proximity to schools in Beijing, China.

**Keywords:** Land market; School valuation; Spatial econometrics

**1. Introduction**

China has experienced substantial land and housing marketisation in the past decade (Liang et al., 2007; Cheshire, 2007; Zheng and Kahn, 2008). This transformation has come alongside massive local public infrastructure investments, booming real estate investments (Zheng and Kahn, 2013). Such rapid but differential spatial expansion in infrastructure and real estate markets will transform the determinants of property prices within cities. In Chinese cities, especially large cities such as Beijing and Shanghai, educational resources are scarce and distributed non-uniformly across space, and thus how school facilities are capitalised into land values have drawn increasing attention of households, policy makers and planners.

As an important source of urban externalities, recent studies have shown that proximity to schools can influence property price premiums and parents’ housing location choice (e.g. Cheshire and Sheppard, 2004; Gibbons and Machin, 2008; Cellini et al., 2010; Gibbons et al., 2013). For example, whilst living near a primary school or middle school may result in commuting time savings for parents and their children, there might also be traffic congestions and noise associated with schools. While mayors in China want to balance the optimal distribution of educational facilities for their cities to achieving an equalisation of educational resources, an important but untested question to optimal education resource allocations is a solid understanding of how the capitalisation effect of proximity to schools varies with the persistence of spatial dependence effects in a land market. This is an important issue because ignoring or mis-specifying spatial dependence in a hedonic land price model is likely to result in biased and inconsistent estimates of the amenity value of proximity to schools (e.g. Brasington and Haurin, 2006).

This paper aims to shed lights on these questions by looking at the distributional effects of proximity to schools on Beijing’s residential land market, by capturing the hierarchical structure underlying the land price data. We improve on the traditional spatial econometric evaluation of proximity to schools by simultaneously modelling two types of unobservables via a Bayesian hierarchical spatial autoregressive model developed in Dong and Harris (2015). More specifically, the property level unobservable effect is modelled by the inclusion of a spatially lagged land price variable as in Brasington and Haurin (2006). The neighbourhood level unobservable impact is modelled as a spatial autoregressive process (see detailed discussions in Section 4). The former corresponds to a horizontal spatial dependence effect—an effect arising from the geographical proximity amongst properties while the latter corresponds to a vertical dependence effect—a top-down effect induced by neighbourhoods (unobservable characteristics such as neighbourhood prestige) upon properties (Dong and Harris, 2015; Dong et al. 2015).

The rest of this paper is organized as follows. Section 2 highlights the limitations of previous studies on the economic valuation of schools. Section 3 describes our econometric models, followed by a summary of the institutional context and data in Section 4. Section 5 presents the results. In the final section, we draw conclusions.

**2. Limitations of previous research**

Most existing research evaluating the captialisation effects of schools assesses their effects on property values and residential sorting patterns. Typically, these studies compare house prices across school catchment zones and boundaries. The empirical results vary across studies. In this section, we highlight some limitations of the existing methodologies on which we try to improve in the present study.

Suppose that we build a statistical model of land price determinants using data where land parcels are nested into neighbourhoods (e.g. census units such as lower layer super output areas in England). A complete model specification would be the one that relates land prices to all land parcel and neighbourhood characteristics. However, not all of the variables are observable or quantitatively measureable and the unobservable part will be embedded into the hedonic equation as a compound model error term. It is reasonable to decompose the unobservable factors to the land parcel level effect denoted as *v* (e.g. the presence of a hazardous landfill near a land parcel) and the neighbourhood level effect denoted as *θ* (e.g. neighbourhood physical and cultural characteristics or prestige). Under this context, if these unobservable factors (either *v* or *θ* or both) are varying systematically across space and correlate with the school variables of interest, the hedonic model residuals will be spatial dependent and the estimates of the school captalisation effect will be biased by using a traditional ordinal least square estimation strategy.

An innovative identification strategy to control for the unobservable factors is the application of boundary fixed effect (or spatial discontinuity) approach (e.g. Black, 1999; Gibbons et al., 2013). In essence, this approach compares house values on the opposite sides of school attendance zone (or catchment area) boundaries in a school district and estimates the value of school quality as differences of the adjusted house prices—controlling for other house characteristics and boundary dummy variables. The key assumption of this approach is that all neighbourhood variables other than school quality remain the same when crossing borders, thus a boundary fixed effect approach can capture both the observed and unobservable neighbourhood characteristics, leaving the property price differentials solely attributable to differences in school quality. While the approach is intuitively straightforward and logically sound, it is not without its problems. For example, neighbourhood characteristics do vary near school attendance zone borders due to a household sorting process (e.g. Brasington and Haurin, 2006; Bayer et al., 2007). If people with high educational status had a stronger preference for schools with high quality than those with low educational status, differences in the neighbourhood-level educational status would exist. This however cannot be effectively controlled by the boundary fixed effect approach.

Brasington and Haurin (2006) proposed to use the spatial econometric approach to tackle the issue of omitted variable bias facing studies of school evaluation. The basic idea is to directly model the unobservable influences (the sum of *v* and *θ*) by adding a spatially lagged dependent variable (house price in this context) into the traditional hedonic price model. The lagged house price variable (constructed through a spatial weights matrix that specifies the connection structure among properties) captures the unobservable influences on house prices and the spatial dependence of them. Technical details of the spatial econometric approach are provided in Anselin (1988) and LeSage and Pace (2009), among others. The spatial dependence in the unobservables is intuitively plausible as the unobservable influences might be similar for properties that are in close geographical proximities (e.g. Brasington and Hite, 2005; Brasington and Haurin, 2006; Wen et al., 2014; Anselin et al., 2010; Lazrak et al., 2014; Leonard et al., 2015).

But a potential problem associated with this econometric approach is the conflation of unobservable influences at the neighbourhood and property levels. It is important to distinguish between the two types of unobservables *u* and *θ* as they might represent different spatial processes—one operating at a property scale and another at a neighbourhood scale. In the methodological term, failing to separate the two sets of unobserved effects could lead to biased estimates of the spatial dependence effect (i.e. the coefficient of the lagged price variable). To explicitly address this concern, we recognise two types of spatial dependence effects in the estimation process. By constructing precisely geo-coded location information and combining this information with GIS maps, we can measure land parcel locations, proximity to school facilities, and other characteristics of local public goods in the spatial context.

**3. Econometric models**

Following the hedonic price modelling literature, land price is related to a series of locational and neighbourhood characteristic variables, as shown in Equation (1):

(1)

where the dependent variable (*LnPriceij*) is the natural logarithm of the price for a residential land parcel *i* in neighborhood *j*. The school variables under investigation are in vectors *Schoolij*, which include proximity to primary and middle schools, and to prestige universities. The quality of nearest primary and middle schools of each land parcel are also included. *Lij* represents locational and structural variables of each land parcel while *Zij* includes neighbourhood level variables. ***β***, ***δ*** and ***φ*** are vectors of regression coefficients to estimate. The vector *θj* are unobserved neighbourhood effects and *εij* are random innovations, following as an independent normal distribution with mean of zero and variance of *σe*2.

An important issue with the standard hedonic model specification is that the horizontal spatial dependence effect of land prices is not captured. A common practice would be incorporating fixed spatial effects in the hedonic price model (*θj* treated as fixed in Equation (1)). It implicitly assumes that dependence in land prices is raised by the neighbourhood level unobservables. This is a fairly strict assumption for real-world land price data as dependence in land prices could also arise from the land parcel level unobservables and spillover effects from one land parcel upon surrounding land parcels and vice versa. By adopting a fixed effect estimation strategy, effects of neighbourhood level variables can no longer be estimated. To address this concern, a typical spatial hedonic price model is usually adopted (Brasington and Haurin, 2006; Anselin et al., 2010),

(2)

where *wi* is a vector of spatial weights, measuring the how closely other observations are related to the *i*th observation. Spatial weights are calculated either by using an inverse distance scheme with a pre-defined threshold distance or based on geographical contiguity (Cliff and Ord, 1981; Anselin, 1988). Multiplying *wi* by the price vector gives a weighted average price of the neighbours of *i* if, as is usually the case, *wi* is normalised so that the sum of its elements equal to 1. Equation (2) allows for an explicit test of whether land price at location *i* is related to land prices at locations to which *i* is connected to (significance of the spatial autoregressive parameter *ρ*), which is a standard simultaneous autoregressive model (SAR) in the spatial econometrics literature.

One potential problem that has not been dealt with by using SAR is the vertical spatial dependence effect in land prices arising from the neighbourhood unobservables. Land parcels in the same neighbourhood are exposed to identical neighbourhood characteristics (*Z*) and unobservables (*θ*). Thus land prices in the same neighbourhood tend to be more similar than land prices across different neighbourhoods (Jones, 1991; Orford, 2000). This vertical type of spatial dependence can be accommodated in a multilevel hedonic price model, which is expressed as,

(3)

where the unobservable neighbourhood effects are treated as random effects, following an independent normal distribution *N*(0, *σu2*). One important feature of the multilevel hedonic price model is that it allows for the estimation of observed neighbourhood characteristics whilst controlling for neighbourhood unobservables. However, it is clear that the horizontal spatial dependence effect is not modelled in Equation (3), leading to the biased estimates of regression coefficients (Anselin, 1988).

To simultaneously capture both the horizontal and vertical spatial dependence in land prices, a hierarchical spatial autoregressive model (HSAR) is adopted in this study:

(4)

where *mj*, similar with *wi*, is a spatial weights vector at the neighbourhood level, measuring how neighbourhood *i* are spatially connected with other neighbourhoods. This model specification has four important characteristics. First, it captures the horizontal spatial dependence in land prices through the term *wi****LnPrice*** and the vertical spatial dependence through the inclusion of unobservable neighbourhood effects (***θ***). Secondly, neighbourhood unobservables are not treated as purely random innovations as in a multilevel hedonic price model. Instead, possible spatial dependence effect in neighbourhood unobservables is also modelled. Lastly, spatial autoregressive parameters at both the land parcel and neighbourhood levels (*ρ* and *λ*) are estimated, and therefore we are able to capture the difference in intensity of spatial dependence at two spatial scales. Failing to model one of them might lead to inaccurate estimation of another (Dong et al., 2015). For instance, ignoring (positive) horizontal spatial dependence, multilevel hedonic models tend to overestimate the vertical dependence effect (overestimated *σu2*). When not controlling for the vertical dependence effect, SAR models tend to overstate the intensity of the horizontal spatial dependence (overestimated *ρ*).

The spatial weights matrix at the land parcel level is calculated by using a Gaussian kernel function with a distance threshold of *h*, as in Equation (5),

(5)

where *dik* represents the Euclidian distance between land parcels *i* and *k*, and *h* is a distance threshold. In this study, *h* was set to 2km. In order to obtain consistent estimates of the proximity effect of schools and universities upon land prices, different threshold distances including 2.5km and 3km were employed in our robustness study. A different spatial weighting scheme, 30 nearest neighbours, was also used for robustness checks of our estimation results. The neighbourhood level spatial weight matrix is measured by using the contiguity of districts. Following the spatial econometric modelling convention, two spatial weights matrices are row-normalized.

**4. Research context and data**

Since the late 1980s, most socialist countries have experienced dramatic socio-economic transitions from a central planned economy system to a market-oriented economy system. Compared with the complete and shock transitions in eastern European countries, the market-oriented transition in China is in a gradual environment (Nee, 1992). As one fundamental part of this transition, urban land reform was first launched in 1987 by the initial practical experiment of land contract leasing in the city of Shenzhen—the first special economic zone in China (*Jingjitequ*). Since then, most Chinese cities have experienced dramatic transformation of its land use system from free allocation toward a leasehold system as the government reinstated the land from a social welfare good to a market commodity.

The context we seek to describe is the urban area of Beijing, one that has experienced dramatic changes in growth and structure over the past two decades. To better understand our datasets, we begin with a brief overview of Beijing and the definition of neighborhood in this study. The metropolitan area of Beijing is divided into several mega-districts, in which six of them are central city districts (Dongcheng, Xicheng, Chaoyang, Haidian, Fengtai, Shijingshan), and others are generally suburban or rural districts. Each mega-district is comprised by a number of neighborhoods (*Jiedao*). Similar as the term of “zone (No.1 to No.6)” in London, Beijing government and previous research have often used the term “ring roads (No.2 to No.6)”—which circled around the CBD[[1]](#footnote-1) to every direction, to define the urban areas of Beijing. We followed this convention to define our study area. A *Jiedao* can be linked to a census block in the sense that it is the primary unit for the organization of the census in urban Beijing, which is defined as neighborhood in the present study.

The primary dataset of this study is the residential land parcel data, collected from the Beijing Land Resources Bureau (BLRB). We have geo-coded all of the 1,210 undeveloped residential land parcels leased to real estate developers during 2003 and 2009. Dropping land parcels without price information, our final sample size is 1,077. Fixed period effects (e.g. external shocks that influence prices of all land parcels) were controlled by year dummy variables.[[2]](#footnote-2) Figure 1 shows the spatial distribution of residential land parcels and land price variations across space.

[Figure 1 about here]

With respect to the school data, we have collected all 652 primary schools, 166 middle schools and 23 prestige universities in Beijing’s main urbanized areas.[[3]](#footnote-3) The school distribution in China is largely exogenous to the current economic conditions since most of schools have been built in the central-planned economy era. Like the UK and other developed countries, Chinese city governments often implement a devolved school policy to allocate eligible school-age children to schools based on the school catchment areas they belong to (especially for primary schools, see Zheng et al., 2015). In general, local primary schools are scarce and have much difference in school quality across neighborhoods in the urban area of Beijing. Households often attach great importance to the primary school quality due to its direct relationship with whether their kids can easily get admission to a better middle school. However, the school quality information (e.g. standard exam scores or spending per student) are not publicly accessible. Following Zheng et al. (2015), we use whether a primary school is a “Key Primary School (KPS)” as a proxy for school quality. These KPS schools were mostly founded in the 1950s by the Beijing Municipal Government. They received much more capital and human resource investment (e.g. modern teaching equipment and high-quality teachers) from the government than other primary schools. With the implementation of the school catchment area policy, significant price premiums have been found for houses within the attendance zones of these KPS compared to those located in attendance zones of other primary schools (Zheng et al., 2015).

But whether primary school quality is capitalised into prices of land parcels is rather unclear. First, residential land parcels were not assigned to any specific school attendance zones when they were leased to real estate developers. The geographical proximity of a land parcel to KPS can, at best, increases its chance of being attaching to the attendance zone of a particular KPS. A further concern is that a land parcel usually consists of several residential complexes, and these complexes could belong to different attendance zones of primary schools. Following a similar logic, we use whether a middle school is rated as Experimental Model Schools (EMS) as a proxy for the middle school quality. Experimental Model Schools are created by the Beijing Municipal Government mainly based on the formerly key middle schools. Given the policy adjustment of school catchment boundaries in Beijing, it is difficult to apply for the standard school catchment zone-based identification strategy to estimate the effects of schools on land values at particular locations without a careful consideration of spatial dependence effects.

Finally, we have calculated important locational variables that have been found to be significantly associated with land prices. For example, the proximity to railway stations is important in influencing the real property values (Bowes and Ihlanfeldt, 2001; Gibbons and Machin, 2005; Liang et al., 2007). From the Beijing Municipal Bureau of Transportation (BMBT), we have acquired the list of Beijing’s well-constructed railway stations before 2003 and calculated distances to the nearest railway station for each land parcel. The proximity to the CBD, nearest green parks, rivers and express roads are also included in our analysis. As suggested by recent studies, any residential sorting arising from neighborhood socio-demographic characteristics will bias the estimates of school values. At the district level (district boundaries in Figure 1), we use the employment and population census data to control for the potential influence of neighbourhood socio-demographic characteristics on residential land values. The employment census data is collected from the 2nd City Employment Census conducted in 2001 by the National Bureau of Statistics of China (NBSC). Our population census data is collected from the City Population Census reported by the National Bureau of Statistics of China (NBSC) in 2000. It includes information on social components of each district such as the median education attainment (Education attainment), the proportion of people renting public housing (Public house renting) and the proportion of buildings built before 1949 (Age of buildings). The indicator of public safety level is measured by the crime rates of the neighborhood (Crime), which is calculated as the number of reported serious crimes (murder, rape, robbery and property crimes including arson, burglary and theft) per 1,000 people in each neighborhood. Table 1 presents the descriptive statistics of variables used in this study.

[Table 1 about here]

**5. Results**

5.1. Baseline estimates of school effects from OLS models

Table 2 presents the estimation results from two standard hedonic price model specifications. Model 1 is our baseline model specification in which only targeted school related variables and fixed period effects are included. We find that proximity to schools and universities are significantly valued in land prices. However, the quality of schools is not significantly associated with land prices, supporting our argument that school quality might not be capitalised into land prices because the geographical proximity of land parcels to a KPS do not guarantee them to be in the school attendance zone of that KPS. Meanwhile, being close to a KPS might be subject to a congestion effect such as traffic-related air and noise pollution. Nonetheless, the basic model specification might be subject to a severe issue of omitted variable bias.

Model 2 further includes land parcel structural and locational characteristics and neighbourhood attributes. In addition, to test whether the proximity effects of primary and middle schools vary with neighbourhood attributes, interaction terms between the distance-to-school variables and the neighbourhood characteristic variables were added.[[4]](#footnote-4)

[Table 2 about here]

After controlling for the structural, locational and neighbourhood variables in Model 2, the effects of proximity to middle schools become insignificant. We also find that the magnitudes of effects of proximity to primary schools and universities decrease considerably in Model 2, as compared to the estimates from Model 1. The large decrease in the estimated effects of proximity to primary schools and universities underlines the importance of a careful consideration of geographical factors in valuing schools. The significant interaction term, Primary school × Age of buildings, suggests that the proximity to primary school tends to be valued more in districts with smaller proportion of buildings built before 1949. A possible explanation is that districts with few very old buildings are suburban areas where primary schools, especially high-quality schools, are much scarcer than that of inner urban areas.[[5]](#footnote-5)

Many of locational variables are significantly associated with land prices. As expected, being closer to the central business district (CBD), railway stations, green parks and express roads would lead to higher land prices. A positive partial correlation between proximity to rivers and land prices implies that moving a land parcel closer to a river tends to decrease its value. The puzzling effect of river accessibility might reflect the situation where most of the rivers in the urban areas of Beijing were severely polluted specifically before the Olympic Games in 2008 (Harris et al., 2013). As for neighbourhood variables, Job density and Age of buildings are two factors found to be significantly related to land prices. Land prices in districts with low job density and high proportion of old buildings tend to be low, everything else equal. Overall, these findings are generally in agreement with previous studies of real estate market in Beijing (e.g. Zheng and Kahn, 2008; Wu et al., 2015; Wu and Dong, 2014). To provide a rationale for the application of the SAR and HSAR as our preferred modelling strategy, we test whether there is spatial dependence left unmodelled in Model 2 by employing the Moran’s *I* statistic based on the spatial weights matrix specified in Equation (5). The results show evidence of significant spatial dependence effects in both Model 1 and 2 that are not modelled (Table 2).

5.2. Estimates of school effects from SAR and HSAR models

Table 3 reports the land price model estimation results using the SAR and HSAR models. For the purpose of model comparison, both HSAR and SAR models were implemented using the Bayesian Markov Chain Monte Carlo (MCMC) simulation approach. The prior distributions and starting values for each parameter in the two models are identical. The 95% credible intervals of each model parameter were reported. Calculations of the credible intervals for each model parameter were based on two MCMC chains, each of which consisted of 60,000 iterations with a burn-in period of 10,000. We further retained every 10-th samples to reduce autocorrelation in each MCMC chain. MCMC diagnostics including trace plots and Brooks-Gelman-Rubin scale reduction statistics (Brooks and Gelman, 1998) indicated rapid convergence of our samplers and efficient mixing of chains for the HSAR model.

[Table 3 about here]

From Model 3 (the SAR model), we find the spatial autoregressive parameter *ρ* is statistically significant, indicating that land price at one location is related to or influenced by that of surrounding locations, i.e. there is significant spatial autocorrelation in land prices. Comparing to our preferred modelling strategy, the HSAR approach (Model 4), the intensity of spatial dependence (*ρ*) amongst land parcels seems to be overestimated in the SAR model by about 50%. The overestimate of the spatial dependence effect is likely due to the conflation of the horizontal spatial dependence at spatial scales of land parcels and districts and the vertical spatial dependence (Dong et al., 2015). This is clearly shown by the estimation results from the HSAR model (Model 4). Horizontal dependence effects at both the land parcel and district levels have been found significant, with *ρ* and *λ* equal to 0.267 [0.157, 0.37] and 0.571 [0.109, 0.842], respectively. The vertical dependence is indicated by the estimated district level variance parameter *σu*2 (0.043), which accounts for about 7% of variations in land prices. By distinguishing between horizontal and vertical spatial dependence, HSAR also improves the model fit by about 5%, as compared to the SAR model (Pseudo-R2 in Table 3).

Turning to the estimation of school variables, we find that proximity to primary schools appears to be significantly capitalised into land values. Regression coefficients can be viewed as significantly different from zero if their 95% credible intervals do not contain zero. The proximity effects of prestige universities are not statistically significant in the HSAR specification. This suggests that previously identified proximity effects of universities might be confounded by spatial dependence effects. The interaction term, Primary school × Age of buildings, is still significant even after controlling for both horizontal and vertical spatial dependence in land prices, suggesting significant differentials in the value of proximity to primary schools across districts. In terms of primary and middle school quality variables, neither of them is statistically significant, suggesting that school quality is not capitalised into land values, ceteris paribus. The differences in the capitalisation effects between primary and middle schools and universities might be related to the varying degrees of benefits that local residents can access from these public goods. That said, primary schools are likely to be local public goods for local residents whereas prestigious universities are more of public goods at the municipality scale or national scale. Living near universities means convenient but not exclusive access to many university facilities and is also subject to potential congestion effects arising from the concentration and large-volume flows of population and traffic.

Looking at the partial marginal effect on land price of a covariate, we need to resort to some scalar summary effect estimates, namely average direct, indirect and total impacts in SAR and HSAR models (LeSage and Pace, 2009; Golgher and Voss, 2015). This is because the partial derivative of land price with respect to a covariate is not equal to the regression coefficient of that variable as long as the spatial autoregressive parameter is significantly different from zero. In a simple case of linear models where the independence assumption for an outcome variable is met, the partial marginal effect of a covariate upon the outcome equals to the regression coefficient of the variable. That is, because observations are assumed to be independent, a one unit change of a variable *xr* at location *k*, *xkr*, will only affect the outcome at location *k*. However, relaxing the assumption of independence as in SAR and HSAR models, the interpretation no longer holds: a one unit change of a variable *xr* at location *k*, *xkr*, will affect the outcome at location *k*, *yk*,and outcomes at all the other locations.

To aid interpreting the parameters from spatial econometrics models, LeSage and Pace (2009) proposed to distinguish the direct, indirect and total impacts on the outcome variable from changes of a covariate. The direct impact is the response of *yk* due to changes of *xkr*, whereas the indirect impact is the sum of responses of outcomes at all the other locations due to changes of *xkr* and thus is also called spatial spillover effects. The total impact is simply the sum of the direct and indirect impacts. The direct and indirect impacts may vary with spatial units, depending on their relative locations in the entire geographical configuration and their neighbours’. Therefore, they further propose to average the direct and indirect impacts arising from changes of *xk* among all locations as a way of interpreting the model parameters. The detailed technical descriptions of the calculation and statistical inference of the averaged direct, indirect and total impacts are provided in LeSage and Pace (2009) and Kirby and LeSage (2009), among others. In addition, Dong and Harris (2015) show that the model parameter in HSAR models can be interpreted in the same way as in SAR model.

The direct, indirect and total impacts of each covariate and the associated 95% credible intervals from SAR and HSAR models are presented in Tables 4 and 5. The direct, indirect and total impacts of the proximity to primary schools in the two models are all significant because the 95% credible intervals do not contain zero. The indirect effects point to spatial spillovers, such that increasing the proximity to primary schools at location *k* leads to higher land prices for neighbouring locations as well. More specifically, in the SAR model (Table 4), the direct, indirect and total effects associated with a 1% decrease in the distance to primary schools would be, on average, around 0.06%, 0.038% and 0.101% increase in land prices, respectively.[[6]](#footnote-6) In contrast, a 1% decrease in the distance to primary schools would be associated with around 0.105% total increase in land prices, which consists of around 0.08% from direct effect and 0.026% from indirect effect. It is useful to note that the direct effect estimates for most variables are quite similar with their regression coefficients in this specific case study, but it is not always true (LeSage and Dominguez, 2012). The spatial spillover effect of the proximity to primary schools on average accounts for about 38% of the total effect in the SAR model, which is about 13% larger than that in the HSAR model. This is likely related to the overestimated spatial autoregressive parameter in the SAR model because of the conflation of spatial dependencies at the land parcel and district scales (Table 3).[[7]](#footnote-7) In summary, we find that proximity to primary schools is significantly capitalised into land prices while proximity to middle schools and prestige universities do not seem to be valued in the land market of Beijing.

5.3. Robustness check

In this section, we provide sensitivity tests for school parameter estimates derived from the HSAR model. We use different threshold distances in the construction of spatial weights matrix at the land parcel level. In addition, one alternative spatial weights matrix based on 30 nearest neighbours of each land parcel was used to implement the HSAR model.

Model estimation results using different spatial weights matrices from HSAR are reported in Table 6. We have reached several robust findings. First, proximity to primary schools is significantly associated with land prices across HSAR models with different spatial weights matrices. The magnitudes of estimated coefficients of primary school accessibility are also quite similar across models. Second, estimates of two types of spatial dependence effect in terms of both spatial autoregressive parameters (*ρ* and *λ*) and variance parameters (*σe*2 and *σu*2) are very similar across model specifications. Therefore, our estimates of the proximity effects of schools and universities using the HSAR model (Model 4) in this study are robust to different specifications of spatial weights matrices.

**5.4 Simulation of the estimated effects**

This section aims to provide a representation of localised variations in values attached to hypothetical changes of the proximity to primary schools. We first interpolate the surface of proximity to primary school in the study area using an ordinary Kriging method (Figure 2(a)). Then, we select land parcels in districts with large average distances to primary schools (i.e. districts with few primary schools) as our experimental or treatment group, indicated by a plus symbol in Figure 2(a). There are 72 land parcels in the treatment group and the average distance to their nearest primary schools is about 3.6 kilometres, which is three times larger than that of the entire sample (about one kilometre). In this simulation scenario, we decrease the distance to primary schools from land parcels in the treatment group to one kilometre. The resulting land price effects are depicted in Figure 2(b). As expected, substantial increases of land prices are identified at locations in the treatment group (from 1.5% to 28.9%). However, locations that are close to the treatment land parcels also experience moderate price increases (from 0.1% to 1.5%). Small effects are found at locations that are far from the treatment locations. Overall, this simulation task offers a more complete understanding of land market gains from the (increasing) provision of educational facilities in areas experiencing least school accessibility.

**6. Conclusion**

Public expenditure on the maintenance and provision of school facilities in Chinese cities involves massive government investments. In order to improve the resource allocation efficiency, it is important to understand the spatial variations in amenity values of proximity to schools. This paper explores the capitalisation effect of proximity to schools in Beijing by simultaneously controlling for vertical and horizon spatial dependence effects.

Our results suggest that the government-funded schools especially primary schools exert significant influences to the residential land markets in Beijing, but the value of proximity to primary schools varies across space. We also find that the quality of nearest primary and middle schools does not appear to be significantly valued in residential land markets. Our study reinforces the existing literature in one important way. That is, there are several reasons to believe that estimates of school values can be improved by a careful consideration of spatial dependence effects. First, there is a plausible priori argument for the persistence of vertical and horizon dependence effects in spatial datasets. The spatial multilevel modelling strategy is potentially useful for getting more reliable estimates of school values. Second, spatial hedonic estimates are relative robust to changes of empirical specifications with the spatial weights matrix for land parcels.

In terms of policy implications, our finding that benefits of proximity to schools are not evenly distributed will make it more difficult for justifying whether the aggregate benefits from adjacent to schools can improve the quality of life for local residents. But the finding that local land values can be affected by proximity to schools and other amenities is an indication that local public goods were increasingly valued in Chinese cities.

**Acknowledgement**

We thank the editor and reviewers for useful comments. We thank Wenzhong Zhang for data support. The authors thank the National Natural Science Foundation of China (Project No. 41230632 and Project No. 71473105) for financial support.

**References**

Anselin, L. 1988. Spatial econometrics: Methods and models. Dordrecht: Kluwer Academic Publisher.

Anselin, L., Lozano-Gracis, N., Deichmann, U., Lall, S., 2010. Valuing access to water-a spatial hedonic approach, with an application to Bangalore, India. *Spatial Economic Analysis* 5, 161-179.

Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* 115, 588–638.

Bivand, R., Piras, G., 2015. Comparing implementation of estimation methods for spatial econometrics. *Journal of Statistical Software* 63, 1-36.

Black, S. E., 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114, 577-599.

Bowes, D., Ihlanfeldt, K., 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics* 50, 1-25.

Brasington, D. M., Hite, D., 2005. Demand for environmental quality: a spatial hedonic analysis. *Regional Science and Urban Economics* 35, 57-82.

Brasington, D. M., Haurin, D. R., 2006. Educational outcomes and house values: a test of the value added approach. *Journal of Regional Science* 46, 245-268.

Brooks, S. P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434-55.

Cellini, S. R., Ferreira, F., Rothstein, J., 2010. The value of school facility investment: evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics* 125, 215-261.

Cheshire, P. C., 2007. Introduction of Price Signals into Land Use Planning: Are they applicable in China?, in Song, Y, and C. Ding (eds) *Urbanization in China: Critical Issues in an era of rapid growth*, Cambridge MA: Lincoln Institute of Land Policy.

Cheshire, P. C., Sheppard, S., 2004. Capitalising the value of free schools: The impact of supply characteristics and uncertainty. *The Economic Journal* 114, 397-424.

Cliff, A., Ord, J. K., 1981. Spatial Processes: Models and Applications. London: Pion.

Dong, G., Harris, R., 2015. Spatial autoregressive models for geographically hierarchical data structures. *Geographical Analysis* 47: 173-191.

Dong, G., Harris. R., Jones, K., Yu, J., 2015. Multilevel modelling with spatial interaction effects with application to an emerging land market in Beijing, China. *PLoS ONE* 10(6), doi: 10.1371/journal.pone.0130761.

Gibbons, S., Machin, S., 2005,. Valuing rail access using transport innovations. *Journal of Urban Economics* 57, 148-169.

Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford Review of Economic Policy* 24, 99-119.

Gibbons, S., Machin, S., Silva, O., 2013. Valuing school quality using boundary discontinuities. *Journal of Urban Economics* 75, 15-28.

Golgher, A. B., Voss, P. R., 2015. How to interpret the coefficients of spatial models: spillovers, direct and indirect effects. *Spatial Demography*, doi: 10.1007/s40980-015-0016-y.

Harris, R., Dong, G. P., Zhang, W. Z., 2013. Using contextualised geographically weighted regression to model the spatial heterogeneity of land prices in Beijing, China. *Transaction in GIS* 17, 901-919.

Jones, K., 1991. Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers* 16, 148-159.

Kirby, D. K., LeSage, J. P., 2009. Changes in commuting to work times over the 1990 to 2000 period. *Regional Science and Urban Economics* 39, 460-471.

Lazrak, F., Nijkamp, P., Rietveld., Rouwendal, J., 2014. The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. *Journa of Geographical System* 16, 89-114.

Leonard, T., Zhang, L., Hoehner, C., 2015. Variations in park facility valuations across neighbourhoods. *Applied Spatial Analysis and Policy* 8, 45-67.

LeSage, J. P., Pace, R. K., 2009. Introduction to Spatial Econometrics. CRC Press/Taylor & Francis Group.

LeSage, J. P., Dominguez, M., 2012. The importance of modeling spatial spillovers in public choice analysis. *Public Choice* 150, 525-545.

Liang, Q.K., Kong, L.Y., Deng, W.B., 2007. Impact of URT on real estate value: the case of Beijing Metro Line 13, *China Civil Engineering Journal* 40, 98-103.

National Statistics Bureau of China (NSBC) 2004–2010 Beijing Statistic Yearbooks, 2004–2010. Beijing, China Statistics Press.

Nee, V., 1992. Organizational dynamics of market transition: forms, property rights, and mixed economy in China. *Administrative Science Quarterly* 37, 1- 27.

Orford, S., 2000. Modelling spatial structures in local housing market dynamics: a multi-level perspective. *Urban Studies* 13, 1643-1671.

Wen, H., Zhang, Y., Zhang, L., 2014., Do educational facilities affect housing price? An empirical study in Hangzhou, China. *Habitat International* 42, 155-163.

Wu,W., Dong, G.P., 2014. Valuing green amenities in a spatial context. *Journal of Regional Science* 54, 569-585.

Wu,W., Dong, G.P., Wang, B., 2015. Does planning matter? Effects on land markets. *Journal of Real Estate Finance and Economics* 50, 242–269.

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

Zheng, S.Q., Kahn, M.E., 2013. Does government investment in local public goods spur gentrification? Evidence from Beijing. *Real Estate Economics* 41, 1-28.

Zheng, S.Q., Hu, W.Y., Wang, R., 2015. How much is a good school worth in Beijing? Identifying price premium with paired resale and rental data. *Journal of Real Estate Finance and Economics,* doi:10.1007/s11146-015-9513-4

**Tables**

**Table 1**.Variable name, definition, and descriptive statistics

|  |  |  |
| --- | --- | --- |
| Variables | Definition | Mean(std.dev)/  proportion |
| Land parcel prices | Log of residential land parcel prices per  square meter (RMB/sq.meter) | 7.429 (1.026) |
| *School accessibility variables* | | |
| Primary school | Log of distance to the nearest primary school | 6.605(0.811) |
| Primary school quality | Binary variable: 1 if the nearest primary school is a KPS | 4.2% |
| Middle school | Log of distance to the nearest middle school | 7.691(0.949) |
| Middle school quality | Binary variable: 1 if the nearest high school is a EMS | 37.5% |
| University | Log of distance to the nearest prestige university | 7.930(0.682) |
| *Land structural and locational variables* | | |
| Parcel size | Log of the size of a land parcel | 9.305 (1.522) |
| CBD | Log of distance to the Central Business District (CBD) | 8.946 (0.668) |
| Railway station | Log of distance to the nearest railway station | 7.151(0.902) |
| Green park | Log of distance to the nearest green park | 7.763 (0.691) |
| River | Log of nearest distance to a river | 7.488(0.924) |
| Express road | Log of nearest distance to a express road | 6.177(1.028) |
| *Neighborhood variables* |  |  |
| Crime | Number of reported serious crimes per 1000 people in each district | 5.321(6.118) |
| Job density | Job density in each district (1000 people/km2) | 12.53 (13.07) |
| Age of buildings | Proportion of buildings built before 1949 in each district (centred) | 0 (0.109) |
| Education attainment | Median educational attainment in each zone:1=junior or lower; | 1.731 (0.553) |
| 2=high school;3=university;4=post graduate |
| Public house renting | Proportion of people who rent public housing | 0.344(0.195) |

Note: RMB = Chinese Yuan RenMinBi

**Table 2**. Land price estimation results from ordinary linear regression models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | | Model 2 | |
|  | Estimates | Std. Error | Estimates | Std. Error |
| Intercept | 11.91**+** | 0.414 | 13.87**+** | 0.804 |
| *School variables* | | | | |
| Primary school | -0.164**+** | 0.041 | -0.099**+** | 0.041 |
| Primary school quality | -0.166 | 0.149 | -0.088 | 0.139 |
| Middle school | -0.312**+** | 0.036 | -0.056 | 0.042 |
| Middle school quality | -0.132 | 0.079 | -0.175**+** | 0.079 |
| University | -0.119**+** | 0.042 | -0.101**+** | 0.044 |
| Primary school × Age of buildings |  |  | 1.258**+** | 0.493 |
| *Land structural and locational variables* | | | | |
| Parcel size |  |  | -0.017 | 0.019 |
| CBD |  |  | -0.251**+** | 0.061 |
| Railway station |  |  | -0.2**+** | 0.037 |
| Green park |  |  | -0.141**+** | 0.051 |
| River |  |  | 0.079**+** | 0.031 |
| Express road |  |  | -0.054**+** | 0.026 |
| *Neighbourhood variables* | | | | |
| Job density |  |  | 0.615**+** | 0.157 |
| Public house renting |  |  | -0.043 | 0.138 |
| Age of buildings |  |  | -8.72**+** | 2.865 |
| Education attainment |  |  | 0.021 | 0.061 |
| Crime |  |  | -0.001 | 0.005 |
| Year dummies | Yes |  | Yes |  |
| *σe*2 | 0.815 |  | 0.659 |  |
| Sample size | 1077 |  | 1077 |  |
| Adjusted R2 | 0.226 |  | 0.374 |  |
| Log-likelihood | -1412 |  | -1292 |  |
| Moran's I | 0.197**+** |  | 0.075**+** |  |

Note. Table reports coefficients and standard errors of each variable from ordinary linear regression models. “**+**” represents significance level at least at 95%. For testing the presence of spatial dependence, the spatial weights matrix was constructed using a Gaussian decay function with a distance threshold of two kilometres. Other weighting schemes such as 30 nearest neighbours, or Gaussian kernels with distance thresholds of 2.5 kilometres and three kilometres, were also tested and the results were quite similar with those reported here.

**Table 3**. Land price estimation results from SAR and HSAR models.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 3 (SAR) | | | Model 4 (HSAR) | | |
|  | Median | 2.5% | 97.5% | Median | 2.5% | 97.5% |
| Intercept | 7.998 | 6.185 | 9.734 | 9.87\* | 7.780 | 12.056 |
| Primary school | -0.061 | -0.126 | -0.002 | -0.077 | -0.153 | -0.004 |
| Primary school quality | -0.071 | -0.297 | 0.137 | -0.03 | -0.271 | 0.218 |
| Middle school | -0.027 | -0.092 | 0.042 | -0.028 | -0.109 | 0.059 |
| Middle school quality | -0.105 | -0.231 | 0.021 | -0.073 | -0.219 | 0.078 |
| University | -0.031 | -0.1 | 0.042 | -0.014 | -0.102 | 0.072 |
| Primary school × Age of buildings | 1.075 | 0.313 | 1.827 | 1.244 | 0.482 | 2.024 |
| Parcel size | -0.02 | -0.05 | 0.012 | -0.023 | -0.054 | 0.009 |
| CBD | -0.136 | -0.237 | -0.032 | -0.212 | -0.379 | -0.063 |
| Railway station | -0.140 | -0.2 | -0.082 | -0.16 | -0.229 | -0.092 |
| Green park | -0.077 | -0.162 | 0.006 | -0.088 | -0.183 | 0.002 |
| River | 0.058 | 0.006 | 0.106 | 0.083 | 0.022 | 0.145 |
| Express road | -0.042 | -0.084 | -0.004 | -0.063 | -0.106 | -0.019 |
| Job density | 0.455 | 0.204 | 0.719 | 0.430 | 0.093 | 0.758 |
| Public house renting | -0.020 | -0.242 | 0.199 | 0.071 | -0.228 | 0.371 |
| Age of buildings | -7.467 | -11.756 | -3.004 | -8.485 | -13.02 | -3.947 |
| Education attainment | 0.036 | -0.06 | 0.133 | 0.034 | -0.107 | 0.161 |
| Crime | -0.001 | -0.008 | 0.006 | 0.001 | -0.012 | 0.013 |
| *ρ* | 0.399 | 0.321 | 0.483 | 0.268 | 0.157 | 0.37 |
| *λ* |  |  |  | 0.571 | 0.109 | 0.842 |
| *σe*2 | 0.618 | 0.577 | 0.664 | 0.582 | 0.54 | 0.627 |
| *σu*2 |  |  |  | 0.044 | 0.02 | 0.081 |
| Year dummies | Yes |  |  | Yes |  |  |
| Sample size | 1077 |  |  | 1077 |  |  |
| Pseudo-R2 | 0.423 |  |  | 0.477 |  |  |
| Moran's I | 0.008 |  |  | 0.007 |  |  |

Note. Both SAR and HSAR models were implemented using the Bayesian MCMC method and the same prior distributions and starting values for model parameters were used. The land parcel level spatial weights matrices used in the two models are same, a Gaussian kernel with a distance threshold of two kilometres. Adjusted R2 for two models were simply calculated as the squared correlation coefficients between predicted outcomes from two models and the observed outcomes.

**Table 4**. Impact estimates of variables from the SAR model.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Direct  impact | Indirect  impact | Total  impact |
| Primary school | -0.061  [-0.128, -0.002] | -0.038  [-0.09, -0.001] | -0.101  [-0.215, -0.002] |
| Primary school quality | -0.071  [-0.3, 0.139] | -0.046  [-0.202, 0.095] | -0.119  [-0.496, 0.23] |
| Middle school | -0.027  [-0.093, 0.042] | -0.017  [-0.064, 0.027] | -0.043  [-0.154, 0.07] |
| Middle school quality | -0.107  [-0.233, 0.021] | -0.067  [-0.166, 0.013] | -0.175  [-0.39, 0.034] |
| University | -0.031  [-0.102, 0.042] | -0.019  [-0.07, 0.027] | -0.05  [-0.172, 0.071] |
| Primary school × Age of buildings | 1.09  [0.318, 1.854 ] | 0.692  [0.195, 1.312] | 1.784  [0.518, 3.083] |
| CBD | -0.138  [-0.24, -0.032] | -0.085  [-0.168, -0.02] | -0.224  [-0.4, 0.053] |
| Railway station | -0.142  [-0.203, -0.08] | -0.09  [-0.15, -0.048] | -0.235  [-0.347, 0.136] |
| River | 0.059  [0.006, 0.107] | 0.037  [0.004, 0.076] | 0.097  [0.01, 0.181] |
| Express road | -0.043  [-0.08, -0.004] | -0.027  [-0.06, -0.003] | -0.071  [-0.144, 0.007] |
| Job density | 0.46  [0.207, 0.73] | 0.291  [0.121, 0.527] | 0.758  [0.337, 1.22] |
| Age of buildings | -7.555  [-11.92, -3.03] | -4.794  [-8.57, -1.842] | -12.46  [-20.12, -4.97] |

Note. Numbers in square brackets are 95% credible intervals of the direct, indirect, and the total impacts for each independent variable. Only school variables and other statistically significant predictors (see Table 3) in the land price equation are reported.

**Table 5**. Impact estimates of variables from the HSAR model.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Direct  impact | Indirect  impact | Total  impact |
| Primary school | -0.077  [-0.154, -0.004] | -0.026  [-0.064, -0.001] | -0.105  [-0.209, -0.006] |
| Primary school quality | -0.03  [-0.273, 0.22] | -0.009  [-0.109, 0.083] | -0.042  [-0.38, 0.304] |
| Middle school | -0.028  [-0.109, 0.059] | -0.009  [-0.042, 0.022] | -0.038  [-0.149, 0.079] |
| Middle school quality | -0.074  [-0.22, 0.078] | -0.023  [-0.09, 0.027] | -0.098  [-0.307, 0.102] |
| University | -0.014  [-0.102, 0.073] | -0.005  [-0.04, 0.026] | -0.019  [-0.142, 0.101] |
| Primary school × Age of buildings | 1.251  [0.486, 2.037] | 0.427  [0.133, 0.881] | 1.695  [0.649, 2.822] |
| CBD | -0.213  [-0.383, -0.064] | -0.074  [-0.155, -0.019] | -0.292  [-0.519, -0.084] |
| Railway station | -0.161  [-0.231, -0.093] | -0.057  [-0.103, -0.024] | -0.218  [-0.32, -0.125] |
| River | 0.083  [0.023, 0.146] | 0.029  [0.007, 0.061] | 0.113  [0.03, 0.197] |
| Express road | -0.064  [-0.107, -0.019] | -0.022  [-0.046, -0.006] | -0.086  [-0.149, -0.027] |
| Job density | 0.434  [0.093, 0.761] | 0.148  [0.026, 0.322] | 0.589  [0.118, 1.057] |
| Age of buildings | -8.548  [-13.11, -3.962] | -2.942  [-5.786, -1.089] | -11.57  [-18.05, -5.27] |

Note. Numbers in square brackets are 95% credible intervals of the direct, indirect, and the total impacts for each independent variable. Only school variables and other statistically significant predictors (see Table 3) in the land price equation are reported.

**Table 6**. Robustness checks for estimated school accessibility effect from HSAR models with different spatial weights matrices.

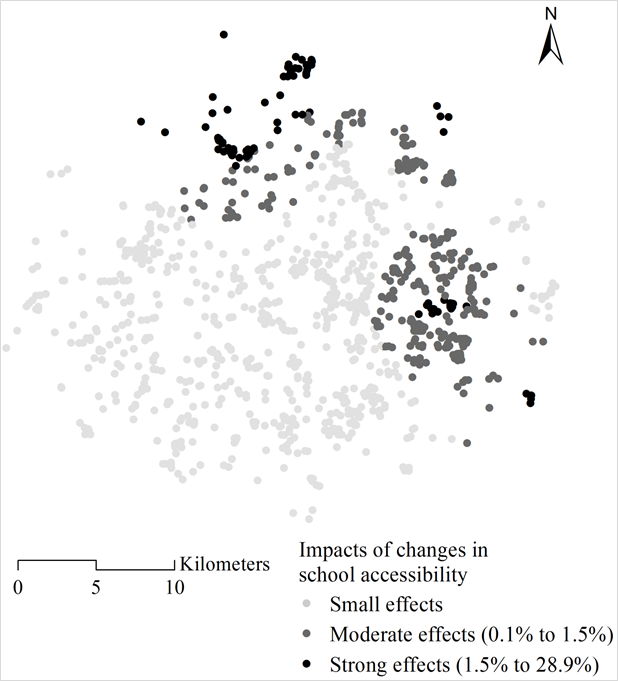
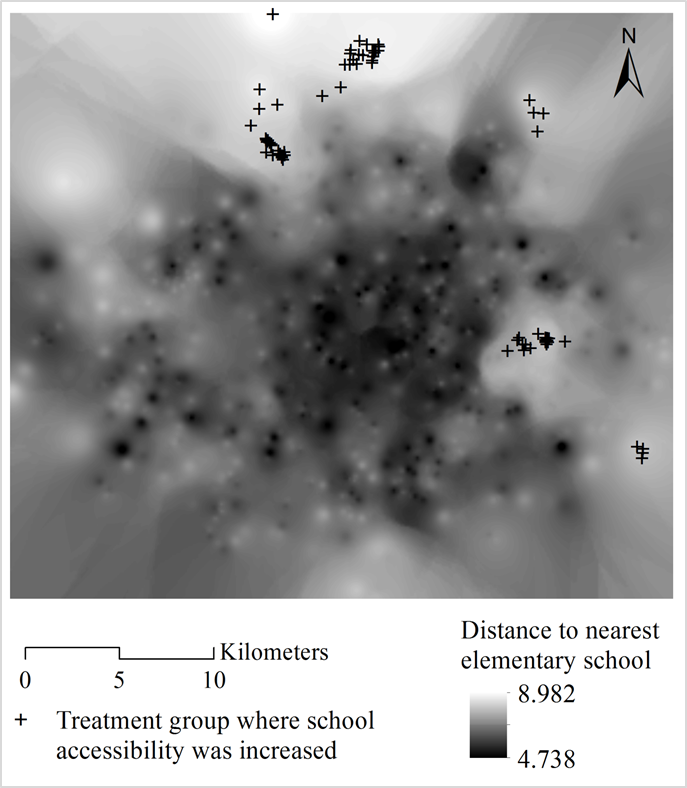
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | HSAR (2.5 kilometres) | | | HSAR (3 kilometres) | | | HSAR (nearest 30 neighbours) | | |
|  | Median | 2.5% | 97.5% | Median | 2.5% | 97.5% | Median | 2.5% | 97.5% |
| Primary school | -0.078 | -0.153 | -0.001 | -0.081 | -0.156 | -0.007 | -0.085 | -0.156 | -0.007 |
| Primary school quality | -0.019 | -0.265 | 0.215 | -0.01 | -0.266 | 0.233 | 0.003 | -0.23 | 0.251 |
| Middle school | -0.032 | -0.119 | 0.053 | -0.035 | -0.118 | 0.054 | -0.039 | -0.122 | 0.047 |
| Middle school quality | -0.074 | -0.225 | 0.081 | -0.077 | -0.238 | 0.077 | -0.074 | -0.228 | 0.082 |
| University | -0.023 | -0.114 | 0.067 | -0.025 | -0.116 | 0.064 | -0.036 | -0.122 | 0.054 |
| Primary school × Age of buildings | 1.272 | 0.488 | 2.079 | 1.299 | 0.49 | 2.083 | 1.301 | 0.494 | 2.047 |
| *ρ* | 0.221 | 0.108 | 0.339 | 0.188 | 0.049 | 0.32 | 0.25 | 0.066 | 0.42 |
| *λ* | 0.594 | 0.129 | 0.867 | 0.604 | 0.157 | 0.857 | 0.627 | 0.16 | 0.981 |
| *σe*2 | 0.583 | 0.54 | 0.63 | 0.582 | 0.539 | 0.629 | 0.584 | 0.544 | 0.63 |
| *σu*2 | 0.053 | 0.023 | 0.091 | 0.059 | 0.029 | 0.103 | 0.052 | 0.022 | 0.095 |

Note. In each HSAR model, land structural and locational characteristics, and neighbourhood variables are also included. Because estimation results for these variables are not quite similar with those in Table 3, they are not reported here.

**Figures**



**Fig. 1.**The price per area of residential land parcels leased between 2003 and 2009 in urban areas of Beijing, China.



(a)

(b)

**Fig. 2.** Spatial distribution of distances to primary schools and the impact of changes in school proximity on land value

1. The CBD in Beijing is located in the east direction of the *TianAnMen* Square, called “*Guomao*,” with a cluster of high-rise office buildings and many international companies’ headquarters. [↑](#footnote-ref-1)
2. We also test whether there is temporal autocorrelation in land prices using the regression coefficients of year dummy variables from a linear regression model. The result from a first-order autoregressive model shows no evidence of statistically significant temporal autocorrelation in the temporal trend of land prices during the study period. [↑](#footnote-ref-2)
3. The number of primary and middle schools are acquired from the website of Beijing Municipal Commission of Education. Our definition of middle school encompasses both middle schools and high schools. Prestige universities were defined as universities that are listed in either the “211 project” or the “985 project” organised by the central education administration of China. These universities receive much more academic and financial support from the central and municipal governments and usually recruit talented undergraduate and postgraduate students in China. [↑](#footnote-ref-3)
4. We have run a series of regression models with interaction terms between the distance to school variables and neighbourhood characteristics. The results show that the interaction term between the proximity to primary schools (Primary school) and the proportion of buildings built before 1949 in each district (Age of buildings) is statistically significant in all model specifications. [↑](#footnote-ref-4)
5. In several districts located in inner urban areas with large concentrations of old buildings, the externality of primary schools switches from positive to negative, a situation we also find in the following SAR and HSAR models. [↑](#footnote-ref-5)
6. Recall that there is an interaction term, Primary school × Age of buildings, in our model. Because the variable Age of buildings is centred before entering our model, the total impact of Primary school is measured at districts with mean levels of proportion of old buildings. [↑](#footnote-ref-6)
7. The spatial Durbin model, a SAR model with spatially lagged independent variables included, is usually more robust to model misspecification than the SAR model (LeSage and Pace, 2009; LeSage and Dominguez, 2012). In the present study, the model misspecification could be the omitting of the district-level random effect and the spatial dependence at this scale. It would seem to be useful to further compare the HSAR model against a spatial Durbin model. However, we choose not to focus on the spatial Durbin model for model comparisons due to the issue of severe multicollinearity. For most of the locational variables including the school-related variables of primary interest are strongly correlated with their spatial lags. For example, the correlation coefficients between the variables of Primary school, Middle school and University and their spatial lags are about 0.87, 0.89 and 0.83 respectively. A further complication is that the correlations between different spatially lagged independent variables are also very strong even though only weak correlations exist between the original variables. These strong correlations among independent variables will cause problems to model parameter estimation and statistical inference. [↑](#footnote-ref-7)