The London School of Economics and Political Science

Empirical Essays on Real Estate, Local Public Goods and Happiness: Evidence from Beijing

Wenjie Wu

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy, London, October 2012

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of around 53193 words.

Abstract

This thesis explores the real estate and happiness consequences of public investment in local public goods improvements by using unique micro-geographical data from Beijing; it focuses on the spatial variations in park amenity values, and on the impact of transport improvements on land prices and homeowners' happiness. Despite intense public interest, little is known about these effects. This thesis aims to fill these gaps.

First, I explore the impact and sources of variations of park proximities as capitalized into the residential land prices. This analysis, using geographically-coded data from Beijing, provides new insights on the ways in which land markets capitalize the values of proximity to parks and suggests that this is highly dependent on the parcel's location and local contextual characteristics.

Next, I examine the real estate consequence of public investment in transport improvements using a rich data set of vacant land parcels in Beijing. I use a multiple intervention difference-in-difference method to document opening and planning effects of new rail stations on prices for different land uses in affected areas versus unaffected areas. Residential and commercial land parcels receiving increased station proximity experience appreciable price premiums, but the relative importance of such benefits varies greatly over space and local demographics.

Finally, I investigate the impact of transport improvements on happiness that altered the residence-station distance for some homeowners, but left others unaffected. My estimation strategy takes advantage of micro happiness surveys conducted before-and-after the building of new rail stations in 2008 Beijing. I deal with the potential concern about the endogeneity in sorting effects by focusing on "stayers" and using non-market housings with pre-determined locations. I find the significantly heterogeneity in the effects from better rail access on homeowners' happiness with respect to different dimensions of residential environment. The welfare analysis results suggest strong social-spatial differentiations.

In combination, the three papers of this thesis make important contributions to a growing literature on public infrastructure, land market and happiness.

Acknowledgment

Many thanks to Steve Gibbons and Paul Cheshire for outstanding supervision, and to my family for love and support.

Thanks also to Olmo Silva, Henry Overman, Brett Day, Francesca Cornaglia, Simona Iammarino, Mark Partridge, Marlon Boarnet, Guanpeng Dong, Wenzhong Zhang, Jie Fan, Fengjun Jin, Yaping Wang, Youqin Huang for helpful comments. I gratefully acknowledge the scholarship support from the London School of Economics and Political Science.

Previous versions of the papers were presented at the following conferences: Annual Conference of European Real Estate Society (2012); International Workshop on Regional, Urban, and Spatial Economics in China (2012); Annual Conference of European Regional Science Association International (2011); International Conference on China's Urban Transition and City Planning (2011); Annual Conference of Regional Science Association International-British and Irish Section (2011). In addition preliminary versions of these papers were discussed at seminars at the LSE, University of Glasgow, Peking University, and the Chinese Academy of Science. Thanks to participants and discussants for kindly suggestions.

Content

List of Tables

List of Figures

I. Introduction

1 Overview

Decades of fast economic growth and urbanisation have significantly changed the urban infrastructure in China. Like other large cities in the BRICS countries[1], Beijing is investing heavily in local public goods, a largely place-based investment process that is of great importance for homeowners, land developers and policymakers. However, not much is known about the real estate and happiness consequences of local public goods improvements. Research on this topic has long been limited by the lack of systemic micro-geographical data and by the lack of the convincing research designs. This thesis aims to fill these important gaps by looking at particular cases of parks and rail transportation in Beijing. More specifically, I focus on the spatial variations in the amenity values of park proximities, and on the impacts of rail access changes on land prices and homeowners' happiness. This thesis comprises three papers, formed as the subsequent chapters.

This introductory chapter proceeds as follows. Section 2 provides a general discussion of the research background. Section 3 outlines the research questions. Sections 4-6 highlight each paper's context and contribution, data and methods, and results. Section 7 provides brief summaries.

2 Background and motivation

There is a tremendous and sprawling literature on valuing local amenities, covering a wide range of factors---from air and water quality, to the accessibility of

[1] BRICS is the title of an association of leading emerging economies (Brazil, Russia, India, China and South Africa). See http://en.wikipedia.org/wiki/BRICS for details.

retail stores, schools, rail stations, parks, and even churches. This thesis is built on two branches of this economic literature: First, urban and real estate economists, at least following seminal works by Oates (1969) and Rosen (1974), have developed a large literature on the ways in which the value of local amenities could be capitalised in urban land markets. This branch of research is often conducted by using a hedonic valuation method based on land/housing price data. A survey of recent examples of the hedonic valuation approach is given by Cheshire and Sheppard (1995), Gibbons and Machin (2008).

The second related body of literature is about happiness studies, which has been one of the promising developments in economics recently (Layard, 2006; Frey, 2008). It is devoted to examine the subjective wellbeing of economic activities. At its heart it argues that economics should be able to determine the question of how economic performances, such as inflation, unemployment, and local public goods accessibility affect human subjective wellbeing. However, some economists have been reluctant to carry out direct happiness tests, partly because of a lack of reliable and consistent survey data, partly because of suspicion about the validity of subjective assessments. This mistrust, however, is unnecessary as the reliability of subjective measures based on survey data has been confirmed by the economic literature in "happiness" (Krueger and Schakde, 2007; Oswald and Powdthavee, 2008; Oswald and Wu, 2010). However, while self-reported happiness assessments have been widely applied by labour and health economists (Clark and Oswald, 1996; Di Tella et al, 2001; Ferrer-i-Carbonell, 2005; Kahneman et al, 2006; Cornaglia et al., 2012), it is not easy to find direct test on how transport improvement program affects homeowners' happiness. So far, this is a mostly unexplored research field.

In order to narrow the scope of my research, I focus on the role of local amenities---in particular, parks and rail transits, within the context of the Beijing urbanized area. To be specific, I explore spatial variations in the amenity values by using parks as an example, and then take advantage of new rail transit constructions as another useful piece of evidence to examine the effects of transport improvement, identified by distance reductions to stations, on land prices and homeowners' happiness in specific residential aspects[2].There is good reason to focus on the land-price premium and happiness benefits attached to the local public goods improvements: rail transits and parks are the important public investment areas and have been a key policy focus for emerging countries like China which has experienced fast urbanisations over the past decade.

2.1 Valuation of local amenities: parks and rail stations

This section focuses on local amenity valuation and its empirical estimation problems relating to my research (paper one and two).

TV news, media documents, and even pub conversations all lend credence to the claim that proximity to parks and railway stations affects local land and housing prices. However, there is a serious research question: to what extent are customers

[2] Throughout this study when I use the term of "residential aspects" or "different dimensions of residential environment" I mean the domains of residents' living conditions relative to the survey questions like residents' happiness about social environment, commuting and living convenience, safety and pollution.

willing to pay for access to parks and rail stations? The answers to these questions, using empirical studies from developed economies, have been well documented in the past; but as micro geographical data becomes available in developing economies this topic has once again attracted international attention.

In the economic literature, local amenity valuation is usually measured either by stated preferences using the contingent valuation method (see Bateman et al, 2006 and Day, 2007) or by revealed preferences using the hedonic valuation method. In this study, I do not attempt to review all these non-market valuation methods on the proximity effects of parks and rail stations on land prices, but rather highlight some of the recent excellent hedonic applications.

The first building block of my research is related to the hedonic valuation of park amenities[3]. Recent studies find that housing prices increase with proximity to parks, but the value of the proximity effect varies by different park types. For example, Cheshire and Sheppard (1995) distinguish between public and private parks in two medium-size British cities. They find that only publicly accessible parks can increase housing value significantly. In addition to varying by park types, they also point out that the amenity values of parks are influenced by a home's location and neighbourhood characteristics (Cheshire and Sheppard, 1998; Barbosa et al., 2007).

[3] Note that other empirical studies use alternative methods to evaluate the effects of parks and related environmental amenities. For example, Schultz and King (2001) examine the effects of green space on average home values by using the aggregated census data; Day and Mourato (1998) and Breffle et al. (1998) use a survey method to investigate people’s willingness to pay for the water quality and open space, respectively. Based on combined physical, census and survey data, Bateman et al (2006) use different welfare measures to model the aggregated amenity benefits of national park wetland and water quality by considering the relationship between distance decay and willingness to pay. See McConnell and Walls (2005) for an extensive review of non-market valuation methods with respect to parks.

Indeed, there are good reasons to expect that the marginal effects of the amenity values of proximity to parks will exhibit spatial heterogeneity in the complex urban real estate markets due to local contextual factors and supply-demand imbalances. For example, as suggested by recent literature, parks tend to be favoured venues for criminal behaviours, so households in high-crime areas may be afraid to engage in outdoor activities in nearby green spaces. Thus, the amenity value of the proximity to parks is likely to be lower in high-crime places. In addition to the interaction effects with localized contextual factors, the supply of a typical local amenity is often distributed unevenly over the urban area due to planning and historical reasons. Demand by households for specific location attributes like access to parks is also known to vary with their socioeconomic characteristics. For instance, if richer households have a greater willingness to pay for parks then in the long run local governments serving such neighbourhoods may supply more park space. As a matter of fact, Freeman (1979) have long suggested that the heterogeneity is predictable, not only because land attributes are heterogeneous across locations, but also because land buyers are heterogeneous in their willingness to pay for certain characteristics and the related location-specific characteristics. This may lead to a spatial imbalance between supply and demand within a fixed geographic area, at least over a short-time period. In a competitive land market, the implicit price of the proximity effects of parks will vary from buyer to buyer, and each buyer, to maximize utility, will seek to balance the marginal implicit price of parks with the marginal willingness to pay. Greater competition for this characteristic at certain locations will result in higher marginal prices than those of other areas. Thus one would expect substantial spatial variations in the amenity values of proximity to parks within a mega-city like Beijing.

Irwin and Bockstael (2001a) summarized two specific estimation issues associated with the application of hedonic techniques to the valuation of parks. First, if parks are privately owned, or can be developed for residential use in the future, then the variables estimating the influence of parks on nearby residential land values are endogenous in the hedonic models. This problem does not occur in my research since all parks in Beijing are publicly accessible and preserved permanently by the city government. The second issue is related to unobserved factors. Although one can control for many localised factors, there would be still a long list of sources of heterogeneity that cannot be observed easily. Again, the decision about what location characteristics to include in model specifications remains largely in the eyes of researchers. Thus the cross-sectional approach is not an attractive way forward if researchers hope to get more reliable causal effects for policy decision-makings. As such my hedonic price regression needs to allow for nonlinearities in known covariates and control for the possibility of omitted variables. While some choose an instrumental variable approach, most studies use local fixed-effects to address this bias source. This is because the instrumental variable approach depends on choosing one or more specific reasons of variation in amenity supply that should be unrelated to land prices. If this assumption holds, then any correlation of land prices with this source of variation in supply is definitely due to variation in the supply of the amenity, and not to other unobserved spatial variables. However, it is very difficult to find proper "instruments" for variation in parks, or other local amenities, and thus, in practice, few studies apply this method alone (Irwin, 2002). Anderson and West (2006) provide a recent good example for controlling of the potential unobserved factors with the census block-group fixed-effects. Nevertheless, this fixed-effect approach has several limitations. First, if block-groups overlap/nest with perceived neighbourhood boundaries then problems with unobserved neighbourhood-level characteristics may still exist. Second, this fixed-effect approach may fail to control for omitted variables that only affect a single land parcel or house. Thirdly, such fixed effects could not effectively control for potential unobserved covariates that influence the amenity values of park proximities. To address this issue, Gibbons and Machin (2003) drive a spatial smoothing technique to control for local effects prior to the estimation. This technique can help avoid the problem of choosing arbitrary neighbourhood boundaries but demand assumptions relating to the choice of smoothing function parameters. Day et al (2007) go further and present the model specification non-parametrically by capturing spatial autocorrelations and considering both spatial coordinates and property characteristics.

Recent progress in spatial econometrics has also focused on developing an alternative approach that would be better able to account for the variations in the estimated values of a local amenity over space[4]. A well-cited method is the locally weighted regression (LWR) approach (Cleveland and Devlin, 1988). The primary

[4] Accompanied with the development of the GIS techniques, empirical hedonic studies have started to consider spatial effects such as spatial dependence and spatial heterogeneity into the estimation process, leading to the so-called spatial hedonic models (McMillen and Redfearn, 2010). While the urban land market tends to be characterized by both, I focus typically on the spatial heterogeneity effect since it has received less attention in the literature.

advantage of a LWR design is that by estimating a vector of implicit prices at each observation, it is able to control for heterogeneity in each house's location. Recently, this LWR technique has been applied intensively in the real estate market to test for local heterogeneity (Leung et al, 2000; Cho et al., 2006; Bitter et al., 2007; McMillen and Redfearn, 2010). Cho et al. (2006) present a first attempt that uses LWR to measure the spatial heterogeneity effects of proximity to parks. They found that the average marginal implicit price of proximity to parks estimated by the OLS model was $172, whereas the LWR model indicated that the marginal implicit prices varied from park to park, ranging from –$662 to $840.

In my first paper, I use Cho's pilot study as a benchmark for departure. But I acknowledge that the direct application of the LWR method could be mostly futile for several reasons. Firstly, this LWR approach, to some extent, can be viewed as a continuum between the OLS model and the completely non-parametric; it can help to maximise the model fit, but this does not mean it is a more useful model than the traditional OLS approach in terms of causal interpretation. One can easily improve on the model, in terms of fit, by making it completely non-parametric and regressing price on a set of house/land specific dummy variables. Secondly, most of the previous LWR-based hedonic studies have not considered the interaction effects between a park and its location-specific characteristics. As such their model estimates are likely to conceal substantial variations among individual parks. Practically, it is quite possible that the benefits derived from proximity to parks would increase when a park is close to subway stations, and would decrease when a park is located in high crime rate areas. Thirdly, existing LWR applications usually present only one model specification with no robustness checks (Redfearn, 2009). Little is known about the stability of the LWR results, that is, how sensitive the LWR parameters of proximity to parks are to the changes in the set of control variables.

To this end, my first paper contributes to the literature in three ways: first, it extends the locally weighted regression (LWR) approach to include the complementary effects between parks and key observable amenities and demographic characteristics; second, it provides a powerful estimation strategy to assess the robustness of the parameters of proximity to parks---estimated by a wide range of LWR model specifications, and therefore, sheds more light on the potential sources of spatial variations in the amenity values; and finally, it visualises spatial variation patterns of the estimated values of local parks. To be clear, I do not attempt to compare the advantages between the LWR and other spatial econometric methods. Instead, this study is mostly looking at how park proximity interacts with other local contextual factors rather than arguing for a specific "optimal" method. The data used in my analysis is a rich geographically-coded dataset that links the location characteristics of land parcels, parks and socio-demographics and other local amenities.

The second building block of my research relates to the literature on urban transport. Governments in a range of international contexts have continued to invest in expanding transport infrastructure. These new rail transit lines have fundamentally increased the transport accessibility over time. However, the traditional cross-sectional hedonic regressions, which ignore the changes in the supply of local amenities, may conceal significant variation in changes in transport access and real property prices (Bowes and Ihlanfeldt, 2001; Debrezion et al, 2011). Unfortunately, few studies have explicitly investigated the consequences of transport improvements on the real estate market, especially in the context of developing countries. Earlier studies, such as Dewees (1976) and Bajic (1983), examined the impact of Toronto's new Spadina subway lines on housing prices. For example, Bajic (1983) found that improved rail access, induced by transport investments increased housing prices by USD 2,237 on average.

Recently, well-cited transport improvement examples include the opening of the Chicago's Midway line and London's Jubilee lines. Using a difference-in-difference approach, Gibbons and Machin (2005) present a pioneering work on measuring how property prices respond to the opening of new stations. Their study uses the new rail transit expansions in London in the late 1990s as a policy background. Their approach captures the effects of a transport improvement which shortened the residence-station distance for the "treated" households on residential property prices[5]. They reported substantial positive impacts, ranging from around 1-4 percent increase, on prices for every 1 km reduction in station-residence distance. Ahlfeldt (2011)'s follow-up work confirmed Gibbons and Machin's (2005) findings using the same style of the difference-in-difference estimation strategy. By comparing before and after outcomes

[5] Recall that this type of analysis usually does not consider the impacts of overall economic climate and financial changes, and thus is likely to be place and time specific. For example, if there is mortgage rationing then would be house owners much less able to response to such changes.

in the building of Chicago's Midway line, McMillen and McDonald (2004) documented the significant anticipation effect associated with transport improvement. They find that there was an appreciable premium on residential land values within half a mile of station locations, even before the new transit line opened.

Despite heavy investment in transport infrastructure in China, there is no direct test on the impact of transport improvement on land prices. This is not surprising, given that empirical analysis is likely to depend heavily on the availability of systemic micro data, which is very hard to obtain in China. After years of data collection and geo-coding, my second paper presents the first attempt to look at the consequences of transport improvements on the prices of vacant residential and commercial land parcels close to stations on new railway lines in Beijing. I improve on previous methods by providing a multiple-intervention difference-in-difference framework that not just exploits the parcel-station distance changes due to the opening of new stations but also highlights the importance of price changes in planned station areas.

2.2 Investigation of residential happiness

The third building block of my research is the growing interest in notion of "happiness", in particular as it relates to the impact of transport improvement program on homeowners' happiness with respect to different dimensions of residential environment.

Happiness is arguably one of the fundamental goals in life. It was originally the subject of socio-psychological and health research, but has recently drawn the attention of economists. The main interest of happiness studies lies in explaining the contributory factors of human subjective wellbeing (loosely named as happiness) [6]. This is highly dimensional, ranging from smoking and obesity (Oswald and Powdthavee, 2007; Katsaiti, 2012), to income and unemployment (Taubman, 1976; Clark and Oswald 1994, 1996; Winkelmann and Winkelmann, 1998; Di Tella el al., 2001; Ferrer-i-Carbonell, 2005), environmental quality (Luechinger, 2009; Frey et al, 2010), crime and terrorism (Frey et al, 2009; Cornaglia and Leigh, 2011), and recently local public goods like schools (Mohan and Twigg, 2007; Permentier et al., 2011; Gibbons and Silva, 2011). The literature relevant to my research includes studies that have investigated the relationship between local public goods accessibility and households' residential happiness.

Unlike objective living conditions that can be easily measured by census and housing price data, residential happiness is more about subjective wellbeing of residents' living experiences and is usually measured through questionnaires (Gruber and Shelton, 1987; Cook, 1988; Lu, 1999; Parkes et al., 2002; Chapman and Lombard, 2006; Hur and Morrow-Jones, 2008). However, despite the wide scope of investigation by researchers from other fields, much is unknown in the economic literature about the relationship between local public goods accessibility and people's residential happiness. This difference is partly because of the lack of survey data, and partly because of a lack of trust for subjective assessments or stated preference measures. For a long time, urban economists in the field of local amenity valuation have developed a strong habit of relying upon objective measures linked with

[6] McGillivray and Clarke (2006) recently summarized the concepts like "satisfaction, and happiness can be used interchangeably with subjective wellbeing without explicit discussion as to their differences."

property price outcomes. However, the simplified assumption that property-price changes provide a sufficient statistic for valuing local amenities is clearly open to scrutiny. In fact, there might be numerous subjective aspects of social and emotional developments caused by local public goods improvements that cannot be observed by price signals, but are observed by households' living experiences (Galster and Hesser, 1981; Baba and Austin, 1989; Lu, 1999; Frey and Stutzer, 2001; Krueger and Schakde, 2007; Oswald and Wu, 2010; Permentier et al., 2011). With the help of rigorously designed and representative surveys, economists can get good indications of households' assessments of their happiness. This can be captured with multi-score survey questions in a straightforward way. In fact, scholars in other fields of economics have made much wider application of subjective/perceived assessments of wellbeing (Baker et al., 2004; Cornaglia et al., 2012).

Encouragingly, a number of recent happiness studies have shown that access to local amenities is significantly correlated with people's residential happiness. Earlier studies like Davis and Fine-Davis (1981) have documented the positive impact of rail access on local residents' overall neighbourhood satisfaction by using nationwide survey data in Ireland. Recent studies have used the reported happiness survey data to examine a wider range of local amenities and disamenities. For example, Van Praag and Baarsma (2005) find a significant effect of noise on individual's life happiness in the Amsterdam Airport area. Frey et al. (2009) discusses the impact of decreased incidents of terrorism on sampled residents' average life happiness changes in the UK, Ireland and France. Cornaglia and Leigh (2011) employ a micro panel data from

Australia to estimate the relationship between changes in crimes and changes in the mental wellbeing of resident non-victims. They find that increases in local crimes (especially the type of violent crime) have strong negative impacts on residents' mental wellbeing. Gibbons and Silva (2011) provide another good example by looking at the linkages between school quality and parents' happiness based on the Longitudinal Survey of Young People in England. They find a strong impact of school quality, measured by test scores, on parental perceptions about education effectiveness. They also find that the estimated happiness effects vary considerably for different individual groups because subjective wellbeing is influenced by not just amenity proximities, but also individual socioeconomic characteristics. Switching the focus to the impact of transport improvement program on happiness, direct tests have been less common. My third paper aims to fill this gap by using aggregated area panel data from Beijing.

Over the years, economists and geographers have developed a variety of frameworks for understanding the relationships between local public goods accessibility and happiness, which provide important foundations for my study. But there are some problems associated with the estimation strategies in the previous happiness literature. First, research on this topic has long been limited by the lack of a reasonable geographical-scale and scientific-designed survey data. Most of existing studies have conducted their analysis by using a general life happiness indicator. Although the general life happiness indicator could reflect people's cognitive assessment of local amenities to a certain degree, behavioural economists suggest that specific questions are more reliable than general questions (Frey and Stutzer, 2001; Alesina et al., 2004). This is clearly the case when my research topic is about transport improvements. Local residents should have direct living experience on how better rail access affects different aspects of their residential happiness. Therefore, the survey measures used in my research are specific questions on people's perceptions about particular residential aspects such as commuting and living convenience, social environment, traffic pollution and safety.

Second, most of the previous happiness studies are often framed by using the traditional cross-sectional type of empirical analysis. Once again, this cross-sectional approach cannot account for changes in the local amenity supply. A good case in point is the creation and expansion of transport infrastructure: local governments have continued to invest in rail transit constructions in order to make them more accessible to residents. I extend the growing happiness literature by providing a direct assessment of the impact of rail access changes at a given local area on homeowners' residential happiness at the same local area. Although I do not have a random experiment, the difference-in-difference style estimation strategy does take advantage of the repeated information about homeowners' happiness to evaluate the direct effect of rail transit development. To my knowledge, no study exists apply this type of analysis to the happiness evaluation in the developing countries.

A third empirical challenge for evaluating the impact of local public goods improvements on happiness is to consider the job searching or residential sorting concerns of the sampled residents. For example, unemployed people may be very happy with a residential location simply because of anticipated job opportunities, rather than changes in rail access. Similarly, households who prefer specific local amenities like good schools will move to places near these targeted local public goods, and this raises the danger of reverse causation effects relating to households' happiness outcomes. In fact, there is also the probability that potential increases in happiness can be offset by rising housing costs for people who do not own but rent their current homes. Given the limitations of available data, it is not possible to globally identify these effects in my study. Instead, I typically focus on examining the consequence of the increased rail access, caused by opening of new stations in 2008 Beijing, on homeowners' happiness with respect to particular dimensions of residential environment. I will look at the sampled residents who are homeowners that worked and held that tenure before the transport was improved. I control for changes in the composition to check for different neighbours moving in and moving out that may have some effect on the resulting estimates. By limiting my research focus onto "stayers" and non-market owners with pre-determined locations and non-market transactional rules, I further avoid contaminating the resulting estimates with changes in different time periods contributing to the estimation strategy.

A final estimation issue is about the choice of geographical boundaries or units. For some public goods, administrative boundary constraints are quite obvious. A typical case is school accessibility (Gibbons et al, 2012). For instance, school authorities usually arrange student admissions on the basis of so-called catchment areas. Suppose that two houses are located on the opposite sides of a particular catchment area boundary and that schools in these different catchment areas provide different quality levels. In this instance researchers who conducted the happiness effects on school quality cannot provide credible estimates by using simple proximity measures (Cheshire and Sheppard, 2004; Gibbons and Silva, 2011). However, this is not a problem for the rail stations given its public accessible characteristic. Still, another potential concern is the choice of aggregated geographical unit. Recall that there are two common ways to estimate the relationships between local public goods and happiness. The first option is to use the individual-level data. The second option is to use the geographically aggregated data to explore average happiness outcomes. This means that regressions are run by using the average happiness responses for certain geographical units as the dependent variable, and other socioeconomic characteristics as independent variables. This is understandable given the limits on data sample sizes. Researchers using the geographically aggregated data to do the analysis should be careful to provide a clear explanation for the rationale behind the aggregation process. See paper three for details.

2.3 China context

China and other BRICS countries are experiencing huge amounts of investment in upgrading the local infrastructure---a process that has significant implications for the land markets and homeowners' happiness. This thesis has been undertaken within this context. This section provides a brief overview of the institutional settings about the urban land market, local public goods provision in China, and some of recent empirical studies related to my thesis.

China, in a period of isolation from the eyes of the world which lasted roughly 30 years (1949-1978) committed most of its valuable recourses to military industry development. No recognisable land or housing market existed during this period. When China launched economic reform and opened up to the world in 1979, accumulated land and housing problems broke out in most Chinese cities (Dowall, 1994; Logan et al, 2010). As a positive outcome, land and housing reform was initiated from the 1990s, which finally give birth to an emerging urban real estate market.

After the land market reform, urban land was still owned by the state. However, urban land is now considered a valuable economic asset, rather than as a non-economically valuable physical space for people to live or work (Wu and Yeh, 1997). Land developers purchased land parcels from the city government, first through government regulations (prior 1999), then mainly through price-negotiations between developers and the city government (1999 to 2003), and recently through completely open auctions (since 2004)---those who offer the highest bid-price can obtain the land parcel (Zhu, 2005). At the macro-level, this remarkable transition of land market reform is representative of the overall economic transition process from a centrally-planned economy to a market-oriented economy. At the micro-level, the re-establishment of the urban land market has made price signals become effective in reflecting the importance of the location characteristics (Cheshire, 2007). It is natural to ask: Whether and to what extent the emerging land market in China exhibits the market characteristics that have been demonstrated in developed economies. As urban land is becoming valuable, it is important to evaluate the benefits of local amenities to households.

The impact of urban housing reform is significant on households' welfare. Before the reform, there was no housing market. All housing was provided by the work units and allocated to residents as aspect of state delivery of social welfare via employers (usually state-owned work units) through the central-planned economy system. All of these houses were owned by either the state or the work units. This meant that urban residents did not have property rights for their housings, and had very limited opportunity to sort themselves into different residential locations according to their income and other background characteristics. Since the late 1980s, most of the work-unit housings have been privatized at very low prices to their employees[7], and often loosely called the non-market (*fang gai*) housing. In the reform era, housing together with urban land markets has been gradually established. Developers have the right to build and sell housing in the real estate market. Within this marketisation context, large amounts of housing in urban areas were built to gain amenity benefits from access to transport infrastructure, green spaces and other local public goods.

It is worth noting that local public goods in urban China were established long ago in the centrally-planned economy and seldom changed their locations after they were

[7] Note that although work units transferred the ownership of the houses they owned to their employees, resale of non-market housing is usually restricted. Such non-transaction rules have been gradually relaxed but with additional limitations like selling the property to other employees in the same work-unit. Despite this, the actual transition of *fang gai* housing into fully market housing in Beijing is restrictively limited in order to forbid 'unreasonable' capital gains.

built. As such, the spatial locations of local public goods are exogenously determined. Meanwhile, Chinese homeowners do not need to pay property tax. Thus the capitalization effects of local public goods should be more significant than places with property taxes since land developers implicitly buy the local public goods when bidding for land parcels (Gyourko et al., 1999). Another thing to note is that, local public goods are financed by the city central government, not by local communities. This is because public facility construction and service provisions are highly centralized and controlled by the city central government. The basic administration units (zone, or *jiedao*) do not have voting rights for public infrastructure construction during the decision-making process. Thus the zone area only functions as a basic geographical unit for data collection, not as a political unit using local revenue to provide local public goods. Although this study seeks a delineation of a geographical unit that has a reasonable degree of homogeneity, the size of zone areas is much larger than the neighbourhood (census-block group) or school district in US and UK cities. Greater precision in geographic delineation can help capture the spatial heterogeneity within zones and improve the explanatory power of the hedonic price functions. However, this usually requires the help and expertise of knowledgeable local market participants such as property tax assessors and residential realtors. Unfortunately, such detailed data set is very difficult to obtain in this large developing country. Given this data limits, my main focus is to allow for differences in the proximity effects of parks or rail stations across local areas (like zones in this case), and the results presented below could be viewed as the best-possible efforts in China.

Recent literature has drawn attention to spatial features and determinants of land price in transitional Chinese cities, in comparison to its counterparts in advanced market economies. For instance, Zheng and Kahn (2008) present the first hedonic application to document the significant local public goods capitalization effects in Beijing. They found that proximity to stations and parks significantly contribute to land and housing prices. Following Zheng and Kahn (2008)'s pioneering work, there have been a small number of hedonic studies using micro-geographical data to evaluate the amenity values in other large Chinese cities (Wang, 2009, Jim and Chen, 2010; Wu et al., 2011). Nonetheless, research on this issue have been limited by the lack of systematic data – especially spatial data –on land leasing parcels as well as other related data sources, and by the limitation of the conventional cross-section hedonic approach in establishing the causal relationship between land price and its determinants. Indeed, there are several serious problems that have not been considered by existing hedonic applications in China. Firstly, previous studies have not explicitly allowed the proximity effects to vary with the local contextual factors that are believed to influence amenity values in the spatial context. It is reasonable to expect that spatial variations in amenity values due to observed and unobserved amenities and their complementarities would make their resulting estimates hard to interpret. Thus their OLS estimated parameters, at best, can only capture the entire urbanized area's average proximity effects. Secondly, most large Chinese cities like Beijing and Shanghai have made enormous investments in building new rail transit lines. These public investments in transport improvements would certainly change the transport accessibility for local areas. But there is still a lack of empirical studies capturing the effect of rail access changes on both of the residential and commercial land markets following the transport improvement programs.

In terms of the happiness research, it is more difficult to find direct test using large-scale micro survey data to examine the changes in amenity supply on people's happiness in China. Some studies concentrated on pre-designated sample areas using a small survey sample (Jiang, 2006). But insufficient information was given about the sampling method of the survey to indicate that whether this is reasonably representative of the urban population. A few recent studies, such as Zhang and Gao (2008), have investigated the general spatial differentiation patterns of traffic satisfaction in Beijing. However, nothing is known about the impact of transport improvement program on homeowners' happiness of any particular residential aspects. Once again, the term of "residential aspects" here means the domains of residents' living conditions relative to the specific survey questions like residents' happiness about the social environment, commuting and living convenience, safety and pollution.

3 Research questions

My main research questions in each paper are:

- ✧ Paper 1: What is the impact of proximity to parks as capitalized into the residential land prices?; and how would this vary according to other conditioning characteristics?
- ✧ Paper 2: What are the consequences of opening and planning new rail stations,

defined by station-distance reductions, on local prices of multiple land uses?

- Paper 3: To what extent are homeowners' happiness in specific residential aspects linked to rail access based on measures of residence-station distance changes?; and to what extent are homeowners' perceptions of better rail access varied based on their different social backgrounds (i.e., income and age)?

This empirical-based research relies on four databases that I have consolidated and geographically-coded over the past few years. Detailed data description can be seen in subsequent papers.

1) The Beijing Land Leasing Parcel Database (1999-2009), which reports the price, size, location and other relevant information for each vacant land parcel.

2) The Beijing Public Facilities and Services Database, which documents the spatial location and quality of local public goods.

3) The Beijing census database, which describes zone-level socio-demographic characteristics like population and employment density, educational attainment, etc.

4) The Beijing micro survey database, which includes two large-scale household surveys conducted in 2005 and 2009 respectively[8]. Each of the survey has about 11,000 respondents, and provides rich information on a household's

[8] The survey research is funded by the National Natural Science Foundation of China. The views expressed in this thesis do not necessarily represent the National Natural Science Foundation of China.

demographic characteristics and happiness evaluations with respect to different dimensions of residential environment.

Empirically, I have paid careful attention to causality when designing research methods and estimation strategies. The papers presented below are some of the first contributions to a growing literature on land markets, local public goods and happiness.

4 Highlight of paper 1

My first paper examines the spatial variations in local parks' capitalized values in the residential land market of Beijing, and how this might be affected by factors conditioning the parcels' location and location-specific characteristics.

4.1 Title

Spatial Variations in Park Amenity Values: Evidence from Beijing

4.2 Context and contribution

Park is a critical part of the urban infrastructure. The importance of being close to a park has been recently recognized by governments and developers in China. An evaluation of the amenity value of parks is useful for planners, enabling them to make better and more evidence-based policy decisions regarding public spending and environmental preservation. Such evaluations also enable real estate developers to know the estimated values of access to individual parks. Given the importance of this insight, there have been surprisingly few direct hedonic studies measuring the proximity effects of parks in a Chinese city context.

This paper explores the extent of land price capitalization of proximity to parks and how this might depend on factors conditioning the land parcel's location and local contextual characteristics. My first contribution to the literature is to allow the proximity effects to vary with a parcel's location and demographic characteristics over the urban area. More specifically, I consider how the effects of proximity to parks vary with park size, population density, educational attainment level, heritage buildings, crime rates, as well as access to other local amenities. Second, I account for the spatial heterogeneous effects in the proximity effect of parks individually by applying a locally weighted regression (LWR) approach. Furthermore, I explicitly exploit the robustness of LWR parameters of proximity to parks to the unobserved amenities and complementarities between amenities, and therefore, shed more light on potential sources of spatial variations in the amenity values. As far as I am aware, this is the first paper of this type of analysis in China, and among other developing countries.

4.3 Data and estimation strategy

This empirical analysis follows the baseline hedonic function but has several novel features. First, it adopts and modifies an existing locally weighted regression (LWR) model to include the complementary effects between the estimated values of proximity to parks and other location characteristics. Second, it suggests a foundation for visualizing the spatial variation patterns for the marginal prices of proximity to parks. The LWR model reveals significantly heterogeneity in the effects of proximity to parks on residential land prices over the urban space. Finally, it provides a powerful estimation strategy to evaluate how sensitive LWR parameters are to changes in the set of control variables. To be clear, the sensitivity of LWR parameters could be induced by a wide range of potential bias sources and this study has just focused on one---the assessment of the presence of omitted variables.

To achieve this, this paper takes advantage of uniquely rich geographically-coded data sets that link the location characteristics of land parcels, parks, local demographics, and other amenities from four micro-geographical datasets: (a) vacant residential land transaction records, which contain detailed information regarding the location, price, and size of each parcel; (b) park amenities data, which indicate the proximity effects of parks; (c) zone-level census data, which describes local socio-demographic characteristics; and (d) the spatial distribution and quality data of other local public goods from relevant government documents, which are used as proximity measures to control for additional location-specific characteristics. These four data sets are all geographically-coded into the GIS shape files. The precise location-matched information makes it possible to characterise detailed capitalization effects on a parcel-by-parcel basis.

4.4 Key results

I have reached two important implications. First, the empirical results show the complex and subtle variations in the estimated amenity values of proximity to parks over space. The point here is that the amenity value, which is being capitalized, varies according to other conditioning characteristics, and, thus, a park on which coal dust always falls is not “the same as” a park with a clean environment beside a beautiful river or lake. Second, I find that heterogeneity in the estimated implicit prices of proximity to parks is sensitive to unobserved amenities and their complementarities. This might not be a surprising technical innovation; however, in applied economics, it is particularly gratifying to identify, model, visualize, and assess the robustness of spatial variation in amenity values. One healthy implication from this is that researchers estimating the amenity value should do a careful plausibility check before directly applying those spatial econometric modelling results for any policy purposes.

5 Highlight of paper 2

The second paper of my research focuses on the consequences of local public goods improvements, in particular, new rail transit constructions on local prices for different land uses. Over 140 billion CNY (1GBP=10CNY) has been spent between 2000 and 2012 in Beijing on the building of new rail lines. This massive investment allows me to examine how residential and commercial land prices respond to changes in parcel-station distances using 1999-2009 vacant land parcel data in Beijing.

5.1 Title

Does Public Investment Spur the Land Markets? Evidence from Transport Improvements in Beijing

5.2 Context and contribution

There is a large volume of literature on the effects of proximity to rail stations on property prices that predominantly focuses on developed countries. However, there have been few studies on valuing rail access in China, and even fewer studies on exploring the opening and planning effects of rail access on local residential and commercial land prices, despite the rapid transport infrastructure changes.

I contribute to this literature in the following ways: First, I compare the impacts of rail access on both commercial and residential land prices. My study is also unique in using vacant land parcel data during 1999 and 2009 in the entire urbanized area of Beijing, rather than pre-designed sample areas. Second, conventional hedonic techniques for estimating amenity values mask the changing nature of geographical links between land parcels and stations induced by rail transit expansions. This paper improves on the previous literature by applying a multiple intervention difference-in-difference model that not only exploits changes in the parcel-station distances that happen when new stations are opened, but also highlights the importance of price changes at planned station areas. Third, I go further and examine the distance decay trend of rail access effect and how it depends on local demographics like employment accessibility, crime rates, and educational attainment.

I believe this is the first empirical study to use a rich vacant parcel sample of multiple land uses data in China, and allows the estimation by accounting for the increasing supply of rail stations. As far as I know, my results are original for the Chinese context.

5.3 Data and estimation strategy

My examination contributes to the small but growing body of literature on valuing rail access based on the difference-in-difference methodology (Gibbons and Machin, 2005; Kahn, 2007; Ahlfeldt, 2011). At its heart it captures the changing nature of geographical links between parcels and stations due to the construction of new lines.

To implement this strategy, I need data on land price changes and changes in access to rail stations. I meet the first data requirement by using a 1999-2009 cross-sectional land parcel data. Of course, the ideal is to use panel data. Given data limitations, my intention is not to claim that such precise price differences occurred to the same land parcel before and after uniquely because of transport improvements, but to identify what happened to the prices of land parcels when their distances to the closest stations were reduced. The second data requirement is easier to meet because of dramatic changes in public transport infrastructure in Beijing. The supply of new rail transit stations increased over time---two railway lines were opened in 2003, four lines were opened around 2008 and another eight lines were planned to open after 2009. These improvements will lead to an increased proximity to stations for a series of subsets of land parcels in my data set after 2003, after 2008, and after 2009 respectively.

I employ geographical information system (GIS) software to derive proximity measures from the Beijing residential and commercial land use dataset. I define the “treatment” as parcels that experience station-distance reductions, and where the outcome distances to the closest station are now less than a certain distance band (0.5km, 1km, 2km, 4km) due to new rail transit constructions. This multiple distance-band design allows me to explore the distance decay trends associated with the station-proximity benefits. I also run a set of sensitivity analyses to test the robustness of main findings across different subsamples, local demographics and analogous econometric specifications.

5.4 Key results

The results suggest that public investments in new rail urban construction have shown to spur spatially targeted land markets. I find that residential and commercial land parcels with better access to both newly opened stations and planned stations experience appreciable price premiums, though the relative benefits are different in magnitudes. I also find that the effect of increased station proximity on residential and commercial land prices varies nonlinearly at different distance ranges from a station, and varies widely with local socio-demographics. Given the huge public investment in the city, the question of who gains is important. Certainly, developers can benefit from appreciating land values in spatially targeted residential and commercial markets. My results may also imply the complementary effects between public investment and private sector investment, as higher levels of economic activity would translate into higher future tax receipts.

6 Highlight of paper 3

The third paper uses two large-scale household surveys conducted before and after the opening of new subway lines in 2008 to examine the heterogeneous impacts of transport improvements on homeowners' happiness in specific residential aspects.

6.1 Title

Does Better Rail Access Improve Homeowners' Happiness? Evidence Based on

Micro Surveys in Beijing

6.2 Context and contribution

In Beijing, four new subway lines were opened around 2008, with the total investment of 42.4 billion CNY (1GBP≅10 CNY). This massive investment provides a favourable setting for examining the consequences of place-based public investment in rail transits on homeowners' happiness nearby new stations.

While a large number of studies provide marginal values for rail access on the real estate market, few studies have focused on examining people's subjective/perceived evaluations of local public goods improvement. The paper makes several contributions to the growing literature on happiness economics. First, it provides new estimates about the subjective benefits of transport improvements based on homeowners' happiness (rather than e.g. house prices or looking at other economic outcomes). Second, it focuses not on general subjective assessments about life happiness, but rather specific questions on perceptions about particular dimensions of residential environment like commuting convenience, living convenience, social environment, traffic pollution and safety. In light of recent literature, this can help to create more reliable results than general questions. It is also noteworthy that this study uses large-scale micro survey data for Beijing's main urbanized area, rather than designated sample areas. Third, it uses a powerful difference-in-difference method that can more reliably assess the casual linkages between rail access changes and homeowners' happiness. My fourth contribution is to monetize the welfare effects of the transport improvement program by comparing the marginal utility of rail access and the marginal utility of income---holding housing prices and other local attributes constant. So far as I am aware, this is the first happiness evaluation of the transport improvement in the developing countries.

6.3 Data and estimation strategy

My estimation strategy takes advantage of two large-scale surveys that have been conducted before-and-after the opening of new rail stations in 2008 Beijing. The survey provided rich information on a household's demographic characteristics and happiness evaluations with respect to different residential aspects. I have geographically-coded the homeowners' place of residence and the newly-opened subway lines/stations with the help of the Geographic Information System technique.

In order to observe rail transit changes before and after 2008, I aggregate homeowners' happiness evaluations to the $1km^2$ cell-unit group. The rationale behind this is that, I will not require repeated individual-level responses of the same household in the difference-in-difference models, but only repeat average-level responses in the same cell unit. This means that my data is not a panel of people but a panel of areas. Empirically, I have tried to control for potentially endogenous changes to the compositions in response to transport improvements by (a) including changes in the average demographics; (b) using long-term residents (they were living there before transport was improved); (c) using non-market housings with pre-determined residential locations.

I use a difference-in-difference style estimation strategy to examine the effects of transport improvement, identified by distance reductions to new stations, on homeowners' happiness. I define the treatment as places that have experienced a fall in station-distance with the opening of new stations in 2008 and the outcome distance in 2008 is now less than 2 km. As a result of the large sample size, I am able to use the 1km and 4km distance bands to select the treatment group as a robustness check. I also run a series of additional robustness checks, testing for distributional effects across social groups and urban areas, the influence of changes in area-level compositions of residents' demographic characteristics, the role of policy-exogenous non-market housing with pre-determined locations, as well as the rail access impacts on different types of commuters' happiness.

6.4 Key results

I reach several novel conclusions. First, I find that better rail access has provided substantial and heterogeneous happiness effects to Beijing homeowners. Places receiving increased access to stations experience higher happiness levels about pollution, commuting and living convenience, and lower happiness levels about social environment and safety. Second, my results suggest that Beijing homeowners place substantial value on the improvements in the rail access brought by the transport investment program. Perhaps more surprisingly, I find that these benefits are not distributed evenly over social groups and geographical areas. Notably, my research has been limited by the lack of long-run and more detailed survey data. However, at the minimum, the results presented in this study provide some "healthy food for thoughts" for the important role of transport improvement program to play on homeowners' happiness, and provide useful implications for further place-based government investments.

7 Brief summary

This thesis provides three spatial applications of real estate, local public goods and happiness based on the new evidence from Beijing. While these empirical essays differ in certain terms, they share important common features. Firstly, all three essays are based on micro-geographical data after years of collection and geo-coding. This provides the basic foundation for the achievements of the insightful results. Secondly, all these essays have followed the restricted estimation strategy to explore the social-spatial differentiations of the relationships between land prices, amenity proximities and happiness, even though from a different perspective of view. Finally, all these essays are applied and policy-focused empirical works: the headline result from the first paper highlights the importance of considering the amenity value not just in terms of its structural characteristics but also how those characteristics interact with local contextual characteristics. The main result of the second paper suggests that new rail transit constructions does spur the spatially targeted land markets, and implies the complementary effects between public investment in transport infrastructure and private sector investment in land development. The third paper provides new evidence on the substantial and heterogeneous benefits of better rail access to homeowners' happiness and shed lights on potential welfare effects of the transport improvement program. In combination, the three papers of this thesis make important contributions to a growing literature on public infrastructure, land market and happiness. See detailed stories of each paper in the following chapters.

Papers

II. Paper 1---Spatial Variations in Park Amenity Values: Evidence from Beijing

1 Introduction

Park is an essential part of the urban infrastructure and is one which contributes to people’s life quality and the sustainable ecological-cities (Chiesura, 2004). The importance of urban parks has been widely recognized by city governments as an important local amenity to affect the land values (Cheshire and Sheppard, 1995).

The identification of the external benefits of urban parks in the form of altered land prices is important for the evaluation of individual parks. However, along with different sources of externalities, a park may simultaneously exert positive and negative benefits to households: while locating in the vicinity of a park will result in recreational access, pleasant landscape vistas, and ecological amenities, proximity to a park could also generate negative externalities linked to noise, congestion and safety concerns. In addition, the size of a nearby park should also be expected to influence the ways of land market capitalization in different places. As hypothesized by Berry and Bednarz (1979), land prices should reflect the complex interactions of amenity and location characteristics relative to a series of local public goods and demographics. This suggests that the land price capitalization effects of proximity to parks might be largely depend on factors conditioning the land parcel’s location and related demographic characteristics. The purpose of this paper is to examine the validity of such a hypothesis in a Chinese urban context.

Decades of urbanization and economic transitions have dramatically spurred the Chinese urban land market (Wu et al., 2011). As urban lands become valuable, planners and land developers have to balance the trade-off between developing and preserving the urban parks and green spaces. Although development could meet additional demands for residential and commercial spaces, proponents of preservation are motivated by major concerns that include environmental awareness and protection, as well as the prevention of social problems within the rapid urbanization context (Jim and Chen, 2010). To this end, an evaluation of the park amenity value is particularly useful for planners, enabling them to make sound policy decisions regarding public investment and related land supply regulations. Such an evaluation may also benefit developers by justifying the expenditures and improving land development efficiency.

This paper explores the impact and sources of variations of park proximities as capitalized into the residential land prices by using the vacant land parcel data in Beijing. Values for proximity to parks are first estimated globally with a traditional ordinary least squares (OLS) model. A locally weighted regression (LWR) model is then used to examine spatial variations in the amenity values for parks individually[9]. I contribute to the literature in three ways. My first contribution is to allow the effects of proximity to park to vary with parcel's location and demographic characteristics over the urban geographical area. To be more specific, I consider how the effects of proximity to parks vary with park size, population density, educational attainment level, heritage building percentage, crime rates, as well as access to other local amenities. Second, I provide a

[9] In essence, LWR is a flexible statistical method that specifies a separate regression at each observation point, thus generating unique coefficients to be estimated at each location (Cleveland and Devlin, 1988). Recent studies have shown that this method can better account for spatial variations in the amenity values in the real estate markets (Leung et al, 2000; Cho et al., 2006; McMillen and Redfearn, 2010). But it is necessary to keep in mind that this study does not attempt to compare advantages between the LWR and other spatial econometric methods or testify all aspects of spatial effects (see Bitter et al., 2007 and Anselin and Lozano-Garcia, 2008 for details).

powerful estimation strategy to assess how sensitive LWR parameters of proximity to parks are to the unobserved amenities and complementarities between amenities, and therefore, shed more light on the potential sources of spatial variations in the amenity values. Thirdly, I suggest a foundation for visualizing spatial variation patterns of the estimated values of proximity to parks over space. To achieve this, I take advantage of rich micro-geographic data that links the specific characteristics of land parcels, parks, local demographics and other amenities. The precise location-matched information makes it possible to characterize detailed capitalization effects on a parcel-by-parcel basis.

The empirical results show the complex and subtle variations in the estimated amenity values of proximity to parks over space. Using the entire urbanized area's average effects might therefore overestimate or underestimate the interpretation of the variations in the amenity values at particular places. Furthermore, the estimated values from LWR models for individual parks document the significantly heterogeneity in the effects of proximity to parks on residential land prices. However, the LWR parameters of proximity to parks are still sensitive to the unobserved amenities and complementarities between amenities. It is important to note that the sensitivity of LWR parameters could be induced by a wide range of potential bias sources and this study has just focused on one---the assessment of the presence of omitted variables. In this complex spatial context, these findings add to the evidence of conceptualizing the “amenity value” not just in terms of its structural characteristics but how those characteristics interact with or are conditioned by social, economic, or other local contextual characteristics.

The remainder of this paper is structured as follows: section 2 discusses the related literature; section 3 describes the econometric models; section 4 introduces the data used in the analysis; section 5 presents the estimation results; and section 6 concludes.

2 Literature review

The literature relevant to my analysis includes studies that have estimated the proximity effects of parks on land or property values by applying the conventional OLS approach and the newly-developed LWR approach.

A large and growing number of studies have estimated the proximity impact of green space and park amenities on property values by using the OLS-based hedonic approach (Gibbons et al., 2011). McConnell and Walls (2005) provide an extensive review for more than 60 published English papers that have examined the external benefits of green spaces or parks by using distance measures. A general conclusion is that, all else being equal, proximity to parks has the significant impact on property values, but the effects vary greatly by types. For example, some studies find that preserved green space usually has the strong positive impact on nearby property values, but developable green space has a weak or insignificant impact on property values (Bolitzer and Netusil, 2000; Irwin and Bockstael, 2001a). Cheshire and Sheppard (1995) examine the proximity impacts of publicly accessible and inaccessible green spaces on residential land prices in two mid-sized UK cities. They find that only publicly accessible green space significantly increases residential land prices. Irwin (2002) summarized the specific estimation issues associated with green space types: If green spaces are privately owned, or can be developed for residential land use in the future, then the variables estimating the influence of green space on nearby residential land values are endogenous in the hedonic models. This should not be a problem in Beijing since all the parks are accessible to the public and preserved permanently by the government.

In addition to varying by types, it is also reasonable to expect the amenity value of parks to depend on its location and surrounding characteristics. Using the OLS approach, Geoghegan et al. (2002) find that the amenity value of parks varies significantly with the distance to the central business district (CBD). Anderson and West (2006) allow the proximity effect of parks to vary with local demographic characteristics and include neighbourhood fixed-effects to control for observed and unobserved location-specific characteristics. They find that the amenity value of proximity to parks is higher in places that are dense, near the central business district, or places with more high-incomes and children. However, their neighbourhood fixed-effect approach is appropriate only when the omitted variables do not vary too much within a neighbourhood like the tax rates. This approach would also fail to control for omitted spatial variables that affect just one single property or a small group of properties within the same neighbourhood. As a result, their estimates should conceal substantial variations among individual parks. Empirically, it is important to specify local fixed effects at a finer geographic scale to control more effectively for omitted variables[10].

[10] See Cheshire and Sheppard (2004) and Redfearn (2009) for a detailed discussion.

Over the past thirty years, spatial econometrics literature has focused on advancing methodologies related to the estimation process of incorporating spatial effects into the model specifications (Anselin, 2010). One brand of this literature has focused on developing an alternative approach that would be better account for the spatial heterogeneity effects of the geographical data (McMillen, 2010). A well-cited candidate method is the LWR approach (Cleveland and Devlin, 1988). The primary advantage of the LWR design is that by estimating a vector of implicit prices at each observation, it is able to control for heterogeneity in each location. This approach has recently been applied intensively in the real estate market to test for local heterogeneity (Leung et al, 2000; Bitter et al, 2007; Redfearn, 2009; McMillen and Redfearn, 2010). Empirically, Cho et al. (2006) presents the first attempt that uses the LWR method to measure the spatial heterogeneity effects of proximity to parks. They find that the average marginal implicit price of proximity to parks estimated by the OLS model was $172 USD, whereas the LWR model indicated that the marginal implicit prices varied from park to park, ranging from –$662 to $840 USD. This paper uses Cho’s study as a useful benchmark of departure, but argues that the direct application of this spatial econometric method is problematic. The key potential concern is that Cho’s seminar work presents only one model specification without any sensitivity analysis for the omitted variable issue. Although the LWR approach can be used to maximize the model fit, this does not demonstrate it is a “correct” model in terms of casual interpretation. Some studies have shown that the LWR estimates are robust to the selection of “optimal” bandwidths (see Farber and Páez, 2007; Redfearn, 2009). But much is still unknown about how sensitive the LWR parameters of proximity to parks are to the unobserved characteristics and their interaction effects associated with the proximity effect of a specific park. Indeed, it is quite possible that the external benefits derived from proximity to a park would increase when a park is close to subway stations, and would decrease when a park is located in high crime rate areas. This paper presents the first application to examine the impact and robustness of spatial variation in the values of proximity to parks in China. The next section spells out the detailed econometric models.

3 Model

Hedonic models are designed to identify the marginal effects of a commodity's differentiated characteristics on its purchase price (See Sheppard, 1999 for a recent review). Land and housing are the most common examples of hedonic application. A hedonic model of residential land prices can be expressed as:

$$P_i = f\left(S_l, N_l, .E_l\right) \qquad (1)$$

where P_l is the market price of the lth residential land parcel; S_l is the land's structural characteristics; N_l is a set of location-specific characteristics; and E_l represents the park amenity attributes. The differentiation of the hedonic price equation with respect to a particular characteristic yields each individual property buyer's marginal willingness to pay, assuming the market spatial equilibrium.[11] Freeman (1979) indicates

[11] Rosen (1974) designed a second-stage hedonic analysis. In the second step, the estimated marginal prices are regressed on a vector of demand variables to identify customers' willingness to pay (see Cheshire and Sheppard, 1998; Day et al, 2007). This study does not attempt to undertake such an analysis due to the lack of high quality

that if the function in equation (1) is a linear relationship, the implicit price of a certain characteristic should be constant for all individual properties. However, if the function shows a heterogeneity relationship, its implicit price will depend on the quantity of that characteristic and its covariates with other attributes. As suggested by Freeman (1979), the heterogeneity is predictable, not only because properties' attributes are heterogeneous in different locations, but also because land buyers are heterogeneous in their willingness to pay for certain characteristics. This leads to the spatial variations in the amenity values, at least over a short-time period.

There is little to say about the choice of functional forms in the hedonic price model. Some empirical studies have shown that a log transformation of land prices and proximity variables performs better than the straightforward linear or the complex Box-Cox functions[12] because it does a good job in accounting for the non-normality of disturbances and capturing the spatial decay trends of the proximity variables in hedonic price models (Cheshire and Sheppard, 1995). By having a number of choices regarding the functional form of the hedonic analysis, a better fit is achieved for the available data and variables. In this study, several flexible-form models were used but were unable to reject a clear log–log relationship between land prices and key explanatory variables. Using the OLS approach, standard hedonic models can be estimated in the following form:

---

demand-side data. Instead, I am mostly looking at the localized externalities of park proximities, as Palmquist (1992) suggested that marginal prices can reasonably measure the external benefits of local amenities.

[12] Though the Box–Cox transformation is more flexible than other methods, the complicated transformation procedures may generate more random errors (Davidson and MacKinnon, 1993).

$$\ln P_{li} = \alpha' \ln X_{li} + \delta' Z_{li} + (\lambda + \theta size_{li} + \mu' Z_{li}) \ln dist_{li} + \gamma_i + \varepsilon_{li} \qquad (2)$$

Where P_{li} is the leasing price of residential land parcel l in zone i; X_{li} is a vector of land parcel structural characteristics and related dummy variables; Z_{li} is a vector of location-specific and demographic characteristics; α and δ are parameter vectors to be estimated; $dist_{li}$ is the distance to the nearest park, and $size_{li}$ is its size; λ, and θ are two parameters, and μ is a parameter vector to be estimated; γ_i is the parcel-specific coordinate location, measured by each parcel's location coordinates (x,y) and its spatial variations (x^2,y^2, xy)[13]; ε_{li} is a residual capturing error term. This OLS model builds up a hedonic functional relationship between the land price and those location characteristics. Two key land structural characteristics included in this study are the parcel size[14] and the median value of surrounding commercial land parcels within 2km. I also try to test the potential spatial autocorrelation effects by including two indicators: the median residential land value of zones and the spatial price lag term, identified by the weighted mean residential land price around each parcel. In addition, the land parcels' coordinate locations and their variations are included as the spatial fixed effects. A set of year dummies is included to capture the potential differences in land prices among different

---

[13] This study also tried to use the area-specific dummies as controls for the fixed effects. However, since the basic geographical scale (i.e. zone) in Beijing is large, it may fail to control for the omitted factors that only affect a single land parcel or a small group of land parcels. To be clear, the application of the parcel coordinate and its variation controls as kinds of spatial fixed effects is not without limitations. In essence, this approach imposes a continuous "dome" pattern on the spatial structure of the real estate market. However, it is widely recognized that some location characteristics that would affect land price heterogeneity are discrete over space. For instance, school districts play a critical role in the determination of land and housing prices (Cheshire and Sheppard, 2004; Gibbons and Machin, 2008). As such one would expect a price-discontinuity pattern when moving from a good-quality school catchment area to a bad-quality area. In this case, it may be more appropriate to use the area-specific fixed effects.

[14] I imposed the quadratic specifications for some structure variables like the parcel size to capture the nonlinear effects but found that the results are virtually similar.

years.

In terms of park amenity variables, I first use the distance to nearest park as the proximity measure. I then use the size of the nearest parks as a proxy indicator to reflect the parks’ quality condition. Some recent studies argue that the proximity effects of parks on land prices may not be observable when the parcel is located at a greater distance from a park (Hoshino and Kuriyama, 2010). I address this issue by including another two variables: the log of the sum of the park areas within a 2 km radius of a residential land parcel, and a dummy variable for a park size larger than 0.5 km^2 within a 2 km radius of a residential land parcel.

Local demographic characteristics were captured primarily by census data on median education attainment level, population density, crime rates, percentage of heritage architectures built before 1949. Due to the lack of income information in the census data, the median education attainment level and crime rates are used to reflect the basic socioeconomic conditions of a zone. Population density is used to measure how population pressure on park amenity affects the land market. Heritage architecture percentage is one interesting local contextual factor that has not been widely examined in previous studies. However, it may play an important role in affecting land prices in countries like China that experienced significant urban renewals due to the fast urbanization process. Other location-specific variables included in this study are distance to CBD, distance to nearest subway station, school and river. These proximity variables are intended to capture their capitalization effects on land prices and their complementary effects with proximity to parks.

The elasticity of residential land prices with respect to park proximities can be expressed as:

$$\partial \ln P_{li} \big/ \partial \ln dist_{li} = \lambda + \theta size_{li} + \mu' Z_{li} \tag{3}$$

When this elasticity is negative, residential land price falls as distance to nearest park increases, so the proximity to parks has a positive effect on residential land price. To simplify the explanation of parameter coefficients, the location-specific and demographic variables are normalized based on the linear transformation: $Z_{li}* = (Z_{li}-Z_{mean})/Z_{mean}$, where Z_{mean} is the sample mean value. The normalization of park size ($size_{li}^{*}$) follows in the same way. Given the normalization of the location-specific and demographic attributes the elasticity in Eq. (3) becomes:

$$\partial \ln P_{li} \big/ \partial \ln dist_{li} = \lambda + \theta size_{li}^{*} + \mu' Z_{li}^{*}, \tag{4}$$

which can further simplify to

$$\partial \ln P_{li} / \partial \ln dist_{li} = \lambda \tag{5}$$

for a park of average size and a land parcel with average location-specific and demographic attributes. Therefore, the coefficient of distance to nearest parks can directly be interpreted as the elasticity of land price with respect to proximity to parks for a land parcel with average local contextual characteristics. It is predicted that the amenity value of the proximity to parks will be lower when it relates to smaller park size. Residential land parcels adjacent to larger parks are more likely to generate substantial external effects and therefore extend this amenity value. In addition, the amenity value of the proximity to parks is expected to be higher when associated with better access to other amenities like schools and subway stations. Meanwhile, the amenity value of the proximity to parks is hypothesized to be lower in places with higher crime rates. As parks are regularly favoured venues for criminal behaviours, households in high-crime areas may be afraid to engage in outdoor activities in nearby green spaces (Gibbons, 2004). Thus, the amenity value of proximity to parks is likely to decrease in those areas. I expect that the amenity value of being close to a park of a given size will increase with high education attainment levels of local residents as well-educated groups may be willing to pay more for the proximity to parks. Heritage architectures in Beijing is more common in the central city with beautiful surroundings due to the recent urban renewal policy, thus the value of being closer to a park of a given size should increase in places with more-heritage architectures. Finally, I expect that the value of proximity to parks will be lower in places with high population density because of the noise, safety and congestion problems.

In the spatial context, the equation (2) can be considered as a global model. The partial derivatives of the OLS hedonic price model with each variables yield an overall marginal implicit price. This marginal implicit price for the nearest park is essentially an average across all parks over space and the willingness to pay for increased proximity to any particular individual park cannot be fully revealed in the OLS model[15]. Therefore I estimate the hedonic price function by using the locally weighted regression (LWR):

[15] Recent studies have shown that the OLS model can also reasonably identify the spatial variations in the effects of amenity proximities after proper modifications, like controlling for the interaction effects between amenities and effective fixed effects (Fik et al., 2003; Gibbons et al., 2011; Gibbons and Overman, 2012). However, some argue that the LWR is a more flexible statistical tool and can perform better than the OLS in terms of modelling fit (Redfearn, 2009). The purpose of applying LWR model here is not just to show its good performance, but also to testify its robustness to the omitted variables. In any case, it is important to emphasis that the key focus of this study is to look at how park proximity interacts with other local contextual factors rather than arguing for a specific “optimal” method.

$$\ln P_{li} = \alpha_l' \ln X_{li} + \delta_l' Z_{li} + (\lambda_l' + \theta_l' size_{li} + \mu_l' Z_{li}) \ln dist_{li} + \varepsilon_{li} \quad (6)$$

Note that each parameter to be estimated in Eq. (6) has a footnote *l* indicating that the locally weighted regression estimates the parameters at each land parcel. Calculation of the locally weighted regression model follows a locally weighted least squares technique. Since Fotheringham et al. (2002), scholars have generally used one specific variant of the LWR---geographically weighted regression (GWR) in hedonic applications (see Bitter et al., 2007 for details). Practically, LWR assigns weights according to their spatial proximity to location *l* to account for the fact that an observation near location *l* has a greater influence on the estimation of parameters than observations located further from *l*. That is,

$$\hat{\beta}(u_l, v_l) = (M^T W(u_l, v_l) M)^{-1} M^T W(u_l, v_l) P \qquad (7)$$

Where (u_l, v_l) denotes the coordinates of the *l*th land parcel in location; $\hat{\beta}$ represents all the estimated parameters; $M = [X_{li}\ Y_{li}\ Z_{li}\ size_{li}]$; and $W(u_l, v_l)$ is an $n \times n$ diagonal spatial weighting matrix. The Gaussian function is used to estimate where d represents the Euclidian distance between the regression point and observation point, and h represents bandwidth as follows:

$$W(u_l, v_l) = \exp(-hd^2) \qquad (8)$$

In the process of calibrating a locally weighted regression, the weighting matrix and *h* should first be decided. Bandwidth *h* can be decided by a cross-validation procedure[16] in order to generate the relatively robust results (Farber and Páez, 2007) as

[16] Note that I have experimented with both of the adaptive bandwidth approach and the fixed bandwidth approach.

follows:

$$\min \sum_{i=1}^{n} [LnP_{li} - Ln\hat{P}_{\neq l}(h)]^2 \qquad (9)$$

where $Ln\hat{P}_{\neq l}(h)$ is the fitted residential land price of LnP_{li} with the observations for point l omitted from the fitting procedure. Sensitivity analysis was conducted for bandwidths relating to both plus and minus 50% of the h selected by the cross-validation approach[17].

Since the LWR model allows each regression coefficient to vary over location by controlling the location-specific characteristics, the spatial variation of the price elasticity of proximity to parks can be then estimated locally. Thus the price elasticity of a residential land parcel with respect to proximity to a specific park can be written as:

$$\partial \ln P_{li} / \partial \ln dist_{li} = \lambda_l + \theta_l size^{*}{}_{li} + \mu_l' Z^{*}{}_{li} \qquad (10)$$

The elasticity calculated from the LWR model depends on λ_l, and the interactions between distance to nearest park, park size and the covariates in vector Z_{li}---a set of local contextual factors that believed to influence the amenity value of parks. A negative sign of this elasticity means that the proximity effect of a specific park will be more valuable with an increase in the corresponding location-specific characteristics. These localized marginal implicit prices of parks are summarized to visualize their spatial variations in amenity values across different parks.

However, the RSS (residual sum square) of the adaptive bandwidth approach is substantially larger than the fixed bandwidth approach, suggesting that the fixed bandwidth approach is more suitable for my spatial datasets.

[17] Recent spatial econometric literature also offers some other techniques than the cross-validation approach for the selection of the optimal bandwidth parameter such as the parametric plug-in method or the semi-variogram analysis (see Anselin and Lozano-Gracia, 2008).

4 Data

Beijing is the capital city of China. It is a largely monocentric city that is more similar to European cities than to American cities, with a few exceptions of historical US cities such as Boston (Brueckner et al., 1999). In Beijing, the CBD (TianAnMen Square and JianGuoMenWai Street) is found to play an important role in the spatial distributions of population density, income, as well as land and housing price (Zheng and Kahn, 2008). This centralized urban form is mainly due to the concentration of employment opportunities and local amenities near the central city. Following the convention, my study area mainly covers four central city districts (Dongcheng, Xicheng, Xuanwu and Chongwen) and four nearby suburb districts (Chaoyang, Fengtai, Shijingshan, Haidian), as other places are predominately rural. Within the Beijing urbanized area, zone (*jiedao*) is a fundamental census administration unit. Zone in Beijing is similar to a very broad census tract in the US cities—it forms the basic geographical unit for data analysis; it is not a political unit using local revenue to provide public services.

This study uses four unique geographically-coded datasets: (a) land parcel records, which contain detailed information regarding the location, price, and size of each parcel; (b) park amenities data, which indicate the proximity effects of parks; (c) zone-level census data, which describes local socio-demographic characteristics; and (d) the spatial distribution and quality data of other local public goods from relevant government documents, which are used as proximity measures to control for additional location-specific characteristics. These four data sets are all geographically-coded into the GIS shapefiles. Table 1.1 provides the descriptive statistics for the variables that I use to estimate the Eq. (2) and (6).

In China, urban land is legally owned by the state government. Since the 1990s, most Chinese cities have experienced dramatically changes in the land allocation system, from the free allocation toward a leasehold system (Zhu, 2005). In practice, the city municipal land authority is responsible for land allocations through sales of leasehold rights (70 years for residential land use and 40 years for commercial land use). To avoid potential corruptions and establish a transparent land market, all land parcel transactions must go through the public competitive auction process since 2004. From the Beijing Land Resource Authority, I have collected specific price and size information on the 685 vacant residential land parcels sold during 2004 and 2008. After excluding incomplete data, the final sample size was 615[18]. The mean residential land price is about CNY 3286.5 per square meter (1GBP equals to approximately 10 CNY).

The data for parks' locations and sizes were collected from the Beijing Municipal Garden Bureau. Using the ArcGIS 9.3 software, the nearest straight-line distance from land parcels to parks were calculated. Geographical information on other location characteristics is taken from a variety of sources for the use of controllable variables in the regression models. Notably, the local public goods were built long ago in the

[18] To mitigate the inflation effect, I have adjusted the land prices by using the CPI index reported by the Beijing Statistical Year Book 2004-2009. All monetary figures are constant in 2008 CNY. Also, I have trimmed the land price distribution by keeping parcels in each year whose price is between the 5th and 95th percentiles of the whole sample price distribution.

central-planning economy and seldom change their locations after they are built. Thus, one advantage of using these local public goods as a set of controllable variables is that the location of public goods is exogenously determined in Beijing. School location and quality[19] comes from the Beijing Municipal Committee of Education. The GIS data on the sites of subway stations and rivers is taken from the Beijing Municipal Transport Bureau and Water Authority respectively. Crime rates for the number of violent crimes taking place in each zone are obtained from the Beijing Public Security and Safety Bureau. The most recent 2000 City Population Census reports the detailed local demographic characteristics like the population density, residents' median education attainment levels and the percentage of heritage architectures built before 1949.

5 Results

The results are reported in Tables 1.2–1.7 with the following objectives. In the first half of this section, I report the estimates of hedonic price functions based on the OLS and LWR model. In particular, I focus on examining the ways local contextual factors interact with the marginal effects of proximity to parks. In the second half of the section, I explore the robustness of LWR parameters of proximity to parks to the unobserved amenities and complementarities between amenities, and thus shed more light on the potential sources of spatial variations in the amenity values.

Table 1.2 summarizes the results of the OLS model and LWR model. The adjusted

[19] The school quality is computed from the Academic Performance Rank Index. This index is measured by both base and growth values of their average scores of Middle School Entry Test and Graduate Test.

R^2 value for the OLS model is 0.39[20], while for the LWR model it is 0.71. A test for significant differences between the LWR and OLS models confirms that the LWR model fits the data better than the OLS model (see Appendix Table 1.1). Of course, my particular interest is not goodness of modelling fit but the impact and sources of park proximities on residential land prices.

The results from the OLS model show that most of the variables are statistical significant with expected signs. The positive signs associated with the variables of parcel size and the surrounding median commercial land values suggest that land price increases with larger land parcel areas and higher values of surrounded commercial lands. I also find that residential land price rises by about 0.63%, 0.14%, and 0.13% for every one percent decrease in distance to CBD, nearest subway station, and school respectively. The positive relationship between distance to rivers and residential land value is counterintuitive because rivers are normally considered an amenity with ecological benefits and pleasing views. The LWR results also show that more than 75% of coefficients have positive signs, suggesting that residential land price increases with increasing distance away from rivers. This is not a surprising finding since most of rivers around Beijing Metropolis usually do not have high water quality (Day and Mourato, 1998). All of the local demographic characteristics are statistically significant at the 5% level. As expected, residential land price falls as population density and crime rates increases, and rises with high education attainment and more heritage

[20] This relatively modest adjusted R-square value was expected given the emerging land market system in China (Zheng and Kahn, 2008, Wu et al., 2011).

architectures.

Switching the focus onto the park amenity coefficients, I find that the value of an average residential land parcel increases with proximity to parks, with benefits of a 0.67% price premium for every one percent decrease in the distance to the nearest park. This effect of proximity to parks is statistically significant and far larger than that for other amenities. Interestingly, I find that the park size and the related dummy variable for adjacent to a larger park have a significantly negative influence on residential land prices. This may be caused by some disamenities associated with large parks, such as noise and crowded population flows.

Nearly all of the interaction terms with the park proximity variable are statistically significant at or near the 1% level. The OLS estimates show that an increase in the size of the nearest park makes the elasticity of land price with respect to distance to nearest parks more negative[21]. Indeed, residential land parcels adjacent to larger parks are likely to provide more leisure spaces and therefore extend the park amenity value. According to the positive coefficients on the amenity accessibility interactions for parks, the amenity value of proximity to parks increases with better access to subway stations and schools. This value falls as population density increases, possibly due to associated noise and congestion effects. As expected, the amenity value of proximity to parks falls as local crime rates increases and rises with education attainment levels of local residents. Perhaps surprisingly, I find that the amenity value of proximity to parks is

[21] Recall that when this elasticity is negative, residential land price falls as distance to nearest park increases, so the proximity to parks has a positive effect on land price.

higher for places with greater proportion of heritage architectures (buildings built before 1949), implying that parks and historical architectures are complements in Beijing. In the preliminary estimation, I have also interacted the park proximity variables with other localized factors such as distance to CBD. As these interactions are not statistically significant and induce a severe multicolinearity problem (Wheeler and Tiefelsdorf, 2005), they are dropped from the final model specification. In place of the parcel location controls[22], I estimated the OLS model by including some new variables like job density, air quality, proximity to highway, and retail establishments. By adding these spatial variables, the significance of park amenity variable is weakly improved but with unexpected signs. Similarly, the interaction terms increased in statistical significance but produced inconsistent signs. This tells a consistent story with other empirical studies that the OLS estimates are very sensitive to unobserved characteristics (Cheshire and Sheppard, 2004; Redfearn, 2009).

As an interesting extension, I also used a Lagrange Multiplier diagnostic test (LM diagnostics) to examine if significant spatial autocorrelation exists in the OLS residuals of the standard hedonic pricing model. At the heart, I estimated the spatial lag and spatial error models[23], and used the R-program to run the LM diagnostics for comparisons between a-spatial model and spatial models. The LM diagnostics treat the a-spatial model (standard OLS model) as the restricted model (null hypothesis), and the spatial model as the unrestricted model (alternative hypothesis). Thus the LM diagnostic

---

[22] Note that although not shown in Table 1.2, the parcel-specific location coordinate effect has significant impact on the land prices.

[23] See Anselin (1990) for the classic exposition, and see Carruthers and Clark (2010) for a recent application

can effectively consider the difference between the spatial and a-spatial models as a result of unobserved variables. Although not shown in the table, I find that in situations where using distance based continuity spatial weight matrix with a threshold of 2km, the LM diagnostics show significant spatial autocorrelation effects in the residuals from the OLS model. Furthermore, the robust versions of Lagrange multiplier tests support the spatial model specifications. This result suggests that the spatial versions of the hedonic pricing model do a reasonable good job for correcting the potential spatial autocorrelation effects.

In the spatial context, one important feature of the LWR analysis is to quantify the localized ways of the spatial heterogeneity effect in the value of proximity to parks individually. To better visualize this, Figure 1.2 shows the locations of the parks and spatial variation in the marginal effects of proximity to parks. Table 1.3 reports the summarized results of the mean marginal implicit prices of proximity to each individual park, which is calculated using a floating circle with 4,000 meters radius[24]. As shown by the figure and table, the marginal effects of proximity to parks in the suburbs (Chaoyang, Haidian, Fengtai and Shijingshan districts) are higher than those located in the central city. Specifically, Diaosu Park and Xiwang Park in Shijingshan District, and Minzu Park and Chaoyang Park in Chaoyang District have the largest capitalization effects. In contrast, some parks (Jingshan Park, Beihai Park and Gugong Park) in the central city

[24] Note that these summarized park values are only used to reflect the spatial variations in amenity values, not to do the precise evaluation. Furthermore, it is necessary to keep in mind that the amenity values generated by the hedonic methods only provide a reasonable measure of marginal economic benefits—hedonic prices cannot fully reflect marginal social-psychological or happiness benefits captured by local residents (see Luechinger, 2009; Gibbons and Silva, 2011).

have the small negative capitalization effects. One possible explanation is the high substitutability effects between parks and other amenities in the central city area. As there are few undeveloped residential lands available in the central city, developers at the downtown areas may value access to employment centres and other local public goods accessibility more than proximity to parks. This variation in the values of proximity to parks may also be explained by the different park functions. For example, some parks in the central city (such as Jingshan Park, Beihai Park and Gugong Park) are world-famous historic attractions, which are likely to generate local concerns for the congestion, safety, and noise disamenities.

In the second half of this section, I present a pioneering work by applying different combinations of LWR model specifications in estimating the proximity effects of parks. My primary goal is to examine the sensitivity of the LWR parameters of proximity to parks to the changes in the set of control variables[25]. Table 1.4 presents the results by using six LWR empirical specifications[26]. These LWR parameter estimates, which vary at each of the 615 observation parcel locations, are displayed as medians and an inter-quartile range (IQR). Model (1) estimates the residential distance to the nearest park, with no additional controls. From models (2) to (3), I estimate the specification to

---

[25] Note that I also test the stability of the LWR parameters to the bandwidth choice, and the results suggest that the median value of the LWR coefficients using a bandwidth of 7.35 km (50% larger than 4.9) and 2.45 km (50% less than 4.9) is fairly close to the median value when the bandwidth is 4.9. However, when the bandwidth widens to 7.35 km, the spatial variations in the LWR estimates of proximity to parks are close to those estimated by the OLS model. One important implication from here is that researchers should balance the tradeoff between the need to capture the very localized variations in the amenity values using the smaller bandwidth and the demand to generate global estimates using the larger bandwidth.

[26] To make these specifications more comparable, all the model specifications are estimated by using the same bandwidth even though this may not be the optimal one for some specifications.

further include related park variables, land structural attributes, and location-specific variables. The final three models in Table 1.4 have increasingly included, as completely as possible, interactive terms in the model specifications.

An assessment of the sensitiveness of LWR results proceeds first by using Pearson correlation and Spearman rank correlation[27] indicators. Tables 1.5 and 1.6 summarize the results of the Pearson correlation coefficients and the Spearman rank correlation coefficients, respectively, for the parameters of proximity to parks estimated by the LWR models. I find that both the Pearson correlation and the Spearman rank correlation coefficients are greater than 0.5 and had statistically significant signs. These results suggest that the estimates for proximity to parks have similar spatial ordering and correlation relationship across different model specifications. With regard to this criterion, it can be concluded that the parameters of proximity to parks estimated by LWR models are generally stable. Nevertheless, these results do not represent a precise test. Correlation coefficients that are greater than 0.5 only provide an indication of shifts that are not considered statistically significant.

Next, I derive a more precise estimation strategy to test the robustness of the LWR parameters of proximity to parks and thus shed more lights on the potential sources of spatial heterogeneity in the park amenity values. Using Eq. (10) and the LWR coefficients in Table 1.4, I first calculate the price elasticity of residential land value

[27] Compared with the linear function illustrated by the Pearson correlation, the Spearman rank correlation describes the monotonic function between parameters (Aitkin and Longford, 1986), and thus it is a more straightforward way to show whether different specifications provide, at least, the same spatial ordering for the LWR parameter estimates at different locations.

with respect to park proximity (park elasticity, thereafter) across different model specifications, and then plot their distribution curves in Figure 1.3. It is apparent that these distribution curves have experienced substantial changes when including additional control variables into the model specifications. To determine whether the observed changes in these distribution curves are statistically significant, a non-parametric test is conducted. Fan and Ullah (1999) proposed a non-parametric statistical test for the comparison of two unknown distribution curves, say f and g—that is, a test of the null hypothesis—H0: f (x) = g(x) for all x, against the alternative, H1: f (x) ≠ g(x) for some x. The rationale behind this test is that if the distribution curve in the subsequent model specification is statistically different from the former model specification, it implies that the newly added control variables (in the subsequent model specification) are the potential sources of spatial variations in park amenity values.

Table 1.7 shows the estimated results. The first column indicates the null hypotheses: first, the inclusion of the variables in the subsequent model specification do not produce a significant difference compared with the previous one; and second, models (1) to (5) do not represent a significant difference compared with the "complete" specification, reported as model (6). The second and third columns on the left of the table are critical parameters in constructing the T statistic given in the fourth column from the left. The final two columns report the corresponding 5% and 1% significance tests. Strikingly, all the null hypotheses are rejected at the 5% level or higher. This finding suggests that the omission of any group of variables from the "complete" specification results in a significantly different distribution curves, and therefore provides important insights on the potential sources of the spatial variations in the parks amenity values.

My third sensitivity test is to examine the stability of spatial patterns of the estimated values of proximity to parks, estimated by using different empirical specifications. Figures 1.4(a–c) visualize the amenity values of proximity to parks based on the LWR results from the simplest, middle and the “full” model specifications (model (1), (3) and (6), respectively). It can be seen from Figure 1.4(a) that price surface varies smoothly over locations when only the park proximity variable is controlled. Price generally declines when moving from the western to the eastern urban regions, mainly due the fact that there are more green spaces distributed in the western city than other regions. The introduction of additional location-specific variables in model (3) had a pronounced effect on the estimated spatial variation patterns, as indicated in Figure 1.4(b). Here the marginal price estimates are based on a model that includes land structural attributes, local amenity measures, and demographic variables. Although the price surface is not tidily shaped, a generally “mono-centric” variation pattern emerges with the high-value areas concentrated in the central city. Nevertheless, a more subtle and complex change is evident when moving to the “complete” model specification (model 6). As shown in Figure 1.4(c), the effects of the inclusion of interaction terms between proximity to parks and relevant demographic variables are reflected significantly in the variations of the amenity values of park proximities.

6 Conclusion

In this paper I use the hedonic analysis of residential land parcel data from Beijing to estimate the proximity effect of parks on land prices. This study builds on previous studies that have examined urban amenities in the land markets of transitional socialist countries (Bertaud and Renaud, 1997). Importantly, I allow the proximity effects to depend on local demographics and other location complementarities expected to influence the amenity values of parks. In addition, my estimation accounts for the robustness of the estimated parameters of proximity to parks, and thus shed more light on the potential sources of spatial variations in the amenity values.

The empirical results yield two important insights. First, the effect of proximity to parks on residential land values largely depends on a parcel's location and local socio-demographics. For example, the value of proximity to a park of a given size is found to be higher in areas with lower population density and more educated residents. The positive signs associated with other amenity proximity measures show complementary effects between proximity to parks and other public goods such as schools and subway stations. There are fewer such benefits in areas with greater crime rates and a larger proportion of heritage buildings. The point here is that the amenity value, which is being capitalized, varies according to other conditioning characteristics, and, thus, a park on which coal dust always falls is not "the same as" a park with a clean environment beside a beautiful river or lake.

Second, my results highlight the capacity of the LWR model in explaining the differentials of the estimated values of proximity to parks individually. But this does not mean that the LWR model can be carelessly applied without robustness checks. By comparing the empirical specifications with and without certain known local amenities and their complementarities, my results suggest that the estimated LWR parameters of proximity to parks still reveal a significant underlying problem with omitted variables. More evidence is needed to explore how informative or robustness of the amenity capitalization effects based on these flexible statistical regressions (Gibbons and Overman, 2012). Nevertheless, at the minimum, my results shed light on the importance of considering the spatial locations in explaining the amenity value differentials that is grounded in the social, economic and other local contextual forces at stake. This finding might not be a surprising innovation; however, in applied spatial economics, unlike in theoretical work, it is particularly gratifying to identify, model, visualize, and assess the robustness of spatial variation in amenity values.

In considering the proximity impacts brought about by local public goods, it is important to emphasize that I have only examined the ways in which residential land markets capitalize the amenity values of proximity to parks. This applied work, however, is subject to several limitations. One underlying concern is that this paper does not provide a unified framework for capturing the spatial sorting effect of household preferences about local amenities. As such the price premium of green space might be overestimated. As suggested by recent literature (Wu et al., 2013), there is clear evidence that residential sorting is going on in Beijing---richer people or those who have Beijing *hukou* with higher preference for positive amenities sort around those high-quality locations. Future work using detailed household surveys in China to collaborate my results are useful. Another issue is also related to the data limitation. This present study is based upon land price data from a certain fixed time period and does not examine the effects of land supply changes or local public goods improvement on the real estate market over time. Here it is noteworthy that one nice aspect of the paper is that it uses vacant land price data in the analysis rather than house prices. However, it would be more interesting to know how does using the housing price data influence the results as opposed to using land price data? Despite these limitations, this study is still useful as it is the first attempt to empirically measure spatial variations in the "greenness" park amenity values among transitional socialist nations, where reliable micro-geographical data are difficult to obtain.

Table list

Table 1.1 Variable name, definition, and descriptive statistics

Variables	Definition	Mean(Std.Dev)
Dependent variable		
PRICE	Residential land parcel price per square meter (CNY/sq.meter)	3286.527(5478.112)
Park variables		
PARK	Distance to the nearest park (meters)	3015.723(2017.358)
PARK AREA_2KM	Summed park area within a 2km radius of a residential land parcel (km^2)	0.252(0.502)
Dummy_PARK	Dummy variable for a park size larger than above 0.5 km^2 within a 2km radius of a residential land parcel	0.17(0.376)
PARK SIZE	The size of the nearest park (km^2)	0.636(0.819)
Structural variables		
PARCEL AREA	The size of a land parcel (m^2)	34504.5(49015.72)
COMMERCIAL	Median price of commercial-use land parcels within 2km radius of a residential land parcel (CNY/sq.meter)	2636.615(1675.821)
Location and demographic variables		
CBD	Distance between a residential land parcel and the CBD (meters)	9409.662(5111.068)
SUBWAY	Distance to the nearest subway station (meters)	2187.467(2097.151)
RIVER	Distance to the nearest river (meters)	2578.607(1639.604)
SCHOOL	Distance to the nearest middle school* the school rank	74.061(72.211)
POPULATION	Population density in each zone (thousand people/km^2)	1.81(2.514)
HERITAGE	Ratio of heritage architectures built before 1949 in each zone (%)	0.052(0.125)
EDUCATION	Education median in each zone:1=junior or lower; 2=high school;3=university;4=post graduate	1.715(0.508)
CRIME	Number of reported serious crimes per 1000 people in each zone	5.335(6.655)
Year Dummies		
YEAR2005	Dummy: Residential land parcels auctioned in 2005	0.077(0.267)
YEAR2006	Dummy: Residential land parcels auctioned in 2006	0.126(0.332)
YEAR2007	Dummy: Residential land parcels auctioned in 2007	0.098(0.297)
YEAR2008	Dummy: Residential land parcels auctioned in 2008	0.077(0.267)

Table 1.2 Estimates of the OLS and LWR models [dependent variable =ln(*PRICE*)]

	OLS Model		LWR Model				
Variables	Coefficient	Std. Error	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Intercept	4.5638	3.4798	-28.883	-2.1683	4.0046	19.895	53.645
Ln(PARK)	-0.6766**	0.2809	-2.6012	-0.9650	-0.8549	-0.5564	0.8817
Ln(PARK AREA_2KM)	-0.1423	0.1287	-0.8421	-0.1844	-0.0690	0.0707	0.4133
Dummy_ PARK	-0.3703**	0.1718	-1.2525	-0.6346	-0.3954	-0.0842	0.2665
PARKSIZE	-0.1281**	0.5702	-0.1850	-0.0212	-0.1369	0.0742	0.1107
Ln(CBD)	-0.6352***	0.2218	-1.7228	-0.8448	-0.7095	-0.4687	0.7061
Ln(SUBWAY)	-0.1493***	0.0478	-0.3519	-0.2215	-0.1788	-0.1406	0.3059
Ln(PARCEL AREA)	0.0386*	0.0212	-0.1840	-0.0212	0.0413	0.0723	0.1107
Ln(COMMERCIAL PRICE)	0.1844*	0.0953	-0.4166	0.0742	0.2613	0.3674	0.8699
Ln(RIVER)	0.1099***	0.037	-0.1644	0.0899	0.1173	0.1677	0.3184
Ln(SCHOOL)	-0.1311**	0.053	0.0024	0.1171	0.1570	0.1806	0.2983
POPULATION	-1.2683***	0.4287	-1.7361	-1.5053	-1.1448	-0.7278	2.1751
HERITAGE	1.0244***	0.3225	-1.2655	-1.1156	-0.9985	-0.9106	0.1721
EDUCATION	6.9121***	2.009	-7.1546	1.5780	3.0206	7.1917	15.528
CRIME	-1.7877***	0.5481	-4.2316	-3.2658	-2.0449	-1.3094	0.0818
PARKSIZE*Ln(PARK)	-0.0189**	0.0076	-0.0411	-0.0384	-0.0259	-0.0055	0.0835
POPULATION*Ln(PARK)	0.1710***	0.0548	-0.1792	0.0996	0.1561	0.2074	0.2924
HERITAGE*Ln(PARK)	-0.1294***	0.0449	0.0354	0.1164	0.1289	0.1389	0.1752
EDUCATION*Ln(PARK)	-0.8735***	0.2495	-1.8854	-0.8565	-0.3234	-0.1443	1.2549
CRIME*Ln(PARK)	0.2233***	0.0678	0.0510	0.1611	0.2438	0.3936	0.5634
Ln(SCHOOL)*Ln(PARK)	0.0169*	0.0096	-0.0076	0.0052	0.0163	0.021	0.0685
Ln(SUBWAY)*Ln(PARK)	0.0297***	0.0095	-0.0051	0.0109	0.0183	0.0281	0.0638
Year dummies	Yes				Yes		
Parcel location coordinates	Yes				-		
Number of observations	615				615		
Adjusted R Square	0.3965				0.7163		

Notes.---*** p<0.01, ** p<0.05, * p<0.1. See Table 1.1 for variable definitions and descriptive statistics. See text for details.

Table 1.3 Mean park value using LWR model

Park Name	Mean marginal effect	Mean residential land price (CNY per sq.meter)	Mean park value (CNY)	N
Diaosu Park	-0.3863	4289.6457	549.4835	17
Xiwang Park	-0.4096	3999.7626	543.2537	11
Minzu Park	-0.2280	7153.2502	540.8126	36
Chaoyang Park	-0.3110	4169.9331	430.0293	89
Yingshan Park	-0.6250	2062.9424	427.5389	2
Shijingshan Park	-0.3024	3664.7698	367.4828	21
Honglingjin Park	-0.3095	3337.0764	342.4801	95
Yudadu Park	-0.1457	5530.4219	267.1938	43
Children Park	-0.2080	3838.5459	264.7516	42
Tuanjiehu Park	-0.1158	5178.9022	198.8634	93
Yuyuantan Park	-0.1118	4157.7372	154.1372	38
Zizhuyuan Park	-0.1083	3579.8380	128.5584	37
Daguanyuan Park	-0.0875	2801.8638	81.2021	55
Lianhuachi Park	-0.0319	3220.4122	34.0652	55
Badachu Park	-0.1457	668.4662	32.2959	3
Yiheyuan Park	-0.0755	1257.9124	31.4924	15
Longtanhu Park	-0.0161	3804.1035	20.3089	61
Yuanmingyuan Park	-0.0145	2987.7417	14.3655	18
Taorantign Park	0.0209	2605.8453	-18.0594	57
Botany institute Park	0.1532	413.9599	-21.0293	5
Wanshou Park	0.0212	3171.1705	-22.2928	63
Xiangshan Park	0.1984	396.6667	-26.0961	3
Youle Park	0.0269	3304.5400	-29.4762	62
Animal Park	0.0198	5742.1188	-37.7004	29
Yuantan Park	0.0630	4173.4861	-87.1863	60
Tiantan Park	0.0752	3578.4309	-89.2317	63
World Park	0.2399	1137.5841	-90.4945	15
Ritan Park	0.0761	4306.1544	-108.6633	103
Botany Park	0.0528	6405.1009	-112.1420	8
Wofosi Park	0.0528	6405.1009	-112.1420	8
Shuangxiu Park	0.2125	3774.9135	-265.9956	38
Liuyinhu Park	0.2416	3733.0476	-299.0674	73
Ditan Park	0.2332	3937.7809	-304.5009	82
Qingnianhua Park	0.2898	3479.2240	-334.3408	69
Dinghu Park	0.3770	2964.0431	-370.5394	55
Biyun Park	0.1494	8819.2754	-436.9101	4
Zhongshan Park	0.4130	3268.9999	-447.6860	87
Renmin Park	0.3791	3589.1283	-451.1815	93
Gugong Park	0.4422	3714.1927	-544.6177	91
Beihai Park	0.4764	3616.5212	-571.3093	82
Jingshan Park	0.4413	4146.6799	-606.7964	91

Notes.---The mean park value is the marginal implicit price for reducing the distance to the nearest park by 1,000 meters, evaluated at the mean residential land value and mean distance to parks. See text for details.

Table 1.4 LWR estimates across different model specifications [dependent variable = *ln(PRICE)*]

	Model (1)		Model (2)		Model (3)		Model (4)		Model (5)		Model (6)	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Variables	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)	(*β*)
Constant	10.921	2.186	-4.902	10.920	2.183	33.985	-0.1644	36.491	4.9060	34.625	4.0046	22.063
Ln(PARK)	-0.4196	0.2852	-0.3183	0.2403	-0.2353	0.1454	-0.1651	0.1564	-0.6121	2.3844	-0.8549	0.4086
Ln(PARK AREA_2KM)			0.0786	0.5221	0.0312	0.4068	0.0027	0.4252	-0.0410	0.332	-0.0690	0.2551
Dummy_PARK			-0.4586	0.8801	-0.3528	0.8169	-0.3967	0.7970	-0.4210	0.6565	-0.3954	0.5504
PARKSIZE			-0.0869	0.2064	-0.1292	0.1782	-1.0730	3.3533	-1.0790	2.3978	-0.1369	0.0954
Ln(CBD)			0.0111	0.3655	-0.2509	0.3194	-0.2412	0.3022	-0.2201	0.2892	-0.7095	0.3761
Ln(SUBWAY)			-0.2353	0.1627	-0.1892	0.1036	-0.1841	0.1037	-0.087	1.2952	-0.1788	0.0809
Ln(PARCEL AREA)			0.0201	0.1087	0.0164	0.0903	0.0173	0.0889	0.0190	0.0698	0.0413	0.0935
Ln(COMMERCIAL)			0.4008	0.5963	0.2668	0.4098	0.2994	0.3918	0.3398	0.3694	0.2613	0.2932
Ln(RIVER)			0.0705	0.1281	0.0931	0.118	0.1053	0.1230	-0.5463	1.2711	0.1173	0.0778
Ln(SCHOOL)			0.0917	0.1046	0.0917	0.1046	0.1256	0.0809	0.0945	1.3028	0.1570	0.0635
POPULATION					-0.0713	0.0643	-0.0843	0.0787	-0.0763	0.0823	-1.1448	0.7775
HERITAGE					-0.0637	0.0715	-0.0628	0.0652	-0.0570	0.0513	-0.9985	0.2050
EDUCATION					0.2326	0.4376	0.2072	0.3952	0.2006	0.4494	3.0206	5.6137
CRIME					-0.1293	0.1934	-0.1555	0.2002	-0.1598	0.2057	-2.0449	1.9564
PARKSIZE*Ln(PARK)							-0.1489	0.4629	-0.1581	0.3212	-0.0259	0.0329
Ln(SUBWAY)*Ln(PARK)									0.1062	0.1638	0.0183	0.0172
Ln(SCHOOL)*Ln(PARK)									0.0062	0.1807	0.0163	0.0158
POPULATION*Ln(PARK)											0.1561	0.1078
HERITAGE *Ln(PARK)											0.1289	0.0225
EDUCATION*Ln(PARK)											-0.3234	0.7122
CRIME*Ln(PARK)											0.2438	0.2325
Year Dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Number of observations	615		615		615		615		615		615	
Adjusted R Square	0.5937		0.6621		0.6682		0.6721		0.6933		0.7163	

Table 1.5 Pearson correlations of the estimated park proximity parameters

	model(1)	model(2)	model(3)	model(4)	model(5)	model(6)
model(1)	1					
model(2)	0.5678 (0.000)	1				
model(3)	0.5227 (0.000)	0.8175 (0.000)	1			
model(4)	0.5152 (0.000)	0.6157 (0.000)	0.6446 (0.000)	1		
model(5)	0.5036 (0.000)	0.5852 (0.000)	0.5648 (0.000)	0.6933 (0.000)	1	
model(6)	0.5081 (0.000)	0.5108 (0.000)	0.5278 (0.000)	0.5399 (0.000)	0.6883 (0.000)	1

Notes.---Beneath the parameter coefficient is the *P*-value for the parameter in parentheses.

Table 1.6 Spearman rank correlations of the estimated park proximity parameters

	model(1)	model(2)	model(3)	model(4)	model(5)	model(6)
model(1)	1					
model(2)	0.6209 (0.000)	1				
model(3)	0.5258 (0.000)	0.8216 (0.000)	1			
model(4)	0.5183 (0.000)	0.6071 (0.000)	0.7656 (0.000)	1		
model(5)	0.5437 (0.000)	0.5808 (0.000)	0.5883 (0.000)	0.7527 (0.000)	1	
model(6)	0.5218 (0.000)	0.5699 (0.000)	0.5546 (0.000)	0.5650 (0.000)	0.6963 (0.000)	1

Notes.---Beneath the parameter coefficient is the *P*-value for the parameter in parentheses.

Table 1.7 Park elasticity distribution hypothesis tests

Null hypothesis (H_0)	***I***	σ^2	T-test statistics	5-Percent significance level	1-Percent significance level
f(model(1))=f(model(2))	450.66	1492.0	15.53	H_0 rejected	H_0 rejected
f(model(2))=f(model(3))	70.65	2552.4	1.86	H_0 rejected	H_0 *not* rejected
f(model(3))=f(model(4))	198.66	1627.3	6.55	H_0 rejected	H_0 rejected
f(model(4))=f(model(5))	226.59	1315.4	8.31	H_0 rejected	H_0 rejected
f(model(5))=f(model(6))	66.00	1382.4	2.42	H_0 rejected	H_0 rejected
f(model(1))=f(model(6))	51.35	1182.1	1.99	H_0 rejected	H_0 *not* rejected
f(model(2))=f(model(6))	469.27	1458.1	16.36	H_0 rejected	H_0 *not* rejected
f(model(3))=f(model(6))	392.38	1203.9	15.05	H_0 rejected	H_0 rejected
f(model(4))=f(model(6))	275.05	1284.8	10.21	H_0 rejected	H_0 rejected

Note.--- Table presents the results of a statistic test to examine the robustness of the park elasticity distribution curves across different model specifications shown in Figure 1.3. If the null hypothesis is rejected, it suggests that the distribution curve in the subsequent model specification is statistically different from the former model specification, and therefore shed more lights on the potential sources of spatial variations in park amenity values.

Figure list

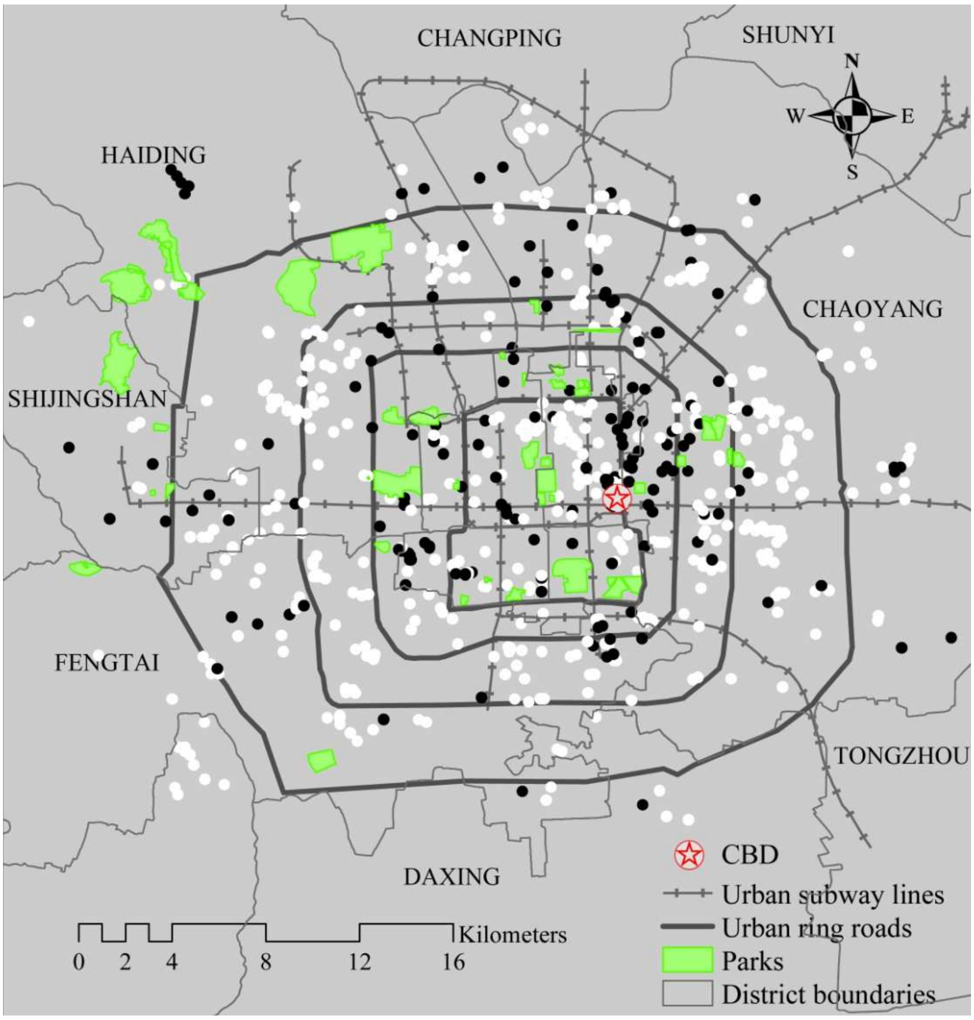


Figure 1.1 Spatial distributions of residential land parcels in Beijing

Notes.--- This figure show the spatial locations of residential land parcel sample by using circle dots. The black/white circle dots represent the prices of residential land parcels that are larger/smaller than the sample mean value, respectively.

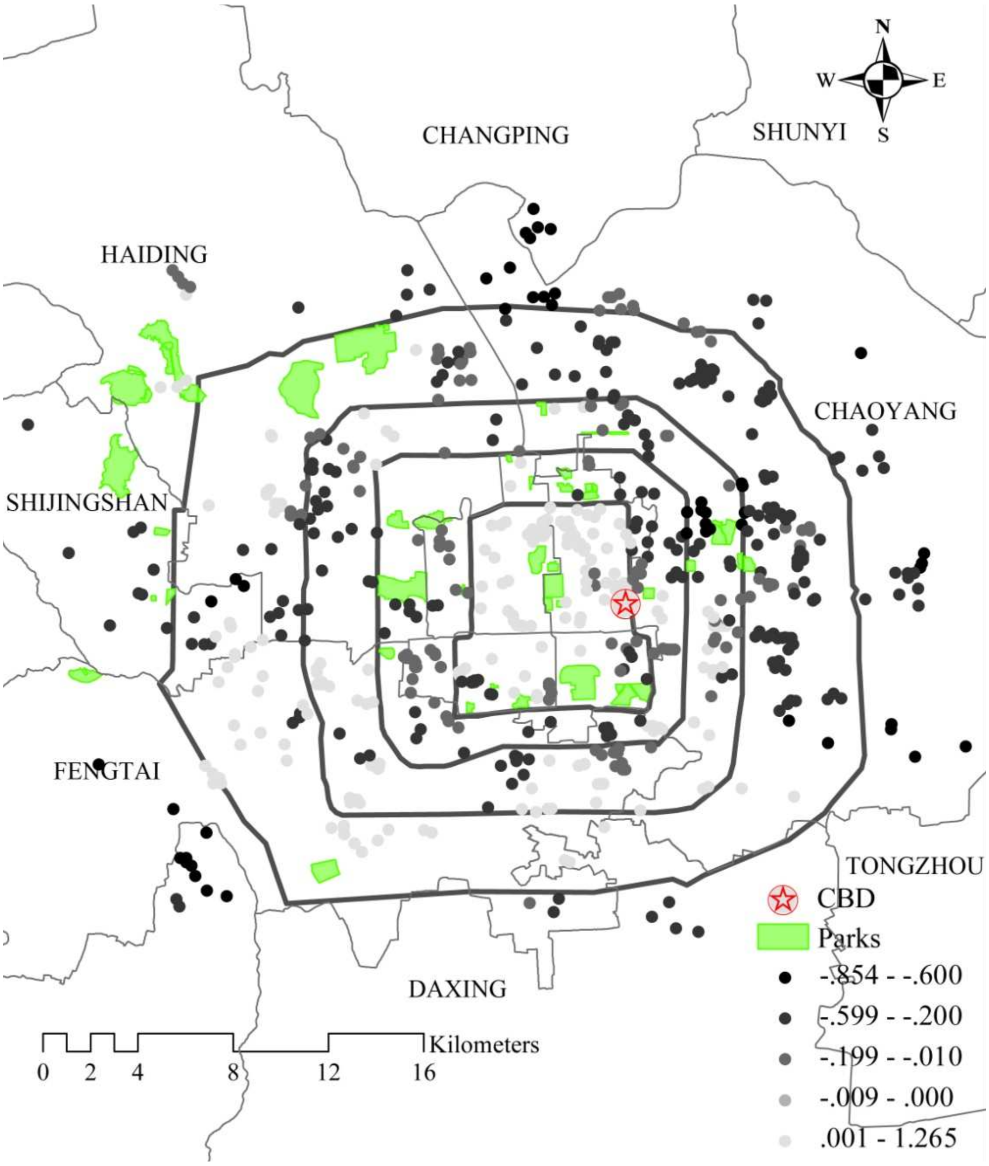


Figure 1.2 Spatial distributions of marginal effects of proximity to parks on land prices

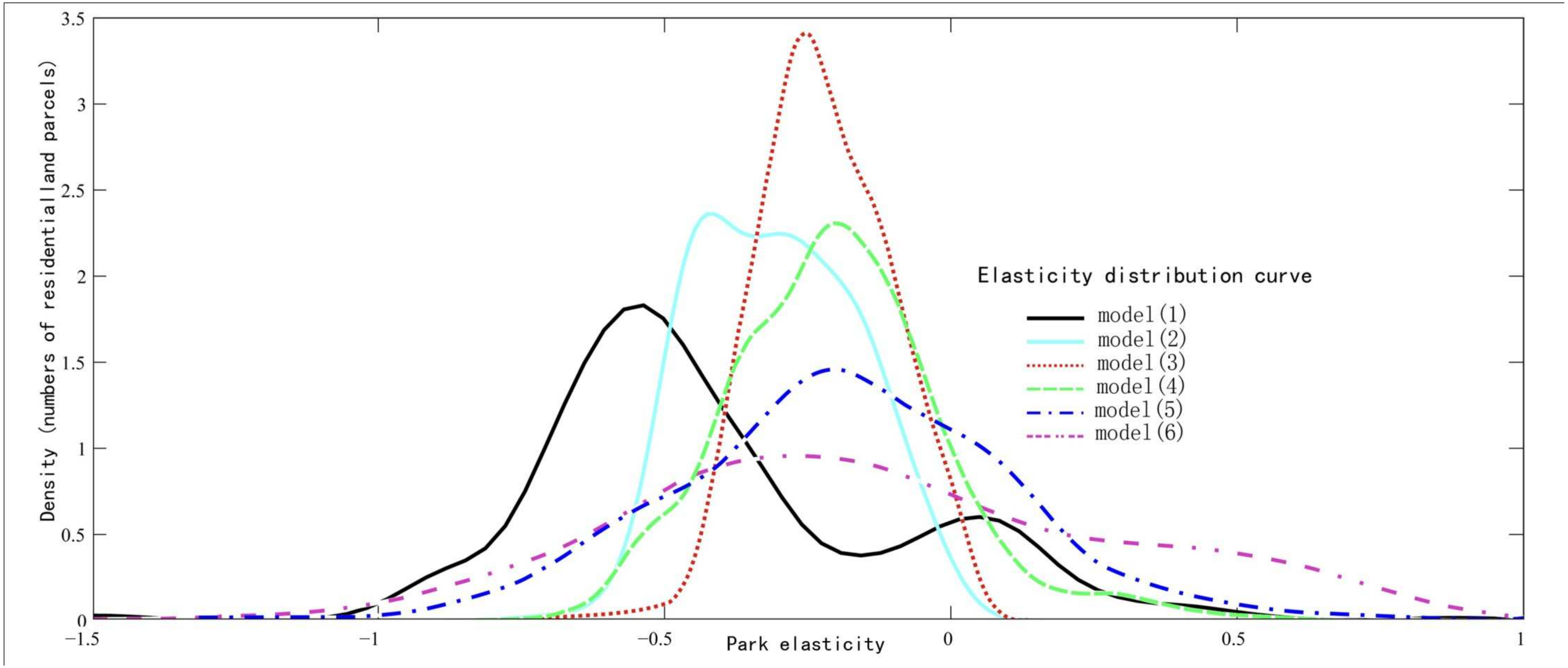
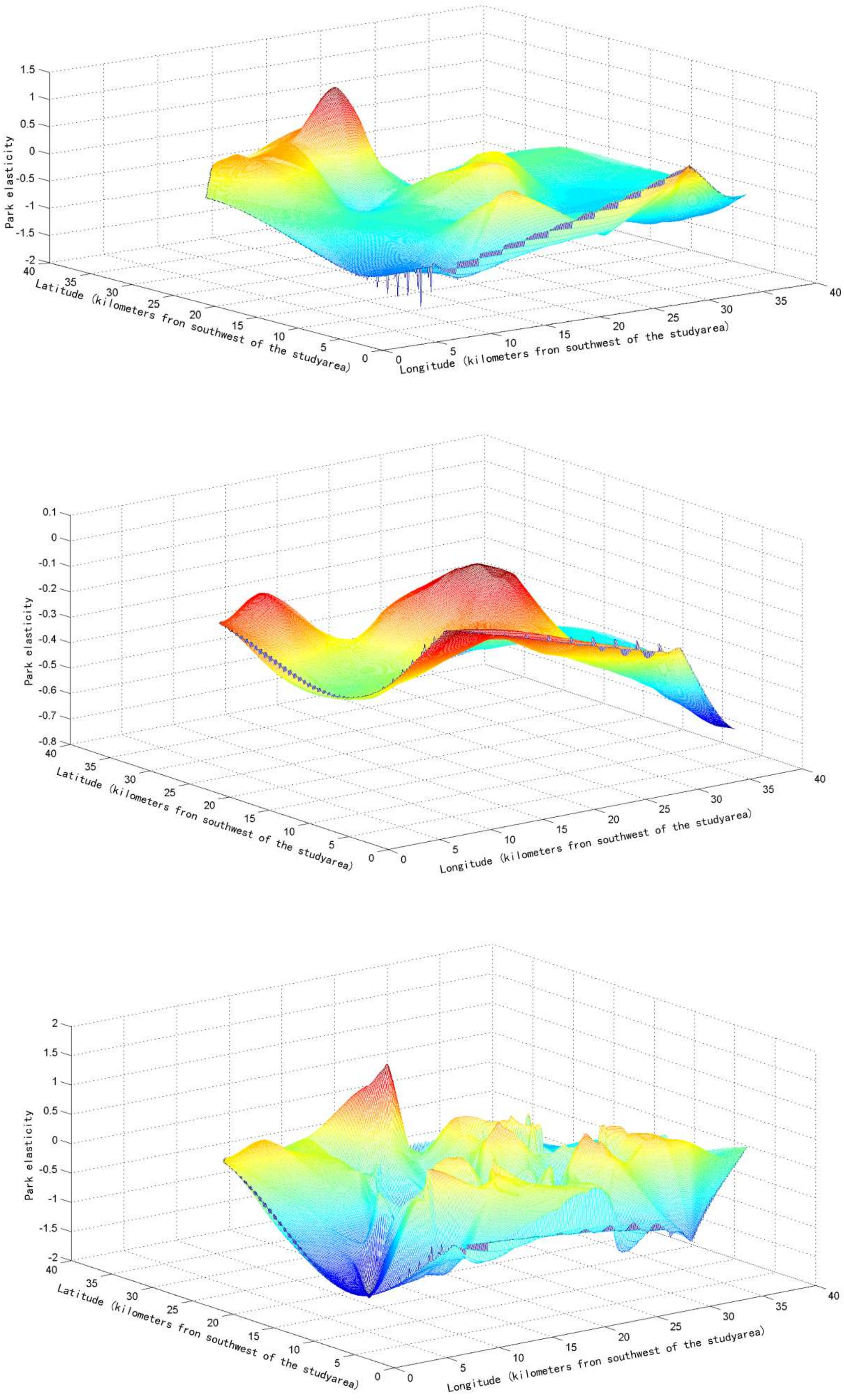


Figure 1.3 Distribution of elasticity effect

Notes.---Distributions are estimated using a non-parametric kernel density estimator.



(a)

(b)

(c)

Figure 1.4 Spatial variations of marginal effects of proximity to parks

Notes.--- (a) model 1; (b) model 3; and (c) model 6

Appendix A.

In order to quantify how much the LWR model perform better than OLS model in terms of modelling fit, an ANOVA test is carried out (Brunsdon et al., 1999). The results of ANOVA test are shown in the table below.

In this table, the first column presents the residual sum of squares of the OLS model (466.22), the LWR model (273.07) and the improvement of LWR model (193.15). The second column gives the corresponding degree of freedom. Mean square gives the results of dividing the sums of squares by their respective degree of freedom. Then, dividing the mean square of LWR model by that of LWR model improvement gives the pseudo-F statistic. Significance of the statistic at the 1% level suggests that the null hypothesis OLS model should be rejected in favour of LWR model.

Appendix Table 1.1 ANOVA test of LWR and OLS results in terms of modelling fit

Source of Variation	Sum Square	Degree of Freedom	Mean Square	F	p-value
OLS Model Residuals	466.22	30			
LWR Model Improvement	193.15	110.55	1.75711		
LWR Model Residuals	273.07	494.45	0.55228	3.1635	0.000

III. Paper 2---Does Public Investment Spur the Land Market?: Evidence from Transport Improvement in Beijing

1 Introduction

Since the late 1990s, the explosive growth of public transport investment has been reshaping the face of most Chinese cities. Between 2000 and 2008, the Beijing government invested about 52 billion CNY[28] on the new rail transit constructions, with a subsequent investment of 105 billion CNY by 2012. This massive investment allows me to examine the consequences of the transport improvements for the price of nearby land parcels.

In this paper I examine how residential and commercial land prices respond to the changes in the parcel-station distance proximities. My purpose is threefold. First, my examination contributes to the small but growing body of literature on valuing rail access based on the difference-in-difference methodology (Gibbons and Machin, 2005 Kahn, 2007; Ahlfeldt, 2011). At its heart it captures the changing nature of geographical links between properties and stations as a result of transport expansion. This study improves on the previous methods by providing a large scale multiple intervention difference-in-difference design that explicitly exploits changes in the parcel-station distances that happen when new stations are opened; it also highlights the importance of price changes in planned station areas. My study is also unique in using vacant land parcel data during 1999 and 2009 in the entire urbanized area of Beijing, rather than pre-designed sample areas. To my knowledge, this is the first attempt to evaluate the impact of transport improvement by applying this type of analysis in China. My results, which focus on vacant land prices, document the

[28] The official exchange rate is around 10 CNY per GBP.

appreciable economic benefits caused by the increased station proximity with the opening and planning of new railway lines.

Second, though frequently discussed in lectures, there have been surprisingly few detailed studies that examined the comparative impacts of rail access on both commercial and residential land prices (Debrezion et al., 2007). I employ the geographical information system (GIS) software to derive proximity measures from the Beijing residential and commercial land use dataset. I define the treatment as parcels that experience the station-distance reductions; and that the outcome distances to the closest station are now less than a certain distance band[29] due to the new rail transit constructions. Such multiple distance-band design allows me to explore the heterogeneous distance decay trends associated with the station-proximity impacts on residential and commercial land prices. Importantly, I also allow the proximity effect of rail stations to depend on employment accessibility, crime rates, educational attainment that believed to influence the value of transport improvement (Bowes and Ihlanfeldt, 2001; Gibbons, 2004; Gibbons and Machin, 2008). Additionally, I control for unobserved spatial characteristics with the local fixed effect. My evidence on new rail transit's effects on land prices, suggests that residential and commercial land developers do value the increased station proximity and these valuation varies widely with local demographics over space. Using the entire urbanized area's average effects might therefore mask the value of proximity to stations in particular spatial location by

---

[29] I use the multiple distance bands (0.5km, 1km, 2km, and 4km) to define the treated parcels in order to exploit which distance band has the most significant impact on local prices. See detailed explanation of treatment groups in Section 3.3.

a substantial margin. For example, the value of proximity to new stations falls as crime rates increases and rises with employment accessibility and local residents' median educational attainment level. With this contribution, I aim to fill an existing gap in the knowledge of missed impacts in the previous empirical analysis and keep my methodology as simple as possible for further applications.

Finally, beyond an obvious academic interest, the question of whether rail transit improvement has a substantial affect on land values has tremendous policy implications: showing complementary effects between public investment and private sector investment. Within the "new urbanism" process, transport-oriented development strategies were designed to gentrify previously depressed areas and reduce congestion in central business and residential districts (Knaap et al., 2001). Classic examples of this include Boston's Big Dig, Chicago's Midway line, Los Angeles's Bay Area subway line, Toronto's Spadina Subway line and London's Jubilee and DLR lines. Given the huge expenditures of transport infrastructure, empirical answers are scarce on whether public investments and private investments are complements that spur the emerging land markets of the BRICS countries. My findings offer a limited support for this by demonstrating that the same "game" plays out in Beijing, where public transport investment has been shown to stimulate spatially targeted residential and commercial land markets[30].

[30] Note that a related body of literature has focused on conducting the cost-benefit analysis of the transport improvements (see Gunn, 2000 for a recent review). Doing this would offer more useful policy implication but collection of such micro commuting data with precise geographical information during 1999-2009 would be very costly. Some studies have shown that traditional cost-benefit appraisal methods, based on travel time savings and other direct cost reductions, can significantly underestimate the actual benefits of transport improvements (Brocker,

This paper proceeds as follows: Section 2 reviews the literature on the rail access effect on property values. Section 3 describes the institutional settings and data. Section 4 presents the econometric models. Section 5 reports the estimation results. Section 6 concludes.

2 Literature review

The literature relevant to my analysis includes hedonic studies that have documented the rail access effects on property prices by using different types of property sample and by using different empirical methodologies (see excellent recent examples in Table 2.1). Findings from each of these dimensions are briefly summarized in this section.

First, existing studies on valuing rail access can be grouped into two broad empirical methodology types. The first approach is a straightforward cross-sectional analysis, in which property price is regressed on accessibility to stations at one specific time whilst controlling for other attributes. Over the past 30 years, a large number of studies have contributed to improving the model specification and the ways in which the values of transport access are capitalised into land values. Recent good examples at least includes Grass (1992), Cheshire and Sheppard (1995), Vessali (1996), Coffman and Gregson (1998), Bowes and Ihlanfeldt (2001), and Debrezion et al (2011). RICS

1998). Whilst recent appraisal approaches have incorporated a wider range of socio-demographic factors into the computable models, there is limited evidence on valuing the ex post effects of multiple transport improvements based on land price outcomes, particularly in China. In this study I am interested in the land price dynamics in treated places that have experienced effective station-distance reductions with the building of new rail transit expansions. My valuation does not attempt to account for the impact of financial and economic climate changes on the real estate market (Deng et al, 2005; Deng and Liu, 2009).

(2002) conducted a detailed review of more than 150 empirical studies on the relationship of land values and public transport in the North American cities, and found largely support for the positive impact of transport access on land values. However, there are some problems associated with this approach. One relates to the omitted variable issue; admittedly, no matter how many control variables can be included in the regression, there are still unobserved characteristics that might be correlated with transport access and land values. Since the affected station areas are relatively small, failure to account for correlated local contextual effects separately would bias the estimated value of the proximity to stations. The second problem is that it cannot take into account the changing nature of rail access, especially when new stations are built. Unlike permanent green spaces and park amenities, the state and local governments have continued to make investments in building public transport infrastructure especially in the developing countries. These new rail transit lines have fundamentally reshaped the evolution of the urban transport network over time and have changed the closest distance from stations to land parcels whilst leaving others unaffected. Thus the estimating results should conceal significant variation in transport access and economic outcomes over time.

Alternatively, the difference-in-difference approach, moved on to use cross-sectional time series data to look at the changes in land values before and after a new rail transit line is in service. By comparing the distance changes in rail access over time, this approach can mitigate most of the problems linked with the cross-sectional applications. Most existing studies, employing before-and-after comparisons, have focused on examining the property price effects of new rail transit lines in North American cities[31]. See, for example, Davis (1970) on the San Fancisco Bay Area subway line, Bajic (1983) on new subway lines in Toronto or McMillen and McDonald (2004) on Chicago Midway Rapid Transit Line. Recent studies, though less common, have exploited changes in the distances between properties and stations as a result of new stations opened and estimated such impacts on property prices (Gibbons and Machin, 2005; Kahn, 2007; Ahlfeldt, 2011). For example, Gibbons and Machin (2005) developed a precise framework for capturing the changes in distances between houses and tube stations in London when the Jubilee line and Docklands Light Railway (DLR) opened in the late 1990s. They highlighted the fact that difference-in-difference regression estimates can better avoid the biases inherent in pure cross-sectional empirical studies. Following Gibbon's and Machin (2005)'s study, Ahlfeldt (2011) re-examined the property price effects of transport network extensions in London using extended housing price data. In particular, he emphasized the importance of employment accessibility on the adjustment of property prices. Kahn (2007) documented the significant heterogeneity in the effects of rail transit expansions across the 14 large US cities. He found that the average housing prices of communities that experienced increased proximity to new stations rise significantly compared to communities that have never experienced improved access to stations. My methods are closest to this approach type, but I improve on previous methods by considering explicit changes in the distances between parcels and stations that occur when new

[31] These studies, however, often uses no control group, only examine how prices respond to travel times, before and after a new subway construction.

stations are opened and planning to opening as a result of urban transport improvement.

Another literature dimension lies in the different types of real property markets. While a large number of studies have focused on examining the residential property market, there have been few studies combined both residential and commercial properties in empirical analysis. Some empirical studies have shown that the affected areas of the rail access effects are larger for residential properties, whereas the effect of proximity to rail stations on commercial properties is concentrated at nearby areas (Cervero and Duncan, 2001; Debrezion et al., 2007). This finding is consistent with the prior expectation that station areas---by gathering large amount of population flow, will attract commercial establishments and thus have greater price premiums for commercial properties at a closer distance range.

By zooming into the urban China literature, it is not easy to find empirical studies on valuing rail access despite the rapid transport infrastructure changes. Research on this issue has been limited by the lack of systemic micro-level land parcel data and related local socio-demographics data. Recent excellent works, however, include Zheng and Kahn (2008), Wang (2009), and Wu et al (2011), among others. For example, Zheng and Kahn (2008) reported the significant impact of the established subway stations access on land and housing prices in Beijing. However, existing empirical studies in China have only focused on the residential property market; nothing is known about the commercial land market. In addition, they don’t capture the increased station access effects as a result of the transport improvement. A further problem is that these studies do not account for the interaction effects between the station access and local socio-demographic characteristics. Thus their resulting estimates are likely to be biased. It is likely, for example, that the net benefits derived from proximity to new stations would decrease when located in high crime rate areas. Empirically, the implications of empirical studies are often difficult to compare because of the heterogeneous local contextual characteristics through which the new transit's impact is thought to operate. In this study I assume that the impact of increased station access on land prices occurred only when parcel-station distance changes due to the transport improvement. The next sections spell out the detailed data and econometric models.

3 Data and Institutional Settings

The focus of this section lies on introducing the land development and transport infrastructure supply within a unique transitional economy context.

To better understand this, it is necessary to provide a brief introduction to the centralized urban governance structure in Beijing. The Beijing municipal administrative system has three levels: Beijing municipality, district and zone (*jiedao*, it will be referred to as zone thereafter in this study). While the Beijing metropolitan area consists of eighteen districts, this study mainly focuses on the eight urbanized districts (*Dongcheng, Xicheng, Xunwu, Chongwen, Chaoyang, Fengtai, Shijingshan, and Haidian*) because the other districts are predominately rural areas. There are five "ring roads (Nos. 2–6)" circled around the central business district (CBD) from the central city to the suburbs (see circle lines in Figure 2.1 below). The Beijing urbanized area is mostly within the No. 5 ring road. Existing empirical studies have shown that the Beijing urbanized area is still quite mono-centric with respect to the spatial distribution of population density, as well as land and housing prices (Zheng and Kahn, 2008). Within the Beijing's urbanized area, *jiedaos* (zones) exist as the fundamental administrative organization and census unit. However, unlike in the US, land supply and public infrastructure construction are highly centralized and controlled by the Beijing municipal government. The zones (*jiedaos*) are only responsible for street cleaning and do not have control over public infrastructure construction and service provision.

This section is divided into three parts. The first part introduces the land parcel data and related micro-geographical data. The second part discusses the transport infrastructure improvement in Beijing. The third part explains the characteristics of "treated" and "control" places.

3.1 Data

The Chinese urban land market is a booming market with vigorous reforms and rapid growth over the past twenty years[32]. Since the 1978 Reform-and-Opening-up policy in China, tremendous changes had happened in this "magic" economy, from a central-planned economy towards a market-oriented economy. Within this context, a land market was reborn in the recent two decades. In 1988, the Chinese Constitution---which had prohibited land transfers before, was amended to permit land

[32] See Wu et al (2011) for a recent evaluation of major Chinese cities' land market.

leasing rights (70 years for residential land use and 40 years for commercial land use) while retaining land ownership. In 1990, the State Council formally affirmed such dramatic transformation of the land use system from free allocation toward a leasehold system. By 1992, local governments in Beijing and Shanghai had begun to practice the land leasing policy, and it quickly spread to other cities in China.

In Beijing, the Municipal Land Resource Authority is responsible for the land allocations and sales of leasehold right, first through negotiation between developers and governments (during 1992 and 1998), then through partly negotiation and partly competitively open auction (during 1999 and 2003), and through the full competitively open auction way since 2004. See Zhu (2005) for more details on the Chinese land market reform policies. From the Beijing Municipal Land Resource Authority, I have collected all the vacant residential and commercial land parcels[33] during 1999 and 2009 within the study area. I have excluded uncompleted land transaction data and the land parcels that were obtained through negotiation because the strong institutional forces could reduce the market price effectiveness (Cai et al, 2009). The final sample size is 2343 and 1341 parcels[34] for residential and commercial land uses respectively.

In this study, the unit of analysis for the hedonic price regressions is a land parcel. Using the Geographical Information System (GIS) software, I have geocoded all the

---

[33] The land supply is exogenous with the public transport planning since it is made independently by Beijing Municipal Land Resource Authority.

[34] To mitigate the inflation effect, I have adjusted the land prices by using the CPI index reported by the Beijing Statistical Year Book 1999-2010. All monetary figures are constant in 2009 CNY *yuan*. Also, I have trimmed the land price distribution by only keeping parcels in each year whose price is between the 5th and 95th percentiles of the whole sample price distribution.

parcels. In order to measure transport infrastructure changes, I map the rail transit network before and after 2003 and calculated the distance from land parcels to the nearest station using the GIS techniques. To implement the transport improvement analysis, I group the parcel-level residential and commercial land parcel data into three time periods: before 2003 ($1999 \leq \text{year} < 2003$); during 2003 and 2008 ($2003 \leq \text{year} < 2008$); after 2008 ($\text{year} \geq 2008$).

Geographical information on other localized characteristics is taken from a variety of sources for the use of controllable variables in the regression models. The local public goods were built long ago in the central-planning economy and seldom change their locations after they are built. Thus, one advantage of using these local public goods as a set of controllable variables is that the location of public goods (such as schools, parks) is exogenously determined in Beijing. School location and quality comes from the Beijing Municipal Committee of Education. The location of bus stops and expressways are used as proxies for the competing commuting modes, and is obtained by a web-based search from the Beijing Municipal Committee of Transport. Additional GIS data on the sites of rivers, parks and green spaces is taken from the Beijing Water Authority and Beijing Municipal Garden Bureau respectively. Air quality is measured by the air pollution index (API) published the Beijing Municipal Environmental Protection Bureau[35]. Crime rates for the number of violent crimes

[35] The Beijing Municipal Environmental Protection Bureau reports daily API by different monitoring station. Instead of including the all-year round data, I only use the spring quarter data because it is the worst air quality season in Beijing. Thus it can reduce the overall noise for the potential impact of air quality on the land market. Following on the conventional way to create the appropriate metric, I assign the average API values of the daily maxima at the monitoring stations to the each parcel using the ordinary Kriging method (Anselin and Le Gallo,

taking place in each zone are obtained from the Beijing Public Security and Safety Bureau (BPSSB). The 2000 City Population Census reports the basic local socio-demographic characteristics such as the population density, resident median education attainment levels, public housing rent ratio, and the percentage of old housings built before 1949. The 2001 City Employment Census provides the necessary information for calculating the employment accessibility[36]. Table 2.2 summarises the descriptive statistics of variables.

3.2 Transport infrastructure improvement

To meet the rapid urbanization process and increasing commuting demand, the Beijing government has invested a huge amount of money into rail transit development during 2000 and 2012. The full set of new rail transit lines data is detailed in Table 2.3[37]. This table highlights that the constructions of rail transit lines differ with respect to their starting time[38] and completion date[39]. This table also provides differential

---

2006).

[36] I use the gravity model to calculate the employment accessibility with respect to each land parcel. The formula can be expressed as: $\mathrm{Employment\ Accessibility}_i = \sum_k \exp\left(-\delta * \mathrm{distance}_{ik}\right) * \mathrm{subcenter}_k$. Where δ is the distance decay parameter over geographical area. The parameter value that provides the best fit would eventually be selected (d=2 in this case). $\mathrm{subcenter}_k$ represents the total job number in the employment sub-center *k* in Beijing, which is identified by the pilot study of Ding et al (2010) based on the non-parametric methods proposed by McMillen (2001).

[37] The Beijing Municipal Committee of Transport's official website http://www.bjjtw.gov.cn/ contains informative details of subway lines in Beijing. This study does not include the subway lines to be completed after 2012, because of the large uncertainties involved with the proposed timetable. As a robustness check, I do test the anticipation effects for subway lines (Line 14 and 16) that had been announced but the exact completion time would be no early than 2015. The coefficients of these estimates are not reported. The insignificant estimating results confirm the prior expectation.

[38] I am unable to test the announcement effect separately because the announcement time of these lines is generally before my study period.

[39] It should be noted that Line 5 was temporarily opened at October 2007, but fully opened at the beginning of 2008.

figures of each line with respect to the construction cost, track length, and station numbers. Figure 2.1 shows the spatial patterns of the Beijing rail transit network before and after the completion of these new rail transit lines. Despite such differences, these new lines share several common characters: First, they are all intended to reduce congestion and meet the rapid growth of the commuting demand in the central city. For instance, a recent internal report by Beijing Municipal Commission of Urban Planning has clarified that subway line 6 and line 7 are constructed to handle the ridership growth of subway line 1 and the road congestion around the CBD areas. Second, they aim to strengthen the connections between the central city and suburbs from different spatial directions. Practically, the rail transit construction can be regarded as a fundamental policy lever to gentrify the less-desirable suburb areas. Therefore most of the new subway lines focus on linking the central city with suburb areas, especially places with emerging super-"bedroom" communities[40] (named as *Tiantongyuan*, *Yizhuang*, *Daxing*, and *Tongzhou*). To facilitate the 2008 Beijing Olympics, new transport infrastructure is also extended to the Olympic Park area. Given the importance of the political economy behind the placed-based investment on rail transit lines, there is a danger of mixing up the Olympics effect and other trends with the station proximity effect. Below, I will control the interactions of time trends with distance to CBD, distance to Olympic Park, and distance to those emerging "bedroom" communities (*Distance to New Residential Area*$_i$) that can affirm the robustness of the

[40] Note that the term of "bedroom communities" represents places where commuters perform most professional and personal activities in another location, maintaining their residence solely as a place to sleep. See http://en.wikipedia.org/wiki/Commuter_town for details.

increased station proximity effects on prices of nearby land parcels.

It is necessary to keep in mind that I use the opening of two lines in 2003, four lines in 2008, and eight planned lines opening after 2009 (to be completed before 2012) as the transport improvement programs. Ideally, I could single out the effects of each of these new lines and even go further by measuring each new station's effect individually[41]. Yet in reality, I simplify the estimation framework by treating them as three nested events (stations open after 2003, after 2008 and after 2009 respectively).

3.3 Balancing test for "treated" and "control" places

The main interest of this part is to answer two questions: what is the treatment?; and whether treatment groups and control groups are balanced in terms of observable pre-treatment demographic characteristics?

This study defines the treatment group by using two selection principles: Specifically, a residential or commercial land parcel will be assigned to a treatment group if: **Criteria 1:** It experienced the station-distance reductions with the stations opening after 2003; **Criteria 2:** And if the outcome distance to the closest station opening after 2003 is now less than 0.5km, 1km, 2km, 4km respectively, it will be assigned to the corresponding treatment group of $(0.5km\_station \geq 2003)$, $(1km\_station \geq 2003)$, $(2km\_station \geq 2003)$, $(4km\_station \geq 2003)$.

Accordingly, my control groups are parcels that have never been experienced distance reductions to the stations opening after 2003 and that the outcome distances to

[41] Below, I will test separately the effect of new simple stations, new cross stations, and new simple-to-cross stations (stations that were converted from simple stops into junctions with the building of new lines).

the closet stations are beyond the distance bands (0.5km, 1km, 2km, 4km respectively).

I impose the second criteria because I want to avoid the estimating noise from the parcels that became closer to a station, but still remain a long distance away from the new station[42]. Notably, the choice of a 2 km threshold is based on most existing empirical literature as well as a reasonable walking distance to a station (about 20 minutes).

Instead of using the fixed distance band such as 2km, this study is also unique by allowing the multiple distance bands (0.5km, 1km, 2km, 4km) to define the treated parcels. As suggested by Gibbons and Machin (2005), the ideal application of a difference-in-difference design would compare the treatment effects using alternative parcel-station distance bands. This comparison would hold everything the same in the model specification and any changes in land prices would be attributable to the difference in the selection of distance bands. As such, I am able to test the marginal effects of each distance band relative to the larger one.

Following the same principles, I further create the treatment groups of $(0.5km\_station \geq 2008/2009)$, $(1km\_station \geq 2008/2009)$, $(2km\_station \geq 2008/2009)$, $(4km\_station \geq 2008/2009)$ when a parcel has been experienced the station-distance reductions with the stations opening after 2008/2009; and the outcome distance to the closest station opening after 2008/2009 is now less than 0.5km, 1km, 2km, 4km respectively. Of necessity, the treatment groups of

[42] It is certainly true that land parcels located more than 2 km away from a new station might also benefit from the opening and planning of such a station. In this study I implicit assume that a 2-kilometer ball around the station is sufficient for defining the impact of rail access at station areas---not at remote places.

$(station \geq 2009)$ are nested within the corresponding treatment groups of $(station \geq 2008)$, and the treatment groups of $(station \geq 2008)$ are nested within the corresponding treatment group of $(station \geq 2003)$.

Figures 2.2 and 2.3 show the spatial distributions of treated residential and commercial land parcels respectively. From the GIS map it can be seen a clear spatial differentiation pattern among parcels in the treatment groups of $(station \geq 2009)$, $(station \geq 2008)$ and $(station \geq 2003)$, which gives some confidences that my results are not sensitive to the potential spillover effects within-treatment groups. Below, I will examine the spillover effects both within and across treatment groups in the robustness check section.

As an initial step towards valuing rail access in the land market, it is worthwhile to do the balancing test to see if treated places would be significantly different from the untreated places in terms of the observable demographic characteristics[43]. I estimate a set of regression models using residential and commercial land parcel sample respectively (see results in Tables 2.4 and 2.5). The dependent variable is the log of initial prices for land parcels sold during 1999 and 2002, educational attainment, public housing rent ratio, population density, old building percentage, employment accessibility, and distance to the CBD respectively. The main independent variables are the treatment groups.

[43] Due to the lack of census panel data, this study has not attempted to measure demographics dynamics in treated places relative to observationally identical control places as a result of transport improvements. In essence, the rationale behind the balancing test is to show that the variation in the 'treatment' variable is as reasonably good as random (see Angrist and Pischke, 2009 for details).

In terms of the perfect treatment-control balancing, it would be expected to see that the estimated coefficients for the treatment groups are not statistically significant. As can be seen from Table 2.4, the treated and control places are not significantly different with respect to their initial residential land prices, population density, old building percentage and public housing rent ratio. However, places with higher educational attainment level are more likely to be treated with rail access under all treatment scenarios (within 4km). Perhaps more informative are the last two columns, which report the results of employment accessibility and distance to the CBD. Lower employment accessibility areas are more likely to be treated with rail access under all treatment scenarios (within 2km and 4km). Places that located further away from the CBD are more likely to be treated with rail access to stations after 2003 and 2008 (within 2km and 4km). I do the same test for commercial land parcel sample in Table 2.5. It shows a balanced pattern for treated and control places in terms of their initial commercial land prices, educational attainment, population density, old building percentage and public housing rent ratio. Though the magnitudes are very small, places with lower employment accessibility are more likely to be treated with rail access under all treatment scenarios (within 2km and 4km). All else equal, places that located further away from the CBD are more likely to be treated with rail access to planned stations opening after 2009 (within 2km and 4km). It is certainly the case that some other pre-treatment characteristics would be unbalanced between treated and control places. Nonetheless, the headline results from Table 2.3 and Table 2.4 suggest that there is limited difference for the treatment groups and control groups in terms of key observable pre-treatment demographics and related spatial characteristics. In the formal modelling analysis, I will include the fixed effect, time trends and a wide range of location-specific factors to further adjust for differences in characteristics, in the regression estimates reported in Section 5.

4 Models

Using a rich geographically-coded dataset, this study estimates the effects of increased station proximity on residential and commercial land prices in Beijing. My transport improvement model builds on the hedonic spirit that is widely used in the evaluation of amenities values[44]. The baseline equation for my analysis is expressed as follows[45]:

$$Ln\,\mathrm{Pr}\,ice_{ilt} = \beta_0 + \sum_{j=1}^{3} \beta_j Lndist_{it} + \sum_{t=1}^{3} \beta_t Y_t + \beta_k X_{ilk} + f_l + \varepsilon \qquad (1)$$

Where $Price_{ilt}$ represents the price of vacant residential or commercial land parcel i located at area l in the period t; $dist_{it}$ is the distance to the nearest station; X_{ilk} is a matrix of land structural and localized characteristics; Y_t presents the time trend effects; f_l indicates area-specific fixed effect; ε is a random error term[46]. Other Greek letters are parameters to be estimated.

This traditional cross-sectional approach is highly successful at capturing long-run relationships between land prices and rail access, but may not recover the

[44] See the seminal work by Rosen (1974). See Hilber (2011) and Gibbons et al. (2011) for recent hedonic reviews.

[45] In this study, I have tried estimating flexible-form models with Box–Cox transformation but could not reject a strong log–log relationship between land prices and key explanatory variables.

[46] Standard errors are clustered at the zone level to allow for heteroscedasticity and spatial and temporal correlation in the error structure within zones.

impact of increased station proximity on local prices before and after a change in transport improvement policy. To explicitly account for this, I adopt a conceptually more attractive approach. By focusing on what happens after the transport improvement, in places affected and unaffected by the change, I can more reliably assess the new rail transit's impact[47] on local land prices.

To achieve this, I need data on land price changes and rail station access changes. In contrast to the systemic repeated sales data and limited transport infrastructure changes in the developed countries, it is easy to observe an opposite scenario in China: an emerging land market system since the 1990s and the rapid urban rail transit development. The first data requirement is met by using a 1999-2009 cross-sectional land parcel transaction data. One limitation here is that I do not have access to repeated observations for the same parcel over time and therefore cannot apply panel-data methods to control for fixed-over-time omitted variables. Thus rail access and price outcomes may both be influenced by a third-party unobserved variable. However, this paper does provide an extremely rich data set which allows me to mitigate this problem (at least partially) by controlling for a wide range of parcel and location characteristics (Ferrer-i-Carbonell and Frijters, 2004; Gibbons and Silva, 2011; Cornaglia et al, 2012). For a more accurate assessment, I am well aware that the model presented above is not effective for eliminating changes in location characteristics as a result of changes in parcel–station distance. For example, if the number of cafe stores increased disproportionately in places treated with a new rail station for exogenous reasons, the

[47] Here and thereafter, the term of "new rail transit's impact" refers to the impact of increased station proximity on local land prices due to the opening and planning of new rail transit lines.

econometrician does not see this but the households do. Thus I would observe the land price premium and would attribute this to the effect of increased station access when in reality it actually accounts for the omitted amenity values. This should not be a problem when researchers have the detailed data and know what variables to include/remove or suitable to be an instrument for the model specification. Given the data limits, I am mainly interested in the whole effect of the new transport infrastructure, including the multiplier effect of the cafes etc. I implicitly assume that the error term is uncorrelated with the explanatory variables, and those time-varying unobserved factors do not spill over space.

The second data requirement is easier to meet because of the recent dramatic changes in public transport infrastructure in Beijing. The supply of new rail transit stations increased over time---two subway lines were opened in 2003, four lines were opened around 2008 and another eight lines were planned to open after 2009. These improvements will lead to the increased proximity to stations for a series of subset of land parcels in my data set after 2003, after 2008, and after 2009 respectively. This means that I can, in principle, estimate the increased station proximity effect in the multi-nested treatment scenarios[48]. The outcome regression equation becomes:

$$Ln\text{Pr}ice_{iljt} = \beta_0 + \sum_{j=1}^{3}\beta_j Treatment_j + \sum_{t=1}^{3}\beta_t Period_t + \sum_{j=1}^{3}\sum_{t=1}^{3}\beta_{jt} Treatment_j * Period_t + \beta_k X_{ilk} + f_l + \varepsilon \quad (2)$$

[48] In the presence of nested treatment groups, my study's estimates provide new insights about each treatment effect conditional on the subsequent treatment scenarios. One major concern is to test whether there are spillover effects among treatment groups when adding all of them into one model specification. As a robustness check, I have tried to add each treatment group subsequently in different model specifications, but the difference between their coefficients won't tell anything about the spillover effect because the sum up value of the treatment coefficients remains the same as when adding all of them into one model specification. To further test this, I will explicitly exploit spatial spillover effects within and across residential/commercial treatment groups in Section 5.2.

In this equation, $Treatment_j$ refers to a specific treatment group (e.g. $(station \geq 2003)$, $(station \geq 2008)$, $(station \geq 2009)$). $Period_t$ is a set of “policy-on” time dummy variables ($(1999 \leq year < 2003)$, $(2003 \leq year < 2008)$, $(year \geq 2008)$). The coefficients β_{jt} then show the various treatment effects ($Treatment_j$*$Period_t$) in different periods[49]. Table 2.6 summarized the underlying meanings and expected signs of these treatment effects.

The rationale behind this multiple intervention research design is that, it allows me to test for heterogeneous new rail transit’s impacts on Beijing’s land market along several dimensions. First, it is expected that these estimates are significantly positive in corresponding periods. For example, the interactions between $(station \geq 2003 * 2003 \leq year < 2008)$ and $(station \geq 2008 * year \geq 2008)$ should be significantly positive and show the opening effect[50] for stations in 2003 and in 2008 respectively. A second dimension is captured by estimates of $(station \geq 2003 * year \geq 2008)$ and $(station \geq 2008 * 2003 \leq year < 2008)$. These two coefficients allow me to test post-opening effect for stations in 2003 and pre-opening effect for station in 2008 respectively. Their expected signs largely depend on the price growth trends during 2003 and 2008 versus after 2008. If the price growth trends after 2008 are greater than that during 2003 and 2008, then their estimates would be less positive and insignificant. A third dimension is to examine the net

[49] β_{j1} represent a set of baseline categories ($Treatment_j$*$Period_1$) that are omitted in the estimating result tables.

[50] Here and thereafter, the opening effect means the estimated amenity benefits from the distance reductions to land parcels that are now within 0.5km, 1km, 2km, 4km respectively from newly-opened stations in 2003/2008.

planning effect[51] for stations opening after 2009 relating to different land market periods. As indicated by recent empirical findings (Knaap et al, 2001), it is reasonable to expect that there would be positive signs associated with estimates of $(station \geq 2009 * 2003 \leq year < 2008)$ and $(station \geq 2009 * year \geq 2008)$.

Empirically, many location factors associated with rail stations would have interaction effects on land prices for reasons other than the benefits of increased station proximities due to the building of new railway lines. For example, stations located near employment-centre could offer more job opportunities and other amenities that might provide additional land values, whereas increasing proximity to station areas with high crime levels may actually decrease the benefits of transport accessibility on land values (Gibbons and Machin, 2008). To help identify such interaction effects, the model specification can be written as:

$$\begin{aligned}\mathrm{Ln}Price_{\mathrm{iljt}} &= \beta_0 + \sum_{j=1}^{3} \beta_j Treatment_j + \sum_{t=1}^{3} \beta_t Period_t \\ &+ \sum_{j=1}^{3}\sum_{t=1}^{3} (\beta_{jt} + \beta_k' X_{ilk})\, Treatment_j * Period_t + \beta_k X_{ilk} + f_l + \varepsilon \quad (3)\end{aligned}$$

5 Results

5.1 Baseline regression estimates

In this section the results obtained from estimating the model described in Eq.(2)

[51] Here and thereafter, the net planning effect means the estimated amenity benefits from the distance reductions to land parcels that are now within 0.5km, 1km, 2km, 4km respectively from stations opening after 2009. It includes a combination of the potential negative construction effect and the positive anticipation effect for planned stations.

using residential and commercial land parcel data are reported in turn. In discussing the baseline regression estimates in Table 2.7-2.8, I focus primarily on heterogeneous effects of increased station proximity on local residential and commercial land prices. In light of recent urban transport improvement literature, the implicit assumptions underlying the interpretation of these estimates are as follows: (i) the measured effects of increased station-distance proximity on land prices happen only through parcel-station distance changes result from new rail transit constructions and expansions; (ii) unobserved characteristics and trends, such as land supply constraints and overall economic climate do not vary greatly; and (iii) the measures for localized characteristics included in the models can effectively explain the impact of transport access on the land market.

Column (1) in both tables shows estimates that include parcel coordinate fixed effects, proximity effects for parcels that are beyond the distance bands (0.5km, 1km, 2km, 4km thereafter), treatment dummies, general time effects[52], but no additional controls. As for the first treatment group $(station \geq 2003)$, the opening effect of stations in 2003 on the residential land prices is found insignificant when treated with the 0.5km distance band, but turns to be significantly positive when using wider distance bands[53]. Parcels that are now within 2km from a station have a significantly

---

[52] To further control the spatial-temporal effect, I also include the interactions between time trends and parcels in each treatment group that only meet the first treatment selection criteria---parcels that experienced distance reductions to the closet stations(Treatment Criteria 1* Time); and interactions between time trends and parcels in each treatment group that only meet the second treatment selection criteria---the parcel-station distance is now within the distance bands(Treatment Criteria 2* Time).

[53] Recall that the distance bands are cumulative, which make the results straightforwardly interpreted. For example, for residential parcels I find a negative effect within 0.5km of a station, which is likely attributes to noise

higher price premium compared to other distance bands. These results suggest that residential land parcels that are very close to stations could be affected by negative externalities, but those at an intermediate spatial range are beyond potential negative externality effects and benefit from the increased station proximity due to the opening of new stations in 2003. There are no statistically significant post-opening impacts from distance reductions to parcels that are beyond 0.5km, 1km, 2km, and 4km spatial contours from new stations in 2003.

When I compare the estimated coefficients on the second treatment group $(station \geq 2008)$, the results are qualitatively similar to those reported in the first treatment group $(station \geq 2003)$. As for the quantitative magnitudes, the price premium paid for being closer to a station opening in 2008 is larger than that of newly-opened station in 2003. This is expected because more new stations were opened in 2008 than in 2003, resulting in obvious parcel-station distance reductions. The pre-opening effects for stations in 2008 are positively significant when treated with the 1km, 2km and 4km distance-bands[54].

Continuing to discuss the results in Column (1), I next focus on the estimated results for the third treatment group $(station \geq 2009)$. This treatment group highlights

externalities emitted by the station. But the next band is within 1km of a station and the results show a positive effect. This result suggests that the proximity impact of rail stations is determined by the mix of properties within 0.5 km and between 0.5km and 1km. Of course, researchers can further disaggregate the distance band selection into the 0.25km range, or choose to define the bands as 0 to 0.5km, 0.5km to 1km, 1km to 2km, and 2km to 4km. The key point here is to shed light on the importance of considering the distributional proximity impacts of rail stations on land prices over the geographical space.

[54] Note that treatment dummies have insignificant signs in Tables 2.7 and 2.8. These results, to some extent, can help explain the pre-opening effect of station in 2008 is not caused by the price-growing trends in the treated places.

the net planning effects for stations opening after 2009. As expected, I find that prices rise significantly in areas affected by planned stations opening after 2009 when treated with both of the $(2003 \leq year < 2008)$ period and the $(year \geq 2008)$ period. When comparing the quantitative nature between different time periods and among different distance bands, the price premiums are greater linked with the 2km distance band, and are much larger during the period after 2008 than that of during 2003 and 2008. This result confirms the possibility that the under-constructed rail transit plans are observed by the developers and increasingly capitalized into land prices when closing to their completion times. Nevertheless, when one is reading the results, it is necessary to keep in mind that the data limits my transport improvement analysis to changes that occurred within about 3 years of the new rail transit development[55]. My estimates might underestimate the whole effect of transport accessibility when the price-lag adjustment process is long before or after the opening of new lines[56], or might overestimate the benefits if negative externalities at station areas evolve with the improved transport accessibility.

For mega-cities like Beijing, part of the increased station proximity effects could be attributed to the spatial effects, like differences in price trends in the central city and suburbs. In Column (2), I estimate the same specification but augmented with a set of spatial measures by allowing the interactions between the time trend and distance to CBD, and by allowing for time trends interacted with the distance to the Olympic park,

[55] Recall that the transport improvement estimates are obtained from price differentials among three time periods: 1999-2003, 2004-2007 and 2008-2009.

[56] See McDonald and Osuji (1995) and McMillen and McDonald (2004) for a detailed discussion.

and the distance to important emerging "bedroom" communities[57]. The rationale behind this is that, during the time period I study, there was a boom in land price growth in Beijing, especially in areas like the central city. Although not shown in the table, this is confirmed by the significantly negative coefficient on the distance to CBD and its interactive terms with the time trend. The key finding here is that whilst the price growth trend effect matters, the increased station proximity effects are still robust and contribute to significantly higher residential land prices.

In Columns (3) and (4), I control for a wide range of land structural and location-specific characteristics (documented in Table 2.2). About 45% of the variation in the log of residential land prices respectively is explained by my transport improvement models. This compares favourably to previous hedonic literature in China. In addition, estimated treatment effect coefficients exhibit reasonable stability over alternative model specifications. After controlling for the full set of localized characteristics and adjusting for different temporal-spatial trends in column (4), I find that the opening effects of station in 2003, on average, are valued at around 0.61%, 1.96%, 1.25% of residential land prices at affected areas (within 1km, 2km, 4km respectively). The opening effects of station in 2008, on average, are valued at about 3.75%, 4.20%, 2.02% of the prices of affected residential land parcels (within 1km, 2km, 4km respectively). The positive and significant signs associated with the pre-opening effect for station in 2008 show that the potential increased station proximity effect is capitalised into local land prices (within 1km, 2km and 4km). In

[57] The estimated coefficients of these interaction terms are not reported. The results remain robust by controlling the interactions between time trends and distance-to-stations.

terms of the net planning effect, prices rise by about 0.23%, 0.58%, 0.48% on average when treated with the time period during 2003 and 2008 (using 1km, 2km, 4km distance band respectively); and prices rise by around 3.01%, 3.79%, 3.51% on average when treated with the time period after 2008 (using 1km, 2km, 4km distance band respectively). The insignificant signs of the increased station proximity effect within 0.5km imply the negative externalities such as noise and congestion effects that reduce the capitalization effect for parcels that are too close to the new stations.

Switching to the commercial land parcel sample in Table 2.8, I find quite similar qualitative patterns with the residential land parcel sample results, expect for the results estimated by using the 0.5 km distance band. There are significantly positive impacts from the opening effect of station in 2003 and 2008 on commercial land prices within 0.5km. This finding is in line with the expectation that commercial land parcels would accrue greater benefit than residential land parcels at a closer distance range from a station---by gathering large population flows and high demand for commercial activities. As for the quantitative nature, I find that station proximity impacts on commercial land prices are slightly lower than those on residential land prices. This is not surprising given that the parcel sample of the commercial land market is relatively thinner than that of the residential land market.

One important implication from the baseline estimates is that these station proximity impacts generally decay with distance in a non-linear trend. For example, the impact from increased proximity to stations on residential land parcels that are now within 2km from new stations is larger than other distance bands' results. The most affected places for commercial land parcels are those that are now within 1km station area (see Tables 2.7 and 2.8 for details). To explore whether the observed differences in proximity impacts on residential and commercial land prices are statistically significant, the Chow statistical test (Chow, 1960) is conducted. The null hypotheses are: the set of coefficients for the treatment effects on the commercial parcels and the corresponding set of coefficients for the treatment effects on the residential parcels are not significantly different from each other. An interesting finding is that, the null hypotheses are rejected at the 5% significance level for those statistically significant treatment effects reported in Tables 2.7 and 2.8. This provides strong evidence of the spatial heterogeneity in the proximity impacts of rail stations across residential and commercial land markets.

5.2 Robustness checks

To test the robustness of main findings, I now examine how sensitive the baseline results are to changes in different data samples and econometric specifications.

The first sensitivity analysis is to adjust spatial selections in the land parcel sample. Because Beijing urbanized area is so large, it may have a large influence on the baseline estimates. I therefore, in model specifications of Table 2.9 report results that only include the land parcel sample located within the central city (within the 3rd ring road) and within the suburb (within the 5th ring road) subsequently[58]. The results, reported in Columns 1-4 of Table 2.9, generally mirror that of the baseline estimates, suggesting that the spatial trimming of parcel sample does not significantly affect the

[58] Recall that the full sample refers to the spatial range within the 6th ring road of Beijing.

new rail transit's impact[59].

Second, I consider a further robustness issue related to the impact on effective proximity of new rail transit lines on existing stations. With the rapid rail transit development in Beijing, it is noteworthy that some new lines convert what were simple stops into new cross-stations. Accordingly, I have re-assigned the new station sample into three sub-categories[60]: new simple stations, new cross stations, and new simple-to-cross stations. Table 2.10 reports that the results are robust relative to the baseline treatment effect estimates[61]. One thing to note is that, the positive impact of distance reductions to the new simple-to-cross stations (within 2km) on commercial land prices are greater than that of the new simple and cross stations. Another interesting finding is that, residential land prices rose slightly higher when treated to a new simple station compared to new cross stations and simple-to-cross stations. Certainly, new simple-to-cross stations are more likely to gather a larger number of population flows and greater demand for retail establishment than purely new cross stations and simple stations. Thus proximity to a new simple-to-cross station is of higher value to commercial land prices than other types of new stations. But residential land prices are more sensitive to the increased negative externalities such as crime and noise emitted by the junction stations, and therefore the effect of proximity to a new

[59] Note that while the qualitative nature of the results is relatively robust, the estimated coefficients of treatment effects within the central city have lower magnitudes than those within the suburbs.

[60] In the preliminary estimation process, I have also divided new station status into underground and over-ground stations, but their results are not statistically significant.

[61] To avoid redundancy, I focus on results using the 2 km distance band here. Though the magnitudes are smaller, treatment effect variables associated with the other distance bands are also robust in terms of qualitative nature relative to the key main results.

simple station benefits more on residential land values.

Next, I consider whether there are significant spillover effects within and across residential/commercial treatment groups. Such test helps to gauge the robustness of the results more fully[62]. As for the within-group spillover effects, I focused on examining whether the parcels in the subsequent treatment group affect the increased station proximity effect on parcels in the prior treatment group. Two steps are involved in measuring the spillover effect from the $Treatment_{j+1}$ onto the $Treatment_j$: first, to calculate the distance between parcels that belong to the $Treatment_j$ (but not belong to the $Treatment_{j+1}$) and parcels in $Treatment_{j+1}$; and second, to make interactions between this distance variable and its corresponding treatment effect ($Treatment_j*Year_t$). The results in Columns 1-2 of Table 2.11 show that the estimated spillover effect coefficients are small in magnitudes and insignificant for both residential and commercial parcel sample[63]. Another natural question is to ask whether the new rail transit's effect on residential land parcels is affected by adjacent commercial land parcels. To this end, the cross-group spillover effect measures are calculated through the interactions of the distance between all treated commercial land parcels[64] and residential land parcels in each treatment group. Estimates from column

[62] This identification method is well-established in the literature. See Irwin and Bockstael (2001b) for further discussion of this issue within the context of land use spillovers.

[63] Table 2.11 only reported the results by using the 2km distance band scenario. There are no statistically significant spillover effects within groups when using the 0.5km, 1km, and 4km distance bands.

[64] Note that I have also interacted the residential land parcels in each treatment group with both of treated and control commercial land parcels. Because the estimating results are not significant, they were dropped from the table. Plus, there are also no statistically significant cross-group spillover impacts when using the interactions of the residential land parcels in $Treatment_j$ with its distance to the commercial land parcels in either $Treatment_j$ or $Treatment_{j+1.}$

(3) in Table 2.11 show that most of the residential treatment effect variables are reassuringly quite robust to the potential spillover impacts from nearby commercial land parcels. The only two exceptions are associated with the treated residential land parcels receiving distance reductions to stations after 2009. The small magnitudes of their coefficients affirm the possibility that residential land parcels could gain slightly positive spillover effects from adjacent commercial land parcels when treated with planned station areas.

Finally, reliance on estimates of amenity benefits for the average sample effect in a metropolitan area would mask rail access values to parcels in particular places. Thus I now turn to the results with interaction terms[65] estimated by using Eq.(3). In Table 2.12, the interactions between treatment effect variables and local residents' median educational attainment level show that residential land price premiums are valued greater for being close to a station in high- than in low-educational attainment areas. Assuming that residents' incomes are positively correlated with their educational attainment level in Beijing[66], this result implies that the greater commuting time savings provided by transport developments enhance the rail access value for well-educated residents. Meanwhile, the commercial land prices are found to be valued higher when treated in high- than in low-educational attainment places, possibly

[65] These estimates do not rely on the within-zone changes induced by the transport improvement. The results are not statistically significant to the inclusion of the interactions between treatment effect variables and other location-specific variables listed in the Table 2.2. Also, I interacted the treatment effect variables with dummies indicating whether the nearest station is underground or over-ground, however, there were no statistically significant signs of these interactions, so they were dropped from the final models.

[66] There is no available information about local residents' income in the urban China census data. Yet in reality, this assumption is consistent with the actual observations in Beijing.

because the larger consumption capability for well-educated residents gentrifies the value they attach to rail access.

The interactions of treatment effect variables with crime rates show that in places within 0.5km of a station, an increase in crime results in less residential and commercial land prices, however, the coefficients are very small in magnitude. There are no statistically significant impacts when treated with residential and commercial land parcels that are within 1km, 2km, and 4km distance contours of a station. The results suggest that the interactive impact of increased station proximity and crime is not strong on both residential and commercial land markets, but clearly the negative effect is dominant close to the station. Estimates from the treatment effect variables and employment accessibility interactions show that the effect of increased station proximity is more valuable in places with higher- than lower-employment accessibility.

Beyond these interaction variables, what other local amenities and disamenities that might have significant complementarities with rail stations were overlooked? The list could be very long, such as climate, social capital and other forces of local heterogeneity that are unlikely to be observed by the econometrician. The key point here is that government investment in transport infrastructure should consider both the direct and the interaction effects of new stations on the gentrifications of nearby land values.

6 Conclusions

Beijing has recently made huge public investments in upgrading its rail transit networks. The investments have created a large number of land parcels that are now closing to new stations. Using rich vacant residential and commercial land parcel data, I examined whether land prices in such treated parcels changed after the geographical distances to the nearest station were reduced, relative to observationally control parcels. The empirical answer is mostly yes.

My results yield several important insights that have not been fully considered in the previous literature. First, residential and commercial land parcels receiving increased proximity to both newly opened stations and planned stations experienced appreciable price premiums, though the relative benefits are different in magnitudes. I further reported that the qualitative pattern of estimation results is remarkably robust across a set of stringent sensitivity analyses. Second, the impacts of transport improvements on land prices play out differently at different distance ranges from a station, and vary widely with local socio-demographics. This finding highlights the importance of considering the significantly heterogeneity in the effects of rail access changes on nearby land values.

Overall the combined empirical findings show that the impact of increased station proximities can be reflected into land price changes. Practically, the impacts of transport improvements on land prices may also serve as an effective means of coordinating public investments with private sector investments in the land markets. For example, the urban spatial structure is likely to change due to such transportation-oriented development strategies. A good case in point is that in order to offset higher land prices, developers are more likely to build high-density constructions in station areas. The public investment in rail transit expansions would also gentrify the commercial activities at station areas and increase metropolitan public transport revenues in the long term. For such strategies to succeed, planners need to create market interventions that discourage low-density land development (such as the town-house) and encourage high-density development. This can be done by using zoning, building floor-area-ratio controls, or other forms of land use planning constraints. Of course, transport-oriented development might not be purely a “free lunch” for all households. Unlike the American cities (Glaeser et al., 2008), it is more likely to observe in Beijing that the poor people, in order to offset the rising rents, are pushed further out to the remote suburbs and bear longer commuting distances to workplaces. Together, these implications provide a rationale for local government to go beyond the real estate consequences, and consider wider aspects of local residents’ wellbeing that may be affected by the public investment in rail transit expansions. Future works using reported survey data to explore the rail access effect on homeowners’ happiness would be useful.

Table list

Table 2.1 Effects of rail access on real property values: recent empirical studies

Effects	Property Types	Context	Methods	Statistical association
Established rail access	Residential	Cheshire and Sheppard (1995): Reading, UK	Cross-sectional approach with controlling for unobserved spatial factors	+
		Zheng and Kahn (2008): Beijing, China	Cross-sectional approach without controlling for unobserved spatial factors	+
		Bowes and Ihlanfeldt(2001): Altanta, US	Cross-sectional approach of direct effects from station proximity and interaction effects based on crime and retail employment	+, vary with distance bands
	Residential+Commercial	Cervero and Duncan(2001): California, US	Cross-sectional approach without controlling for unobserved spatial factors	+; proximity impact on commercial properties is greater than it is on residential properties within closer distance band
		Debrezion et al (2011): 3 cities in the Netherlands	Cross-sectional approach without controlling for unobserved spatial factors	
Opening effect of new stations	Residential	Baum-Snow and Kahn (2000): 5 large cities, US Kahn (2007): 14 large cities, US	Difference-in-difference approach based on average census tract housing prices and the distance reduction in census tract-station arising from new lines	+
		Ahlfeldt (2011); Gibbons and Machin (2005): London, UK	Difference-in-difference approach based on changes in station-home distance arising from new lines, based on repeated sales data	+
Planning effect of new stations	Residential	Knaap et al(2001): Oregon, US	Traditional cross-sectional analysis about the planning effect	+
		McMillen and McDonald (2004):Chicago, US	Comparison analysis among the proposal and opening effect	+

Table 2.2 Descriptive statistics of variables

Variables	Definition	Residential land sample Mean/ (Std.Dev)	Commercial land sample Mean/ (Std.Dev)
Dependent Variable			
Land Price	Ln (Land parcels' leasing price per square meter (CNY/sq.meter))	7.45(1.08)	7.76(1.42)
Locational-specific Variables			
CBD	Ln (Distance between a land parcel and CBD (meters))	9.03 (0.64)	8.85(0.75)
Land parcel size	Ln (The area of a land parcel (m^2))	9.06 (1.34)	7.59(1.78)
Park	Ln (Distance to the nearest park (meters))	7.77 (0.72)	7.61(0.81)
River	Indicator of proximity to rivers (<500 meter)	0.18 (0.38)	0.11(0.31)
Air quality	Indicator of Air pollution index to each parcel	1.93 (0.87)	1.99(0.88)
Bus	Ln (Distance to the nearest bus stop (meters))	6.03(0.82)	6.12 (1.06)
Expressways	Ln (Distance to the nearest expressway (meters))	6.43(1.14)	6.36 (0.98)
School	Ln (Distance to the nearest middle school*school rank)	25.01 (5.68)	24.34(6.34)
Employment Accessibility	Indicator of employment accessibility to each parcel	0.04(0.05)	0.06(0.07)
Population Density	Population density in each zone (1,000 people per km^2)	2.37 (3.35)	2.76(4.35)
Old Building	Ratio of buildings built before 1949 in each zone (%)	0.03(0.09)	0.07(0.14)
Education Attainment	Median resident educational attainment in each zone:1=middle school or lower;2=high school;3=university;4=post graduate	1.715(0.508)	1.91(0.46)
Crime	Number of crimes per 1000 people in each zone	5.335(6.655)	4.08(5.15)
Public Housing	Percentage of people renting public housing in each zone	0.31(0.20)	0.33(0.21)

Table 2.3 New rail transit constructions in the urbanized area of Beijing

Line	Start by (year)	Open by (year)	Cost (CNY billions)	Length (kilometre)	Station (number)
13	2000	2003	6.6	40.5	16
Batong	2001	2003	3.4	19	13
4	2004	2008	15.2	28	24
5	2003	2008	11.9	27.6	23
10A	2004	2008	12.8	24.6	22
8A	2005	2008	2.5	15.8	4
Daxing	2008	2010	6.0	22	12
Yizhuang	2008	2011	11.0	23.2	14
8B	2009	2012	10.1	17	11
6	2007	2012	18.2	39	30
7	2009	2012	15.1	24	21
9	2007	2012	8.8	16.4	13
10B	2007	2012	18.5	32.9	23
15A	2009	2012	18.1	20.2	13

Notes.---The information on the rail transit lines that have not been completed yet may be changed. See the updated information on the Beijing Municipal Committee of Transport's official website http://www.bjjtw.gov.cn/

Table 2.4 Balancing test results based on residential land parcel sample

	Land Price	Education Attainment	Public housing	Population density	Old Building	Employment Accessibility	Distance to CBD
0.5km_$(station \geq 2003)$	0.023	0.086	-0.021	0.255	0.013	-0.001	-0.044
	(0.291)	(1.365)	(-1.235)	(0.823)	(1.182)	(-0.333)	(-1.343)
0.5km_$(station \geq 2008)$	-0.011	0.122	0.027	-0.141	-0.012	0.003	0.040
	(-0.129)	(1.371)	(1.421)	(-0.429)	(-1.338)	(1.500)	(1.143)
0.5km_$(station \geq 2009)$	0.054	0.029	-0.001	-0.327	-0.010	0.001	-0.021
	(0.635)	(0.592)	(-0.048)	(-0.991)	(-1.250)	(0.503)	(-0.603)
1km_$(station \geq 2003)$	0.009	0.139	0.005	1.082	-0.005	0.003	0.021
	(0.083)	(1.495)	(0.208)	(1.497)	(-0.556)	(1.502)	(0.467)
1km_$(station \geq 2008)$	0.161	0.123	0.028	-1.188	0.006	-0.004	-0.071
	(1.626)	(2.158)	(1.273)	(-1.344)	(0.667)	(-1.333)	(-1.392)
1km_$(station \geq 2009)$	-0.104	0.009	-0.009	0.093	0.012	-0.002	0.004
	(-1.268)	(0.196)	(-0.501)	(0.292)	(1.200)	(-1.010)	(0.121)
2km_$(station \geq 2003)$	0.173	0.013	-0.038	-0.972	0.015	-0.013	0.197
	(0.935)	(0.121)	(-0.950)	(-1.358)	(0.938)	(-3.250)	(2.592)
2km_$(station \geq 2008)$	-0.235	0.163	-0.005	1.398	-0.039	-0.024	0.232
	(-1.343)	(1.598)	(-0.132)	(1.431)	(-1.560)	(-6.001)	(3.222)
2km_$(station \geq 2009)$	-0.072	0.019	-0.019	0.198	-0.010	-0.004	0.037
	(-1.075)	(0.487)	(-1.267)	(0.759)	(-1.429)	(-2.021)	(1.370)
4km_$(station \geq 2003)$	-0.316	0.312	-0.027	1.640	-0.053	-0.027	0.491
	(-1.430)	(2.403)	(-0.551)	(1.534)	(-1.359)	(-4.513)	(5.337)
4km_$(station \geq 2008)$	0.083	0.419	-0.015	1.035	0.026	-0.004	0.157
	(0.483)	(4.190)	(-0.395)	(1.549)	(1.368)	(-2.008)	(2.211)
4km_$(station \geq 2009)$	0.111	0.429	-0.039	-1.602	0.016	-0.019	0.078
	(0.631)	(4.206)	(-1.083)	(-1.689)	(1.067)	(-4.508)	(1.083)
Constant	0.304	1.011	-0.402	-3.016	-0.106	-0.272	8.680
	(0.749)	(4.284)	(-4.568)	(-2.811)	(-3.029)	(-6.727)	(5.667)
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1054	1054	1054	1054	1054	1054	1054
Adjusted R-squared	0.416	0.34	0.207	0.222	0.153	0.577	0.698

Notes.---Each column reports estimates of the balancing tests from a separate regression. The dependent variable for each regression is listed in the first row of the table (initial residential land prices, educational attainment, public housing rent ratio, population density, old building percentage, employment accessibility, distance to CBD), as described in the text. The data sample is used residential land parcels sold during 1999 and 2002. t-statistics in parentheses, clustered on zone unit.

Table 2.5 Balancing test results based on commercial land parcel sample

	Land Price	Education Attainment	Public housing	Population density	Old Building	Employment Accessibility	Distance to CBD
0.5km_($station \geq 2003$)	-0.252	0.126	0.009	-1.001	0.039	-0.003	-0.132
	(-1.120)	(1.370)	(0.265)	(-1.053)	(1.393)	(-0.750)	(-1.361)
0.5km_($station \geq 2008$)	0.381	0.082	0.048	1.781	-0.022	0.002	0.191
	(1.371)	(0.719)	(1.143)	(1.516)	(-0.846)	(0.401)	(1.619)
0.5km_($station \geq 2009$)	0.009	0.145	-0.015	1.605	-0.009	0.002	-0.021
	(0.032)	(1.261)	(-0.349)	(1.354)	(-0.346)	(0.400)	(-0.219)
1km_($station \geq 2003$)	0.455	0.126	-0.037	2.871	0.003	0.009	-0.015
	(1.458)	(0.977)	(-0.805)	(1.670)	(0.103)	(1.501)	(-0.139)
1km_($station \geq 2008$)	0.073	0.006	0.018	-2.567	-0.016	0.011	-0.112
	(0.213)	(0.043)	(0.346)	(-1.389)	(-0.503)	(1.222)	(-0.949)
1km_($station \geq 2009$)	-0.083	0.039	0.086	-0.899	0.035	-0.002	-0.054
	(-0.219)	(0.250)	(1.509)	(-0.562)	(0.971)	(-0.333)	(-0.412)
2km_($station \geq 2003$)	0.323	0.242	-0.028	-1.373	0.057	-0.023	0.143
	(0.441)	(0.804)	(-0.252)	(-0.454)	(0.838)	(-1.769)	(0.565)
2km_($station \geq 2008$)	-0.312	0.117	0.097	3.737	-0.143	-0.056	0.305
	(-0.429)	(0.391)	(0.875)	(1.217)	(-1.607)	(-4.308)	(1.215)
2km_($station \geq 2009$)	0.681	0.303	-0.145	-4.506	0.081	-0.039	0.813
	(0.799)	(0.866)	(-1.124)	(-1.253)	(1.025)	(-2.610)	(2.765)
4km_($station \geq 2003$)	-0.976	0.317	0.003	1.519	-0.112	-0.036	0.297
	(-1.310)	(1.036)	(0.027)	(0.483)	(-1.623)	(-2.769)	(1.156)
4km_($station \geq 2008$)	-0.229	0.121	-0.116	-3.978	0.136	-0.043	-0.179
	(-0.318)	(0.409)	(-1.064)	(-1.310)	(1.563)	(-3.308)	(-0.722)
4km_($station \geq 2009$)	-0.592	0.261	0.051	5.043	-0.096	-0.039	0.852
	(-0.743)	(0.796)	(0.423)	(1.499)	(-1.280)	(-2.786)	(3.098)
Constant	1.357	0.428	-0.536	-5.312	-0.263	-0.190	7.061
	(1.182)	(0.909)	(-3.045)	(-2.006)	(-2.430)	(-9.048)	(-10.823)
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	466	466	466	466	466	466	466
Adjusted R-squared	0.277	0.278	0.199	0.278	0.148	0.538	0.682

Notes.---Each column reports estimates of the balancing tests from a separate regression. The dependent variable for each regression is listed in the first row of the table (initial commercial land prices, educational attainment, public housing rent ratio, population density, old building percentage, employment accessibility, distance to CBD), as described in the text. The data sample is used commercial land parcels sold during 1999 and 2002. t-statistics in parentheses, clustered on zone unit.

Table 2.6 Underlying meanings and expected signs of the treatment effects

Treatment effects	Underlying Meaning	Expected signs
$(station \geq 2003) * (2008 > year \geq 2003)$	Opening effect of stations in 2003	+
$(station \geq 2003) * (year \geq 2008)$	Post-opening effect of stations in 2003	+/-
$(station \geq 2008) * (2008 > year \geq 2003)$	Pre-opening effect of stations in 2008	+/-
$(station \geq 2008) * (year \geq 2008)$	Opening effect of stations in 2008	+
$(station \geq 2009) * (2008 > year \geq 2003)$	Net planning effect of stations after 2009	+
$(station \geq 2009) * (year \geq 2008)$	Net planning effect of stations after 2009	+

Table 2.7 Baseline estimates of rail transit's effect on residential land parcel sample

Distance band	Variables	Model 1	Model 2	Model 3	Model 4
0.5 km	$(station \geq 2003) * (2008 > year \geq 2003)$	-0.014	-0.011	-0.012	-0.006
		(-0.115)	(-0.089)	(-0.103)	(-0.051)
	$(station \geq 2003) * (year \geq 2008)$	-0.185	-0.162	-0.151	-0.137
		(-0.387)	(-0.336)	(-0.330)	(-0.297)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.116	0.105	0.094	0.083
		(0.678)	(0.618)	(0.573)	(0.509)
	$(station \geq 2008) * (year \geq 2008)$	0.811	0.765	0.661	0.619
		(1.542)	(1.457)	(1.317)	(1.231)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.117	0.115	0.071	0.053
		(0.713)	(0.706)	(0.452)	(0.340)
	$(station \geq 2009) * (year \geq 2008)$	0.382	0.323	0.249	0.213
		(0.737)	(0.620)	(0.504)	(0.428)
1 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.664	0.642	0.621	0.611
		(1.829)	(1.778)	(1.876)	(1.746)
	$(station \geq 2003) * (year \geq 2008)$	0.383	0.196	0.183	0.199
		(0.834)	(0.422)	(0.416)	(0.449)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.653	0.592	0.584	0.575
		(0.351)	(0.333)	(0.312)	(0.319)
	$(station \geq 2008) * (year \geq 2008)$	4.532	4.218	3.992	3.750
		(4.263)	(3.957)	(3.580)	(3.378)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.298	0.256	0.242	0.239
		(1.776)	(1.631)	(1.779)	(2.025)
	$(station \geq 2009) * (year \geq 2008)$	3.276	3.134	3.188	3.009
		(2.884)	(2.796)	(3.107)	(3.067)
2 km	$(station \geq 2003) * (2008 > year \geq 2003)$	2.337	2.148	2.026	1.968
		(2.452)	(2.201)	(2.034)	(1.977)
	$(station \geq 2003) * (year \geq 2008)$	1.121	1.799	1.206	1.053
		(1.045)	(1.598)	(1.193)	(1.020)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.695	1.675	1.509	1.281
		(2.042)	(2.204)	(2.219)	(2.100)
	$(station \geq 2008) * (year \geq 2008)$	4.661	4.427	4.221	4.206
		(4.099)	(3.900)	(3.765)	(3.862)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.699	0.634	0.601	0.584
		(3.344)	(3.268)	(3.284)	(3.281)
	$(station \geq 2009) * (year \geq 2008)$	4.206	4.052	4.001	3.799
		(3.456)	(3.289)	(3.143)	(2.954)
4 km	$(station \geq 2003) * (2008 > year \geq 2003)$	1.664	1.459	1.361	1.259
		(2.956)	(2.727)	(2.638)	(2.596)
	$(station \geq 2003) * (year \geq 2008)$	1.992	1.750	1.518	1.332
		(1.515)	(1.345)	(1.248)	(1.213)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.449	1.225	0.912	0.941
		(1.723)	(1.690)	(1.737)	(1.860)

$(station \geq 2008) * (year \geq 2008)$	2.589	2.297	2.129	2.023
	(2.631)	(2.441)	(2.234)	(2.168)
$(station \geq 2009) * (2008 > year \geq 2003)$	0.538	0.496	0.485	0.481
	(1.724)	(1.664)	(1.792)	(1.979)
$(station \geq 2009) * (year \geq 2008)$	4.179	4.156	4.043	3.511
	(4.053)	(4.194)	(4.092)	(4.388)
Distance to CBD*Trends	No	Yes	No	Yes
Distance to Stations>0.5/1/2/4KM	Yes	Yes	Yes	Yes
Parcel Characteristics	No	No	Yes	Yes
Treatment dummies	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Treatment Criteria 1*Time	Yes	Yes	Yes	Yes
Treatment Criteria 2*Time	Yes	Yes	Yes	Yes
Distance to OlympicPark*Time	No	Yes	No	Yes
Distance to New Residential Area$_i$*Time	No	Yes	No	Yes
Station-distance*Time	No	Yes	No	Yes
Fixed effect	Yes	Yes	Yes	Yes
Location-specific characteristics	No	No	Yes	Yes
Observations	2,343	2,343	2,343	2,343
Adjusted R-squared	0.384	0.393	0.437	0.456

Notes.---Dependent variable is log residential land price. Data is the disaggregated parcel-level data for three periods: pre-2003, 2003-2007 and after. The baseline omitted category is *Treatment$_j$*Period$_1$(pre-2003)*. Regressions include control variables detailed in Table 2.2. t-statistics in parentheses, clustered on zone unit.

Table 2.8 Baseline estimates of rail transit's effect on commercial land parcel sample

Distance band	Variables	Model 1	Model 2	Model 3	Model 4
0.5 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.669	0.615	0.571	0.535
		(1.962)	(1.825)	(1.757)	(1.720)
	$(station \geq 2003) * (year \geq 2008)$	0.336	0.531	0.277	0.513
		(0.723)	(1.137)	(0.602)	(1.108)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.431	0.442	0.415	0.382
		(1.014)	(1.046)	(1.007)	(0.905)
	$(station \geq 2008) * (year \geq 2008)$	1.057	1.042	0.987	0.958
		(1.752)	(1.734)	(1.648)	(1.666)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.231	0.207	0.158	0.149
		(0.542)	(0.489)	(0.375)	(0.356)
	$(station \geq 2009) * (year \geq 2008)$	0.364	0.281	0.223	0.181
		(0.599)	(0.464)	(0.370)	(0.301)
1 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.889	0.763	0.662	0.625
		(2.102)	(2.079)	(2.181)	(2.097)
	$(station \geq 2003) * (year \geq 2008)$	0.805	0.621	0.556	0.511
		(1.214)	(0.944)	(0.832)	(0.768)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.605	1.469	1.185	0.871
		(2.439)	(2.652)	(2.319)	(1.919)
	$(station \geq 2008) * (year \geq 2008)$	2.788	2.376	1.828	1.663
		(3.584)	(3.337)	(2.653)	(3.035)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.795	0.668	0.622	0.582
		(1.944)	(1.663)	(1.709)	(1.921)
	$(station \geq 2009) * (year \geq 2008)$	1.463	1.298	1.165	1.081
		(1.701)	(1.728)	(1.686)	(1.941)
2 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.736	0.687	0.675	0.622
		(1.669)	(1.789)	(1.843)	(1.891)
	$(station \geq 2003) * (year \geq 2008)$	0.753	0.501	0.499	0.389
		(0.506)	(0.334)	(0.341)	(0.268)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.359	1.135	0.986	0.766
		(1.810)	(1.736)	(1.680)	(1.662)
	$(station \geq 2008) * (year \geq 2008)$	1.913	1.616	1.567	1.449
		(2.142)	(1.973)	(2.040)	(2.153)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.706	0.616	0.588	0.516
		(1.709)	(1.735)	(1.861)	(1.823)
	$(station \geq 2009) * (year \geq 2008)$	1.493	1.346	1.211	1.014
		(1.868)	(1.775)	(1.670)	(1.684)
4 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.646	0.627	0.552	0.396
		(1.755)	(1.923)	(1.890)	(1.692)
	$(station \geq 2003) * (year \geq 2008)$	1.268	1.007	0.939	0.604
		(0.856)	(0.679)	(0.644)	(0.586)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.251	1.208	1.024	0.876
		(1.700)	(1.808)	(1.769)	(1.708)

$(station \geq 2008) * (year \geq 2008)$	2.494	2.111	1.988	1.834
	(2.269)	(2.359)	(2.513)	(2.327)
$(station \geq 2009) * (2008 > year \geq 2003)$	0.885	0.756	0.688	0.582
	(1.667)	(1.662)	(1.707)	(1.813)
$(station \geq 2009) * (year \geq 2008)$	1.651	1.489	1.439	1.322
	(1.705)	(1.686)	(1.725)	(1.737)
Distance to CBD*Trends	No	Yes	No	Yes
Distance to Stations>0.5/1/2/4KM	Yes	Yes	Yes	Yes
Parcel Characteristics	No	No	Yes	Yes
Treatment dummies	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Treatment Criteria 1*Time	Yes	Yes	Yes	Yes
Treatment Criteria 2*Time	Yes	Yes	Yes	Yes
Distance to OlympicPark*Time	No	Yes	No	Yes
Distance to New Residential Area$_i$*Time	No	Yes	No	Yes
Station-distance*Time	No	Yes	No	Yes
Fixed effect	Yes	Yes	Yes	Yes
Location-specific characteristics	No	No	Yes	Yes
Observations	1,341	1,341	1,341	1,341
Adjusted R-squared	0.297	0.331	0.365	0.388

Notes.---Dependent variable is log commercial land price. See notes to Table 2.7 for additional details.

Table 2.9 Regression estimates of rail transit's effect on selected sample, sensitivity analysis

Distance band	Variables	Residential land parcel sample		Commercial land parcel sample	
		(1)	(2)	(3)	(4)
0.5 km	$(station \geq 2003) * (2008 > year \geq 2003)$	-0.012	-0.003	0.287	0.541
		(-0.064)	(-0.021)	(0.663)	(1.663)
	$(station \geq 2003) * (year \geq 2008)$	-0.095	-0.121	0.324	0.411
		(-0.135)	(-0.138)	(0.573)	(0.853)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.041	0.072	0.262	0.357
		(0.214)	(0.483)	(0.483)	(0.828)
	$(station \geq 2008) * (year \geq 2008)$	0.307	0.556	1.352	1.021
		(0.506)	(0.981)	(2.067)	(1.916)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.038	0.041	0.034	0.108
		(0.251)	(0.287)	(0.058)	(0.242)
	$(station \geq 2009) * (year \geq 2008)$	0.162	0.184	0.066	0.102
		(0.241)	(0.332)	(0.094)	(0.160)
1 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.565	0.597	0.568	0.592
		(1.652)	(1.860)	(1.656)	(1.935)
	$(station \geq 2003) * (year \geq 2008)$	0.093	0.135	0.377	0.425
		(0.178)	(0.288)	(0.475)	(0.613)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.328	0.551	0.798	0.889
		(0.818)	(1.662)	(1.659)	(1.912)
	$(station \geq 2008) * (year \geq 2008)$	2.851	3.031	1.392	1.556
		(2.021)	(2.403)	(1.891)	(2.542)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.216	0.225	0.501	0.614
		(1.649)	(1.844)	(1.176)	(1.878)
	$(station \geq 2009) * (year \geq 2008)$	1.949	2.352	0.981	1.016
		(1.633)	(2.277)	(1.657)	(1.648)
2 km	$(station \geq 2003) * (2008 > year \geq 2003)$	1.763	1.835	0.579	0.605
		(1.676)	(1.829)	(1.662)	(1.790)
	$(station \geq 2003) * (year \geq 2008)$	0.825	1.027	0.172	0.225
		(0.621)	(0.973)	(0.089)	(0.142)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.146	1.162	0.628	0.791
		(1.654)	(1.793)	(1.244)	(1.750)
	$(station \geq 2008) * (year \geq 2008)$	2.202	2.911	1.185	1.295
		(1.661)	(2.281)	(1.676)	(1.986)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.456	0.512	0.578	0.603
		(2.151)	(2.653)	(1.656)	(2.003)
	$(station \geq 2009) * (year \geq 2008)$	2.271	2.862	1.552	1.068
		(1.706)	(2.273)	(1.685)	(1.687)
4 km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.981	1.148	0.278	0.423
		(1.654)	(2.199)	(0.921)	(1.652)
	$(station \geq 2003) * (year \geq 2008)$	0.967	1.201	0.212	0.278
		(0.729)	(1.055)	(0.113)	(0.168)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.695	0.852	0.732	0.813

	(1)	(2)	(3)	(4)
	(1.221)	(1.661)	(1.386)	(1.886)
$(station \geq 2008) * (year \geq 2008)$	1.793	1.916	1.663	1.735
	(1.732)	(1.912)	(1.691)	(2.070)
$(station \geq 2009) * (2008 > year \geq 2003)$	0.381	0.458	0.481	0.545
	(1.180)	(1.665)	(1.033)	(1.548)
$(station \geq 2009) * (year \geq 2008)$	1.889	2.878	1.026	1.211
	(2.373)	(4.389)	(1.177)	(1.821)
Distance to CBD*Trends	Yes	Yes	Yes	Yes
Distance to Stations>0.5/1/2/4KM	Yes	Yes	Yes	Yes
Parcel Characteristics	Yes	Yes	Yes	Yes
Treatment dummies	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Treatment Criteria 1*Time	Yes	Yes	Yes	Yes
Treatment Criteria 2*Time	Yes	Yes	Yes	Yes
Distance to Olympic Park*Time	Yes	Yes	Yes	Yes
Distance to New Residential Area$_i$*Time	Yes	Yes	Yes	Yes
Station-distance*Time	Yes	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes	Yes
Location-specific characteristics	Yes	Yes	Yes	Yes
Observations	1181	1826	707	1036
Adjusted R-squared	0.389	0.431	0.322	0.346

Notes.---The dependent variable is the log of land prices. This table reports the estimates of treatment effects from spatially selected data samples. Specifications 1-2 are based on the residential land parcel sample within the central city and suburb respectively. Specifications 3-4 are based on the commercial land parcel sample within the central city and suburb respectively. t-statistics in parentheses, clustered on zone unit.

Table 2.10 Regression estimates of effective proximity impacts of new lines, sensitivity analysis

Station sample	Variables	Residential land parcel sample (1)	(2)	Commercial land parcel sample (3)	(4)
Cross_station	$(station \geq 2003) * (2008 > year \geq 2003)$	0.258	0.251	0.819	0.694
		(1.870)	(1.832)	(1.777)	(1.689)
	$(station \geq 2003) * (year \geq 2008)$	0.066	0.086	0.649	0.531
		(0.402)	(0.534)	(0.994)	(0.800)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.915	0.863	1.242	0.988
		(1.743)	(1.672)	(2.168)	(1.743)
	$(station \geq 2008) * (year \geq 2008)$	3.321	3.179	1.684	1.297
		(2.232)	(2.267)	(2.190)	(1.671)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.454	0.444	-0.829	-0.794
		(2.009)	(1.991)	1.946	1.873
	$(station \geq 2009) * (year \geq 2008)$	2.544	2.282	1.904	1.907
		(1.889)	(1.715)	(2.159)	(2.172)
Simple_station	$(station \geq 2003) * (2008 > year \geq 2003)$	1.738	1.069	0.598	0.585
		(3.201)	(2.056)	(1.718)	(1.671)
	$(station \geq 2003) * (year \geq 2008)$	0.140	0.101	0.318	0.311
		(0.893)	(0.669)	(0.779)	(0.766)
	$(station \geq 2008) * (2008 > year \geq 2003)$	1.662	1.528	0.660	0.636
		(2.537)	(2.344)	(1.875)	(1.797)
	$(station \geq 2008) * (year \geq 2008)$	4.807	4.044	0.951	0.916
		(3.371)	(2.799)	(1.645)	(1.699)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.719	0.683	0.383	0.406
		(1.858)	(1.798)	(1.079)	(1.150)
	$(station \geq 2009) * (year \geq 2008)$	4.693	4.455	1.342	1.511
		(4.245)	(3.942)	(1.525)	(1.670)
Simple_to_Cross station	$(station \geq 2003) * (2008 > year \geq 2003)$	1.197	0.992	1.131	1.035
		(2.196)	(1.830)	(1.827)	(1.677)
	$(station \geq 2003) * (year \geq 2008)$	1.609	1.449	0.768	0.533
		(0.753)	(0.679)	(1.180)	(0.811)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.875	0.864	1.437	1.403
		(1.953)	(1.942)	(2.123)	(2.091)
	$(station \geq 2008) * (year \geq 2008)$	4.158	3.849	2.305	2.070
		(3.194)	(2.988)	(2.333)	(2.151)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.634	0.612	1.003	0.870
		(2.120)	(2.054)	(1.990)	(1.723)
	$(station \geq 2009) * (year \geq 2008)$	3.832	2.901	1.360	1.286
		(3.840)	(2.994)	(1.843)	(1.752)
Distance to CBD*Trends		No	Yes	No	Yes
Distance to Stations>0.5/1/2/4KM		Yes	Yes	Yes	Yes
Parcel Characteristics		Yes	Yes	Yes	Yes
Treatment dummies		Yes	Yes	Yes	Yes

Time dummies	Yes	Yes	Yes	Yes
Treatment Criteria 1*Time	Yes	Yes	Yes	Yes
Treatment Criteria 2* Time	Yes	Yes	Yes	Yes
Distance to Olympic Park* Time	No	Yes	No	Yes
Distance to New Residential Area$_i$*Time	No	Yes	No	Yes
Station-distance* Time	No	Yes	No	Yes
Fixed effect	Yes	Yes	Yes	Yes
Location-specific characteristics	Yes	Yes	Yes	Yes
Observations	2343	2343	1341	1341
Adjusted R-squared	0.423	0.441	0.311	0.334

Notes.---The dependent variable is the log of land prices. Specifications 1-2 are based on the residential land parcel sample. Specifications 3-4 are based on the commercial land parcel sample. The sample sizes are the same as the baseline resulting tables. All specifications are based on treated parcels that experienced distance reductions and the outcome distance to the nearest stations are now within the 2km distance band. t-statistics in parentheses, clustered on zone unit.

Table 2.12 Regression estimates of spatial spillover effects, sensitivity analysis

Variables	(1)	(2)	(3)
Dist$*(station \geq 2003) * (2008 > year \geq 2003)$	0.011	0.020	-0.021
	(1.222)	(0.153)	(0.375)
Dist$*(station \geq 2003) * (year \geq 2008)$	0.033	0.080	-0.095
	(1.031)	(0.320)	(0.429)
Dist$*(station \geq 2008) * (2008 > year \geq 2003)$	0.020	0.010	-0.024
	(1.001)	(0.250)	(0.381)
Dist$*(station \geq 2008) * (year \geq 2008)$	0.040	0.010	-0.029
	(0.801)	(0.166)	(1.223)
Dist$*(station \geq 2009) * (2008 > year \geq 2003)$	0.040	0.010	-0.011
	(1.000)	(0.250)	(1.911)
Dist$*(station \geq 2009) * (year \geq 2008)$	-0.080	0.010	-0.021
	(-1.143)	(0.142)	(2.131)
Distance to CBD*Trends	Yes	Yes	Yes
Distance to Stations>0.5/1/2/4KM	Yes	Yes	Yes
Parcel Characteristics	Yes	Yes	Yes
Treatment dummies	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes
Treatment Criteria 1* Time	Yes	Yes	Yes
Treatment Criteria 2* Time	Yes	Yes	Yes
Distance to Olympic Park* Time	Yes	Yes	Yes
Distance to New Residential Area$_i$ * Time	Yes	Yes	Yes
Station-distance*Time	Yes	Yes	Yes
Fixed effect	Yes	Yes	Yes
Location-specific characteristics	Yes	Yes	Yes
Observations	2343	1341	2343
Adjusted R-squared	0.428	0.278	0.439

Notes.---This table reports the estimates of spillover effects. The within-group spillover effects estimates are shown on model specification 1 and 2 based on residential and commercial land parcel sample respectively. The sample sizes are the same as the baseline resulting tables. Estimates of cross-group spillover effects from commercial parcels to residential parcels are shown on specification 3. In specifications 1-2, *Dist* represents a series of distance (in kilometre) interactions between parcels in the subsequent treatment group and parcels in the prior treatment group, as described more details in the text. In specification 3, *Dist* means the interactions of the distance (in kilometre) between treated commercial parcels and treated residential parcels with each residential treatment effect. All specifications are based on treated parcels that experienced distance reductions and the outcome distance to the nearest stations are now within the 2km distance band. t-statistics in parentheses, clustered on zone unit.

Table 2.12 Regression estimates of interaction effects, sensitivity analysis

Distance band	Variables	Residential land parcel sample			Commercial land parcel sample		
		Educational attainment	Employment accessibility	Crime	Educational attainment	Employment accessibility	Crime
0.5km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.003	0.011	-0.076	0.143	0.038	-0.059
		(0.044)	(0.408)	(2.235)	(0.177)	(0.975)	(-2.565)
	$(station \geq 2003) * (year \geq 2008)$	-0.273	0.154	-0.256	0.154	0.077	-0.135
		(-1.079)	(0.526)	(-1.939)	(0.726)	(0.681)	(-0.912)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.044	0.059	-0.096	0.034	0.039	-0.079
		(0.611)	(1.475)	(-3.01)	(0.213)	(0.848)	(-1.491)
	$(station \geq 2008) * (year \geq 2008)$	0.222	0.138	-0.244	0.362	0.325	-0.287
		(1.187)	(0.484)	(-2.103)	(1.716)	(1.593)	(-2.009)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.052	0.023	-0.005	0.058	0.054	-0.009
		(0.788)	(0.639)	(-0.278)	(0.503)	(1.176)	(-0.221)
	$(station \geq 2009) * (year \geq 2008)$	0.087	0.028	-0.013	0.096	0.001	-0.443
		(0.323)	(0.092)	(-0.157)	(0.382)	(0.007)	(-3.852)
1km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.138	0.064	-0.007	0.296	0.085	-0.016
		(1.816)	(1.685)	(-0.121)	(2.176)	(1.667)	(-0.262)
	$(station \geq 2003) * (year \geq 2008)$	0.937	0.239	-0.201	0.054	0.293	-0.225
		(1.583)	(1.067)	(-1.142)	(0.185)	(1.296)	(-1.271)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.191	0.005	-0.026	0.197	0.058	-0.009
		(1.073)	(0.172)	(-0.473)	(1.225)	(0.925)	(-0.148)
	$(station \geq 2008) * (year \geq 2008)$	2.071	4.837	-0.782	2.268	2.676	-0.036
		(3.277)	(2.072)	(-1.367)	(2.187)	(1.988)	(-0.165)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.032	0.028	-0.016	0.131	0.075	-0.053
		(0.481)	(0.622)	(-0.941)	(0.824)	(1.019)	(-1.104)
	$(station \geq 2009) * (year \geq 2008)$	0.148	1.156	-0.063	2.082	0.966	-0.145
		(0.534)	(1.883)	(-0.488)	(2.511)	(1.845)	(-0.879)

2km	$(station \geq 2003) * (2008 > year \geq 2003)$	2.114	0.105	-0.004	0.266	0.378	-0.166
		(4.161)	(2.283)	(-0.058)	(1.750)	(1.979)	(-0.933)
	$(station \geq 2003) * (year \geq 2008)$	0.191	0.325	-0.292	0.292	0.712	-0.809
		(1.073)	(1.109)	(-0.598)	(0.861)	(0.698)	(-0.967)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.281	0.026	-0.028	0.208	0.942	-0.322
		(1.965)	(0.116)	(-0.444)	(0.504)	(0.661)	(-0.578)
	$(station \geq 2008) * (year \geq 2008)$	3.007	4.660	-0.353	1.751	2.115	-2.793
		(1.755)	(2.149)	(-1.587)	(1.689)	(1.779)	(-1.623)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.068	0.022	-0.012	0.568	1.213	-0.211
		(1.243)	(0.688)	(-0.800)	(1.303)	(0.719)	(-1.148)
	$(station \geq 2009) * (year \geq 2008)$	2.382	2.521	-0.061	1.979	1.185	-0.389
		(4.436)	(2.942)	(-0.457)	(2.213)	(1.787)	(-1.154)
4km	$(station \geq 2003) * (2008 > year \geq 2003)$	0.462	0.646	-0.039	0.332	0.236	-0.163
		(3.756)	(1.737)	(-0.582)	(1.677)	(2.165)	(-0.896)
	$(station \geq 2003) * (year \geq 2008)$	0.672	0.316	-0.202	0.311	0.642	-2.751
		(1.566)	(1.295)	(-1.270)	(0.816)	(0.633)	(-1.597)
	$(station \geq 2008) * (2008 > year \geq 2003)$	0.186	0.070	-0.027	0.356	0.901	-0.935
		(1.420)	(0.731)	(-0.519)	(0.866)	(0.632)	(-0.962)
	$(station \geq 2008) * (year \geq 2008)$	0.969	3.046	-0.218	1.071	1.439	-1.782
		(2.612)	(1.765)	(-0.965)	(2.052)	(2.129)	(-1.129)
	$(station \geq 2009) * (2008 > year \geq 2003)$	0.039	0.044	-0.012	0.255	1.162	-0.144
		(0.283)	(0.201)	(-0.185)	(0.593)	(0.689)	(-1.321)
	$(station \geq 2009) * (year \geq 2008)$	2.064	1.636	-0.106	0.654	3.395	-0.456
		(4.291)	(3.752)	(-0.404)	(2.003)	(2.066)	(-0.889)
	Observations		2343			1341	

Notes.---This matrix table can be viewed as two parts with respect to the residential and commercial land parcel sample respectively. The sample sizes are the same as baseline results. Each part of the table reports the estimates of the interactions between treatment effect variables and educational attainment, employment accessibility, crime rates from one single regression. The regressions shown in the table also include a full set of controls. t-statistics in parentheses, clustered on zone unit.

Figure list

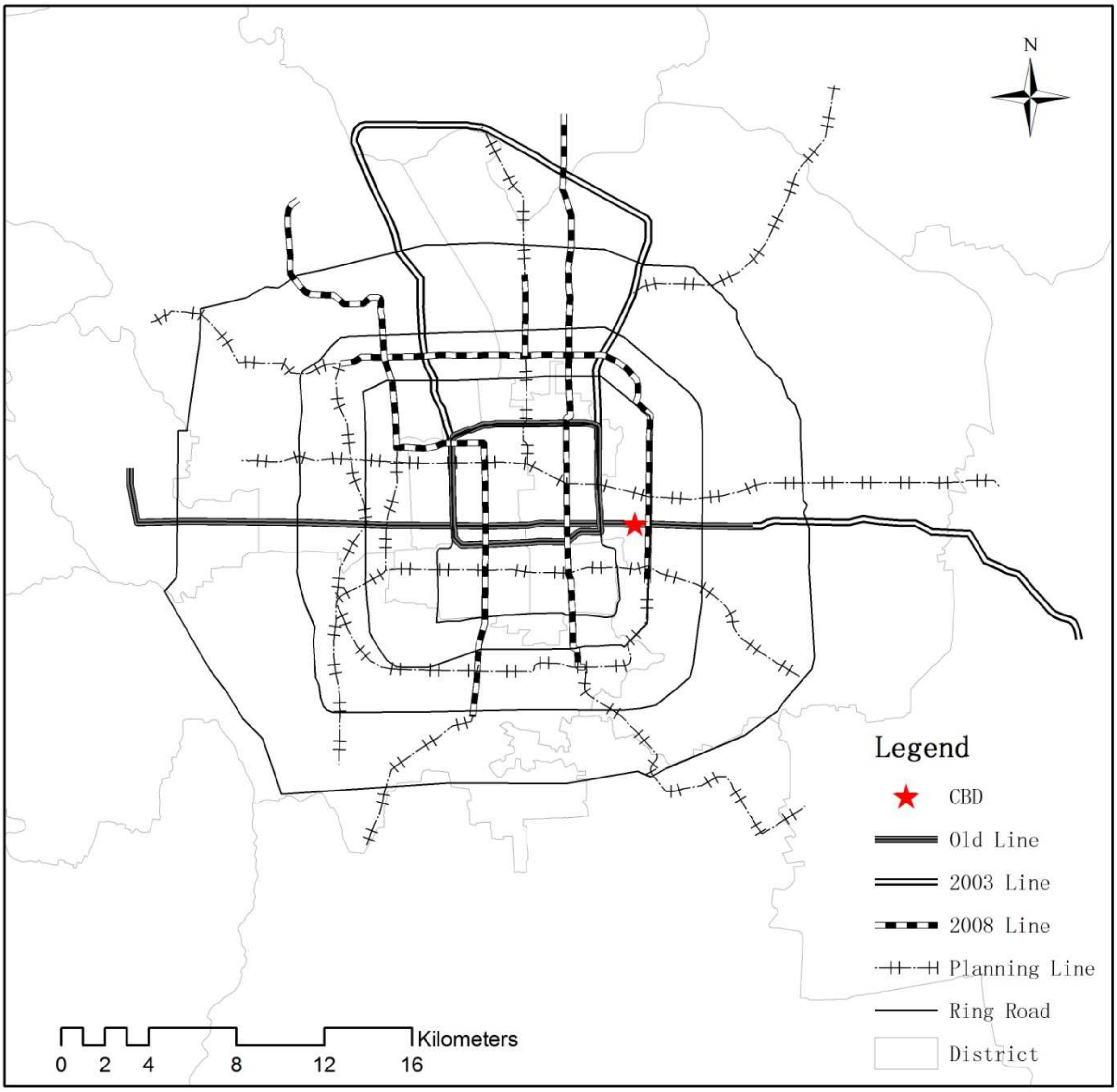


Figure 2.1 Rail transit network in the Beijing urbanized area

Notes.---Old lines were opened before 2003; 2003 lines were opened in 2003; 2008 lines were opened in 2008; Planned lines will open after 2009. See detailed explanation in Section 3.2.

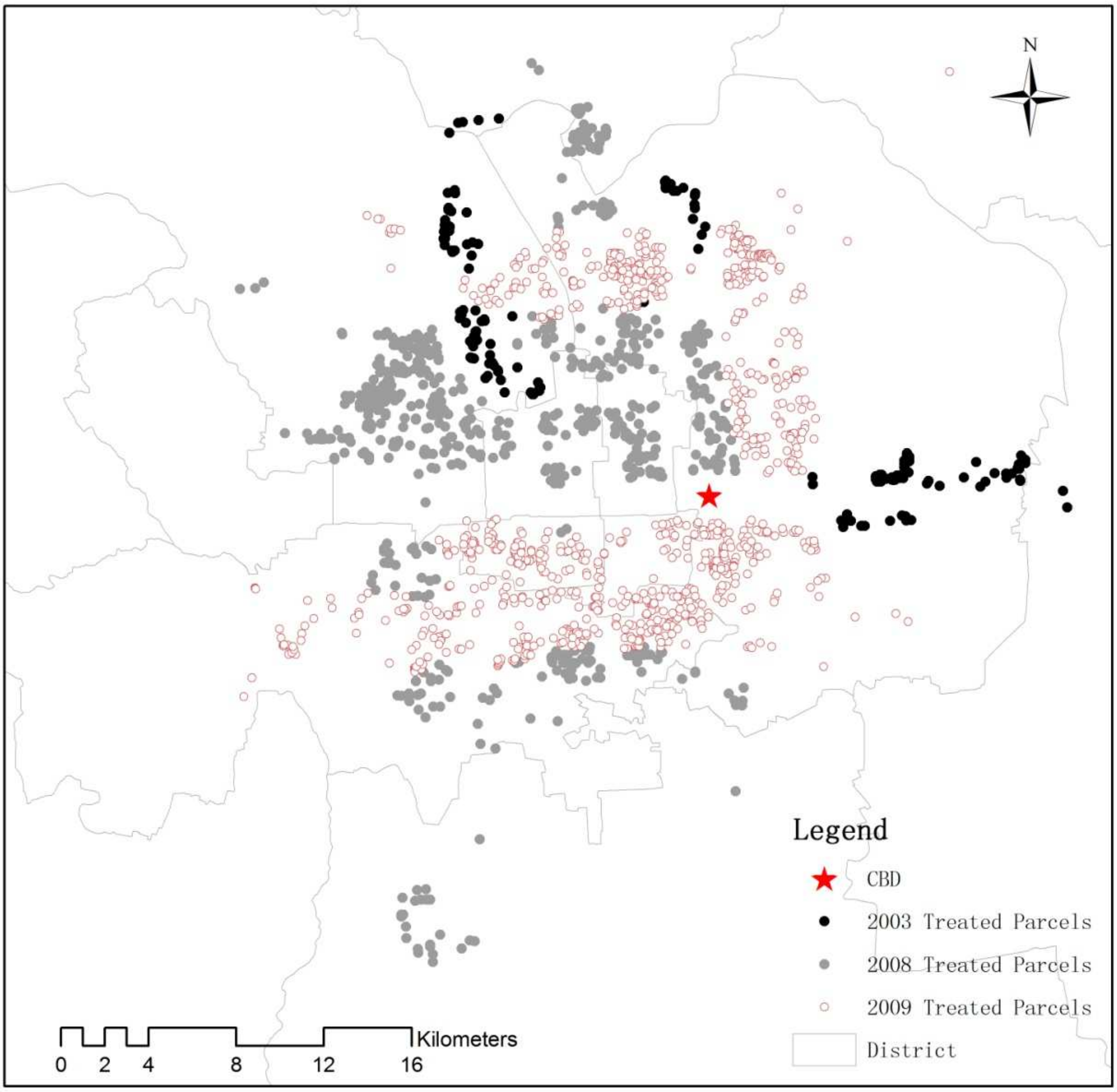


Figure 2.2 Spatial distributions of treated residential land parcels

Notes.---“2009 Treated Parcels” refer to the parcels in the *Treatment*$_3$ (station $\geq$ 2009); In comparison to the *Treatment*$_3$, “2008 Treated Parcels” are the additional parcels that belong to the *Treatment*$_2$ (station $\geq$ 2008). In comparison to the *Treatment*$_2$, “2003 Treated Parcels” are the additional parcels that belong to the *Treatment*$_1$ (station $\geq$ 2003). All treated parcels are selected using the 2km distance band.

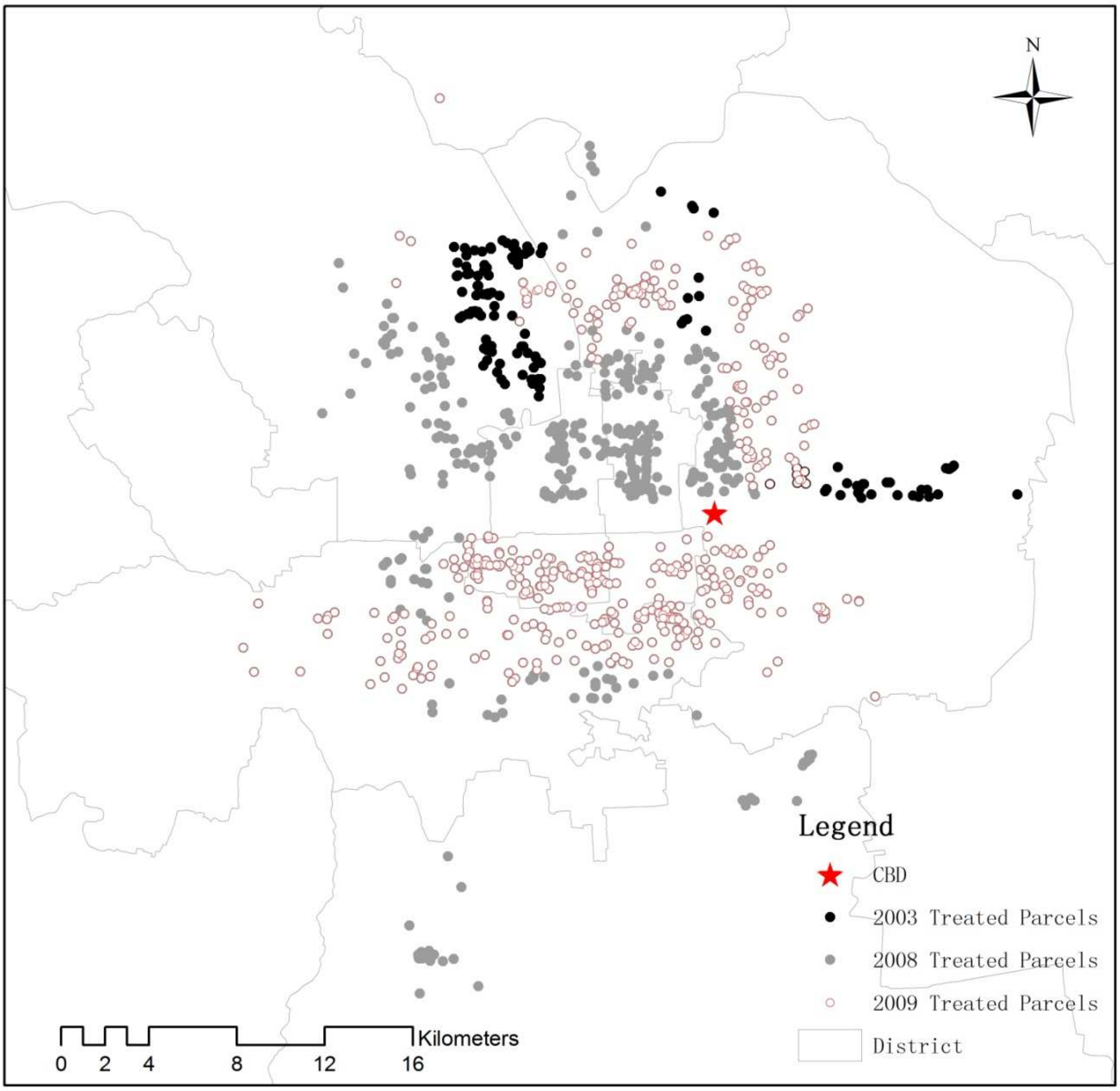


Figure 2.3 Spatial distributions of treated commercial land parcels

See notes to Figure 2.2 for details.

IV. Paper 3---Does Better Rail Access Improve Homeowners' Happiness?: Evidence Based on Micro Surveys in Beijing

1 Introduction

Transport infrastructure is undoubtedly believed as an important part of government investment program and is of great importance for homeowners' living experiences. Many developing countries like China are implementing transport policies to invest in new rail transit constructions. Recently, four new rail transit lines were opened around 2008 Beijing, with the total investment of 42.4 billion CNY (1GBP≅10 CNY). From a policy perspective, this transport improvement program is conducted against the backdrop of broad public conversations of residents' happiness in Beijing through the "Towards Livable City" initiative since 2004. This agenda has driven important changes in transport services in order to reflect several key outcomes for people's residential happiness with respect to "commuting convenience", "living convenience", "traffic pollution", "traffic safety", and "social environment" (Zhang et al, 2006). The question then arises as to whether these kinds of objectives can be met in the 2008 transport improvement context, where policymakers tend to judge the success of transport investment program solely on the basis of economic census data. While most researchers value the amenity benefits of rail access in the real estate markets (Gibbons and Machin, 2008), little is known about whether this is mirrored in higher levels of happiness with respect to these different dimensions of the residential environment.

In this paper, I provide an alternative (direct) way of estimating the impact of the transport improvement program, identified by rail access changes, on homeowners' happiness (rather than e.g. house price or looking at other economic outcomes) [67]. My

[67] Recall that this paper does not attempt to identify the anticipation effects of new stations on people's happiness and related housing price changes, residential mobility or neighbourhood dynamics. Instead, this paper typically focuses on examining the direct impact of the increased station proximities on homeowners' happiness, as opposed to the indirect effect from the fact that local residents may become wealthier because of the increased values of their homes.

outcome measures are based on detailed and repeated survey responses that allow specifications about happiness with respect to different dimensions of residential environment---commuting and living convenience[68], social environment, traffic pollution and safety, rather than only one general life happiness indicator. My main goal is to consider two related research questions, that is: i) To what extent are happiness in specific residential aspects, amongst homeowners, linked to rail access based on measures of residence-station distance changes? ii) To what extent are homeowners' perceptions of better rail access varied based on their different social backgrounds (i.e., income and age)? To answer these questions I aggregate the micro surveys into an area panel, which contains a rich set of repeated happiness responses and individual background characteristics, and which I have matched to rail access on homeowners' places of residences. To my knowledge, this is the first application of this type of analysis to the happiness studies in the developing countries. I deal with the central problem of the potential endogeneity in sorting effects by focusing on "stayers" [69] and by using the non-market housing---a legacy from the socialist welfare housing system with pre-determined locations and non-market transactional rules. Using the difference-in-difference style estimation strategy, I will first run the regressions for the whole sampled "stayers", and then for the non-market housing sampled homeowners, to further verify the impact of rail access changes on happiness before-and-after the building of new rail transit lines.

Another contribution of this paper is to explore the potential welfare benefits of improvements in rail access on the Chinese homeowners' happiness with respect to

[68] Note that the living convenience indicates residents' happiness about the convenience to use public transits to do non-working activities, whilst the commuting convenience indicates residents' happiness about the convenience to use public transits to work. See detailed description of the definition of each happiness indicator in the appendix table 3.1.

[69] Note that the term of "stayers" here means homeowners who were living at their homes before the transport was improved.

different dimensions of residential environment. By measuring the marginal utility of rail access and the marginal utility of income, compensating variation between income and rail access can be calculated. This welfare measure has recently been used elsewhere in the literature to evaluate subjective benefits of air quality based on reported happiness data and have useful implications in the benefit-cost analyses for evaluating public policies (see Luechinger, 2009; Frey et al., 2010). This paper improves on previous studies by quantifying both of the average and distributional benefits of the transport improvement program.

The remainder of this paper is structured as follows. Section 2 reviews related literature. Section 3 describes the data and institutional background. Sections 4 explains the methodology. Section 5 presents the main findings on the impact of rail access changes and homeowners' happiness with respect to different dimensions of residential environment. Section 6 monetises the welfare effects of transport improvement program. Section 7 concludes.

2 Literature review

Researchers of sociology and geography have often used survey data to elicit household preferences for transport facilities (Lu, 1999; Parkes et al., 2002; Mohan and Twigg, 2007; Adriaanse, 2007; Hur and Morrow-Jones, 2008; Permentier et al., 2011), but analysis of the perceived assessments for transport accessibility remains a relatively unexplored research field in urban economics. In fact, a large volume of economic literature has focused on examining the net benefits of rail access by using the reveal preference techniques---like the hedonic valuation approach. Assessment of these net benefits from changes in rail access is usually valued based on nearby housing prices (some excellent hedonic applications include Cheshire and Sheppard, 1995; Bowes and Ihlanfeldt, 2001; McMillen and McDonald, 2004; Gibbons and Machin, 2008). However, one potential concern of this reduced-form approach is that one cannot separately identify the direct and indirect benefits associated with rail access. Whilst it is true that property-price outcomes matter, there may be wider aspects of socio-psychological developments that are at least as important as price premiums in evaluating the amenity benefits.

Using perceived happiness survey questions, economists can better single out the direct relationship between local public goods and people's subjective wellbeing (often loosely called as happiness[70]). For example, Luechinger (2009) finds the negative effect of air pollution on happiness based on individual survey data in Germany. Cornaglia and Leigh (2011) use an area panel data from Australia to estimate the direct impact of changes in crimes on mental wellbeing of resident non-victims. They find that crime---especially the type of violent crime rates have a negative impact on people's mental wellbeing. Gibbons and Silva (2011) find a strong impact of school quality, measured by test scores, on parents' happiness about education effectiveness based on the longitudinal survey of young people in England. Indeed, recent literature in happiness economics also point out that the estimated happiness effects can avoid some problems inherent in the hedonic method (see Frey et al., 2009 for a review). For example, the assumptions of the happiness approach can be less restrictive than the hedonic approach since it is not based on observed behaviours. Recall that the hedonic approach is based on the underlying assumption that housing and labour markets are in fully spatial equilibrium. To meet this assumption, households should have enough market information, the land and housing

[70] Happiness is considered as a fundamental measurement of human subjective wellbeing (Campbell et al, 1976). It is naturally the topic of socio-psychology, medicine, and health research, and has recently expanded its focus on people's happiness about residential environment. See Layard (2006) and Frey et al (2008) for details.

supply should be sufficient, and the moving costs in the housing and labour markets should be very low (Freeman, 2003). Yet in reality, these assumptions associated with the hedonic approach cannot be fully meet at certain local contexts. Conversely, the happiness approach can explicitly capture utility gains or losses even without such market equilibrium assumptions. Though the self-reported happiness data may not as accurate as housing transaction data, it is still an effective way to evaluate local public goods in utility terms (Blanchflower and Oswald, 2004; Krueger and Schakde, 2007; Oswald and Wu, 2010). By measuring the marginal utility of a specific local amenity and the marginal utility of income, the trade-off ratio between income and that particular local amenity can be calculated. Indeed, this happiness approach has recently become one of the promising development in economics and has been used elsewhere to evaluate a wide range of public goods like air quality (Luechinger, 2009; Frey et al., 2010) and slum improvements (Takeuchi et al., 2008). My analysis adds to this growing literature by providing new evidence on the direct effect from rail access changes to homeowners' happiness in particular residential aspects [71].

3 Institutional Settings and Data

In this section I first outline the institutional background about the housing reform in China. I then go on to explain the micro data involved into the empirical analysis.

3.1 The housing reform in urban China

To better understand the exogenous nature of non-market (*fang gai*) housing, this section briefly introduces the housing reform policy background, with the key focus

[71] It is certainly the case that combined estimates from both of the hedonic valuation approach and the happiness approach would offer more precise information about the rail access effects, but collection of micro housing transaction data with precise geographical characteristics would be very costly and not publicly accessible in Beijing. Some comparisons of hedonic and wellbeing measures can be found in Van Praag and Baarsma (2005), Gibbons and Silva (2011), among others.

on the non-market (*fang gai*) housing in China.

Before the housing reform policy launched in the 1990s, no housing market was existed and housing was not a commodity in China (see Logan et al., 2010 for details). All housing units were provided by work unit (*Danwei*) to their employees as employee welfare. Under the centralized planning-economy era, urban lands were owned by the state and were allocated to work units. A work unit typically constructed housing units on its allocated lands, and then assigned them to its employees based on their job ranking and working life length, etc (Fu et al., 2000). All work-unit housings are owned by the employers, and their employees did not have to pay or only paid very low fees for renting. All urban workers did not need to choose their residential locations.

In the reform era, housing market has been gradually established. Real estate developers began to construct and sell market housing to households (Zheng et al., 2006). Meanwhile, the central government of China stopped to offer the lands for constructing work-unit housings based on the 1998 housing reform policy (see Huang and Clark, 2002). But most work units continued to provide heavy subsidized housings through the “internal housing market” (Sato, 2006). All of these work-unit housings were privatized by selling to their employees at low prices and were commonly called the non-market (*fang gai*) housings. Due to the historical policy reasons mentioned above, the pre-determined location nature of non-market (*fang gai*) housings can be regarded as exogenous. Thus the baseline robustness test examined in this context is to use the sampled non-market housing homeowners to account for potential endogeneity in residential sorting. One thing to note is that, the effect of work-unit housing privatization may not impose an immediate wealth transfer. This is because that although work units transferred the ownership to their employees, resale of non-market (*fang gai*) housings is not allowed. Recently, this non-transaction rule has been gradually relaxed in some Chinese cities, however, the actual transition of *fang gai* housings into fully market housings in Beijing is restrictedly limited[72]. Notably, homeowners who hold the non-market housing tenure may not actually live in their non-market housings. Thus in this study, I only focus on the sampled homeowners who hold the non-market housing tenure and currently living in the non-market housing during my study period.

3.2 Data

My analysis is estimated using households' happiness data from Beijing, China. Beijing is largely a mono-centric city in terms of population density, land and housing price gradients (Zheng and Kahn, 2008). The *JianGuoMenWai* area is conventionally viewed as the central business district (CBD). The main Beijing' urbanized area is within the No.5 ring road, with a small proportion located outside the No. 5 ring road in the north and east directions. This area comprises more than 60% percent of the metropolitan population with just over 10 million residents in 2000.

This study adopts a unique micro survey dataset of Beijing residents that include two large-scale surveys conducted in 2005 and in 2009 respectively. The data samples for each surveys is about 11,000 respondents[73]. The surveys provide rich information on a household's demographic characteristics and residential happiness conditions[74]. For each member of the household roster, the survey reports age, income[75], education, family size, job rank, place of residence, commuting time and modes. The

[72] In some cases, the sale of former work-unit housings had additional limitations like the owner can sale the property back to the work-unit or other employees in this work-unit.
[73] The effective response rate is about 79% in the 2005 survey and 72% in the 2009 survey.
[74] The happiness survey questions are shown in the appendix A table.
[75] Note that I have converted the categorized income information into the mid-point value of the respective categorical interval. Since the highest income category is open-ended, I predict the mid-point value of this category by using the sample's normalizing distribution. All monetary values are adjusted by the Beijing consumer price indices and reported as CNY (1GBP≅10CNY).

household's ownership identity is given[76]. In addition, the surveys document detailed living conditions such as the housing's type (non-market housing or market-housing), the duration living in this residential location, housing size, as well as local residents' happiness in specific residential aspects, such as commuting convenience, traffic safety, social environment, traffic pollution, and living convenience. There are several key characteristics of this survey dataset: i) It has large sample size that covered Beijing's main urbanized area instead of selected sample areas; ii) Its samples were selected randomly and proportional to the population at each zone (*jiedao*). Zone is the fundamental administrative organization and census unit in China. While zones generally are aggregations of residential places, they do not reflect the boundaries of political jurisdictions like the developed countries. Zones are intended to be similar areas with respect to general socio-demographic characteristics; iii) The unit of the survey was Beijing households, excluding the floating population or travellers who had been in Beijing for less than six months. Such sampling strategy enables all the respondents to be familiar with their living environment; iv) This micro data appears to be reasonably representative. A comparison of 2005 sampled household demographics with data from the 2005 Beijing Population Survey revealed no significant differences[77]. In the empirical analysis, I will use the sample of the homeowner head ages 18-65 who work and have lived in the current residences for at least five years. The underlying reasons are that: As 2008 new subway lines have mainly been started to construct since 2003, this sample restriction can help to guarantee that these homeowners are not unemployed or new movers into the current places of residence due to their preference for the expected job opportunities and

[76] As for housing property types, about 53.6% households own non-market housing unit in the 2005 survey, and this figure remains stable in the 2009 survey (52.1%). The other households own market housing units. The survey's non-market housing ownership ratio tells a consistent story with the overall non-market housing ownership ratio in Beijing.

[77] One potential source of bias resulted from oversampling employees, in order to get households' commuting characteristics.

improvements in rail access[78]. As such it allows me to focus on the homeowners independent of job searching and residential sorting concerns. In addition, I am well aware that my sampled homeowners include both public transport users and non-public transport users. It is expected that public transport users might gain more benefits from the new rail transit constructions. However, it would be useful to know if the happiness effects for public transport users were offset by the nuisance effects for everyone else. I will look at the public transport users' happiness results as a special case in the sensitivity analysis.

The measure of the happiness in specific residential aspects, is based on responses to the survey question[79]: "How well do you satisfy your residential location" with respect to "its particular local (neighbourhood) characteristics, such as commuting convenience, social environment, etc" on a scale from "1 being very unhappy" to "5 being very happy". One alternative answer was "not familiar" and this was discarded for the purpose of this research (less than 1% of respondents gave such a response). Recent literature in happiness economics has often assumed that respondents are able and willing to answer the happiness questions; and there is the significant difference between a respondent with a happiness score of 5 and the one with a score of 4 (see Ferrer-i-Carbonell, 2005 for details). Aragonés et al. (2002) also find that implicit wording can help minimize social-desirability bias by pushing people to report higher happiness levels. I find that responses not to be right-skewed and the distribution patterns of key happiness questions have no significant differences across 2005 and 2009 survey samples.[80] Another issue relates to the

[78] This sample restriction can also help to avoid the fact that any potential increases in happiness brought by the local public goods improvements may be offset by rising housing costs for residents facing market housing costs.
[79] Note that both of the two surveys have the same happiness questions.
[80] To better visualize this, I plot out the happiness distributions of commuting convenience and living convenience across 2005 and 2009 surveys and find quite similar patterns (see appendix B figure for details). Also note that the Pearson Chi-squared tests show that the distributions of all happiness questions have no significant differences

question wording is that, it does not specify what is the definition of the term "local (neighbourhood)"? This means that, even for households living in the same geographic place, they may have different concepts in mind when answering questions about their happiness of local (neighbourhood) characteristics. However, this question-and-answer formulation largely holds as the standard in the happiness studies (Kahneman and Krueger, 2006). It is expected that the concept of local (neighbourhood) to be consistent throughout the survey respondents (Galster and Hesser, 1981; Lu, 1999).

In order to look at the rail transit changes before and after 2008, I aggregate homeowners' happiness evaluations to the 1km^2 cell-unit groups[81] in two-survey time periods: 2005 and 2009. Then I geographically-coded the newly-opened subway stations in 2008 with the help of the GIS software. The spatial straight-line distance from a cell unit's central point to the closest station is defined as this cell unit's rail access[82]. The rationale behind this is that, it allows me to use the repeated average responses in the same geographical unit, as opposed to repeated individual responses of the same household given the data sample size limitation[83]. Ideally it would be perfect to find a geographical space that can yield perfectly homogeneity in the characteristics of each location. But further disaggregation would not provide enough sampled residents for the empirical analysis.

In a nutshell, my data is not a panel of people but a panel of areas, and I try to control for potentially endogenous changes to the compositions in response to

across 2005 and 2009 surveys (see appendix B table). This result is important because it simultaneously supports the consistency of the surveys.

[81] It is necessary to emphasis that I have also tried to aggregate data to higher geographical-unit level like 2km^2 and even the zone (*jiedao*) level to explore the robustness to the choice of aggregation. The results do not make a markedly difference.

[82] In practice, I have taken care of this measurement to ensure that the closet stations are not inaccessible—for example if separated by the river or expressway, where few crossing points are available.

[83] Another underlying reason is that by matching area rail access changes to repeated area happiness responses, I am therefore able to mitigate the problem of the potential bias from the inconsistent individuals' perceptions about the local geographical area.

transport improvements by (a) including changes in the average demographics; (b) using long-term “stayer” sample (homeowner who were living there before transport was improved); (c) using non-market housing homeowner sub-sample with pre-determined residential locations. Once again, the baseline motivation of focusing on the “stayers” sample is to try to identify homeowners in the 2009 sample that have not selected themselves into the area as a result of the transport improvement. When one is reading the results, it is important to keep in mind that this identification strategy cannot fully ensure the people who moved out of the area are being representative. Indeed, there may be concerns that the selected sample are the most or the least responsive to the transport changes. For example, if the people who moved out were the ones who expected to be made unhappy by the transport improvements, the selected “stayers” sample may provide potentially an upper bound to the transport impacts.

Geographical information on location characteristics is taken from a variety of sources as additional controls. School location and performance data comes from the Beijing Municipal Committee of Education. The location of bus stops and expressways are used as proxies for the competing commuting modes, and is obtained by a web-based search from the Beijing Municipal Committee of Transport. Geographical data on the sites of rivers and parks is taken from the Beijing Water Authority and Beijing Municipal Garden Bureau respectively. Crime rates for the number of violent crimes taking place in each zone are obtained from the Beijing Public Security and Safety Bureau. The 2001 City Employment Census provides local employment density. The 2000 City Population Census reports the detailed local demographic characteristics such as population density, education attainment, public housing rent ratio, and the percentage of heritage buildings built before 1949.

3.3 Transport improvement in Beijing

In Beijing, the largest public infrastructure investment project that has taken place recently is the new rail transit constructions. As discussed above, I use the opening of Line 4, 5, 8A, and 10A in 2008 as the transport improvement program[84]. 10 old stations experienced substantial upgrade, but I consider only the 59 new stations here[85]. Indeed, these new subway lines were viewed as the most significant improvement in the Beijing subway network since the 1980s. Figure 3.1 shows the map of the Beijing subway network before and after the transport improvement. It is expected that the 2008 transport improvement program has altered the residence-station distance for some households, whilst left others unaffected. This provides me with an exogenous change in the distance between homeowners' residential locations and their nearest rail station, from which I can examine the impact of effective rail access changes on homeowners' happiness.

These place-based investments were not chosen randomly[86]. In order to better reflect the sitting process, it is necessary to overview the urban governance structure in Beijing. As the capital city of China, Beijing has three levels of its administrative system: Beijing municipality, district and zone (*jiedao*, it will be referred to as zone thereafter in this study). Following the convention, my study area mainly focuses on the eight urbanized districts (*Dongcheng, Xicheng, Xunwu, Chongwen, Chaoyang, Fengtai, Shijingshan, and Haidian*) since other districts are predominately rural areas with no rail transit lines. Public investment is highly centralized and controlled by the

[84] The construction of these new subway lines started mainly since 2003, and was completed in and around 2008. It should be noted that Line 5 was temporarily opened at October 2007, but fully opened at the beginning of 2008. To facilitate the interpretation, I treat all the four railway lines opened in 2008. As a robustness check, the results are identical when excluding station sample of the Line 5.

[85] Except for 6 over-ground stations, all the other new stations are in underground status. The results are robust to excluding the over-ground stations and to the inclusion of those 10 upgraded old stations.

[86] In section 3.4, I will test to what extents do the treatment and control places are balanced in terms of the pre-treatment characteristics.

Beijing municipal government. The zones (*jiedaos*) are only responsible for street cleaning and do not have the voting power for deciding the public infrastructure construction. In other words, the zone functions as a basic geographical area for data collection, not as a political unit using local revenue to offer public goods.

Based on a broad historical document search, the motivations behind the place-based investment decision can be summarized as follows: The primary reason for constructing new rail transit lines is to reduce congestion and meet the rapid growth of the commuting demand. The second aim is to strengthen the connections between the central city and the suburb, especially those emerging super-"bedroom" residential communities in the suburbs (such as *Tiantongyuan*, *Yizhuang*, *Daxing*, and *Tongzhou*). Finally, the Beijing municipal government has decided to built one short subway line (Line 8A, with only four stations) to connect the Olympic park with the main rail transit network. I could, in principle, examine the effects of these four new subway lines separately and go further by looking at individual-level new station effect. Nevertheless, I simplify the analysis by treating them as one single event since they occurred at the same time-period in Beijing. Given the importance of the political economy behind the transport improvement, it is important to control the distance to CBD, Olympic Park, large "bedroom" areas as well as other location characteristics that would contribute to the robustness of the rail access effects.

3.4 Characteristics of "treated" and "control" places

In this descriptive analysis section, I show results based on differences in the average happiness changes between affected places and unaffected places by the transport improvement. The results are based on the aggregated dataset, where aggregations are to the cell unit pre-/post the opening of new rail transit lines. There is no significant variation in cell unit-to-station distance within cells.

To be clear from the outset, I term groups of cell units as “treatment” and “control” groups, namely those affected by the transport improvement and those not affected by it. A cell unit is assigned to the treatment group[87] if:

1) It experienced a fall in station-distance with the opening of new rail stations in 2008;

2) The outcome station-distance in 2008 is now less than 2 km.

I impose the second selecting condition because this study has not attempted to measure how entire metropolitan areas’ residents are affected by new rail transit constructions. The choice of a 2 km distance band is based on existing empirical literature and a reasonable walking distance to a station (about 20 minutes) in Beijing. I am implicitly assuming that homeowners’ residences that are more than 2km away from rail stations are not affected by the treatment. The rationale behind this is fairly reasonable: even though homeowners’ from remote places (no distance reductions or larger than 2km station distance) might also become happier, the main impacts of new stations on homeowners’ happiness should be in places near the stations. Owing to the large sample size, I am able to use the 1km and 4km distance band to select the treatment group as a robustness check.

Table 3.1 summarizes the results of descriptive statistics. I have restricted attention to the whole sampled homeowners. I have also restricted attention to those cell units that are represented in the sample both before and after the transport improvement. Columns (1)–(6) of the table show the average distances to stations,

[87] Ideally each 1km^2 cell unit represents the 1*1km geographical area. However, in a few “treatment” cell units, they also include some homeowners’ places of residences that belong to the “control” group. To eliminate this overlap issue, I have used the Thiessen-polygon method to create the cell unit with relative flexible boundaries like the “jigsaw puzzle” based on the GIS software. Of necessity, this method has kept the whole area of each cell unit as 1km^2 and no spaces among cell units. An alternative strategy is to assign a probability for those “treatment” cell units that contained “controls”. To be specific, I define this probability according to proportion of homeowners that would be in each group. For example, if a cell unit contains 15 sampled homeowners, and if 10 out of 15 are the “treatments” and the left are the “controls”, then I will assign a probability of 0.75 in this treatment cell unit. As a robustness check, the results are virtually similar by applying this alternative strategy.

and happiness of five residential aspects, for the full sample, the "treatment" group and the "control" group[88], before and after the transport changes occurred. Column (7) presents the difference-in-difference estimates based on the raw data[89].

In line with all quasi-experimental works like this, a natural question is to ask: did the transport improvement really do what I expect to do, namely increase proximity to stations? As can be seen from the first row of Table 3.1, the answer is yes. The opening of new rail transit lines did provide distance reduction to stations of 1.2 km for the treatment group, whilst the controls also became slightly closer. This is because my "controls" include residences that had received distance reductions and were still beyond 2 km from the nearest station. From row 2 to row 6 of the Table 3.1, I report the mean value of happiness of each residential aspect before and after the transport improvement. The headline finding is that homeowners' happiness with respect to different dimensions of residential environment in the treatment group experienced effective changes relative to the control group. For example, homeowners are found to become happier about commuting and living convenience, on average, in affected areas after the transport improvement. Homeowners at treated places tend to show less satisfaction about social environment and traffic safety with the building of new rail stations. These results provide preliminary descriptions on the various channels through which the transport improvements might affect happiness. Column (7) tests this more formally by using a diff-in-diff based *t*-test estimator of the

[88] Recall that my "controls" are places that have never been experienced station-distance reductions and places that may have experienced station-distance reductions but the nearest station distance is still larger than 2km threshold. This research design allows me to identify how happiness changes for places experiencing big station-distance changes compared to happiness changes in places with smaller station-distance changes. Intuitively, there is a danger for mixing up the new rail transit's impact by including the places that are within 2km station-distance ball both before-and-after the building of 2008 rail stations into the controls. This can lead to the bias of the results that may only capture the average variations in the happiness changes of the controls. As an additional robustness check, it is necessary to re-run the empirical regressions by dropping out this group of control samples (See appendix C). While the qualitative nature remains the same, doing this does bring improvements in the treatment effects in terms of quantitative nature.

[89] This is the estimate $(x_1^{treatment} - x_0^{treatment}) - (x_1^{control} - x_0^{control})$ where x is the variable, period 1 and period 0 represent post-/pre-transport improvement, respectively.

differences in the average changes of happiness[90]. The difference in happiness of commuting convenience is strongly significant at the 5% level, showing that homeowners' happiness towards commuting convenience growth to be roughly 6.6% (100*[exp (0.064)-1]) higher, on average, in areas affected by the transport improvement. The relative happiness changes in other residential aspects, though slightly less, are still significant in statistical terms.

Figures 3.2-3.3 provide more evidence on this, which take the happiness about commuting convenience as an example to quantify this variation. To begin with I present a simple plot of the whole sampled homeowners' median happiness value before-against-after the transport improvement, within 2km of a new station (see Figure 3.2). The triangle-dots are those new stations at the central city, and star-dots are new stations at the suburb. The solid line is the 45 degree line. In Figure 3.3, I use the vertical deviation of each dot in Figure 3.2 from the 45 degree line to visualise the spatial variation of the median happiness changes at each new station area[91]. Perhaps surprisingly, most of new station areas---primarily at the suburb, lie well above the 45 degree line implying that they are relative high happiness improvement areas. In contrast, some central new stations lie slightly below the 45 degree line implying that they are relative negative happiness improvement places. One possible explanation is that homeowners living in the station areas of the central city may have experienced less distance reductions to new stations than those who live in the suburb station areas[92]. This could also be explained by the dilemma between the heavy transport demand in the central city and inadequate rail transit capacities and frequencies during

[90] I define a "treatment" group dummy and a "post" dummy and regress the log-happiness on the 'treatment' dummy, "post" dummy, their interactions, and cell unit fixed effects. Here the "post" dummy is a dummy variable that equals 1 in the 2009 survey time period following the opening of new subway lines in 2008.

[91] See full results of the vertical deviation of the happiness changes within 2km of each new station area in the appendix D.

[92] Below, I will do formal regression test for the differential impacts of rail access on happiness living in the suburbs versus the entire urbanized area.

the rush-hours.

The visualization of happiness changes shown in the above figures[93] is essentially the complimentary descriptions for the table results. One should not read too much into these tables and figures at this stage because I have not examined whether the differences in key observable pre-treatment characteristics of treated and control areas are statistically significant. For the most part, a *t*-test in the mean difference between column (3) and column (5) shows no obvious differences at the 5% significance level[94]. The only two imperfect variables on which the treated and control places do not appear to be well balanced are indicators about station-distance and happiness about traffic safety. For example, station-distance is relatively lower and happiness about traffic safety is relatively higher in the treatment group. One potential concern about the imperfect balancing is if the place-based transport investment and consequent rail access differentials, encourage sorting of households for places with higher happiness about traffic safety. In this sense, it is likely that I might do better in terms of control-treatment balancing by considering a restricted sub-sample of non-market housing homeowners whose pre-determined residential locations can be regarded as exogenous. I test this in Table 3.2, which uses the same treatment selection principles but focus typically on non-market housing homeowners. Doing this does bring improvements in the treatment-control balancing conditions, where a *t*-test of the differences in mean between column (3) and column (5) shows no differences at the 5% significance level. Importantly, it does not make significant difference to the main results, showing that these descriptive statistics are not

[93] I have also investigated the median happiness changes relative to commuting convenience by using the 1 km and 4 km distance bands. The results mirror the 2km distance band results (see appendix F).
[94] Note that repeating this exercise for either 2km or zone-level cell unit cluster sizes, tends to improve the balancing conditions in terms of pre-treatment characteristics, but I report the "worst scenario" so that the reader can judge for themselves the scientific reliability of the results.

sensitive to the sample choice. In Figures 3.4-3.5 I move to non-market housing sample but apply the same method described in Figures 3.2-3.3. Again I see the general result patterns are reassuringly robust to this sample change, though fewer new stations lie below the 45-degree line. In any case, I will test formally the impact of rail access changes on homeowners' happiness using the model specified in the following section.

4 Model

Using the survey data, I examine what happens to homeowners' happiness before and after the transport infrastructure changes. Then, by observing what happens in "treated" versus "control" places, I can more reliably assess the effects from rail access changes on happiness.

The starting point for my analysis is a basic regression model[95] relating homeowners' happiness to rail access---measured by the nearest distance to the station:

$$LnHappy_{it} = \alpha + \beta \cdot dist_{it} + \delta \cdot \ln(income_{it}) + \theta \cdot X'_{it} + f_i + g_t + \varepsilon_{it} \qquad (1)$$

$Happy_{it}$ in Eq. (1) is the average happiness of a particular residential aspect (commuting convenience, traffic safety, social environment, living convenience and traffic pollution) in cell unit i in period t, $dist_{it}$ is the nearest-station distance, $income_{it}$ is the sampled households' average monthly income[96], X_{it} is a vector of other household and location characteristics (see variable definitions in the appendix table 3.3), f_i represents place-specific fixed effects, and g_t indicates a time effect that would better capture changes in happiness over time (that are not accounted for by changes

[95] Searching over a number of choices of the functional forms it was determined that a function with the log transformation provided the best fit to the data.

[96] For the evaluation in monetary terms, estimates for the marginal utility of household income need to be considered.

observable characteristics).

This model specification can be easily generalized. For example, I would expect a 100 meter distance reduction to stations within 2 km distance ball to be much more highly valued than a 100 meter reduction at 20 km distance. My empirical model specifications allow for such differences between place of residences that are within 2 km of a station and place of residences that are beyond 2 km from the nearest station. Defining $r_{it} = I\{\mathrm{dist_{it}} \le 2\mathrm{km}\}$ an indicator that distance is less than 2 km, then I can have:

$$LnHappy_{it} = \alpha + \beta_1 \cdot r_{it} dist_{it} + \beta_2 \cdot (1 - r_{it}) dist_{it} + \delta \cdot \ln(income_{it}) + \theta \cdot X'_{it} + f_i + g_t + \varepsilon_{it} \quad (2)$$

Estimation of a model specification like Eq.(1) and (2) can provide estimates of happiness for a wide range of determinants associated with the location of a homeowner's place of residence. However, some factors may have indirect effects on homeowners' happiness for reasons other than the benefits of increased rail access. For instance, stations may be located in street corners that offer fancy pubs, retail outlets, churches and other local amenities that might bring additional residential happiness for households.

To account for these factors, one can always control for as many as local characteristics in the regressions. But some factors like air quality cannot be observed easily. As such, the model in (1) and (2) assumes that unobserved factors are fixed over time (f_i). However, the estimation results are still likely to be biased if these unobserved attributes are correlated with the station-distance variable. The difference-in-difference strategy based on time differences would eliminate pre-existing location characteristics and provide more reliable estimates on the net happiness effects of the transport improvement program. The final underlying model becomes:

$$\begin{aligned}(LnHappy_1 - LnHappy_0) = (g_1 - g_0) + \beta_1 \cdot r_{i1}(dist_1 - dist_0) + \beta_2 \cdot (1 - r_{i1})(dist_1 - dist_0) \\ + \delta \cdot [\ln(income_{i1}) - \ln(income_{i0})] + \theta \cdot (X_{i1} - X_{i0})' + (\varepsilon_{i1} - \varepsilon_{i0})\end{aligned} \quad (3)$$

I estimate this model using micro data on individual respondents, aggregated to cell-unit-period level. The two time periods are post-transport improvement (t =1) if year=2009, and pre-transport improvement (t =0) if year=2005. Since I have only two survey samples, the parameters β_1 and β_2 therefore provide difference-in-difference estimates for the impact of rail access changes on homeowners' happiness at affected places before and after the building of new rail transit lines. In the result section, regression estimates are measured by the specification form in Eq. (3).

There are at least three limitations to the models presented above. The first limitation is the common time-trend assumption. In general, one would expect observed and unobserved characteristics to be evolved with the transport improvement. My results might therefore underestimate the rail access effect if homeowners' happiness adjustment process is long before or after the building of new subway lines, or might overestimate the amenity benefits if other local externalities at station areas evolve with the increased rail access. This problem is not unique here. Ideally, one could control for a number of things (i.e. crime, shops, cafes, travel time) change together as a result of the stations opening if those detailed data is accessible. However, to the best of my knowledge, there are no publicly available data sources in which I can merge systemic information on localized changes with detailed data on residents' happiness and characteristics. When one is reading the results, it is important to keep in mind that the cell-unit level happiness measures might capture the additional impact of variation at the local areas. Practically, I do check the resulting estimates by using different data sample to make sure that they appeared reasonable. I also conduct the analysis disaggregated by households' income and age[97]. As such I can better capture the heterogeneous effects from rail access changes on happiness across different social groups.

Secondly, empirical studies like this have often faced the difficulty of the joint choice of transportation modes and residential locations. Households who live near rail stations may be more likely to travel by rail transits. But there are several explanations underlying this observed correlation. On the one hand, better rail access is expected to encourage residents who were not public transport users to commute by rail transits now. To this end, I control for the proportions of public transport users in every cell unit before and after the transport improvement. On the other hand, households who prefer this transit mode will choose to live near rail stations. To address this issue, I focus solely on the homeowners ("stayers") who have already lived in the current residences for at least five years---that is, the period before the new rail transit constructions. Since my data is a panel of areas, I also test for potentially endogenous changes to the compositions of residents in response to transport improvements. Specifically, I have examined changes in the composition of residents in affected places but found little evidence by comparing the cell unit composition of 2005 sampled homeowners and 2009 sampled homeowners, and by comparing the composition of those in the locations with greatest accessibility improvements who had recently moved-in with those living there more than 5 years. Finally, I take advantage of policy-exogeneity nature of non-market (*fang gai*)

[97] I use the sampled residents' median income and median age as the cut off points to create four social groups, and I find that there are no significant happiness variations within each group. However, I recognize that this classification method is not the only way to group households' characteristics. Other household characteristics would also contribute to their happiness evaluations. Ideally, I can match all household characteristics and further create a large number of social groups. But I simplify the analysis by only matching income and age because they are believed to be the two important factors that affect people's happiness (Lu, 1999).

housing with pre-determined locations as an additional robustness check. Below, I will run the regressions for the whole sampled homeowners first, and then for the non-market (*fang gai*) housing sampled homeowners.

Thirdly, the results of my analysis will depend on the validity of the survey responses. Before moving to a discussion of the results, it may be important to answer questions like: how informative of the subjective measures of residential happiness, and should we trust these measures? For example, the interviewers had clearly stated that this survey aimed at reflecting residents' current happiness levels, however, it is not possible to fully identify whether the responses embedded residents' anticipation effects on transport improvements in future. Some economists tend to be suspicious of the validity of subjective survey data. Again, recent empirical evidence have confirmed that the subjective measures of happiness are valid and trustable (Frey and Stutzer, 2001; Blanchflower and Oswald, 2004; Krueger and Schakde, 2007; Oswald and Wu, 2010). Subjective happiness data used in this study have passed through a series of validation exercises (Zhang et al., 2006). Note also that the measures I use here are not general subjective assessments about life happiness, but specific questions on perceptions about particular residential aspects. The economic and psychological studies have suggested that specific questions are more reliable than one general question to reflect changes in households' subjective wellbeing (Alesina et al., 2004; Kahneman and Krueger, 2006). This gives me more confidences on the reliability of the estimation results. However, another potential source of bias may arise from the conducted timing of the surveys. It is worth noting that in 2005 when the new metro lines and stations were being constructed, accessibility for residents at station areas might be in fact lowered by localized congestion---which could lead to lower residents' happiness level in 2005 survey. When new stations were opened, the changes in their happiness levels would reflect not just the commuting time savings, but also the disappearance of the noise or congestion effects at the station areas. Despite these limitations, I believe that difference-in-difference modeling the happiness consequences of transport improvements is an important step in better understanding the benefits of government investment policy.

5 Results

The rail access changes induced by the building of new rail transit lines allow me to estimate how homeowners' happiness with respect to several dimensions of residential environment changes for places experiencing station-distance reductions to within 2km. In Tables 3.3-3.7, I report regression estimates of the model in Eq. (3) using the cell unit panels, both for the whole sample and for the non-market housing subsample. The only variation in residence-station distance is before and after the building of new stations in 2008, in places affected by the rail access changes. Thus any measured effects of the transport improvement program on happiness occur through station distance changes due to the building of new rail transit lines.

5.1 Baseline estimates

Columns (1)-(2) in Table 3.3 are for the whole sample. Column (1) shows estimates that allow for household income and other characteristics, as well as cell unit fixed effects. Happiness about commuting convenience is found to rise in treated places by around 6.18% (=100*[exp (0.060)-1]), on average, for every kilometre reduction in distance close to the stations (within 2 km)[98]. There is no statistically significant impact from distance reductions to places that are beyond 2 km from the new stations.

[98] Note that happiness changes would rise more than proportionately with station proximities. As shown in Table 3.7, there is a bit of a non-linear happiness elasticity effect going from those within 4kms to those within 1km of a station.

Part of the increased rail access effect could be attributable to local contextual effects. One reason to do this is that, for the time period I study, new subway lines are likely to extend into the 2008 Olympic Park area and important "bedroom" communities (*Tiantongyuan*, *Yizhuang*, *Tongzhou*, *Daxing*). Thus I control for a long list of location characteristics such as the distance to CBD, Olympic Park, bedroom areas, etc (documented in appendix E). The main result is robust to this model specification[99], showing that better rail access can lead to the higher levels of happiness about commuting convenience. One thing to note is that, in the specifications from columns (1) to (2), I also find that homeowners' happiness about commuting convenience, though not statistically significant, rise slightly with distance reductions in the "control" group (places that is beyond the 2 km distance band). This result suggests that the impact of new stations on homeowners' happiness about commuting convenience is higher closer in.

Switching to the sample for non-market (*fang gai*) housing homeowners in columns (3)-(4), I find the same qualitative pattern, though the increased rail access effect are estimated to be larger than that in the whole sample. After controlling for all the characteristics in column (4), there is a 9.19% (=100*[exp(0.088)-1]) happiness rise per kilometre distance reduction to stations. Importantly, this result confirm the possibility that the rail access impact on homeowners' happiness largely holds after considering for the potential endogeneity in residential locations by using the non-market (*fang gai*) housing subsample.

Continuing to discuss about the rail access impact on happiness of commuting convenience, I next break down such impact by using four social groups: high income*high age, high income*low age, low income*high age, low income*low age.

[99] The results are also robust to the inclusion of the cell-unit happiness value of traffic safety, social environment, living convenience and traffic pollution as additional controls.

This comparison highlights the significant heterogeneity happiness effects across different social groups. Estimates in columns (1) and (6) of Table 3.6 show that, for homeowners in the high-income groups, the happiness effect relative to commuting convenience is about two times higher than the average, whilst such effect is very small for the low-income groups. I also find that young residents gain more happiness than elderly people when treated with new stations. The results are robust across the whole sample and the non-market housing sample. This is consistent with the expectations that high-income and low-age residents who attach great value to their works, are likely to be much more happier about the commuting time savings provided by the improved transport accessibility.

Table 3.4 reports the results of the impact of increased rail access on homeowners' happiness about traffic safety. In the whole sample specifications (columns 1-2), homeowners' traffic safety happiness significantly decrease with the distance reductions to stations. Specifications (columns 3-4) of the non-market housing sample share the same pattern of results. This result implies that the increased rail access may alter the distribution of traffic safety happiness around the station areas by increasing the local residents' safety concerns.

Comparing the coefficients on different social groups (documented in columns (2) and (7) of Table 3.6) provides estimates of the bias associated with the sample mean results in Table 3.4. Estimates from the high income*high-age group show the highest traffic safety happiness declination when treated with new stations. This is expected because the higher opportunity costs of safety issues at the station areas may enhance high-income residents' dissatisfaction about the rail transit expansions. Perhaps interestingly, I also find that the increased rail access impact does not significantly influence the traffic safety happiness for the low age*low income group.

In Table 3.5, I find that the presence of the new stations slightly improves homeowners' happiness about living convenience and traffic pollution in places that received effective distance reductions to stations. However, homeowners become less happy about social environment when their residences are treated with new stations. This tells a consistent story with actual observations in Beijing, where the original homeowners are not satisfied about the growing population flows and noise at the station areas. In the specifications linked with different social groups (Table 3.6), high-income groups show significant improvements in their happiness levels of living convenience and traffic pollution, but they become less happy about local social environment when treated with new stations. In the low-income groups, there are no strong evidence of better rail access effects on their happiness of living convenience, social environment and traffic pollution.

My purpose here is twofold: my first purpose is to shed light on what is known and unknown about homeowners' happiness with respect to different residential dimensions that may be affected by the transport improvement program. My results suggest that rail access effects on the various happiness dimensions of residential environment might tend to offset each other. Using the overall life happiness indicator would therefore mask the interpretations about the impact of the transport improvement program at particular residential aspects of households' living experiences. Second, I clarify the importance of considering the heterogeneous happiness effects on different social groups. For a more accurate assessment, I conduct the Chow test (Chow, 1960) to examine whether the key coefficients in each of the two regressions on different social group data sets are equal. This means that, for each of the happiness indicator, I use the Chow test to examine whether the coefficient of station-distance reductions (within 2km) in one specific social group is statistically equal to that in another social group. Perhaps surprisingly, I find that the null hypotheses are rejected at the 5% significance level, suggesting that the observed differences in the effects from rail access changes on happiness for different social groups are statistically significant. To the extent that this type of exercise is a significant tool informing this argument, my results show that the amenity benefits of rail access are highly dependent on residents' background characteristics. Which social group should be the "most representative"? This is certainly debatable. But clearly, beyond income-and-age groups presented in this study, there should be a long list of individual characteristics like education attainment, occupation that would further disaggregate residents into a large number of social groups. Researchers estimating the benefits of transport improvement program should take care to consider social differentiations at a reasonable geographical scale.

5.2 Sensitivity analyses

Tables 3.7 shows various sensitivity analyses for the baseline results presented in Tables 3.3-3.5. The first test focuses on whether my conclusions are sensitive to issues regarding the distance band selections. This test would hold everything the same in the model specification and any changes in rail access effects would attribute to the difference in the valuation of distance bands. In Table 3.7, estimates from the specification A overviews the baseline estimates. Specifications B-C show estimates for the variations in how I define the distance bands. The rationale behind this is that, homeowners' happiness would change more than proportionately with station proximity. Recall that the hedonic studies tend to have found capitalization effects of rail stations is localized with a strong distance decay effect (see Cheshire and Sheppard, 1995; Bowes and Ihlanfeldt, 2001). First, I use a 1km distance band instead of the previous 2 km distance band. This modification results in little changes, with stronger evidence of positive happiness effects associated with commuting convenience. This is in line with my prior expectations that the substantial increase in happiness about commuting convenience with better accessibility of those living near the new stations. This may also be partly due to the disappearance of negative construction impacts at the very localized station areas. Second, I revert to the 4km distance band. While the commuting and living convenience happiness effects are generally robust, happiness effects relative to other residential aspects turn to be insignificant. This implies that a 2-kilometre ball around the station is suitable for defining the happiness impacts of improvements in rail access at station areas---not at remote places[100].

Because the Beijing urbanized area is very large, it may have a substantial impact on the baseline estimates. I therefore, in the specification D, report results based on the 2km distance band but excluding the central city sample[101]. For the specification with the suburb sample, there is more sizable positive association between rail access and sub-urbanites' happiness about commuting convenience. There are several explanations for this: on the one hand, it is likely that the discomfort station facilities, insufficient capacities and frequencies of new rail transits may reduce the happiness improvement about commuting convenience for central city homeowners; on the other hand, it is well understood that suburbanites, faced with long commuting distance to work, are more easier to gain happiness towards commuting convenience due to the building of new rail transit lines. I also find slightly higher negative traffic safety happiness outcomes compared to the baseline results. This is possibly because of the high crime rates in the suburb areas. The

[100] Note that homeowners who resided more than 4 km away from a new station might also benefit from the improvements in rail access and would be far enough from the localized congestion nuisances at the station areas. I have tested this hypothesis and find no evidence to support this claim.

[101] Following the convention, the central city is defined as the areas within the No.3 ring road of Beijing.

happiness results relating to other residential aspects are similar to the baseline estimates.

Finally, I consider another issue related to different commuting modes. Recall that the baseline results are estimated by the sampled homeowners no matter whether they commute to work by using public transportation or not. In the specification E, I have restricted the attention to the public commuters sample only. I examine whether and to what extent the impact of station distance reduction (to within 2km) affects the happiness outcomes on public transport users. The results show the same qualitative nature as the baseline estimates with respect to all happiness indicators. In terms of quantitative differences, I find that public commuters have gained are more significant and sizable happiness improvements with respect to commuting and living convenience than all commuters. This is expected because public transport users are more likely to get direct time savings with better access to stations. And although not shown in the table, the estimated coefficients between public commuters and all commuters are statistically different from each other.

6 Monetization

One of the primary goals of the transport improvement program in Beijing is to upgrade households' living experience with respect to different dimensions of residential environment. In this section, the monetised welfare effect of implementing the transport improvement program is measured by the compensating variation (CV). That is, I examine the homeowners' average willingness-to-pay at the aggregated cell unit level for changes in rail access at their residence, holding housing prices and other local attributes constant[102]. For a transport policy which leads to rail access

[102] Note that I also assume that the housing supply is constant. This implies that the computed welfare estimates are essentially partial equilibrium measures. Given the data limits, I find little evidence of general equilibrium

changes from $dist_{i0}$ to $dist_{i1}$ the CV estimate can be implicitly defined as:

$$F_i\left(income_{i0};dist_{i0}\right)=F_i\left(income_{i1}-CV;dist_{i1}\right) \quad (4)$$

Where *F(*)* represents the indirect utility function. The subscript zero denotes originally household income and station-distance attributes at cell-unit *i*, whereas the subscript one indicates these attributes at cell-unit *i* after the transport improvement program. With the estimated coefficients of the econometric happiness equation (3) for changes in rail access (β_1), and income (δ), the CV can be calculated as follows:

$$CV = income_1 - e^{\frac{\beta_1 \cdot (rail\_access_0 - rail\_access_1) + \delta \ln(income_0)}{\delta}} \quad (5)$$

In light of recent literature, this CV welfare measure implicitly assumes that public investment programs do not immediately affect housing prices in the urbanized area (Takeuchi et al, 2008; Frey et al, 2009). Hence, this CV welfare measure could be interpreted as the monetary benefits of the transport improvement program over and above housing costs. Since the CV calculation also holds the housing supply fixed, my results therefore only reflect the benefits of the transport improvement program in the short run.

Table 3.8 reports the mean welfare effects of the improvements in rail access on homeowners' different happiness aspects. All CV estimates are measured on the basis of cell unit aggregated data, before-and-after the building of new subway lines in 2008. In terms of happiness about commuting convenience, I find that the improvements in rail access are worth, on average, about CNY 1,136 (approximately 100 GBP) per month to the whole sampled homeowners in Beijing. This means that the welfare benefit represents roughly 17.3 percent of the monthly average income of

happiness effects from the transport improvements in Beijing. Thus this study does not attempt to identify general equilibrium benefit measures that account for anticipated housing price effects. See detailed comparisons between partial and general equilibrium welfare measures in Sieg et al. (2004) and Tra (2010), among others.

a sampled Beijing homeowner. The happiness results for living convenience and traffic pollution show that the average sampled homeowners would be willing to pay around 9.5% and 6.6% respectively of their monthly income for the distance reduction to stations. The mean welfare measure is CNY-489 per month for the happiness about traffic safety, compared to an average benefit of CNY-378 per month for the happiness about social environment. In general, these welfare estimates are robust across the whole sample and the non-market housing sample homeowners.

Interestingly, I also find that the benefits of the transport improvements vary considerably across income groups. For example, the mean welfare measure relative to happiness of commuting convenience is about CNY 716 per month for the 20th income percentile homeowners, compared to a mean monthly benefit of CNY 1,828 for the 80th income percentile homeowners. In addition, the effects of increased rail access are not distributed evenly across the urban space. For example, the welfare results for happiness about commuting and living convenience show that suburb homeowners experience, on average, relative higher welfare gains compared to homeowners living in the central city. However, central city homeowners experience higher benefits relating to happiness about traffic pollution than suburb homeowners under the transport improvement program. Such variations among urbanites and suburbanites are also obvious in term of social environment and traffic safety happiness.

I do not want to over-emphasize these findings, however, as there are some problems underlying this happiness valuation approach. One relates to the survey-reported income. Most of surveys employed in happiness studies provide implicit information on income. For example, the micro survey data applied in this study only recorded households' income in categorical terms rather than actual income money figures. Thus this measurement would lead to imprecise the estimated welfare measures. Another issue is that the causes and consequences of household income changes will vary across places and in some situations might vary systemically within a certain place (Clark and Oswald, 1996; Frijters et al., 2004; Ferrer-i-Carbonell, 2005). Indeed it is highly possible that the rising income itself would provide additional life enjoyment in all residential aspects. This makes the valuation of happiness consequences of exogenous income adjustments an interesting topic in the economic literature that I leave to future research. Further, I am well aware that there is a detailed discussion in the happiness literature about the reverse causation (happy people are less unemployed and earn more) and unobserved factors (there may be no link between happiness and income---it's all driven by resilience and non-cognitive abilities). However, this paper cannot fully explore these effects without long-run surveys and more detailed census data. Finally, the estimated welfare benefits here are measured by using the aggregated cell-unit data. Thus the resulting monetary estimates are likely to be biased and would conceal variations among individuals. I note, however, that most of these issues can be addressed when better data become available; and they do not fully invalid the happiness valuation approach. Despite of these limitations, this monetization analysis is still a useful exercise that could shed light on potential welfare benefits of the transport improvement program.

7 Conclusions

In this paper I consider links between rail access and homeowners' happiness, providing new evidence that better rail access does affect homeowners' happiness with respect to different dimensions of residential environment. I implement the difference-in-difference model based a recent transport infrastructure change in Beijing. The change I consider referred to the building of new stations, so I can use repeated survey data to examine what happened to homeowners' happiness in particular residential aspects when residence-station distances were reduced.

My results yield three important insights. First, I find that homeowners' happiness about commuting convenience rise significantly in places affected by the building of new stations, relative to places that were unaffected. I also find that homeowners' residences receiving increased station proximities experience improvements in happiness about traffic pollution and living convenience. On the flip side, the impacts of station-distance reductions are found to decrease the homeowners' happiness towards traffic safety and social environment nearby station areas. These results pass through a series of sensitivity tests and remain robust in terms of qualitative nature.

Second, the effect of rail access changes on happiness depends on geographical locations and socioeconomic characteristics. Broadly speaking, suburbanites gain greater happiness about commuting convenience than urbanites when their places of residences experienced distance reductions to new stations. High-income homeowners in areas affected by the transport improvement place substantial happiness value on commuting convenience and other residential aspects, whilst low-income homeowners do not appear to value the increased rail access highly. These findings are robust after controlling for the potential endogeneity in residential locations by using the non-market (*fang gai*) housing sample. One important implication here is that researchers estimating the rail access effects should take care to do data mining and empirical specifications that allow the inclusion of targeted social groups over urban areas.

Third, the welfare evaluation results suggest that Beijing homeowners place substantial value on improvements in the rail access. I estimate the average willing-to-pay by homeowners for the improvement in the rail access at their residence, holding housing prices and other location factors constant. I find that the happiness effects of transport improvements on homeowners' perceptions about commuting convenience, are worth, on average, about CNY 1,136 per month, or roughly 17.3 percent of the monthly income in 2005 to the whole sampled homeowners in Beijing. However, the welfare benefits vary considerably relative to different happiness aspects, income groups and urban areas. All of these pieces of evidence support the claim that planners and policymakers need to consider social-spatial differentiations when evaluating the happiness consequences of government investment in local infrastructure.

In considering the happiness consequences brought about by transport improvements, it is important to note that I have only examined the direct effects from rail access changes on homeowners' happiness with respect to different dimensions of residential environment. There is considerably debate with respect to the inter-connected reflections on changes in happiness, housing price capitalization, and self-selection. More evidence is needed to strengthen our knowledge of the interrelationship between changes in happiness and expected changes in housing prices, and how would such changes affect homeowners' decision to stay or move.

My future work in this line of research would include several pieces. First, I expect to obtain more detailed income and systemic housing transaction data in the appropriate years and locations. Indeed, it would be really interesting not just to back out the "value" of happiness via the sample incomes but to directly relate the happiness measures to the hedonic estimation of capitalization effects. Presumably the changes in happiness are reflected in changes in effective housing demands so in some way capitalized into housing prices. Second, I would link the real estate consequences directly to my findings on the distribution of changes in happiness for high income/young compared to low income/old resident group, using detailed residential mobility information at the individual level. Specifically, I will test the extent to which the changes in happiness are linked to changes in housing prices and in turn linked to differential residential mobility with an inflow of those most benefiting in happiness terms from the improvements in transport accessibility. In so far as this occurred then there would be policy implications for neighbourhood dynamics and also for the long-term impact on the social welfare. Even those who do not value the transport improvements would be compensated if property owners have experienced price premiums and have the ability to turn that into money if they move to an area where accessibility has not improved. Given that they appear to value good access less than young/high wage then their welfare would be improved by trading more money for less transport accessibility. The third piece of my future work hopes to learn more about the self-selection and anticipation effects of transport improvements. I clarify the importance of considering the interrelationship between changes in happiness and neighbourhood demographic dynamics as a result of transport improvement. It is interesting to know: Is there a change in the composition of residents in locations benefiting most from the transport improvements with a differential increase in the representative groups rating the transport improvements highly in terms of happiness? Future happiness studies using long-run survey data in different contexts to corroborate the robustness of my findings would be useful.

Table list

Table 3.1 Descriptive statistics of rail access and happiness: the whole sample

	Full sample		Treatments		Controls		Estimates
	Before	After	Before	After	Before	After	Raw
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Station Distance	1.871	1.173	1.720	0.513	1.931[†]	1.399	-1.015*
	(0.610)	(0.621)	(0.372)	(0.236)	(0.682)	(0.655)	(0.571)
Ln (Commuting convenience)	1.423	1.476	1.451	1.526	1.413	1.460	0.064**
	(0.171)	(0.179)	(0.120)	(0.142)	(0.186)	(0.187)	(0.033)
Ln (Traffic safety)	1.428	1.408	1.446	1.418	1.421[†]	1.405	-0.056**
	(0.182)	(0.175)	(0.145)	(0.152)	(0.194)	(0.181)	(0.028)
Ln (Social environment)	1.533	1.504	1.548	1.513	1.527	1.501	-0.043*
	(0.192)	(0.178)	(0.182)	(0.117)	(0.196)	(0.192)	(0.025)
Ln (Traffic pollution)	1.360	1.388	1.372	1.393	1.355	1.386	0.031*
	(0.211)	(0.216)	(0.213)	(0.164)	(0.211)	(0.226)	(0.018)
Ln (Living convenience)	1.443	1.475	1.461	1.521	1.437	1.460	0.044**
	(0.163)	(0.165)	(0.131)	(0.143)	(0.173)	(0.172)	(0.021)
Sample size	883	750	252	191	631	559	1633

Notes.--- The whole sample refers to the sampled homeowners who work and hold the tenure before the transport improvement happened. Treatment refers to cell units for which distance to rail station was less in year 2009 than in 2005, and where distance in year 2009 was less than 2 km. Data units are before/after cell units. Columns (1)-(6) show means and standard deviations (in parentheses). Column (7) shows the simple difference-in-difference estimated coefficients based on the raw data (Standard errors corrected for clustering at the cell unit level are reported in parentheses). [†] denotes that the control group is significantly different from the treatment group in terms of the pre-treatment characteristic at the 5% level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.2 Descriptive statistics of rail access and happiness: non-market housing sample

	Full sample		Treatments		Controls		Estimates
	Before	After	Before	After	Before	After	Raw
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Station Distance	1.601	1.142	1.569	0.420	1.611	1.363	-0.991**
	(0.584)	(0.595)	(0.361)	(0.238)	(0.652)	(0.630)	(0.386)
Ln (Commuting convenience)	1.446	1.488	1.451	1.556	1.443	1.468	0.108**
	(0.171)	(0.212)	(0.115)	(0.166)	(0.187)	(0.223)	(0.051)
Ln (Traffic safety)	1.463	1.430	1.469	1.428	1.460	1.431	-0.052**
	(0.192)	(0.206)	(0.155)	(0.150)	(0.202)	(0.217)	(0.026)
Ln (Social environment)	1.540	1.516	1.543	1.501	1.538	1.520	-0.051*
	(0.185)	(0.207)	(0.183)	(0.183)	(0.187)	(0.213)	(0.029)
Ln (Traffic pollution)	1.369	1.388	1.362	1.398	1.371	1.385	0.045**
	(0.225)	(0.253)	(0.221)	(0.228)	(0.227)	(0.260)	(0.023)
Ln (Living convenience)	1.445	1.489	1.468	1.533	1.435	1.476	0.048**
	(0.166)	(0.192)	(0.135)	(0.148)	(0.175)	(0.211)	(0.021)
Sample size	751	587	235	137	516	450	1338

Notes.---The non-market housing sample refers to the sampled homeowners who work and hold the tenure of the non-market (*fang gai*)housings before the transport improvement happened. The estimation accounts for the endogeneity residential sorting by using this non-market housing sub-sample with pre-determined residential locations. Treatment refers to cell units for which distance to rail station was less in year 2009 than in 2005, and where distance in year 2009 was less than 2 km. Data units are before/after cell units. Columns (1)-(6) show means and standard deviations (in parentheses). Column (7) shows the simple difference-in-difference estimated coefficients based on the raw data (Standard errors corrected for clustering at the cell unit level are reported in parentheses). The *t*-test in mean difference between columns (3) and (5) shows no differences at the 5% level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.3 Rail access and homeowners' happiness of commuting convenience

	(1)	(2)	(3)	(4)
	The whole sample		Non-market housing sample	
Rail access				
station distance <2km	-0.060**	-0.051**	-0.105**	-0.088**
	(0.029)	(0.026)	(0.046)	(0.040)
station distance >2km	-0.035	-0.028	-0.051	-0.039
	(0.062)	(0.046)	(0.045)	(0.032)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Within R^2	0.426	0.445	0.567	0.581
Sample size	1633	1633	1338	1338
Fixed effects variance share	0.693	0.686	0.733	0.725

Notes.---Dependent variable is log happiness of commuting convenience. Columns (1)-(2) is estimated using the whole sampled residents. Columns (3)-(4) is estimated based on the non-market housing sample. The whole sample refers to the sampled homeowners who work and hold the tenure before the transport improvement happened. The non-market housing sample means the long-term tenure homeowners who work and lived in non-market (*fang gai*) housings with pre-determined locations. Data is aggregated to cell unit level for two snapshots: 2005 and 2009. Regressions include control variables detailed in appendix E table. The constant terms are omitted for simplicity. Standard errors corrected for clustering at the cell unit level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4 Rail access and homeowners' happiness of traffic safety

	(1)	(2)	(3)	(4)
	The whole sample		Non-market housing sample	
Rail access				
station distance <2km	0.053**	0.046*	0.048**	0.038**
	(0.026)	(0.023)	(0.021)	(0.018)
station distance >2km	0.031	0.025	0.026	0.019**
	(0.022)	(0.018)	(0.023)	(0.011)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Within R^2	0.413	0.419	0.523	0.526
Sample size	1633	1633	1338	1338
Fixed effects variance share	0.651	0.645	0.693	0.685

Notes.---Dependent variable is log happiness of traffic safety. See other notes in table 3.3.

Table 3.5 Rail access and homeowners' happiness of other residential aspects

	(1)	(2)	(3)	(4)
	The whole sample		Non-market housing sample	
A. Happiness about living convenience				
Rail access				
station distance <2km	-0.037**	-0.033*	-0.045*	-0.036**
	(0.016)	(0.018)	(0.023)	(0.015)
station distance >2km	-0.022	-0.019	-0.015	-0.011
	(0.026)	(0.030)	(0.023)	(0.023)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Within R^2	0.349	0.356	0.436	0.448
Sample size	1633	1633	1338	1338
Fixed effects variance share	0.577	0.571	0.669	0.662
B. Happiness about social environment				
Rail access				
station distance <2km	0.048**	0.033**	0.053**	0.045**
	(0.023)	(0.016)	(0.025)	(0.021)
station distance >2km	0.032	0.015	0.024	0.011
	(0.021)	(0.022)	(0.015)	(0.023)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Within R^2	0.326	0.331	0.335	0.356
Sample size	1633	1633	1338	1338
Fixed effects variance share	0.619	0.612	0.653	0.646
C. Happiness about traffic pollution				
Rail access				
station distance <2km	-0.035**	-0.031*	-0.045**	-0.039*
	(0.015)	(0.018)	(0.019)	(0.022)
station distance >2km	0.012	0.008	0.016	0.013
	(0.025)	(0.022)	(0.027)	(0.018)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Within R^2	0.332	0.351	0.396	0.382
Sample size	1633	1633	1338	1338
Fixed effects variance share	0.611	0.602	0.636	0.631

Notes.---The dependant variable in the specifications A-C is the log of happiness of living convenience, social environment, and traffic pollution, respectively. Each specification is a separate set of regressions. Columns (1)-(2) is estimated using the whole sampled residents. Columns (3)-(4) is estimated based on the non-market housing sample. See other notes in table 3.3.

Table 3.6 Rail access and homeowners' happiness disaggregated by income and age groups

	Commuting (1)	Safety (2)	Social (3)	Living (4)	Pollution (5)	Commuting (6)	Safety (7)	Social (8)	Living (9)	Pollution (10)
	The whole sample					Non-market housing sample				
Group 1 (low age*low income)										
station distance <2km	-0.016*	0.013	0.016	-0.029	-0.033*	-0.023*	0.027	0.021	-0.042	-0.031*
	(0.009)	(0.018)	(0.011)	(0.018)	(0.019)	(0.012)	(0.034)	(0.015)	(0.038)	(0.018)
station distance >2km	-0.011	0.010	0.006	-0.007	0.011	-0.006	0.025	0.006	-0.014	0.009
	(0.008)	(0.009)	(0.010)	(0.005)	(0.008)	(0.010)	(0.018)	(0.014)	(0.018)	(0.016)
Within R^2	0.411	0.427	0.302	0.352	0.380	0.461	0.409	0.332	0.355	0.353
Sample size	496	496	496	496	496	380	380	380	380	380
Fixed effects variance share	0.906	0.795	0.758	0.594	0.640	0.884	0.983	0.932	0.516	0.614
Group 2 (low age*high income)										
station distance <2km	-0.109**	0.048*	0.039**	-0.031**	-0.056*	-0.149**	0.052**	0.042*	-0.056**	-0.055**
	(0.045)	(0.027)	(0.018)	(0.014)	(0.032)	(0.58)	(0.021)	(0.019)	(0.025)	(0.028)
station distance >2km	-0.062	0.023	0.010	-0.013	0.029	-0.061	0.021	0.025	-0.019	0.017
	(0.058)	(0.022)	(0.012)	(0.035)	(0.018)	(0.083)	(0.013)	(0.018)	(0.025)	(0.015)
Within R^2	0.432	0.312	0.326	0.311	0.302	0.503	0.486	0.403	0.345	0.305
Sample size	425	425	425	425	425	335	335	335	335	335
Fixed effects variance share	0.954	0.930	0.798	0.687	0.885	0.896	0.697	0.616	0.790	0.776
Group 3 (high age*low income)										
station distance <2km	-0.039**	0.033*	0.021*	-0.026	-0.012	-0.033*	0.026*	0.026*	-0.016	-0.027
	(0.023)	(0.019)	(0.011)	(0.023)	(0.009)	(0.018)	(0.014)	(0.014)	(0.028)	(0.019)
station distance >2km	-0.022	0.008	0.003	-0.055	0.008	-0.015	0.010	0.007	-0.006	0.020
	(0.043)	(0.035)	(0.005)	(0.057)	(0.013)	(0.036)	(0.029)	(0.033)	(0.010)	(0.046)
Within R^2	0.439	0.304	0.316	0.234	0.309	0.428	0.410	0.306	0.383	0.334
Sample size	411	411	411	411	411	360	360	360	360	360
Fixed effects variance share	0.515	0.690	0.570	0.625	0.936	0.768	0.823	0.712	0.829	0.559
Group 4 (high age*high income)										
station distance <2km	-0.132**	0.067*	0.038**	-0.048**	-0.040***	-0.151**	0.072**	0.056**	-0.069**	-0.065**
	(0.061)	(0.035)	(0.017)	(0.023)	(0.015)	(0.066)	(0.035)	(0.024)	(0.035)	(0.030)
station distance >2km	-0.078	0.015	0.012	-0.013	0.021	-0.023	0.009	0.025	-0.015	0.025
	(0.083)	(0.010)	(0.071)	(0.009)	(0.022)	(0.093)	(0.013)	(0.072)	(0.028)	(0.021)
Within R^2	0.413	0.321	0.218	0.382	0.411	0.530	0.513	0.401	0.412	0.387
Sample size	301	301	301	301	301	263	263	263	263	263
Fixed effects variance share	0.779	0.603	0.882	0.750	0.901	0.931	0.756	0.870	0.679	0.688

Notes.--- Each column and group is a separate regression with full set of controls (see appendix E table). Dependent variable in columns (1)–(5) and (6)-(10) is the log of different happiness measures. Groups 1-4 are classified by using sample median income and age level. Standard errors corrected for clustering at the cell unit level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.7 Regression estimates of rail access effects, sensitivity analyses

	Commuting (1)	Safety (2)	Social (3)	Living (4)	Pollution (5)	Commuting (6)	Safety (7)	Social (8)	Living (9)	Pollution (10)
	The whole sample					Non-market housing sample				
A. Baseline estimates										
station distance <2km	-0.051**	0.046**	0.033**	-0.033*	-0.031*	-0.088**	0.038**	0.045**	-0.036**	-0.039*
	(0.026)	(0.023)	(0.016)	(0.018)	(0.018)	(0.040)	(0.018)	(0.021)	(0.015)	(0.022)
station distance >2km	-0.028	0.025	0.015	-0.019	-0.008	-0.039	0.019**	0.011	-0.011	-0.013
	(0.046)	(0.018)	(0.022)	(0.030)	(0.022)	(0.032)	(0.011)	(0.023)	(0.023)	(0.018)
B. 1km distance band										
station distance <1km	-0.058***	0.056*	0.030*	-0.035*	-0.025*	-0.093**	0.048**	0.041**	-0.030**	-0.043**
	(0.022)	(0.030)	(0.018)	(0.019)	(0.015)	(0.041)	(0.021)	(0.016)	(0.014)	(0.021)
station distance >1km	-0.026	0.012	0.003	-0.024	-0.005	0.052	0.014*	0.006	-0.012	-0.005
	(0.027)	(0.008)	(0.037)	(0.035)	(0.023)	(0.038)	(0.008)	(0.022)	(0.016)	(0.011)
C. 4km distance band										
station distance <4km	-0.049*	0.051	0.016	-0.055*	-0.023	-0.078*	0.039	0.032	-0.045**	-0.036
	(0.028)	(0.044)	(0.011)	(0.032)	(0.019)	(0.045)	(0.032)	(0.029)	(0.024)	(0.028)
station distance >4km	-0.035	0.023	0.009	-0.031	-0.012	-0.069	0.018*	0.016	-0.022	-0.007
	(0.081)	(0.018)	(0.043)	(0.035)	(0.034)	(0.052)	(0.010)	(0.018)	(0.031)	(0.023)
D. Dropping central city sample										
station distance <2km	-0.193***	0.056**	0.031**	-0.036*	-0.026**	-0.231***	0.046**	0.038*	-0.032**	-0.032**
	(0.031)	(0.023)	(0.015)	(0.020)	(0.013)	(0.027)	(0.023)	(0.022)	(0.014)	(0.016)
station distance >2km	-0.048	0.019	0.008	-0.021	-0.008	-0.052*	0.013**	0.041	-0.011	-0.006
	(0.055)	(0.012)	(0.026)	(0.038)	(0.021)	(0.031)	(0.006)	(0.065)	(0.020)	(0.008)
E. Using public commuter sample										
station distance <2km	-0.056**	0.043*	0.013	-0.039**	-0.028*	-0.097**	0.036**	0.026*	-0.043**	-0.048*
	(0.023)	(0.024)	(0.012)	(0.016)	(0.016)	(0.058)	(0.017)	(0.014)	(0.018)	(0.025)
station distance >2km	-0.022	0.017	0.005	-0.025	-0.007	-0.042	0.015**	0.008	-0.010	-0.006
	(0.050)	(0.012)	(0.028)	(0.021)	(0.022)	(0.035)	(0.007)	(0.029)	(0.029)	(0.015)

Notes. --- Dependent variable in columns (1)–(5) and (6)-(10) is the log happiness of commuting convenience, traffic safety, social environment, living convenience, and traffic pollution, respectively. Specification A shows the baseline estimates reported in Tables 3.3-35. Specifications B-C use different distance bands to select the treatment group, as described in the text. Specification D is similar to specification A except for dropping the central city sample. The sample used in specification E only includes homeowners who use public transport to work and hold the tenure before the transport was improved. Each column and specification is a separate regression. All regressions shown in the table include the full set of controls. Standard errors corrected for clustering at the cell unit level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.8 Benefits of the transport improvements in the Beijing urbanized area (CNY/month)

			Welfare measures for improved rail access														
			Commuting convenience			Traffic safety			Social environment			Living convenience			Traffic pollution		
	Average monthly household income (CNY)	Distance reduction (km)	Mean	P20th income	P80th income	Mean	P20th income	P80th income	Mean	P20th income	P80th income	Mean	P20th income	P80th income	Mean	P20th income	P80th income
Entire urbanized area																	
Whole sample	6533	1.15	1136	716	1828	-489	-184	-675	-378	-256	-711	622	356	1261	433	387	866
non-market housing sample	5946	1.03	1095	464	1579	-595	-395	-906	-489	-380	-816	737	251	1325	558	368	963
Central city only																	
Whole sample	6601	0.78	881	653	1287	-521	-234	-772	-590	-325	-912	516	327	1021	568	465	920
non-market housing sample	6180	0.72	796	498	1016	-615	-458	-1134	-677	-469	-1126	575	212	1138	685	483	1040
Suburb only																	
Whole sample	6494	1.20	1368	845	2196	-418	-131	-556	-228	-169	-542	654	381	1293	391	106	726
non-market housing sample	5911	1.13	1165	778	1831	-557	-368	-834	-316	-230	-608	808	288	1396	445	211	752

Note.--- Welfare estimates are calculated by using the equation (4) and (5), as described in the text. The whole sample represents the sampled homeowners who work and hold the tenure before the transport improvement happened. The non-market housing sample means the long-term tenure homeowners who work and lived in non-market (*fang gai*) housings with pre-determined locations. "P20th income" and "P80th income" represent the 20th and 80th income percentile, respectively. 1GBP= around 10 CNY.

Figure list

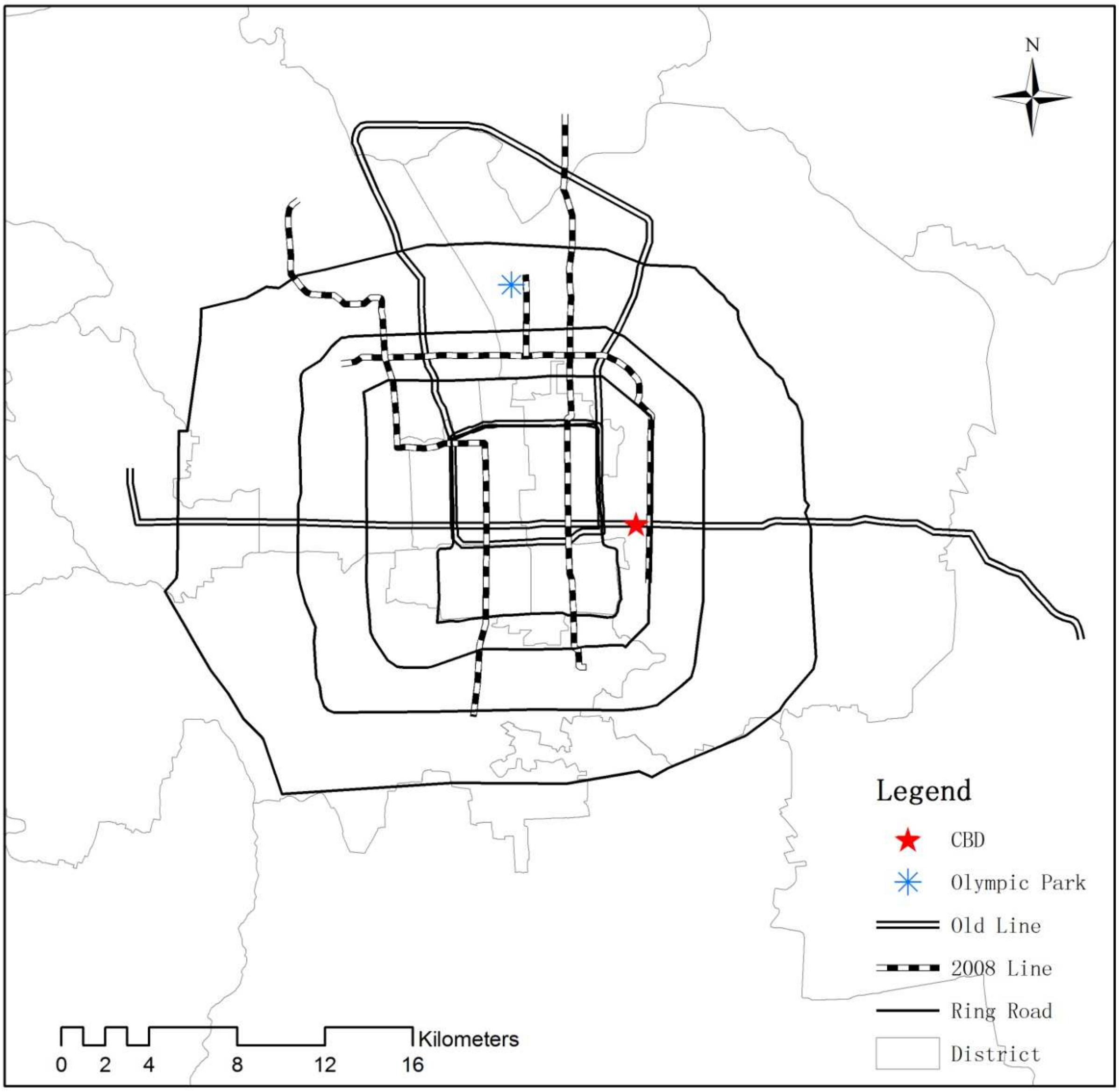


Figure 3.1 New rail transit constructions in 2008 Beijing

Notes.---Old Line means the subway lines built before 2008; 2008 Line means the newly-opened subway lines in 2008.

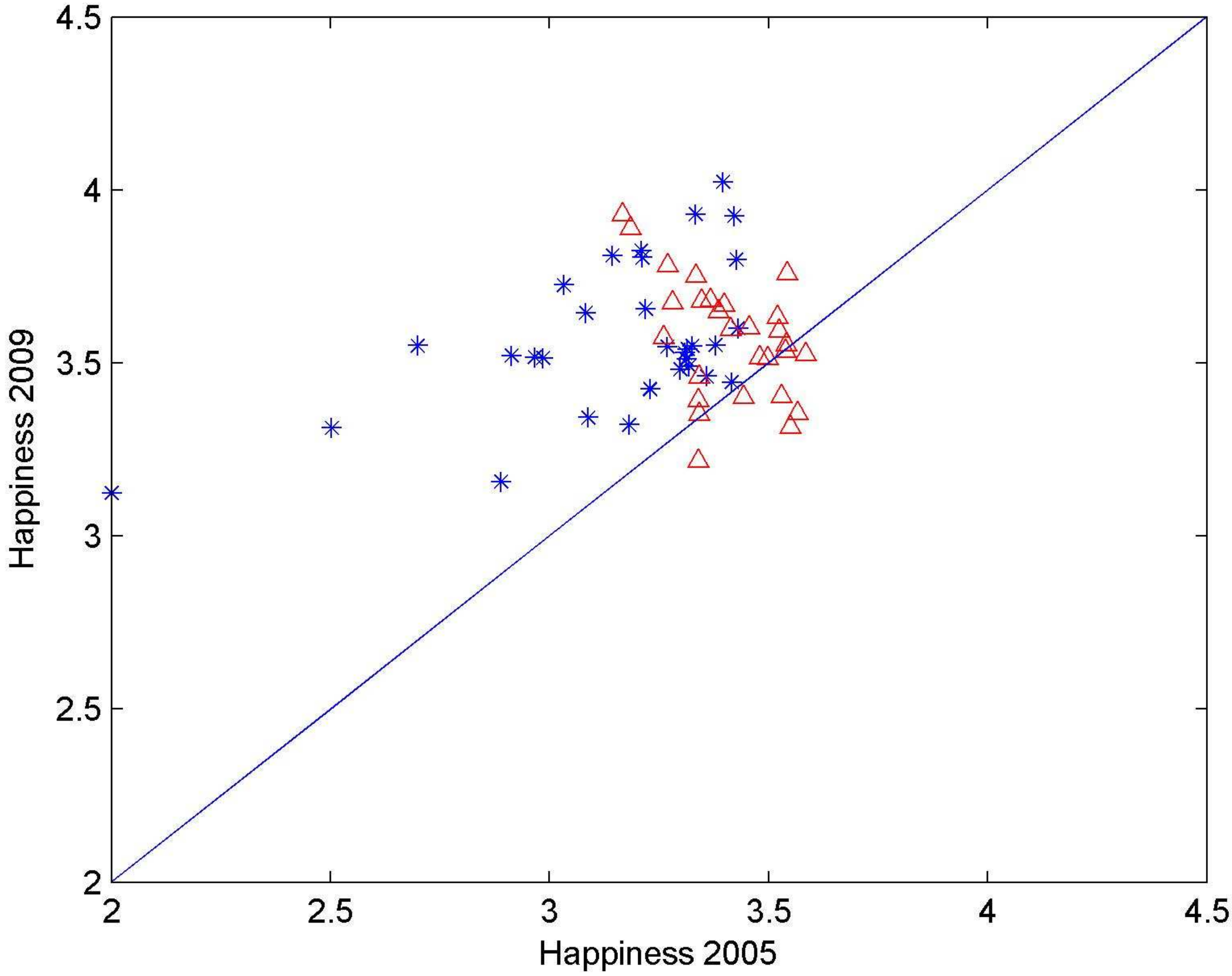


Figure 3.2 Happiness changes in commuting convenience: the whole sample

Notes.---Each triangle-dot represents the median happiness value within 2km of a new station located at the central city. Each star-dot represents the median happiness value within 2km of a new station located at the suburbs. The solid line is the 45 degree line. Happiness value is measured on a scale from "1 being very unhappy" to "5 being very happy". The horizontal axis is the median 2005 happiness value of commuting convenience. The vertical axis is the median 2009 happiness value of commuting convenience.

Figure 3.3 Spatial distributions of happiness changes: the whole sample

Notes.---Each circle label represents the vertical deviation of each dot in Figure 3.2 from the 45 degree line, as described in the text.

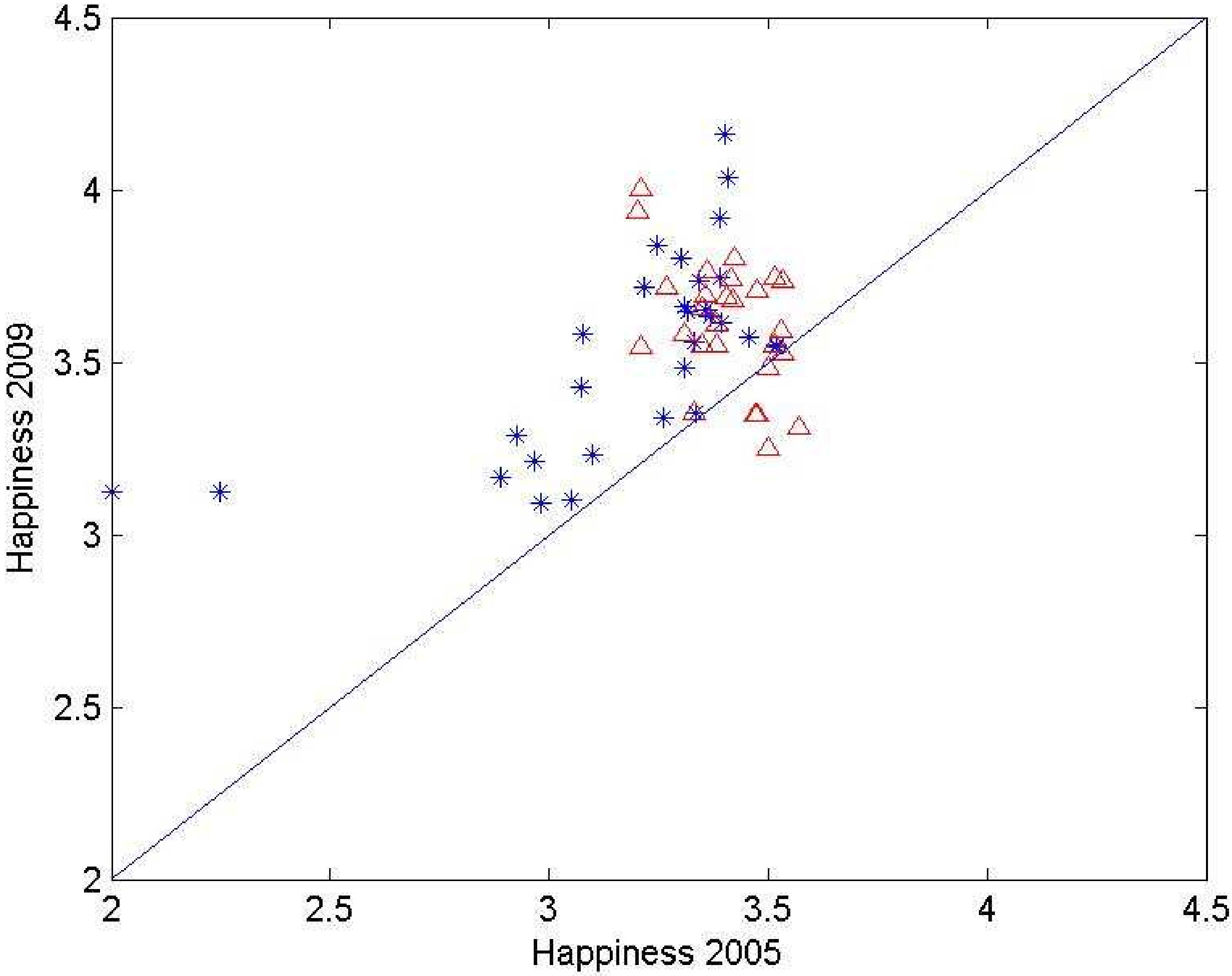


Figure 3.4 Happiness changes in commuting convenience: non-market housing sample

Notes.--- Each triangle-dot represents the median happiness value within 2km of a new station located at the central city. Each star-dot represents the median happiness value within 2km of a new station located at the suburbs. See other notes in Figure 3.2.

Figure 3.5 Spatial distributions of happiness changes: non-market housing sample

Notes.---Each circle label represents the vertical deviation of each dot in Figure 3.4 from the 45 degree line, as described in the text.

Appendix A.

Appendix Table 3.1 Happiness survey questions

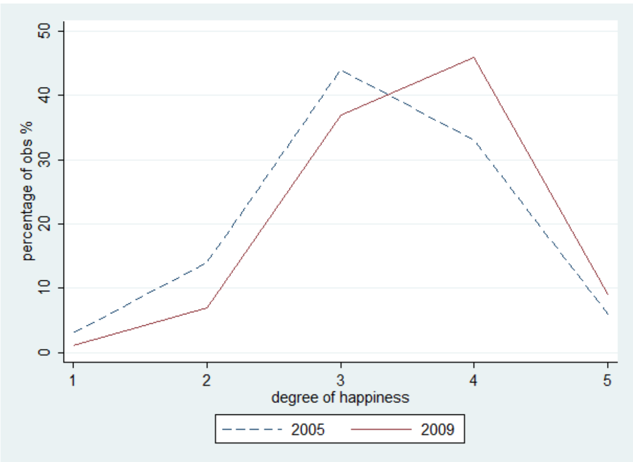
Happiness indicator	Original survey question	Measurement	Expected signs after transport improved	Possible reasons
Commuting convenience	How well do you satisfy your residential location about its local (neighbourhood) area's convenience to use rail transit to do work-related activities?	0= not familiar ; 1 = very unhappy; 2 = unhappy; 3=normal; 4 = happy; 5 = very happy	Happiness rise	Commuting time-savings by living closer to stations
Living convenience	How well do you satisfy your residential location about its local (neighbourhood) area's convenience to use rail transit to do non-working related activities?	0= not familiar ; 1 = very unhappy; 2 = unhappy; 3=normal; 4 = happy; 5 = very happy	Happiness rise	Time-savings for doing life activities by living closer to stations
Traffic pollution	How well do you satisfy your residential location about its local (neighbourhood) area's traffic pollution conditions (including automobile gas emission and other concerns about the pollution induced by traffic facilities)?	0= not familiar ; 1 = very unhappy; 2 = unhappy; 3=normal; 4 = happy; 5 = very happy	Unclear	A positive impact could be due to the reduced local road traffic and cleaning station conditions compared with before; A negative impact could be caused by crowded traffic and dirty parking spaces at station areas
Traffic safety	How well do you satisfy your residential location about its local (neighbourhood) area's traffic accidents and station areas' safety conditions?	0= not familiar ; 1 = very unhappy; 2 = unhappy; 3=normal; 4 = happy; 5 = very happy	Happiness fall	Safety concerns caused by growing population flows at station areas
Social environment	How well do you satisfy your residential location about its local (neighbourhood) area's social environment (including social culture, social capital, common-sense of worth and other related concerns about social environment)?	0= not familiar ; 1 = very unhappy; 2 = unhappy; 3=normal; 4 = happy; 5 = very happy	Happiness fall	Noise and congestion effects caused by growing population flows at station areas

Appendix B.

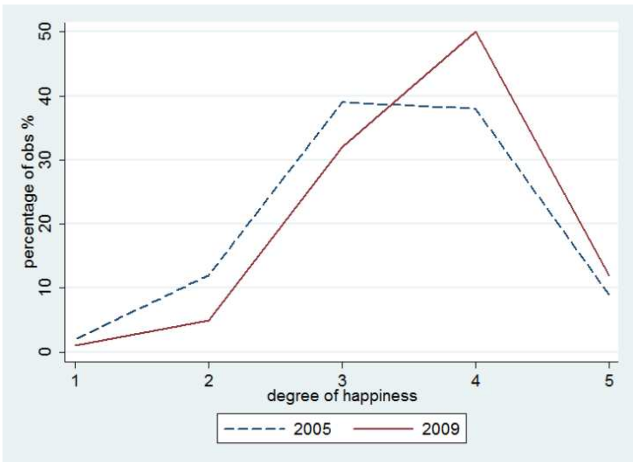
Appendix Figure 3.1 shows that the happiness distribution curves of commuting convenience (living convenience) between 2005 and 2009 surveys share quite similar distribution patterns. To test whether there are any significant differences in the frequency distributions of the happiness measures between 2005 and 2009 surveys, I performed the Pearson Chi-square test. As shown by the table below, the null hypothesis that there is no significant difference in the distributions of the happiness measures between 2005 and 2009 surveys cannot be rejected at the 5% significance level.

Appendix Table 3.2 Pearson Chi-squared test results

	Commuting convenience	Safety	Social environment	Living convenience	Pollution
Pearson chi2 (df=4)	4.6775	4.8113	2.9327	4.9708	5.5615
5% significance level	H_0 accepted	H_0 accepted	H_0 accepted	H_0 accepted	H_0 accepted



Happiness about commuting convenience



Happiness about living convenience

Appendix Figure 3.1 Distributions of key happiness measures across 2005 and 2009 surveys

Appendix C.

Appendix Table 3.3-3.4 shows the adjusted descriptive statistics and main regression results by dropping out the control samples that are within 2km station-distance both before-and-after the building of 2008 rail stations. The headline finding is that doing this does bring improvements in the treatment effects in terms of quantitative nature.

Appendix Table 3.3 Adjusted descriptive statistics of rail access and happiness

	Full sample		Treatments		Controls		Estimates
	Before	After	Before	After	Before	After	Raw
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: the whole sample							
Station Distance	4.192	2.758	2.628	0.946	4.962†	3.227	-1.075**
	(3.415)	(2.476)	(1.371)	(0.487)	(4.616)	(3.093)	(0.437)
Commuting convenience	3.231	3.391	3.281	3.504	3.207	3.340	0.075**
	(0.577)	(0.636)	(0.502)	(0.696)	(0.703)	(0.670)	(0.033)
Traffic safety	3.448	3.285	3.523	3.372	3.415†	3.242	-0.067**
	(0.676)	(0.666)	(0.604)	(0.657)	(0.792)	(0.762)	(0.028)
Social environment	3.723	3.573	3.709	3.593	3.730	3.564	-0.048*
	(0.624)	(0.592)	(0.606)	(0.565)	(0.687)	(0.709)	(0.027)
Traffic pollution	3.126	3.210	3.011	3.222	3.183†	3.205	0.056**
	(0.793)	(0.738)	(0.779)	(0.755)	(0.912)	(0.896)	(0.022)
Living convenience	3.329	3.570	3.419	3.664	3.286	3.528	0.053**
	(0.590)	(0.617)	(0.542)	(0.643)	(0.764)	(0.662)	(0.025)
Sample size	764	613	252	191	512	422	1377
Panel B: the non-market housing sample							
Station Distance	3.346	2.136	2.596	0.942	3.965	2.763	-1.182**
	(3.020)	(2.120)	(1.351)	(0.476)	(3.243)	(2.769)	(0.542)
Commuting convenience	3.250	3.385	3.269	3.504	3.239	3.323	0.126**
	(0.587)	(0.702)	(0.482)	(0.751)	(0.725)	(0.854)	(0.061)
Traffic safety	3.325	3.246	3.438	3.340	3.266	3.187	-0.063*
	(0.690)	(0.683)	(0.634)	(0.636)	(0.842)	(0.881)	(0.037)
Social environment	3.638	3.550	3.701	3.599	3.606	3.518	-0.056*
	(0.622)	(0.645)	(0.597)	(0.627)	(0.709)	(0.782)	(0.030)
Traffic pollution	3.128	3.072	2.959	3.084	3.067	3.235	0.078**
	(0.835)	(0.779)	(0.810)	(0.812)	(1.016)	(0.972)	(0.036)
Living convenience	3.317	3.569	3.405	3.712	3.262	3.495	0.068**
	(0.586)	(0.653)	(0.554)	(0.678)	(0.694)	(0.810)	(0.031)
Sample size	608	398	235	137	373	261	1006

Notes.--- While the "treatment" remains the same as described in the text, the "controls" here means cell units for which the nearest station distance is still larger than 2km in year 2009. Columns (1)-(6) show means and standard deviations (in parentheses). Column (7) shows the simple diff-in-diff estimated coefficient based on the raw data. † denotes that the *t*-test in mean difference between columns (3) and (5) is significance at the 5% level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 3.4 Adjusted main regression results

	The whole sample		Non-market housing sample	
A. Happiness about commuting convenience				
Rail access				
station distance <2km	-0.071**	-0.066**	-0.118**	-0.102**
	(0.035)	(0.027)	(0.052)	(0.043)
station distance >2km	-0.028	-0.022	-0.045	-0.035
	(0.037)	(0.032)	(0.035)	(0.031)
B. Happiness about living convenience				
Rail access				
station distance <2km	-0.061**	-0.054**	0.069**	0.063**
	(0.032)	(0.026)	(0.034)	(0.031)
station distance >2km	-0.023	-0.022	-0.033	-0.030
	(0.037)	(0.032)	(0.025)	(0.023)
C. Happiness about social environment				
Rail access				
station distance <2km	0.053**	0.046**	0.061**	0.055***
	(0.025)	(0.021)	(0.024)	(0.019)
station distance >2km	0.016	0.013	0.018	0.015
	(0.015)	(0.012)	(0.016)	(0.011)
D. Happiness about traffic safety				
Rail access				
station distance <2km	0.059**	0.042*	0.065**	0.061**
	(0.032)	(0.025)	(0.028)	(0.024)
station distance >2km	0.028	0.017	0.032	0.025
	(0.018)	(0.016)	(0.021)	(0.017)
E. Happiness about traffic pollution				
Rail access				
station distance <2km	-0.058**	-0.052**	-0.082**	-0.075**
	(0.027)	(0.023)	(0.036)	(0.030)
station distance >2km	0.038	0.013	0.025	0.017
	(0.025)	(0.015)	(0.024)	(0.011)
Household income	Yes	Yes	Yes	Yes
Other household characteristics	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Location characteristics	No	Yes	No	Yes
Sample size	1377	1377	1006	1006

Notes.---The dependant variable in the specifications A-E is the log of different happiness measures. Each specification is a separate set of regressions. Columns (1)-(2) is estimated using the whole sampled residents. Columns (3)-(4) is estimated using the non-market housing sub-sample. Data is aggregated to cell unit level for two snapshots: 2005 and 2009. The constant term of each regression is omitted for simplicity. All regressions shown in the table include the full set of controls. Standard errors corrected for clustering at the cell unit level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix D.

Appendix Table 3.5 Vertical deviation of happiness changes within 2km new station area

Station Name	Vertical deviation (1)	(2)	Location (3)	Station Name	Vertical deviation (1)	(2)	Location (3)
tiantongyuanbei	0.80	0.80	1	suzhoujie	0.02	0.16	1
tiantongyuan	0.57	0.80	1	bagou	0.07	0.08	1
tiantongyuannan	0.60	0.62	1	yuanmingyuan	0.16	0.19	1
lishuiqiaonan	0.43	0.20	1	beigongmen	0.12	0.01	1
beiyuanlubei	0.18	0.08	1	xiyuan	0.13	0.12	1
datunludong	0.37	0.04	1	beijingdaxuedongmen	0.20	0.28	1
huixinxijiebeikou	0.49	0.36	1	zhongguancun	0.14	0.23	1
huixinxijienankou	0.47	0.42	1	renmindaxue	-0.09	0.01	0
hepingxiqiao	0.36	0.21	0	weigongcun	0.04	0.19	0
hepinglibeijie	0.29	0.16	0	guojiatushuguan	0.10	0.18	0
anzhenmen	0.44	0.35	1	dongwuyuan	-0.03	0.12	0
mudanyuan	0.44	0.54	1	ciqikou	0.13	0.27	0
jiandemen	0.36	0.37	1	tiantandongmen	0.08	0.23	0
beitucheng	0.26	0.25	1	puhuangyu	0.01	0.14	0
xitucheng	0.42	0.44	1	liujiayao	0.16	0.02	1
aolinpikezhongxin	0.42	0.36	1	songjiazhuang	0.16	0.02	1
aolinpikegongyuan	0.40	0.25	1	caishikou	0.23	0.32	0
senlingongyuannanmen	0.39	0.18	1	taoranting	0.22	0.24	0
beixinqiao	0.03	0.02	0	beijingnanzhan	0.22	0.24	0
dongsi	-0.09	-0.09	0	majiabao	0.14	0.06	1
zhangzizhonglu	-0.15	-0.09	0	jiaomenxi	0.10	0.10	1
dengshikou	-0.17	-0.18	0	gongyixiqiao	0.19	0.26	1
xinjiekou	0.01	-0.18	0	jintaixizhao	0.19	0.14	0
pinganli	0.01	-0.01	0	hujialou	0.19	0.16	0
xisi	-0.04	-0.01	0	tuanjiehu	-0.01	-0.01	0
lingjinghutong	0.05	0.02	0	nongyezhanlanguan	0.08	0.04	0
shuangjing	0.54	0.56	0	liangmaqiao	0.15	0.16	0
jinsong	0.50	0.52	0	sanyuanqiao	0.12	0.21	1
haidianhuangzhuang	0.12	0.25	1	taiyanggong	0.31	0.16	1
				yonganli	0.28	0.28	0

Notes.---Columns (1) and (2) report the vertical deviation of the median happiness of commuting convenience within 2km of each new station area from the 45 degree line shown in the Figures 3.2 and 3.4 respectively. Column (3) indicates whether a new station is located in the suburb or not (suburb stations=1, central city stations=0).

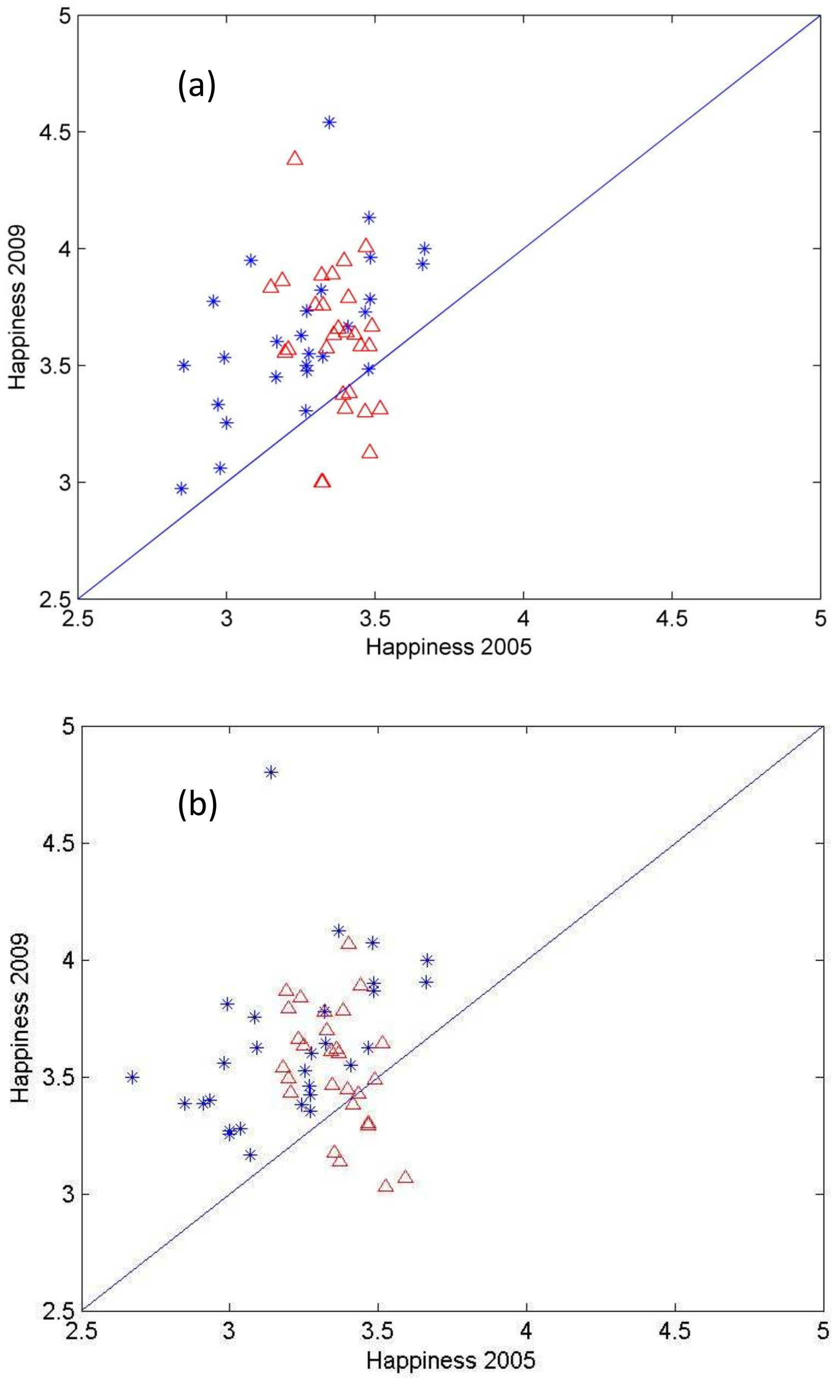
Appendix E.

Appendix Table 3.6 Variable name and definitions

Variable name	Definition
Household characteristics	
Income	Monthly household wages (in CNY 1000): 1 = 30 and less; 2 = 31–50; 3=51–100; 4 = 101–150; 5 = 151–200; 6 = 200 and above
Age	Age of the respondent (years). 1=young age:18-39; 0=others
Family size	Number of the family members in each household
Housing size	in m^2
Job rank	The job rank status: 1=entry-level job rank and below; 2=middle-level job rank; 3=high-level job rank and above
Education attainment	Highest education level:1 = primary school and lower; 2 = high school; 3 = undergraduate; 4 = postgraduate and above
Commuting time	one-way commuting time to work in minutes
Location characteristics	
CBD distance	Distance to the Beijing's central business district (CBD) in kilometres
School distance	Distance to the nearest middle school*school rank in kilometres
Park distance	Distance to the nearest park in kilometres
Bus stop distance	Distance to the nearest bus stop in kilometres
River	Indicator of proximity of cell unit to rivers (<500 meters)
Expressway	Indicator of proximity of cell unit to the expressway, ring road and primary road (<500 meters)
Airport	Indicator of proximity of cell unit to airport (<5 kilometre)
Olympic	Indicator of proximity of cell unit to the Olympic park (<2 kilometre)
Bedroom Area$_i$	Indicator of proximity of cell unit to the bedroom communities of *Yizhuang, Tiantongyuan, Tongzhou, Daxing* respectively (<2 kilometre)
Commuting mode	Proportion of public transport users in cell unit (%)
Employment Density	Total employment density in each zone (employees per km^2)
Population Density	Total population density in each zone (persons per km^2)
Old Building	Ratio of buildings built before 1949 in each zone (%)
Education Attainment	Median educational attainment in each zone:1=middle school or lower;2=high school;3=university;4=post graduate
Crime	Number of crimes per 1000 person in each zone
Public Housing	Percentage of people renting public housing in each zone

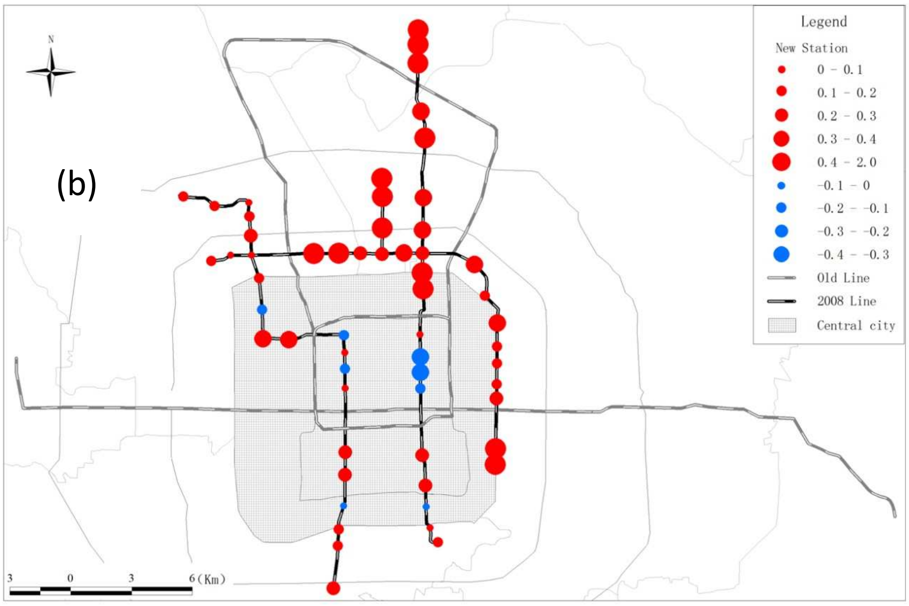
Notes.---All variables are aggregated to cell-unit, pre-post of the transport improvement, and used in regressions as controls.

Appendix F.



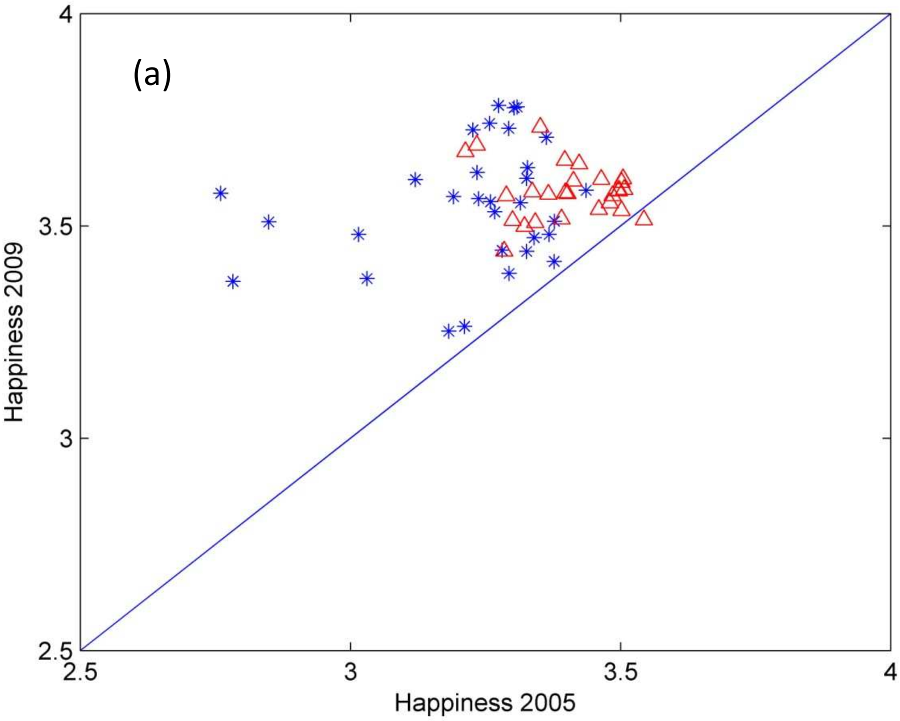
Appendix Figure 3.2 Happiness changes within 1km new station area

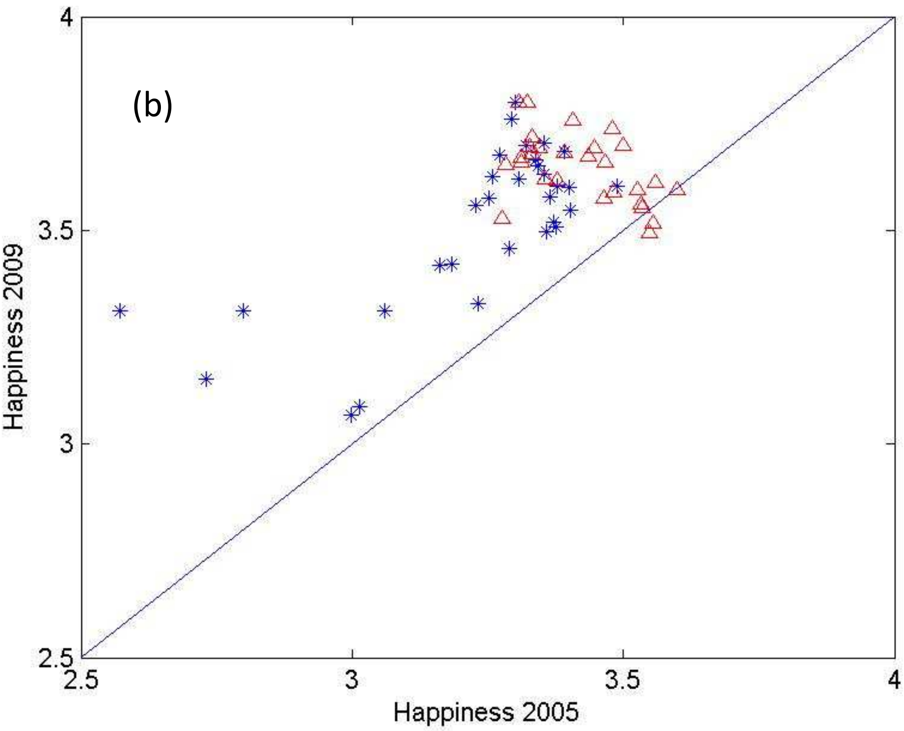
Notes.---Figure (a) shows the pattern of the whole sampled homeowners' median happiness value of commuting convenience within 1km of a new station. Figure (b) shows the pattern of non-market housing homeowners' median happiness value of commuting convenience within 1km of a new station.



Appendix Figure 3.3 Spatial distributions of happiness changes within 1km new station area

Notes.--- Figures (a) and (b) show the spatial distributions of changes in median value of happiness towards commuting convenience by using the whole sample and the non-market housing sample, respectively. Each circle label represents the vertical deviation of each dot in Appendix Figure 3.2 (a-b) from the 45 degree line accordingly.





Appendix Figure 3.4 Happiness changes within 4km new station area

Notes.---Figure (a) shows the pattern of the whole sampled homeowners' median happiness value of commuting convenience within 4km of a new station. Figure (b) shows the pattern of non-market housing homeowners' median happiness value of commuting convenience within 4km of a new station.

Appendix Figure 3.5 Spatial distributions of happiness changes within 4km new station area

Notes.--- Figures (a) and (b) show the spatial distributions of changes in median value of happiness towards commuting convenience by using the whole sample and the non-market housing sample, respectively. Each circle label represents the vertical deviation of each dot in Appendix Figure3.4 (a-b) from the 45 degree line accordingly.

IV. Conclusion

Conclusion

China and other BRICS countries have been investing heavily in urban infrastructure over the past decade. My research explores the real estate and perceived happiness consequences of local public goods improvements in Beijing. Despite significant public interest, there are surprisingly few studies on these important issues, especially in the context of a Chinese city. Many aspects of and findings from the papers are China-“firsts”.

The first paper, using the local parks as an example, clarifies the importance of conceptualizing amenity values not just in terms of their structural characteristics but how those characteristics interact with or are conditioned by social, economic, and spatial characteristics. I also point out that researchers estimating the amenity value should do a careful robustness check before directly applying the spatial econometric modelling results for any policy purposes. Paper two looks at how large local public goods improvement might affect local land prices, in particular, links to new rail transit constructions. The results suggest that public investment did spur the spatially targeted land market. Residential and commercial land parcels receiving increased station proximity have experienced appreciable price premiums, but that the relative importance of such benefits varies significantly over space. In paper three, I switch focus onto examining the direct effect from rail access changes on homeowners’ happiness, using repeated micro surveys conducted before and after the building of new rail transit lines in 2008 Beijing. My evidence shows that new rail transit developments not only provide sizeable happiness on commuting convenience to homeowners, but also affects homeowners' happiness in other dimensions of residential environment. The welfare estimates suggest the substantial benefits of non-marginal rail access improvements to homeowners' happiness. Perhaps more surprisingly, I find that these happiness effects have strong social-spatial differentiations. These findings add to the evidence that the government investment program in transport infrastructure has an important role to play in influencing homeowners' living experience.

This research has been undertaken against the backdrop of intense public and planning concerns. Over the past decade, on-going land and housing reform have finally given birth to a vibrant real estate market in urban China. To create a low-carbon urban environment, the Beijing municipal government has been investing heavily in local public goods---where parks and rail transits have consistently scored as the two largest public investment areas. On local parks, policymakers have had to spend huge maintenance fees for gardening and cleaning in order to strengthen their amenity benefits. On rail transits, lagging public transport development has long faced criticism, and policymakers have recently placed greater emphasis on increasing station proximities through new rail transit creations. Urban policymakers would gain substantial benefits from a better understanding of the impact of public investment in local public goods on land prices and homeowners' happiness with respect to different dimensions of residential environment.

Overall, my results go beyond popular narratives about the straightforward "positive" or "negative" effects associated with local amenities. The empirical findings quantify new evidence on the complex and subtle ways in which land markets capitalise on the value of local amenities such as parks and rail stations, and suggest that this is highly contingent upon local contextual factors. Of course, it is expected that the amenity values can be not just reflected by price premiums, but also contributes to people's subjective wellbeing. To this end, I documented the significantly heterogeneity in the effects from better rail access on influencing homeowners' happiness perceptions with respect to different dimensions of residential environment. This is a promising research field. Future works using long-run happiness data in different contexts to corroborate the robustness of my results would be useful.

Importantly, the results presented in this thesis would provide healthy policy implications for local governments and planners. While British politicians have recently argued about infrastructure, Chinese policymakers have been laying it out. My evidence from Beijing has shown that public investment in infrastructure programs can have significant capitalization effects on land markets. However, it should be noted that such capitalization effects may further evolve within the rapid urbanization process in China. Thus policy initiatives regarding public goods provision and land use planning should be tailored to fit the local contexts. Meanwhile, as the city government invested in the new rail transits, local homeowners' living experience changed. My evidence from the transport improvement supports the claim that the public investment program and residents' subjective wellbeing are not without connections. Indeed, Beijing homeowners' happiness can be significantly affected by increased rail access to their residence brought by the building of new rail transit lines. However, my results suggest that these welfare benefits vary considerably relative to different residential aspects, social groups and urban areas. To the extent that these results hold more broadly, these pieces of evidence provide direct implications for local governments to consider social-spatial differentiations when launching the place-based investment programs. Empirically, it is expected that the changes in happiness can be reflected in changes in housing demand so in some way can be capitalized into house prices. This will further result in differential residential mobility among residents with an inflow of those most benefiting in happiness terms from the improvements in transport accessibility. In so far as this occurred then there would be policy implications for neighbourhood dynamics and also for the long-term impact on social welfare. Thus policymakers should take effective steps to help maximize welfare, for example by offering sufficient affordable housing with reasonable distances to local amenities, by considering households' subjective assessments, and by making sound plans and appropriate government interventions that could help to gentrify the depressed areas.

References

Adriaanse, C. C. M., 2007. Measuring residential satisfaction: A residential environmental satisfaction scale. Journal of Housing and Built Environment 22, 287–304.

Ahlfeldt, G. M., 2011. If we build, will they pay? Predicting property price effects of transport innovations. SERC Discussion Papers, SERCDP0075. Spatial Economics Research Centre (SERC), London School of Economics and Political Sciences, London, UK.

Aitkin, M., Longford, N. T., 1986. Statistical modelling issues in school effectiveness studies. Journal of the Royal Statistical Society, series A 149, 1–43.

Alesina, A., Di Tella, R., MacCulloch, R., 2004. Inequality and happiness: are Europeans and Americans different?. Journal of Public Economics 88, 2009–2042.

Anderson S. T. West, S. E., 2006. Open space, residential property values, and spatial context. Regional Science and Urban Economics 36, 773–789.

Angrist, J., Pischke, J., 2009. Mostly harmless econometrics: An empiricist's companion. London: Princeton University Press.

Anselin L. 1990. Spatial dependence and spatial structural instability in applied regression analysis. Journal of Regional Science, 30, 185–207.

Anselin L, Gallo, J., 2006. Interpolation of air quality measures in hedonic house price models: Spatial aspects. Spatial Economic Analysis 1, 31–52.

Anselin, L., Lozano-Gracia, N., 2008. Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics 34, 5–34.

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89, 3–25.

Aragonés, J.I., Francescato, G., Gärling, T., 2002. Evaluating residential environment. In: *Residential environment: Choice, Satisfaction, and Behavior.* Aragonés, J.I., Francescato, G., Gärling, T. (Eds.), Westport: Bergin and Garvey.

Baba, Y., Austin, D.M., 1989. Neighborhood environmental satisfaction, victimization, and social participation as determinants of perceived neighborhood safety. Environment and Behavior 21, 763–780.

Baum-Snow, N., Kahn, M.E., 2000. The effects of new public projects to expand urban rail transit. Journal of Public Economics 77, 241-263.

Bajic, V., 1983. The effects of a new subway line on housing prices in metropolitan Toronto. Urban Studies 20, 147-158.

Baker, M., Stabile, M., Deri, C., 2004. What do self-reported, objective, measures of health measure?. Journal of Human Resources 39, 1067–1093.

Barbosa, O., Tratalos, J.A., Armsworth, P.R., et al., 2007. Who benefits from access to green space? A case study from Sheffield, UK. Landscape and Urban Planning 83, 187–195.

Bateman, I. J., Day, B. Georgiou, H. S., Lake, I. R. 2006. The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. Ecological Economics 60, 450–460.

Berry, B. J. L. Bednarz, R. S., 1979. The disbenefits of neighbourhood and environment to urban property, in *The economics of neighbourhood*, D. Segal(eds.), New York: Academic Press.

Bertaud, A., Renaud, B., 1997. Socialist cities without land markets. Journal of Urban Economics 41, 137–151.

Bitter, C., Mulligan, G.F., Dall'erba, S., 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems 9, 7–27.

Blanchflower, D., Oswald, A., 2004. Wellbeing over time in Britain and the USA. Journal of Public Economics 88, 1359–1386.

Bolitzer, B., Netusil, N.R., 2000. The impact of open spaces on property values in Portland, Oregon. Journal of Environmental Management 59, 185–193.

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

Breffle,W.S., Morey, E.R., Lodder, T.S., 1998. Using contingent valuation to estimate a neighborhood's willingness to pay for undeveloped urban land. Urban Studies 35, 715–727.

Brocker, J. 1998. Operational spatial computable general equilibrium modeling, Annals of Regional Science 32, 367–387.

Brueckner, J.K., Thisse, J.-F., Zenou, Y., 1999. Why is downtown Paris so rich and Detroit so poor?: An amenity based explanation. European Economic Review 43, 91–107.

Brunsdon, C., Fotheringham, A.S., Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression. Journal of Regional Science 39, 497–524.

Cai, H., Henderson, J.V., Zhang, Q., 2009. China's land market auctions: Evidence of corruption. NBER Working Paper, No. 15067.

Campbell, A., Converse, P. E., Rodgers, W. L., 1976. The quality of American life: Perceptions, evaluations and satisfactions. New York: Russell Sage.

Chapman, D.W., Lombard, J.R., 2006. Determinants of neighborhood satisfaction in fee-based gated and nongated communities. Urban Affairs Review 41, 769–799.

Carruthers, J.I., Clark, D.E., 2010. Valuing environmental quality: A space-based strategy. Journal of Regional Science 50, 801–832.

Cervero, R., Duncan, M., 2001. Rail transit's value added: effect of proximity to light and commuter rail transit on commercial land values in Santa Clara County California. Paper prepared for the Urban Land Institute National Association of Realtors, Washington, DC.

Cheshire, P.C., Sheppard, S. 1995. On the price of land and the value of amenities. Economica 62, 247–267.

Cheshire, P.C., Sheppard S., 1998. Estimating demand for housing, land, and neighbourhood characteristics. Oxford Bulletin of Economics and Statistics 60, 357–382.

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.

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

Chiesura, A., 2004. The role of urban parks for the sustainable city. Landscape and Urban Planning 68, 129–138.

Cho S. H., J. M. Bowker, Park, W. M., 2006. Measuring the contribution of water and green space amenities to housing values: An application and comparison of spatially weighted hedonic models. Journal of Agricultural and Resource Economics 31, 485–507.

Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.

Clark, A., Oswald, A., 1994. Unhappiness and Unemployment. The Economic Journal 104, 648–659.

Clark, A., Oswald, A., 1996. Satisfaction and comparison Income. Journal of Public Economics 61, 359–381.

Cleveland W.S., Devlin, S. J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association 83, 596–610.

Coffman, C., Gregson, M.E., 1998. Railroad development and land value, Journal of Real Estate Finance and Economics 16, 191–204.

Cook, C.C., 1988. Components of neighborhood satisfaction: Responses from urban and suburban single-parent women. Environment and Behavior 20, 115–149.

Cornaglia, F., Leigh, A., 2011. Crime and Mental Wellbeing. CEP Discussion Papers, No.1049. Centre for Economic Performance (CEP), London School of Economics and Political Sciences, London, UK.

Cornaglia, F., Crivellaro, E., McNally, S., 2012. Mental Health and Education Decisions. CEP Discussion Papers, No.0136. Centre for Economic Performance (CEP), London School of Economics and Political Sciences, London, UK.

Davidson, R., MacKinnon, J. G., 1993. Estimation and inference in econometric. New York:

Oxford University Press.

Davis, E.E., Fine-Davis, M., 1981. Predictors of satisfaction with housing and neighbourhood: a nationwide study in the Republic of Ireland. Social Indicators Research 9, 477–494.

Davis, F.W., 1970. Proximity to a rapid transit station as a factor in residential property values. The Appraisal Journal 38, 554–572.

Day, B. H., Mourato, S., 1998. Willingness to pay for water quality maintenance in Chinese rivers. CSERGE Working Paper, WM 98-02. University College London, UK.

Day, B. H., 2007. Distribution-free estimation with interval-censored contingent valuation data: Troubles with Turnbull?. Environmental and Resource Economics 37, 777–795.

Day, B. H., Bateman, I. J., Lake, I. R., 2007. Beyond implicit prices: Recovering theoretically consistent and transferable values for noise avoidance from a hedonic property price model. Environmental and Resource Economics 37, 211–232.

Debrezion, G., Pels, E., Rietveld, P., 2007. The impact of railway stations on residential and commercial property value: A meta-analysis. Journal of Real Estate Finance and Economics 35, 161–180.

Debrezion, G., Pels, E., Rietveld, P., 2011. The impact of rail transport on real estate prices: An empirical analysis of the Dutch housing market. Urban Studies 48, 997–1015.

Deng, Y., Zheng, D., Ling, C., 2005. An early assessment of residential mortgage performance in China. Journal of Real Estate Finance and Economics 31, 117–136.

Deng, Y., Liu, P., 2009. Mortgage prepayment and default Behaviour with embedded forward contract risks in China's housing market. Journal of Real Estate Finance and Economics 38, 214–240.

Dewees, D. N., 1976. The effect of a subway on residential property values in Toronto, Journal of Urban Economics 3, 357–369.

Di Tella, R., MacCulloch, R., Oswald, A., 2001. Preferences over inflation and unemployment: Evidence from surveys of happiness. American Economic Review 91, 335–341.

Ding, W., Zheng, S., Guo, X., 2010. Value of access to jobs and amenities: Evidence for new residential properties in Beijing. Tsinghua Science and Technology 15, 595–603.

Dowall, D.E., 1994. Urban residential redevelopment in the People's Republic of China. Urban Studies 31, 1497–1516.

Fan, Y. Q., Ullah, A., 1999. Asymptotic normality of a combined regression estimator. Journal of Multivariate Analysis 71, 191–240.

Farber, S., Páez, A., 2007. A systematic investigation of cross-validation in LWR model estimation: empirical analysis and Monte Carlo simulation. Journal of Geographic Systems 9, 371–396.

Ferrer-i-Carbonell, A., 2005. Income and wellbeing: an empirical analysis of the comparison income effect. Journal of Public Economics 89, 997–1019.

Ferrer-i-Carbonell, A., Frijters, P., 2004. How important is methodology for the estimates of the determinants of happiness? The Economic Journal 114, 641-659.

Fik, T., Ling, D. C., Mulligan, G. F., 2003. Modeling spatial variation in housing prices: a variable interaction approach. Real Estate Economics 31, 623–646.

Fotheringham, A. S., Brunsdon, C., Charlton, M. E., 2002. Geographically weighted regression: the analysis of spatially varying relationships. Chichester: John Wiley.

Freeman, A. M., 1979. The hedonic price approach to measuring demand for neighbourhood characteristics, in *The economics of neighbourhood*, David Segal (eds.). New York: Academic Press, Inc..

Freeman, A.M.III., 2003. The measurement of environmental and resource values: Theory and Methods. Washington, D.C.: Resources for the Future.

Frey, B., 2008. Happiness: A revolution in economics. Cambridge, MA and London: MIT Press.

Frey, B., Luechinger, S., Stutzer, A., 2009. The life satisfaction approach to the value of public goods: the case of terrorism. Public Choice 138, 317–345.

Frey, B., Luechinger, S., Stutzer, A., 2010. The life satisfaction approach to environmental valuation. Annual Review of Resource Economics 2, 139-160.

Frey, B., Stutzer, A., 2001. What can economists learn from happiness research?. Journal of Economic Literature 40, 402–435.

Frijters, P., Haisken-DeNew, J., Shields, M., 2004. Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification. American Economic Review 94, 730–740.

Fu, Y.M., Tse, D.K., Zhou, N., 2000. Housing choice behaviour of urban workers in China's transition to a housing market. Journal of Urban Economics 47, 61–87.

Galster, G., Hesser, G. W. 1981. Residential satisfaction: Compositional and contextual correlates. Environment and Behavior 13, 35–758.

Geoghegan, J., 2002. The value of open spaces in residential land use. Land Use Policy 19, 91–98.

Gibbons, S., Machin, S., 2003. Valuing English primary schools. Journal of Urban Economics 53, 197–219

Gibbons, S., 2004. The costs of urban property crime. Economic Journal 114, 441–463.

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 4, 99–119.

Gibbons, S., Mourato, S., Resende, G. 2011. The amenity value of English nature: A hedonic price approach. SERC Discussion Papers, SERCDP0074, Spatial Economics Research Centre (SERC), London School of Economics and Political Sciences, London, UK.

Gibbons, S., Silva, O., 2011. School quality, child wellbeing and parents' satisfaction. Economics of Education Review 30, 312–331.

Gibbons, S., Overman, H. G., 2012. Mostly pointless spatial econometrics?. Journal of Regional Science 52, 172–191.

Gibbons, S., Machin, S., Silva, O., 2012. Valuing school quality using boundary discontinuities. CEE discussion papers, CEE DP 132. Centre for the Economics of Education, London, UK

Glaeser, E.L., Kahn, M.E., Rappaport, J., 2008. Why do the poor live in cities? The role of public transportation. Journal of Urban Economics 63, 1–24.

Grass, R. G., 1992. The estimation of residential property values around transit station sites in Washington, DC. Journal of Economics and Finance 16, 139–146.

Gruber, K.J., Shelton, G.G., 1987. Assessment of neighborhood satisfaction by residents of three housing types. Social Indicators Research 19, 303–315.

Gunn, H. 2000. An introduction to the valuation of travel time savings and losses, Chapter 26 in: *Handbook of transport modelling*, D.A. Hensher, K.J. Button (Eds.), Amsterdam: Elsevier Science.

Gyourko, J., Kahn, M.E., Tracy, J., 1999. Quality of life and the environment. , in: *Handbook of regional and urban economics vol. 3: applied urban economics.*, P. Cheshire, E. Mills (Eds.), Amsterdam: Elsevier-North Holland.

Hilber, C.A.L., 2011. The economic implications of house price capitalization: A survey of an emerging literature. SERC Discussion Papers, SERCDP0091. Spatial Economics Research Centre (SERC), London School of Economics and Political Sciences, London, UK.

Hoshino, T., Kuriyama, K., 2010. Measuring the benefits of neighbourhood park amenities: Application and comparison of spatial hedonic approaches. Environmental and Resource Economics 45, 429–444.

Huang, Y., Clark, W.A.V., 2002. Housing tenure choice in transitional urban China: a multilevel analysis. Urban Studies 39, 7–32.

Hur, M., Morrow-Jones, H., 2008. Factors that influence residents' satisfaction with neighborhoods. Environment and Behavior 40, 619–635.

Irwin, E. G., Bockstael, N. E., 2001a. The problem of identifying land use spillovers: Measuring the effects of open space on residential property values. American Journal of Agricultural Economics 8, 698–704.

Irwin, E .G., Bockstael, N. E., 2001b. Interacting agents, spatial externalities, and endogenous evolution of residential land use pattern. Journal of Economic Geography 2, 31-54.

Irwin, E. G., 2002. The effects of open space on residential property values. Land Economics 78, 465–480.

Jiang, L., 2006. Living Conditions of the Floating Population in Urban China. Housing Studies 21, 719–744.

Jim, C. Y., Chen, W. Y., 2010. External effects of neighbourhood parks and landscape elements on high-rise residential value. Land Use Policy 27, 662–670.

Kahn, M. E., 2007. Gentrification trends in new transit-oriented communities: Evidence from 14 cities that expanded and built rail transit systems. Real Estate Economics 35, 155–182.

Kahneman, D., Krueger, A.B., 2006. Developments in the measurement of subjective wellbeing. Journal of Economic Perspectives 20, 3–24.

Kahneman, D., Krueger, A.B., Schkade, D., etc, 2006. Would you be happier if you were richer? A focusing illusion. Science 312, 1908–1910.

Knaap, G.J., Ding,C., Hopkins, L.D., 2001, Do plans matter?: Effects of light rail plans on land values in station areas. Journal of Planning Education and Research 21, 32–39.

Krueger A., Schkade, D., 2007. The reliability of subjective wellbeing measures. NBER Working Paper No. 13027.

Layard, R., 2006. Happiness and public policy: a challenge to the profession. The Economic Journal 116, 24–33.

Leung, Y., C. Mei, W. Zhang., 2000. Statistical tests for spatial nonstationary based on the geographically weighted regression model. Environment and Planning A 32, 9–32.

Logan, J.R., Zhang, W.W., Xu, H.W., 2010. The winners in China's urban housing reform. Housing Studies 25, 101–118.

Lu, M., 1999. Do people move when they say they will? Inconsistencies in individual migration behavior. Population and Environment 20, 467–488.

Luechinger, S., 2009. Valuing air quality using the life satisfaction approach. Economic Journal 119, 482–515.

McConnell, V., Walls, M., 2005. The value of open space: evidence from studies of non-market benefits. Working Paper. Resources for the Future, Washington, DC.

McDonald, J.F., Osuji, C.I. 1995. The effect of anticipated transportation improvement on residential land values. Regional Science and Urban Economics 25, 261-278.

McGillivray, M., Clarke, M., 2006. Human wellbeing: Concepts and measures. In: *Understanding Human Wellbeing*, Mark McGillivray and Matthew Clarke (eds.). Basingstoke: Palgrave MacMillan.

McMillen, D.P. 2001. Nonparametric employment subcenter identification. Journal of Urban Economics 50, 448–473.

McMillen, D.P., McDonald, J., 2004. Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983-1999. Real Estate Economics 32, 463–486.

McMillen, D. P., 2010. Issues in spatial data analysis. Journal of Regional Science 50, 119–141.

McMillen, D.P., Redfearn, C.L.,2010. Estimation and hypothesis testing for nonparametric hedonic price functions. Journal of Regional Science 50, 712–733.

Mohan, J., Twigg, L., 2007. Sense of place, quality of life and local socioeconomic context: evidence from the survey of English housing, 2002/03. Urban Studies 44, 2029–2045.

Oates, W., 1969. The effects of property taxes and local public services on property values: An empirical study of tax capitalisation and the Tiebout hypothesis. Journal of Political Economy 77, 957–971.

Osland, L., Thorsen, I. 2008. Effects on housing prices of urban attraction and labor market accessibility. Environment and Planning A 40, 2490–2509.

Oswald, A., Wu, S., 2010. Objective confirmation of subjective measures of human wellbeing: Evidence from the U.S.A. Science 327, 576–579.

Palmquist, R. B., 1992. Valuing localized externalities. Journal of Urban Economics 32, 40–44.

Parkes, A., Kearns, A., Atkinson, R., 2002. What makes people dissatisfied with their neighbourhoods?. Urban Studies 39, 2413–2438.

Permentier M., Bolt G., Ham, M. van., 2011. Determinants of neighbourhood satisfaction and perception of neighbourhood reputation. Urban Studies 48, 977–996.

Redfearn, C., 2009. How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. Regional Science and Urban Economics 39, 297–306.

RICS Policy Unit. 2002. Land value and public transport: stage 1-summary of findings. RICS, London.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy 82, 34–55.

Sato, H., 2006. Housing inequality and housing poverty in urban China in the late 1990s. China Economic Review 17, 37–50.

Schultz, S.D., King, D.A., 2001. The use of census data for hedonic price estimates of open-space amenities and land use. Journal of Real Estate Finance and Economics 22, 239–252.

Sheppard, S., 1999. Hedonic analysis of housing markets, in: *Handbook of regional and urban economics vol. 3: applied urban economics.*, P. Cheshire, E. Mills (Eds.), Amsterdam: Elsevier-North Holland.

Sieg, H., V. K. Smith, H. S. Banzhaf, R. Walsh., 2004. Estimating the general equilibrium benefits of large changes in spatially delineated public goods. International Economic Review 45, 1047-1077.

Takeuchi, A., Cropper, M., Bento, A., 2008. Measuring the welfare effects of slum improvement programs: The case of Mumbai. Journal of Urban Economics 64, 65–84.

Tra, C. 2010. A discrete choice equilibrium approach to valuing large environmental changes. Journal of Public Economics 94, 183–196.

UN-Habitat., 2010. The State of Chinese Cities 2010/2011. Beijing: Foreign Languages Press.

Van Praag, B.M.S., Baarsma, B.E., 2005. Using happiness surveys to value intangibles: The case of airport noise. Economic Journal 115, 224–246.

Vessali, K. V., 1996. Land use impacts of rapid transit: a review of empirical literature. Berkeley Planning Journal 11, 71-105.

Wang, R., 2009. The structure of Chinese urban land prices: Estimates from benchmark land price data. Journal of Real Estate Finance and Economics 39, 24–38.

Winkelmann, L., Winkelmann, R., 1998. Why Are the Unemployed So Unhappy? Evidence from Panel Data. Economica 65, 1–15.

Wheeler, D.C., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographic Systems 7, 161–187.

Wu,J., Gyourko, J., Deng, Y. 2011. Evaluating conditions in major Chinese housing markets. NBER Working Paper, No. 16189.

Wu, W.J., Zhang, W.Z., Dong, G.P. 2013. Determinant of residential location choice in a transitional housing market: Evidence based on micro survey from Beijing. Habitat International, 39, 16–24.

Zhang, W., Gao, X., 2008. Spatial differentiations of traffic satisfaction and its policy implications in Beijing. Habitat International 32, 437–451.

Zhang, W., Yin, W., Zhang, J., etc., 2006. A Study of Livable Cities in China (Beijing). Beijing: Social Sciences Academic Press (China).

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., Fu, Y.M., Liu, H.Y., 2006. Housing-choice hindrances and urban spatial structure: Evidence from matched location and location-preference data in Chinese cities. Journal of Urban Economics 60, 535–557.

Zhu, J., 2005. A transitional institution for the emerging land market in urban China. Urban Studies 42, 1369-1390.