



Valuing the urban hukou in China: Evidence from a regression discontinuity design for housing prices

Yu Chen^a, Shaobin Shi^{b,c}, TangYugang^{b,c,*}

^a Department of Economics, University of Graz, Universitätsstraße 15/F4, 8010, Graz, Austria

^b Center for Public Economy and Policy Research (CPEPR), Shandong University, 27 Shanda Nanlu, Jinan, 250100, PR China

^c School of Economics, Shandong University, 27 Shanda Nanlu, Jinan, 250100, PR China

ARTICLE INFO

JEL classification:

P25
H75
R38
J61

Keywords:

Hukou system
Willingness to pay for hukous
Acquiring hukous by purchasing houses
Regression discontinuity design

ABSTRACT

This paper explores the demand side of *hukou* (household registration) acquisition in China by estimating the market valuation of urban hukou. Based on Jinan City's *acquiring the hukous by purchasing houses* policy, this paper uses houses with a floor area slightly larger and slightly smaller than the minimum required as the treatment and control groups, respectively, to implement a regression discontinuity design. The results show that residents' willingness to pay for urban hukou in Jinan City was approximately 90,000–126,000 yuan in 2017. We also find great heterogeneity in different housing submarkets; the value of hukou is much higher in immigrant-dominated housing markets and top primary school districts. Our findings are robust to parametric and nonparametric estimates and different model specifications. We perform falsification tests by assuming a false policy introduction date and placebo tests based on rental data. Our analysis offers insights for hukou system reform and public services provision.

1. Introduction

The hukou system, which is China's household registration system, is not only the method for population registration but also the instrument by which a city government grants access to a variety of exclusive rights—most importantly, the right to enjoy certain public services and welfare. The hukou system was established in 1958 to control the influx of rural residents into cities and to modernize the development of capital-intensive heavy industries (Lin et al., 1994; Cheng and Selden, 1994). Since then, the urban hukou has conferred access to residential welfare benefits. Substantial public services continue to be closely related to the urban hukou and cannot be enjoyed equally by “floating” populations (i.e., citizens seeking to relocate in an urban area). These services mainly consist of compulsory education for children, registration for local examinations, social insurance subsidies for nonworking individuals, urban minimum subsistence allowances, and affordable housing benefits (Wu and Zhang, 2010; Tao et al., 2011). From the perspective of individuals who seek an urban hukou, its value can be measured as an individual's willingness to pay (WTP) for the hukou. From the perspective of hukou suppliers (governments), they must bear the costs of these public services and welfare benefits, such as the construction and operating costs of schools and social security subsidies for nonworking individuals.

Furthermore, relevant analyses on both the demand and supply sides can be of significance for the policy design of hukou granting of local governments.

Many studies estimate the supply (cost) aspects of granting hukous. China's Research Group of the Ministry of Construction (2006) finds that for every additional individual in an urban population, it is necessary to increase the support fees for municipal public facilities by 20,000 yuan (RMB) in small cities, 30,000 in medium cities, 60,000 in large cities, and 100,000 in mega cities. The China Development Research Foundation (2010) finds that the average cost of rural workers' citizenship in China is approximately 100,000 yuan. In terms of government expenditures for rural workers' citizenship, based on investigations in Chongqing, Wuhan, Zhengzhou, and Jiaying, the The Research Group of the Development Research Center of the State Council (2011a, 2011b) calculates that the cost is between 77,000 yuan and 85,000 yuan. Qu and Cheng (2013), Zhang and Wu (2013), and Ding and Xu (2014) estimate the costs of granting the urban hukou from different perspectives.

This paper fills an important gap by exploring the demand side of hukou granting in China. We aim to estimate the market valuation of the urban hukou. To the best of our knowledge, previous studies have not provided a quantitative method for estimating the market WTP for the urban hukou. In fact, hukou “transactions” have been recorded in China.

* Corresponding author. CPEPR and School of Economics, Shandong University, 27 Shanda Nanlu, Jinan, 250100, PR China.

E-mail addresses: yu.chen@uni-graz.at (Y. Chen), shishaobin@email.sdu.edu.cn (S. Shi), ygtang@email.sdu.edu.cn (Y. Tang).

As mentioned by Lu (2002), in the 1990s, governments in Lai'an and Qianjiang Counties in Chuzhou, Anhui Province, publicly sought investment at a rate of 5000 yuan for one urban hukou; within only a few days, the public income of these counties had increased by more than 3 million yuan. *The People's Daily* reported that the Beijing hukou could cost as much as 300,000 yuan on the black market in 2013.¹

This paper uses the clause that addresses “acquiring the hukou by purchasing a house” in the settlement policies of certain large and medium-sized cities to perform discontinuity regression and to empirically estimate the market value of the urban hukou. These clauses stipulate the minimum floor area of a house required to apply for the local hukou. The basic idea is that assuming the minimum floor area stipulated in the clause is measured in square meters (sq.m.), any migrant who intends to acquire a hukou by purchasing a house must buy a house of that size or larger. Otherwise, he or she will not be able to obtain the local hukou or be granted access to the rights and benefits enjoyed by local residents. Therefore, we take houses with a floor area slightly larger than the minimum required as the treatment group and houses with a floor area slightly smaller than the minimum required as the control group to implement a regression discontinuity design (RDD).

Our analysis belongs to an active research strand that uses a combination of the hedonic pricing model (HPM) and natural experimental methods to evaluate policy effects or the value of nonmarket goods or services. Black (1999) uses boundary fixed-effects models to remove the variation in neighborhoods, taxes, school spending, etc., and she precisely infers parents' WTP for better school quality. Gibbons et al. (2013) improve the methodology of boundary discontinuity regression by matching and weighting, and they find that a one-standard-deviation change in either a school's average value added or prior achievement raises prices by approximately 3%. Andreyeva and Patrick (2017) use the hedonic difference-in-differences model to separate households' WTP for a charter school. Chay and Greenstone (2005) exploit the structure of the Clean Air Act to provide new evidence on the capitalization of total suspended particulates in air pollution into housing values. Moreover, a large strand of research focuses on different kinds of nonmarket goods and services. For example, Pope (2008a) estimates the value of airport noise using data from the housing market; Linden and Rockoff (2008) and Pope (2008b) estimate the value of crime reduction; and Bui and Mayer (2003) and Gayer et al. (2000) estimate the value of hazardous waste and toxic emissions.

This paper uses specifications for the floor area cutoff point (90 sq.m.) for housing purchases in Jinan City's urban hukou management policy and implements an RDD to evaluate residents' WTP for the urban hukou. Both the parametric and nonparametric estimations indicate that hukou rights cause the unit price of residential property to rise by approximately 1000–1400 yuan. In other words, residents' WTP for the Jinan City hukou will be between approximately 90,000 and 126,000 yuan. This method and these empirical findings complement research on the cost of citizenization by examining the demand side.

The remainder of the paper is organized as follows. Section 2 presents the (institutional) background of the hukou-granting policy and theoretical considerations. Section 3 describes the data, empirical strategy, and summary statistics. In Section 4, we present the empirical results of the parametric estimation. Section 5 conducts robustness tests. Section 6 are some extensions and Section 7 concludes.

2. Policy background and theoretical considerations

First, we present the institutional background of the policy from the national (i.e., the central government) perspective. Several time points are critical:

- (1) Under the People's Republic of China Household Registration Regulation system, which was established in 1958, China began to mandate a strict restriction and regulation of free relocation by the citizenry. Moreover, the government divided residents into two categories: the agricultural hukou and the nonagricultural, or urban, hukou. People with an agricultural hukou are also called peasants. Because the central government prioritized industrialization, state welfare programs—which were tied to hukou status—heavily favored urban residents; holders of agricultural hukous were unable to access these benefits and were saddled with inferior welfare policies. The conversion of hukou status from agricultural to nonagricultural was highly restricted. With the economic reform and opening up in China, the market economy in rural areas revived. Township and village enterprises developed rapidly, and an increasing number of people with agricultural hukous moved to towns for jobs or business, hoping to change their social status from rural to urban citizenship with a nonagricultural hukou.
- (2) In October 1984, the State Council issued the “Notice on Peasants Entering Towns and Settling with Local Household Registration” to allow peasants to take their own rations into rural towns (but not county towns).
- (3) In July 1985, the Ministry of Public Security promulgated the “Provisional Regulations on Urban Population Management,” which set the quota for granting nonagricultural status to agricultural residents at 0.02%. However, the quota system generated considerable corruption. At the same time, the government adopted a more open attitude toward the settlement of rural workers in cities.
- (4) In June 1997, the State Council approved the Ministry of Public Security's “Pilot Scheme for the Reform of the Household Registration System in Small Towns and Opinions on Improving the Rural Household Registration Management System.” This scheme clearly stipulated that rural workers (and their direct relatives) who found jobs in small cities and towns or bought real estate would be allowed to apply for a regular permanent hukou in these towns and cities.
- (5) In July 1998, the Ministry of Public Security issued the “Guiding Opinions on Solving Several Conspicuous Problems in Current Hukou Management.” According to these opinions, newborns could acquire hukous along with their fathers, and elderly individuals could live with their children. Those who invested in cities, set up businesses, or purchased houses could acquire urban hukous, as long as they had a fixed residence and a legal and stable occupation or source of income or had lived there for a certain number of years and complied with relevant local government regulations. Their direct relatives were also allowed to acquire urban hukous.
- (6) In March 2001, to accelerate the development of urbanization, the State Council approved the “Opinions on Promoting the Reform of the Household Registration System in Small Towns” of the Ministry of Public Security. Presently, the government no longer specifies a quota for granting nonagricultural hukous to those who have permanent residence in small cities and towns, and it effectively safeguards their legal right to settle in small towns.

From the evolution of the abovementioned national-level hukou regulations, we see that household registration has been linked with the purchase of a house since the second half of the 1990s. Since the turn of the 21st century, with the rapid development of China's real estate market, the persistence of high housing prices and the emergence of an asset bubble have prompted local governments to consider their own cases and to introduce a number of restrictions on house purchases and mortgages, which has loosened the relationship between household registration and the purchase of a house.

Since the end of the 20th century, in many large and medium-sized

¹ “What's the Value of Hukou in China's Mega Cities?” in *People's Daily*, Oct.11, 2013.

cities, household registration—as a mechanism for screening desirable talents and providing exclusive rights to public service consumption—has been tied to the purchase of a house. Tenants, in contrast, do not have equal access to public services such as preschool education, compulsory education, higher education, and social security, or they have only comparatively weak public service consumption priorities. Clearly, there is a disparity in rights between renters and homeowners.

Furthermore, to acquire a hukou in some cities, it is necessary not only to purchase a house but also to purchase a house with the minimum floor area required by hukou management regulations. Only buyers who purchase a house of at least the minimum area prescribed can acquire a local hukou. Otherwise, buyers can obtain only the same citizen rights as tenants. For example, as early as 1996, Shanghai's city government stipulated that those who purchased newly built homes with a floor area of no less than 80 sq.m. and a value of no less than 400,000 yuan would be eligible for a blue-stamped hukou.² Additionally, in 2004, the city government of Nanjing specified a cutoff point for floor area (more than 60 sq.m.) required to acquire the hukou.

Next, we focus on the evolution of home purchase policies for the hukou management system in Jinan since 2000, as our empirical analysis is based on data from Jinan. According to Wikipedia,³ “Jinan is the capital of Shandong province in Eastern China. The area of present-day Jinan has played an important role in the history of the region from the earliest beginnings of civilization and has evolved into a major national administrative, economic, and transportation hub.” Its urban area is 3304 square kilometers, and its urban population is 4,693,700 (as of 2010). Among large cities in China, it plays a significant role.

The critical time points in the case of Jinan are as follows:

- (1) In 2001, the “Notice of the People's Government of Jinan Municipality on Reforming the Urban Household Registration System” set the basic requirements for the hukou as “a legal fixed residence and a stable occupation and source of income.” It stipulated five ways to obtain a hukou: purchasing a home, introducing talents, making business investments, being a husband or wife who is seeking to reunite, and being a rural resident whose farmland was urbanized. Those who meet one of the above conditions are allowed to acquire a hukou in Jinan's urban area. Here, we focus on the specific terms of home purchase. The notice stipulates that for domestic residents, the home buyer and his or her spouse and their minor children are allowed to acquire hukous in the home's location if (i) the floor area of a purchased house in Jinan City is no less than 100 sq.m., the total housing price is no less than 500,000 yuan, and the buyer obtains a real estate certificate; or (ii) the floor area of the purchased house is no less than 80 sq.m., the total price of the house is no less than 250,000 yuan, and the buyer has obtained a real estate certificate and lived in the house for more than three years. The notice also emphasizes that in reforming the household registration system, one of the basic principles is to “require strict control over the entry of low-quality immigration and to prevent low-level expansion of the urban population and aging of the migrating population.” These guidelines reflect the local government's strong fiscal motivations in the process of urbanization and citizenization.
- (2) In 2005, the Jinan City government slightly strengthened the conditions for acquiring a hukou by purchasing a house. Only an individual who purchases a house with a floor area of no less than 100 sq.m. in the urban area, acquires a real estate certificate with full payment, actually resides in the purchased house, and is legally employed is eligible to apply for a hukou in the location of

the house purchased, together with his or her spouse and minor children.⁴

- (3) In 2008, to stabilize the economy and boost the recovery of the real estate market after the global financial crisis, the Jinan City government relaxed the conditions for home purchase. In November 2008, the General Office of the People's Government of Jinan Municipality released the “Notice on Maintaining the Steady and Healthy Development of the Real Estate Market,” which was jointly issued by the Municipal Construction Committee and nine other city government bureaus. This notice stipulates the conditions for hukou granting by home purchase. “In the urban area, the hukou will be granted if the homeowner purchases a newly built house with a floor area of 90 sq.m. or more and he or she obtains a real estate certificate. With regard to second-hand houses, the hukou will be granted if the floor area is over 90 sq.m. and a real estate certificate has been held for more than two years. ... Each dwelling can enjoy the policy of acquiring a hukou by purchasing a house only once within a five-year period.”
- (4) Since 2009, both central and local governments have mainly focused on controlling housing prices nationwide. More restrictions have been imposed on qualifying house purchases, e.g., a sufficiently long history of paying personal income tax or social security contributions in Jinan City. The “Implementation Opinions of the People's Government of Jinan Municipality on Further Deepening the Reform of the Household Registration System” (2017) promulgated in August 2017 abolished quantitative restrictions on acquiring hukous based on investment, tax revenue contributions, and house purchase. The threshold floor area requirements were cancelled. For ordinary immigrating workers, purchasing a house is still a necessary condition for settling within the city's central urban jurisdictions; however, the city government no longer imposes mandatory requirements for the size of the house. Therefore, the floor area requirements for house purchases prior to the reform appear to be a valuable but uncommon case that will enable us to study the value of the hukou and the underlying right to enjoy public services and welfare in cities during the evolution of the hukou system in China.

We can further derive a simple theoretical background for the value of the hukou that is similar to the models for public goods provision in the theory of mechanism design. We consider the government's policy concerning floor area requirements for acquiring hukous by purchasing houses as an institutional parameter. A representative individual (without an urban hukou) has a private valuation (WTP for each sq.m.) t for the hukou. Let s denote the area of the house and x denote the other (aggregate) characteristics of the house. The Chinese government provides a hukou-granting rule $D(s) = 0, \text{ or } 1$, where 0 denotes “a homeowner has no right to apply for a hukou” and 1 denotes “a homeowner has the right to apply for a hukou.” Such an indicator (binary-valued) function normally takes the form of a step function with a cutoff point s^* . If $s \geq s^*$, then $D(s) = 1$, and if $s < s^*$, then $D(s) = 0$. In the case of Jinan City, the cutoff point is 90 sq.m., and the cost function of hukou granting is $C(D(s))$.

Once the granting rule is announced, the individual will generate some best responses. Specifically, an individual with private valuation t will choose area s and other characteristics that the purchase decision is also based on x .⁵ Market demand will be a total price function of the area and whether the hukou is granted, $P(s, D(s), x)$, where x represents a

² The blue-stamped hukou is probationary and grants the same public rights as those of local residents with a regular hukou. A probationary hukou can be upgraded to a regular hukou after three years.

³ See <https://en.wikipedia.org/wiki/Jinan>.

⁴ See the “Notice of the People's Government of Jinan Municipality on Printing and Distributing the Interim Measures for Deepening the Household Registration System Reform in Jinan City” (2005).

⁵ Note that in practice, the evaluation or purchasing decision may be based on the household level but not on the individual level.

vector of housing characteristics—that is, the individual must pay $P(s, D(s), x)$ to obtain a house with floor area s depending on whether the homeowner has the right to apply for a hukou. The individual must also pay a unit tax, a , based on his or her purchase of the house. The total for purchasing a house with floor area s will be $(1 + a)P(s, D(s), x)$. The individual has a utility function, $tD(s) + B(s) - (1 + a)P(s, D(s), x)$, where $tD(s)$ denotes his or her private valuation for acquiring the hukou and the attendant right to enjoy public services and welfare in the city; $B(s, x)$ denotes his or her consumption utility of area s and other characteristics x ; and $(1 + a)P(s, D(s), x)$ denotes his or her total payment for purchasing a house with floor area s and other characteristics x .

Thus, the best response of the individual to a given granting rule should be $s^*(t)$ and $x^*(t)$, solving $\max tD(s) + B(s) - (1 + a)P(s, D(s), x)$ for each t . However, the private valuation t is normally unobservable; in addition, both the potential buyer and the government want to identify the market demand function for facilitating their decision-making—for instance, the government must know P so that it can design an optimal hukou access policy.

Therefore, the main task of this paper is to estimate the market demand function P and, therefore, the market WTP for acquiring the hukou in terms of the coefficient of D if we focus on the estimated functional form in which P is linear with D . This reflects the market value (pricing) of the house with floor area s and hukou acquisition. Since D is an indicator function, the difference in P between two values of D can yield the market value of the hukou.

Notably, our theoretical considerations and empirical research focus on the market demand driven by actual residential demand and demand for the hukou. Since 2016, most local governments, including the Jinan City government, have strengthened control over housing purchases, which has crowded out housing demand driven by investment and speculation. In 2016, the Jinan City government issued the “Notice of the Jinan City Urban and Rural Construction Committee, the Jinan Housing Security and Housing Management Bureau, and the Jinan Municipal Bureau of Land and Resources on the Implementation of the Housing Restriction Purchase Policy,” which imposed more restrictions on housing purchases. For instance, any family without a Jinan City hukou could purchase only one house in the urban area of Jinan City. This fact strengthens the rationale for our focus on the market demand driven by actual residential demand and hukou acquisition.

3. Data, empirical strategy, and summary statistics

We use listed housing transaction data for Jinan City from June 2017 to July 2017 to implement an RDD and to estimate the value of the urban hukou in Jinan City or the marginal WTP for the hukou. These data are from the largest second-hand housing trading platform in China, Fang.com. From the platform, we collected many characteristics of every house for sale and of the communities in which the houses were located. We also obtained the rental price dataset for the same period. Over the one-month period we examined, the housing market, surrounding facilities, and relevant public policies underwent almost no change. Hence, the datasets we use can be viewed as cross-sectional in essence. The listed housing transaction dataset contains 26,031 observations after trimming to eliminate possibly spurious outliers. The rental dataset contains 11,059 observations after similar data preprocessing.

The HPM is the main regression model used in this study. The dependent variable is housing price P per sq.m., and the independent variable is X , which represents a series of house and community characteristics, including total price, the floor area, the number of rooms, orientation, the floor number, decoration, the floor area ratio, the afforestation rate, the building height, and supplementary facilities. We pay particular attention to floor area S . According to the policy of Jinan City, when the floor area is at least 90 sq.m., the buyer is eligible to acquire the local hukou. Otherwise, the buyer purchases only the residential rights to the house and not the additional rights attached to the

hukou. We present a full list of the summary statistics in Table 1 for observations near the cutoff point.

We implement the RDD using a parametric estimation method and robustness tests using a nonparametric estimation method. Area S is the running variable in the RDD. Let $S'_i = S_i - 90$. We define $D_i = \begin{cases} 1 & \text{if } S_i \geq 90 \\ 0 & \text{if } S_i < 90 \end{cases}$; that is, we take houses with an area no less than 90 sq.m. as the treatment group and houses with an area less than 90 sq.m. as the control group.

The basic model of parametric estimation is set as equation (1) below:

$$P_i = \alpha + \beta_0 \cdot D_i + g(S'_i) + \pi' \cdot X_i + \varepsilon_i \quad (1)$$

where D_i is the treatment variable and $g(S'_i)$ denotes a polynomial function of running variables S' , which may include the linear term and quadratic term of the running variable and the interaction of these terms and the treatment variable D_i . X_i is the covariate and includes all house and community characteristics except for area. The coefficient β_0 for treatment assignment represents the marginal impact of the policy at the cutoff point, which is the appreciation of housing prices caused by the hukou-granting policy and represents the estimation of the market WTP to acquire the hukou. The coefficient π' represents the implicit prices of a set of housing and community characteristics. α is the constant term, and ε_i is a random error term for observation i , which is assumed to be independently and identically distributed. The standard errors in all specifications are heteroscedasticity robust.

Specifically, we estimate four models:

$$\text{Model 1 } P_i = \alpha + \beta_0 \cdot D_i + \beta_1 \cdot S'_i + \pi' \cdot X_i + \varepsilon_i \quad (2)$$

$$\text{Model 2 } P_i = \alpha + \beta_0 \cdot D_i + \beta_1 \cdot S'_i + \beta_2 \cdot D_i \cdot S'_i + \pi' \cdot X_i + \varepsilon_i \quad (3)$$

$$\text{Model 3 } P_i = \alpha + \beta_0 \cdot D_i + \beta_1 \cdot S'_i + \beta_2 S_i^2 + \pi' \cdot X_i + \varepsilon_i \quad (4)$$

$$\text{Model 4 } P_i = \alpha + \beta_0 \cdot D_i + \beta_1 \cdot S'_i + \beta_2 S_i^2 + \beta_3 \cdot D_i \cdot S'_i + \beta_4 \cdot D_i \cdot S_i^2 + \pi' \cdot X_i + \varepsilon_i \quad (5)$$

Table 1 summarizes the statistics for two subsamples that center on the cutoff point with a bandwidth of 10, including the effective observations, means, and standard errors. The last two columns test the balance between the control group and the treatment group from two perspectives: the means and standard errors. SMD denotes the standardized mean difference between the two groups of data. The more closely this value approaches 0, the more balanced the two groups are. The variation ratio denotes the variance difference between the two groups. Again, the more closely this value approaches 0, the more balanced the two groups are.

Fig. 1 depicts the house price bins around the floor area of 90 sq.m. in the range of ± 20 sq.m., where we follow the data-driven procedure to select the number of bins.⁶ The graph indicates that there is a significant jump in the average house price on each side around the cutoff point and in the first bins closest to the cutoff point; the average house price on the right side is almost 1000 yuan higher than that on the left side.

Will the policy discontinuity point generate any abnormal distributions on either side of the discontinuity? Is there any possibility of data manipulation? It is necessary to test whether the density function of the

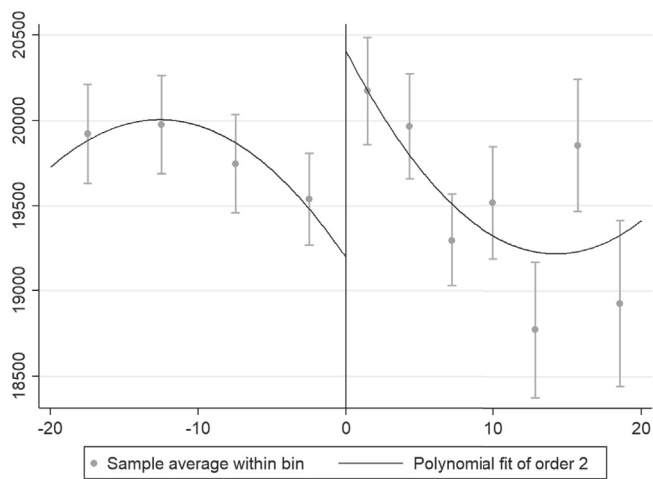
⁶ In Fig. 1, we specify the integrated mean squared error (IMSE)-optimal evenly spaced method using spacing estimators to select the number of bins. Other specifications such as the IMSE-optimal evenly spaced method using polynomial regression and the IMSE-optimal quantile-spaced method using spacing estimators or polynomial regression or arbitrarily setting 10 bins on both sides of the cutoff point, will lead to a similar quantitative relationship between the housing price and floor area.

Table 1

Summary statistics and balance test.

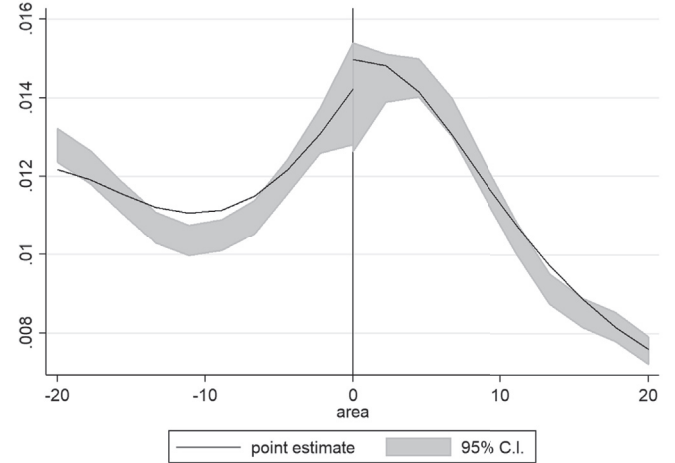
	Control group—area $\in [80, 90]$			Treatment group—area $\in [90, 100]$			SMD	Variance ratio
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.		
House price	3257	19,564	5338	3936	19,838	5093	−0.053	1.099
Area	3257	84.74	3.014	3936	94.82	3.1	−1.715	0.945
Rooms	3257	2.350	0.487	3936	2.357	0.497	−0.014	0.960
Floor	3257	2.147	0.860	3936	2.199	0.815	−0.062	1.113
Decoration	3256	2.036	1.525	3935	1.927	1.480	0.073	1.062
Building height	3257	9.13	6.817	3936	10.940	7.618	−0.247	0.801
Building age	3254	17.978	7.158	3929	15.104	7.181	0.393	0.994
Elevator	3254	1.269	0.443	3930	1.372	0.483	−0.220	0.841
Green	3253	0.288	0.117	3927	0.330	0.118	−0.351	0.983
FAR	3091	1.862	0.745	3741	1.872	0.801	−0.013	0.865
Total buildings	3214	36.51	36.12	3912	34.32	35.51	0.061	1.035
Total houses	3214	2311	2405	3912	2446	2983	−0.049	0.650

Notes: House price is the unit price per sq.m.; area represents the floor area of the house; rooms represents the number of rooms in the house; floor = 1, 2, or 3 represent lower, middle, or higher floors, respectively; decoration = 1, 2, 3, 4, or 5 represent no decoration, simple decoration, medium decoration, good decoration, or luxury decoration, respectively; elevator = 1 or 2 represents with or without an elevator in the building; FAR represents the floor area ratio.

**Fig. 1.** Regression discontinuity (RD) plot: Average housing price and floor area.

running variable (floor area) is continuous on both sides of the discontinuity. McCrary (2008) developed a test based on the nonparametric local polynomial density estimator, which requires a prebinning of the data. Recently, Cattaneo et al., 2018 developed a set of manipulation tests based on a local polynomial density estimator that does not require a prebinning of the data and is constructed based on easy-to-interpret kernel functions. Following Cattaneo et al., 2018, we implement a regression discontinuity (RD) manipulation test using local polynomial density estimation (local quadratic approximation) and plot the corresponding density function. As shown in Fig. 2, the confidence intervals on the two sides of the discontinuity will overlap greatly; thus, there is no evidence of systematic manipulation of the running variable.

Formally, we use two subsamples, $\{s_i : s_i < \bar{s}\}$ and $\{s_i : s_i \geq \bar{s}\}$, to test the continuity of density function $f(s)$ on the two sides of the discontinuity. The H_0 to be tested is set to be $\lim_{s \downarrow \bar{s}} f(s) = \lim_{s \uparrow \bar{s}} f(s)$. According to the new estimator developed by Cattaneo et al., 2018, we specify different orders of the local polynomial used to construct the density point estimators and assume the same bandwidth on both sides of the discontinuity. Table 2 reports the results under different model specifications. T tests and their corresponding p-values indicate that H_0 cannot be rejected; that is, there is no statistical evidence of systematic manipulation of the running variable.

**Fig. 2.** Manipulation test plot.

4. Empirical results of the parametric estimation

In parametric estimation, the first step is to select the functional form. In this section, we report the regression results of parametric estimations under Models 1–4. In addition, RDD requires the selection of a reasonable bandwidth. Considering the size of the sample and the balance of data between the two sides of the cutoff point, we choose 10 sq.m. as the reference bandwidth in this section; that is, we take ± 10 sq.m. around the cutoff point of 90 sq.m. of floor area. Furthermore, we choose bandwidths of 5 and 20 sq.m. to conduct robustness tests. In Section 5,

Table 2
Manipulation test.

	Bandwidth		Observations		t-Test	
	left	right	left	right	t-test	p-value
$T_1(h_1)$	3.504	3.504	1052	1485	−0.0530	0.9578
$T_2(h_2)$	11.699	11.699	3459	4194	0.6144	0.5390
$T_3(h_3)$	18.726	18.726	5513	5761	0.6153	0.5383

Notes: $T_p(h)$ is the manipulation test statistic, where h indicates the bandwidth and p indicates the order of the local polynomial used to construct the density point estimator. For density estimation, we assume an equal cumulative distribution function (c.d.f.) and higher-order derivatives for both subsamples. The optimal bandwidth selection methods are based on asymptotic mean squared error (AMSE) minimization. A triangular kernel function is used to construct the local polynomial estimator.

Table 3

Flexible parametric RD methods: Model 1.

	Model 1: linear					
	h = 5	h = 5	h = 10	h = 10	h = 20	h = 20
RD treatment effect	1097	871	717	1031	578	1176
Covariates	No	Yes	No	Yes	No	Yes
Parametric 95% CI	[457, 1737]	[290, 1452]	[246, 1189]	[599, 1464]	[232, 924]	[856, 1496]
Parametric <i>p</i> -value	0.001	0.003	0.003	0.000	0.001	0.000
$N_w^- N_w^+$	1846 2238	1726 2126	3257 3936	3055 3725	6341 6304	5948 5978

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates include house characteristics, such as the number of rooms, the floor area, decoration, whether there is an elevator, and the height and age of the building, and community characteristics, such as the FAR, the afforestation rate, the total number of buildings, and the total number of houses; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

we employ the data-driven procedures developed by [Calonico et al. \(2014\)](#) to obtain the optimal bandwidth.

[Table 3](#) reports RD parametric estimations with bandwidths of 5, 10, and 20 sq.m. under the linear model. The first row of [Table 3](#) shows the treatment effect. The second row shows whether the covariate exists. The third row shows the 95% confidence interval. The fourth row shows *p*-values that represent statistical significance. The last row indicates the number of observed values within each specified bandwidth around the discontinuity.

When the bandwidth is 5 sq.m. and there are no covariates, the RD treatment effect is 1097 yuan—that is, the implicit price of the hukou is 1097 yuan per sq.m. When the bandwidth increases to 10 sq.m., the RD treatment effect without covariates is 717 yuan per sq.m. However, after controlling the main covariates, the RD treatment effect is 1031 yuan per sq.m. When the bandwidth increases to 20 sq.m., the results are similar. According to the *p*-values, all of the RD treatment effects are statistically significant at the 1% level. Moreover, we find that the models with covariates result in narrower confidence intervals. In other words, the introduction of covariates improves the precision of the estimation. Meanwhile, the spans of the confidence intervals vary across models with different bandwidths, which implies the importance of optimal bandwidth selection. We discuss the data-driven optimal bandwidth selection in [Section 5](#).

In parametric regression, a major challenge is setting the functional form to the estimate. If the functional form is correctly specified, then the RD estimator will be unbiased in estimating the policy impact at the cutoff point; if the functional form is incorrectly specified, then the treatment effects will be estimated with bias. For instance, if the true functional form is highly nonlinear, then a simple linear model can produce misleading results. To test the robustness found in [Table 3](#), we report the respective RD estimation results for the linear interaction model, quadratic model, and quadratic interaction model. Regardless of which functional form is considered, the RD treatment effect attains both economic significance and statistical significance. For instance, with the bandwidth of 10 sq.m. and controlling all covariates in the regression equation, the treatment effect of the linear regression estimation is 1031; that of the linear interaction model is 987; that of the quadratic model is 1045; and that of the quadratic interaction model is 1418. Therefore, the RD treatment effect is basically stable in the economic sense—that is, houses with an urban hukou (no less than 90 sq.m. in floor area) cost more than 1000 yuan (per sq.m.) more than houses with less than 90 sq.m.

5. Robustness tests

Since specific functional forms will affect parametric estimation, we also adopt a nonparametric method of RDD as the robustness test. An important feature of nonparametric RD is that the bandwidth is not selected arbitrarily; instead, it is calculated based on the data properties themselves. The mean squared error (MSE)-optimal method is frequently used to calculate bandwidth ([Imbens and Kalyanaraman, 2012](#)). The

main idea of this method is that one can make a tradeoff between estimation bias and variance. Narrow bandwidth can reduce estimation bias, but the reduction in sample volume will increase estimation variance. The opposite conditions are present in the case with wide bandwidth. Let $MSE = Bias^2 + Variance$; then, the MSE-minimized bandwidth will be the optimal bandwidth. Based on the above idea, [Calonico et al. \(2014\)](#) further extend this method to include covariate adjustment, heteroscedasticity-robust or cluster-robust variance, etc.

[Table 7](#) reports the estimation results for the RD treatment effects using the local linear interaction model ($p = 1$) and local quadratic interaction model ($p = 2$) under different bandwidth selection methods.⁷ Suppose that the left and right bandwidths are identical. For the linear interaction model, the one-side optimal bandwidths are 7.704 and 4.634, which are obtained using the MSE and coverage error rate (CER) bandwidth selectors, respectively; for the quadratic interaction model, the one-side optimal bandwidths are 14.623 and 8.18, respectively, again using the two bandwidth selectors. To compare our results with previous results, we also report the results of nonparametric estimation with a bandwidth of 10 sq.m. The local linear estimation treatments effects are 1,411, 1,495, and 1409. These numbers are very close; however, the effects for the local quadratic model are 1,402, 1,259, and 925. Regarding significance, we expect that the treatment effect with a bandwidth of 10 sq.m. in the local quadratic model has relatively weak statistical significance, and the other estimation results are significant at least at the 5% level. Compared to parametric estimation, although the RD treatment effects differ slightly in magnitude, the findings are robust in terms of both economics and statistics.

We have estimated both parametric and nonparametric models to test the significance of the unit price difference that is present around a floor area of 90 sq.m. and have attributed the difference to the WTP for the hukou. However, is the treatment effect of the above estimations caused solely by the physical property of the floor area? To answer this question, we explore two strategies to run a falsification test. First, we assume that policy discontinuity occurs at either 70 sq.m. or 110 sq.m. and perform a placebo test. Second, we use rent data to test whether the policy discontinuity at 90 sq.m. has a significant treatment effect. The rationale is that tenants do not have rights equal to those of homeowners—whose rights allow access to local public services such as education and public health care—although tenants can still enjoy the residential utility of the house or apartment. If the treatment effect of RD based on rental data does not reach statistical significance or reaches statistical significance but little economic significance, we infer that the treatment effect of RD found in the housing price dataset is caused by the difference in citizenship (hukou) rather than by the physical properties of dwellings.

[Table 8](#) reports the results of a placebo test that assumes a policy cutoff point at 70 or 110 sq.m. Clearly, regardless of whether the local linear regression model or local quadratic model is used, the treatment

⁷ This section only reports regression results without covariates. When controlling for all covariates, the main findings remain unchanged.

Table 4

Flexible parametric RD methods: Model 2.

	Model 2: linear interaction					
	h = 5	h = 5	h = 10	h = 10	h = 20	h = 20
RD treatment effect	1134	879	651	987	576	1155
Covariates	No	Yes	No	Yes	No	Yes
Parametric 95% CI	[485, 1783]	[293,1466]	[177,1125]	[553,1421]	[229,923]	[834,1476]
Parametric p-value	0.001	0.003	0.007	0.000	0.001	0.000
$N_w^- N_w^+$	1846 2238	1726 2126	3257 3936	3055 3725	6341 6304	5948 5978

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates include house characteristics, such as the number of rooms, the floor area, decoration, whether there is an elevator, and the height and age of the building, and community characteristics, such as the FAR, the afforestation rate, the total number of buildings, and the total number of houses; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 5

Flexible parametric RD methods: Model 3.

	Model 3: quadratic					
	h = 5	h = 5	h = 10	h = 10	h = 20	h = 20
RD treatment effect	800	813	645	1045	528	1153
Covariates	No	Yes	No	Yes	No	Yes
Parametric 95% CI	[134, 1467]	[207,1419]	[168,1123]	[609,1481]	[183,874]	[834,1472]
Parametric p-value	0.019	0.009	0.008	0.000	0.003	0.000
$N_w^- N_w^+$	1846 2238	1726 2126	3257 3936	3055 3725	6341 6304	5948 5978

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates include house characteristics, such as the number of rooms, the floor area, decoration, whether there is an elevator, and the height and age of the building, and community characteristics, such as the FAR, the afforestation rate, the total number of buildings, and the total number of houses; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 6

Flexible parametric RD methods: Model 4.

	Model 4: quadratic interaction					
	h = 5	h = 5	h = 10	h = 10	h = 20	h = 20
RD treatment effect	1705	1061	1965	1418	1086	1270
Covariates	No	Yes	No	Yes	No	Yes
Parametric 95% CI	[648, 2763]	[119,2004]	[1248, 2683]	[771,2066]	[571, 1601]	[799,1741]
Parametric p-value	0.002	0.027	0.000	0.000	0.000	0.000
$N_w^- N_w^+$	1846 2238	1726 2126	3257 3936	3055 3725	6341 6304	5948 5978

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates include house characteristics, such as the number of rooms, the floor area, decoration, whether there is an elevator, and the height and age of the building, and community characteristics, such as the FAR, the afforestation rate, the total number of buildings, and the total number of houses; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 7

Robust bias-corrected local polynomial methods.

	Linear Model (p = 1)			Quadratic Model (p = 2)		
	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$
RD treatment effect	1411	1495	1409	1402	1259	925
Robust 95% CI	[718,2104]	[681,2309]	[626,2191]	[688,2118]	[377,2140]	[-176,2025]
Robust p-value	0.000	0.000	0.000	0.000	0.005	0.100
$N_w^- N_w^+$	2345 2958	1431 1807	2856 3599	4250 4801	2594 3376	2856 3599
h	7.704	4.634	10	14.623	8.18	10

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p-values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) $\hat{h}_{MSE}$ corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al. \(2017\)](#); $\hat{h}_{CER}$ corresponds to the coverage error rate (CER) criterion used to calculate the optimal bandwidth; $\hat{h}_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

effects are statistically nonsignificant. This result primarily indicates that the jump in housing prices at 90 sq.m. is caused by the difference in citizenship and not by the physical properties of dwellings.

Furthermore, we still set the cutoff point at 90 sq.m., but we use the rental data to perform falsification tests. [Table 9](#) reports the RDD results under the local linear model and local quadratic model. The RD

treatment effect does not reach statistical significance; however, the RD treatment effect under the local linear model with the optimal bandwidth of 9.378 sq.m. is statistically significant at the 5% level. This result indicates that, once again, the discontinuity at 90 sq.m. is caused by the difference in citizenship or hukou and not by the physical properties of dwellings.

Table 8

Placebo test: Assuming false policy discontinuity.

False discontinuity	Linear Model (p = 1)		Quadratic Model (p = 2)	
	Area = 70 sq.m.	Area = 110 sq.m.	Area = 70 sq.m.	Area = 110 sq.m.
RD treatment effect	879	445	1202	520
Robust 95% CI	[-824,2582]	[-175,1065]	[-489,2893]	[-300,1339]
Robust p-value	0.312	0.160	0.163	0.214
$N_w^- N_w^+$	684 1148	4134 2724	1686 2098	4784 2951
h	2.921	15.838	5.614	17.956

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p-values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) h_{MSE} corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al., 2017](#); h_{CER} corresponds to the coverage error rate (CER) criterion used to calculate the optimal bandwidth; $h_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^- = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^+ = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Finally, the Jinan City government implemented the reform in its hukou-granting policy in August 2017, after which the “acquiring the hukou by purchasing a house” regulation was abandoned. Therefore, we can predict that the policy discontinuity will no longer be present at 90 sq.m. We use the data for house bidding prices on [Fang.com](#) to further test whether this policy effect still exists at 90 sq.m.⁸ [Tables 10 and 11](#) show that this treatment effect was negative or not statistically significant generally after the policy for acquiring hukou by purchasing a house was abandoned.

6. Extensions

The migrant population is a demander of hukous; therefore, it would be more reasonable for us to use submarkets of the migrant community to estimate migrants' WTP for a local hukou. We use street-level population data from the most recent census (the 2010 population census of the People's Republic of China) to categorize streets into two submarkets: (1) a submarket dominated by migrants, which means that in that street, the percentage of the nonmigrant population is less than 50% of the residential population, and (2) a submarket dominated by nonmigrants, which means that the proportion of the nonmigrant population for the street is higher than 50%. The nonmigrant population is defined as residents who already have the local hukou. The geographical distribution of immigrants can be seen in [Fig. A1](#).

We perform an RD analysis (linear and quadratic model) to estimate the hukou value in a submarket dominated by migrants, and the results are reported in [Table 12](#). The empirical results show that the estimated value of the hukou under various specifications is statistically significant at the 1% level, except when the bandwidth is 5 sq.m. with the quadratic model setting. Under this setting, the RD treatment effect is not statistically significant. Furthermore, the RD treatment effect reported in [Table 12](#) is significantly higher than the estimated result reported in [Tables 3–6](#).

For comparison, [Table 13](#) reports the RD results of the market dominated by the nonmigrant population. We find that the estimated hukou value is generally low and not statistically significant under most model specifications, which provides a stark contrast to the findings in [Table 12](#). This result suggests that for Jinan City as a whole, in areas where the migrant population percentage is higher than the nonmigrant

population percentage, the hukou value is more easily reflected by the price of a house with a floor area of 90 sq.m., the cutoff point for this hukou policy. More specific, taking an extreme case as an example, i.e., the demanders in the housing market all are migrants, the housing price difference between houses with 89 and houses with 91 sq.m. is not only the housing price for 2 sq.m. but including the price for the right to apply for hukous. While for areas where nonmigrants dominate the housing submarket, the housing price mechanism is less influenced by hukou policies. When the demanders in the housing market all are nonmigrants, the housing price difference between houses with 89 sq.m. and houses with 91 sq.m. might reflect the housing price for two sq.m. It is because, in this extreme case, demanders in the housing market are nonmigrants who already have local hukous and do not demand local hukous; therefore, the right to apply for hukous might be not valuable for homebuyers in this submarket.

Another important consideration is that migrants are granted the right to apply for a local hukou by purchasing a house with a floor area slightly larger than 90 sq.m. However, for nonmigrants who already have the local hukou, it appears that there is no sufficient incentive for them to purchase this kind of house; otherwise, these hukou holders would pay extra money on the right to apply for hukou that seems useless for them.

In fact, nonmigrant residents who purchase a house with a floor area slightly larger than 90 sq.m. also have the right to apply for a local hukou, although they might not exercise this right. This right will be capitalized into the house price. When a homeowner sells his or her property, the right to apply for the hukou is also bundled and sells at the same time. Therefore, nonmigrant residents, especially those with strong financial ability, are also motivated to purchase a home with a floor area slightly larger than 90 sq.m. To further support our argument, we choose high-quality school districts with active housing transactions as research objects. Specifically, we select the housing in Jinan City's socially recognized top ten primary school districts (shown in [Fig. A2](#)) as a sample.⁹

[Table 14](#) shows that for high-quality top school districts, regardless of whether the estimation involves a linear model or a quadratic model, the RD treatment effects are statistically significant; furthermore, the estimated hukou value is higher than that in the previous benchmark model. In contrast, [Table 15](#) reports the empirical results from the RD based on the sample within ordinary school districts, which shows that the estimated hukou value is lower and that the statistical significance is not very robust. This result indicates that in a housing market in a high-quality school district with active housing trading, the capitalization of hukou rights is more prominent. Therefore, nonmigrant residents also have the underlying motivation to purchase a house with a floor area slightly larger than 90 sq.m.

Here, we also employ a nonparametric method for RD as a robustness test of the heterogeneity analysis. Data-driven optimal bandwidths are calculated using the MSE and CER bandwidth selectors. We also report the results of nonparametric estimation with a bandwidth of 10 sq.m. to facilitate comparison with previous results. To save space, we report only the estimations of the local linear interaction models and assume that the bandwidths on both sides are identical. [Table 16](#) shows findings that are consistent with those of [Tables 12 and 14](#), and the value of the hukou is more than 2000 yuan per sq.m. in both the immigrant-dominated housing markets and the top primary school districts. [Table 17](#) presents results that are similar to those of [Tables 13 and 15](#), where the estimated

⁸ The data were preprocessed by deleting duplicate observations and abnormal values. We truncated price and area variables with 5% tails. Ultimately, we have 31,334 effective observations. Our general findings are not affected by whether covariates are present.

⁹ The top ten primary schools are as follows: Shengli Street Primary School, Jingwu Road Primary School, Dianliu No.1 Primary School, Shandong Normal University-affiliated Primary School, Lixia Experimental Primary School, Shandong Experimental Primary School, Jingshiyi Road Primary School, Oriental Bilingual Experimental School, Jiefang Road No.1 Primary School, and Jiefang Road No.2 Primary School. Shandong Normal University-affiliated Primary School does not have a specific school district, but we use the remaining nine school districts in our analysis. We attach the map of these school districts in the [Appendix A](#).

Table 9

Placebo tests on rents.

	Linear Model (p = 1)			Quadratic Model (p = 2)		
	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$
Treatment effect	1.564	1.233	1.030	0.996	1.179	-0.096
Robust 95% CI	[0.14, 3.00]	[-0.58, 3.05]	[-0.89, 2.95]	[-0.51, 2.50]	[-0.92, 3.28]	[-3.22, 3.03]
Robust p-value	0.032	0.183	0.293	0.194	0.271	0.952
$N_w^- N_w^+$	883 2195	692 1448	887 2206	1714 2983	1230 2531	887 2206
h	9.378	5.887	10	17.068	10.026	10

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p-values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) h_{MSE} corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al., 2017](#); h_{CER} corresponds to the coverage error rate (CER) criterion used to calculate the optimal bandwidth; $h_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 10

Flexible parametric RD method: After the “acquiring the hukou by purchasing a house” policy was abolished.

	Linear Model			Quadratic Model		
	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
treatment effect	-453.6	-73.423	269	-554	-574	-176
95% CI	[-984, 77]	[-450, 303]	[-0.258, 538]	[-1548, 439]	[-1165, 17]	[-583, 232]
p-value	0.094	0.702	0.050	0.274	0.057	0.398
$N_w^- N_w^+$	2093 2685	3753 4816	7129 7921	2093 2685	3753 4816	7129 7921

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates include the number of rooms, age of the house, towards of the house, floor and total floors of the building; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 11

Robust bias-corrected local polynomial method: After the “acquiring the hukou by purchasing a house” policy was abolished.

	Linear Model (p = 1)			Quadratic Model (p = 2)		
	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$
Treatment effect	-583.8	-624	-666	-601	-725	-781
Robust 95% CI	[-1170, 2.5]	[-1335, 87]	[-1352, 21]	[-1226, 24]	[-1554, 105]	[-1882, 319]
Robust p-value	0.051	0.086	0.057	0.06	0.087	0.164
$N_w^- N_w^+$	2731 3575	1598 2150	3307 4351	4595 5676	2731 3575	3307 4351
h	7.58	4.518	10	13.89	7.689	10

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p-values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) h_{MSE} corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al., 2017](#); h_{CER} corresponds to the coverage error rate (CER) criterion used to calculate the optimal bandwidth; $h_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 12

Flexible parametric RD method: (migrants > 50%).

	Linear Model			Quadratic Model		
	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
treatment effect	2801	2875	2306	681	2757	2936
95% CI	[1753, 3849]	[2045, 3705]	[1683, 2929]	[-781, 2144]	[1624, 3890]	[2059, 3814]
p-value	0.000	0.000	0.000	0.361	0.000	0.000
$N_w^- N_w^+$	385 502	631 857	1151 1395	385 502	631 857	1151 1395

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates are not included in the above models, but the basic findings will not be changed when they are included; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 13

Flexible parametric RD method: (migrants ≤ 50%).

	Linear Model			Quadratic Model		
	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
treatment effect	467	99	97	1030	1247	642
95% CI	[-285, 1219]	[-444, 642]	[-297, 492]	[-221, 2281]	[414, 2080]	[53, 1231]
p-value	0.223	0.722	0.629	0.107	0.003	0.033
$N_w^- N_w^+$	1461 1736	2626 3079	5190 4909	1461 1736	2626 3079	5190 4909

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates are not included in the above models, but the basic findings will not be changed when they are included; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 14

Flexible parametric RD method: Top school districts.

	Linear Model			Quadratic Model		
	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
treatment effect	2283	1472	1341	2480	2778	1460
95% CI	[1061,3506]	[586,2358]	[728,1954]	[464,4497]	[1430,4125]	[537,2384]
p-value	0.000	0.000	0.000	0.016	0.000	0.002
$N_w^- N_w^+$	288 360	529 547	1197 911	288 360	529 547	1197 911

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates are not included in the above models, but the basic findings will not be changed when they are included; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 15

Flexible parametric RD method: Ordinary school districts.

	Linear Model			Quadratic Model		
	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
treatment effect	690	266	401	745	1292	676
95% CI	[55,1326]	[-196,727]	[63,739]	[-320,1809]	[583,2001]	[170,1182]
p-value	0.033	0.259	0.02	0.170	0.000	0.009
$N_w^- N_w^+$	1558 1878	2728 3389	5144 5393	1558 1878	2728 3389	5144 5393

Notes: (i) All estimates are constructed using linear ordinary least squares estimators with heteroscedasticity-robust standard errors; (ii) the covariates are not included in the above models, but the basic findings will not be changed when they are included; and (iii) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

hukou value is much lower and, in some cases, statistically insignificant.

7. Conclusion

It is well known that the Chinese urban hukou system not only performs population registration functions but also has economic value. To the best of our knowledge, this paper is the first to study the market value of urban hukou using the case of acquiring a hukou by purchasing a house. Using the urban hukou access policy of Jinan City as the context of our study, we employ an RDD and present a way to disentangle the underlying value of the hukou through estimation based on house prices. We show that in 2017, urban hukou rights in Jinan City result in an increase in residential unit prices of approximately 1000–1400 yuan and that the market WTP for the urban hukou of Jinan City will be between approximately 90,000 and 126,000 yuan. Meanwhile, we find great heterogeneity in different housing submarkets and that the value of the hukou is much higher in migrant-dominated housing markets and top primary school districts.

This study also has policy implications for further analysis of public policies related to hukou reform in China and public goods provision. First, this paper complements previous studies of citizenization, which adopt a supply (cost) perspective. In contrast, this paper uses a demand perspective regarding relevant practices. Furthermore, based on this analysis, we are able to consider a local government's fiscal capacity and tax system in designing optimal provision policies for public services and welfare, as well as equality of public rights.

Our study has several potential limitations. First, the hukou is an instrument that prevents “floating” individuals from enjoying some aspects of social welfare because hukou certificates are required for these kinds of services. Although other public goods or services—such as urban infrastructures—are also vital to the welfare of citizens, local governments cannot exclude people from consuming them by using a hukou requirement, regardless of whether they are renters or homeowners, local or foreign. Therefore, the value of the hukou actually stems from a set of exclusive public services or welfare. While the right to enjoy nonexclusive public services or welfare may still be capitalized into housing prices, the treatment effect of our RD analysis does not address this aspect.

Second, acquiring hukous is now more complicated due to the many factors that the local government considers—e.g., local fiscal reforms, China's national goals, or real estate market reform. Chinese cities are reforming their policies regarding the acquisition of a hukou by purchasing a house, and many local governments have also adopted a scoring system to comprehensively rate hukou applications. Although hukou acquisition channels are being reformed, the hukou persists as a screening mechanism for selecting desirable applicants. Therefore, the economic value of the hukou still exists but perhaps in a more complicated way; researchers may find other experimental opportunities to estimate its value in the future.

Lastly, further empirical work would be desirable to directly estimate the optimal provision of public services and welfare in the design of hukou-granting rules. Based on this study and previous studies that examine the supply (cost) side of hukou granting, it is now possible to test or estimate the entire process of the optimal design of hukou-granting

Table 16Robust bias-corrected local polynomial methods: Local linear model ($p = 1$).

	Migrants > 50%			Top School Districts		
	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$
RD treatment effect	2932	2873	1706	2101	2441	2630
Robust 95% CI	[2071,3794]	[1939,3809]	[575,2837]	[985, 3217]	[1159,3723]	[1117,4144]
Robust p-value	0.000	0.000	0.003	0.000	0.000	0.001
$N_w^- N_w^+$	759 1027	539 806	542 817	533 550	367 429	465 498
h	13.7	9	10	10.8	7.1	10

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p-values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) $\hat{h}_{MSE}$ corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al. \(2017\)](#); $\hat{h}_{CER}$ corresponds to the coverage error rate (CER) criterion used to calculate the optimal bandwidth; $\hat{h}_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

Table 17Robust bias-corrected local polynomial methods: Local linear model ($p = 1$).

	Migrants $\leq 50\%$			Ordinary School Districts		
	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$	$h = \hat{h}_{MSE}$	$h = \hat{h}_{CER}$	$h = \hat{h}_{FP}$
RD treatment effect	724	790	689	706	794	748
Robust 95% CI	[-35,1483]	[-61,1641]	[-213,1,591]	[173, 1239]	[167,1420]	[-39,1536]
Robust p -value	0.061	0.069	0.134	0.009	0.013	0.063
$N_w^- N_w^+$	2268 2755	1478 1760	2314 2782	3149 3812	1964 2517	2391 3101
h	9.1	5.5	10	12.37	7.5	10

Notes: (i) The point estimators are constructed using local polynomial estimators with a triangular kernel function; (ii) “robust p -values” are constructed using bias correction with robust standard errors, as derived from [Calonico et al. \(2014\)](#); (iii) h_{MSE} corresponds to the second-generation data-driven MSE-optimal bandwidth selector proposed by [Calonico et al. \(2014\)](#) and [Calonico et al. \(2017\)](#); h_{CER} corresponds to the coverage error rate (CER) criterion used to calculate optimal bandwidth; $h_{FP} = 10$ to facilitate comparisons with previous parametric estimations; and (iv) $N_w^+ = \sum_1^n 1(\bar{s} \leq S_i < \bar{s} + h)$, $N_w^- = \sum_1^n 1(\bar{s} - h \leq S_i < \bar{s})$.

rules.

Acknowledgement

We would like to acknowledge helpful comments from Mengwei Chen, Jie Chen, Qiang Chen, Hao Li, Jie Mao, Ming Li, Ziyang Fan, Wenlin Fu, Hongsheng Fang, Yongyou Li, Marko Köthenbürger, Cuicui Lu, Stefan Borsky, Michael Alexeev, participants in the Public Economics Workshop at Zhejiang University and Qi Lu Public Economics Workshop at Shandong University. This paper also benefited from helpful comments from

the anonymous referees and Andrew Foster. All errors are our own. This work was supported by the National Natural Science Foundation of China (grant number 71573160, 71973080), National Social Science Foundation of China (grant number 16BJY151, 16ZDA080), Social Science Foundation of Shandong Province (grant number 15CJJJ05), and Shandong University Youth Team Project (grant number IFYT17034). The first author, Yu Chen, deceased unfortunately during the revision of this paper. We'd like to dedicate this article to him and may he rest in peace.

Appendix A

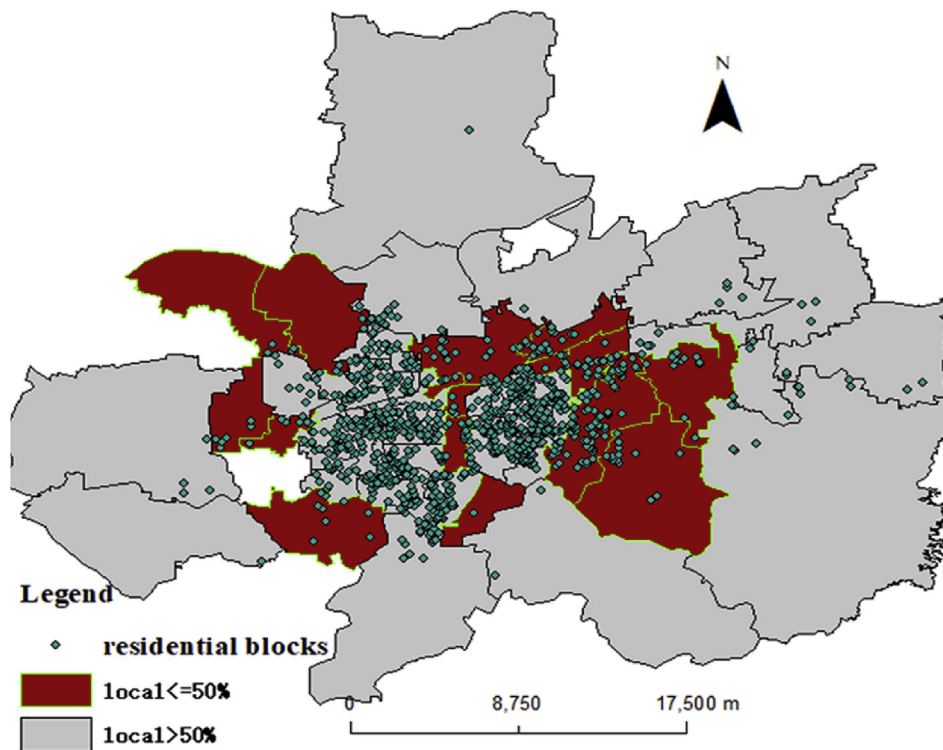


Fig. A1.

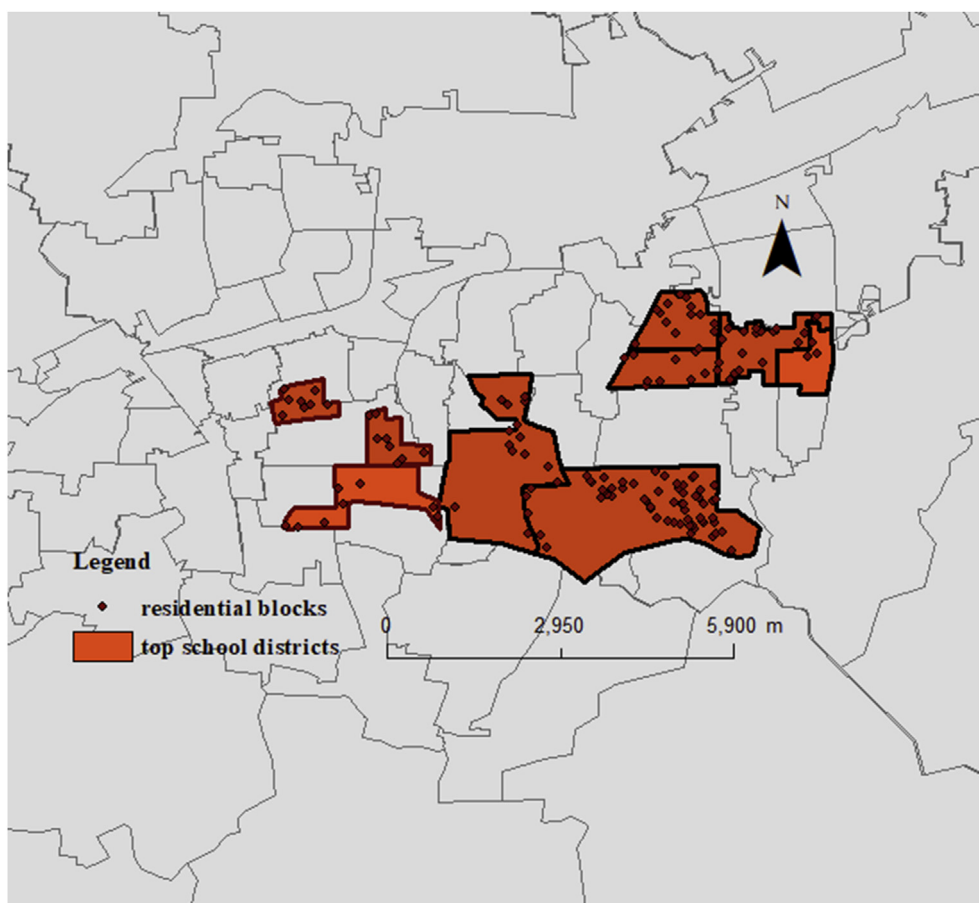


Fig. A2. 2

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jdeveco.2019.102381>.

References

- Andreyeva, E., Patrick, C., 2017. Paying for priority in school choice: capitalization effects of charter school admission zones. *J. Urban Econ.* 100, 19–32.
- Black, S.E., 1999. Do better schools matter? Parental valuation of elementary education. *Q. J. Econ.* 114 (2), 577–599.
- Bui, L.T., Mayer, C.J., 2003. Regulation and capitalization of environmental amenities: evidence from the toxic release inventory in Massachusetts. *Rev. Econ. Stat.* 85 (3), 693–708.
- Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2017. rdrobust: Software for regression discontinuity designs. *STATA J.* 17 (2), 372–404.
- Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82 (6), 2295–2326.
- Cattaneo, M.D., Jansson, M., Ma, X., 2018. Manipulation testing based on density discontinuity. *STATA J.* 18 (1), 234–261.
- Chay, K.Y., Greenstone, M., 2005. Does air quality matter? Evidence from the housing market. *J. Political Econ.* 113 (2), 376–424.
- Cheng, T., Selden, M., 1994. The origins and social consequences of China's hukou system. *China Q.* 139, 644–668.
- China Development Research Foundation, 2010. China Development Report 2010: China's New Urbanization Strategy to Promote Human Development. People's Publishing House, Beijing.
- Ding, M., Xu, D., 2014. On the cost estimation of migrant workers' citizenization. *Econ. Perspect.* 2, 36–43.
- Gayer, T., Hamilton, J.T., Viscusi, W.K., 2000. Private values of risk tradeoffs at superfund sites: housing market evidence on learning about risk. *Rev. Econ. Stat.* 82 (3), 439–451.
- Gibbons, S., Machin, S., Silva, O., 2013. Valuing school quality using boundary discontinuities. *J. Urban Econ.* 75, 15–28.
- Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Rev. Econ. Stud.* 79 (3), 933–959.
- Lin, J.Y., Cai, F., Li, Z., 1994. The China Miracle: Development Strategy and Economic Reform. SDX Joint Publishing Company, Shanghai.
- Linden, L., Rockoff, J.E., 2008. Estimates of the impact of crime risk on property values from Megan's laws. *Am. Econ. Rev.* 98 (3), 1103–1127.
- Lu, Y., 2002. Chinese household registration system after 1949: structure and changes. *J. Peking Univ.* 2, 123–130 (Philosophy and Social Sciences Edition).
- McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: a density test. *J. Econom.* 142 (2), 698–714.
- Pope, J.C., 2008a. Buyer information and the hedonic: the impact of a seller disclosure on the implicit price for airport noise. *J. Urban Econ.* 63 (2), 498–516.
- Pope, J.C., 2008b. Fear of crime and housing prices: household reactions to sex offender registries. *J. Urban Econ.* 64 (3), 601–614.
- Qu, X., Cheng, J., 2013. Relevance of regional disparities, urbanization promotion and household registration reform costs. *Reform* 3, 37–44.
- Tao, R., Shi, C., Wang, H., Zhuang, G., 2011. Lewis Turning Point Paradox and China's household registration-land-tax system reform. *Int. Econ. Rev.* 5, 120–149 (in Chinese).
- The Research Group of the Development Research Center of the State Council, 2011b. Migrant Workers' Citizenship: Institutional Innovation and Top-Level Policy Design. China Development Press.
- The Research Group of the Development Research Center of the State Council, 2011a. The general situation and strategic orientation of the process of citizenization of rural migrant workers. *Reform* 5, 5–29.
- The Research Group of the Ministry of Construction, 2006. China's Migrant Workers Research Report. China Yan Shi Press, Beijing.
- Wu, K.Y., Zhang, L., 2010. Developmental government and urban settlement threshold: reflections on the reform of the household registration system. *Sociol. Res.* 6, 58–85.
- Zhang, L., Wu, K.Y., 2013. An analysis of the local public finance pressure on the city's free settlement. *Chin. Popul. Sci.* 6, 17–26+126.