ESSAYS ON DEVELOPMENT TOPICS ON CHINA

A Dissertation

by

RUICHAO SI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Li Gan
Committee Members,	Yonghong An
	Jason Lindo
	Ximing Wu
Head of Department,	Timothy Gronberg

May 2019

Major Subject: Economics

ABSTRACT

China has experienced fast development in the past decades. This dissertation studies several topics related to the recent development issues in rural and urban China. Section 2 studies China's urbanization since the 2010s. In Section 3, I study the factor utilization in the agricultural sector in rural China while Section 4 studies the education investment and human capital accumulation for Chinese high school students.

First, urbanization is a key path for countries to achievement development. China's rapid economic growth over the past 40 years has been accompanied by an increasingly rapid rate of urbanization, from less than 20% in the late 1970s to almost 60% in the 2010s. In addition to natural population growth, rural-urban migration is generally believed to be a dominant driving force of the urbanization process. In Section 2, however, I demonstrate that a large share of urban population growth comes from community reclassification. I find that these in situ migrants (from communities which were reclassified from rural to urban) accounted for almost 35% of total urban population growth in the first half of the 2010s. Households in reclassified communities share similar characteristics with rural village communities, particularly in their ownership of housing. Furthermore, I provide evidence that the scale of in situ migrants is significantly related to residential land supply at the prefecture level.

Moreover, the development of agricultural sector is also of great significance for a highly populous developing country like China. Secure property rights are key determinants of economic development. In developing countries where agricultural activities still play a major role, property rights in rural land become critical. In Section 3, I evaluate the impact of land certification program in rural China. Employing a panel dataset of over 1,000 villages all across the country and exploiting the variations in years of certification, I find that less agricultural land is left idle in villages where the land has been certified. This result is robust to various checks including using a more comparable control group from the propensity score matching method. Such effects are more significant in high-income villages or those from provinces in which the economy highly

depends on the agricultural sector. Additionally, I document evidence that the effect of land certification program on improving land utilization might be a result that households with land being certified are more likely to transfer their cultivated land to new agricultural business entities that might possibly increase land utilization.

Finally, the development of human capital and investment in education play a fundamental role for long-run development. Household investment in children's education is among one of the largest shares of inter-generational investments. In Section 4, I study whether changes in opportunity costs of re-taking a high-stake exam caused by an exogenous policy result in different levels of household investment in children's education. Specifically, I employ the curriculum reform of senior high school in mainland China in 2000s and demonstrate that students in the last cohort using old curriculum are faced with a more difficult situation in terms of re-taking College Entrance Examination since their re-taking costs are much higher due to the curriculum switch. Using province-level administrative data, I confirm the pattern that the adoption of the new curriculum significantly reduces the number of re-takers at its introduction year. I then study the effect of this changing opportunity cost on household investments in senior high school students. Empirical results indicate that households with students facing higher re-taking opportunity costs are more likely to have equipments and inputs that help improve their learning efficiency. However, I do not find any long-term effect after students are enrolled into college.

DEDICATION

To my mother, my father, who have been a great source of inspiration and support, and all my friends without whom none of the work can be achieved.

ACKNOWLEDGMENTS

I would like to thank my advisor, all members in my committee, as well as all the faculty members in the Department of Economics in Texas A&M University for their guidance in my progress for the degree. I am also grateful to my classmates and all the Ph.D. students in the Department of Economics in Texas A&M University who have been offering great support. I would also thank all members in the Survey and Research Center of China Household Finance for conducting the survey and collecting the data.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Li Gan (Committee Chair) and Professors Yonghong An and Jason Lindo of the Department of Economics and Professor Ximing Wu of the Department of Agricultural Economics from Texas A&M University.

The main data analyzed for Section 2 was from the China Household Finance Survey. The analyses depicted in this section were conducted in part by Dr. Li Gan of the Department of Economics from Texas A&M University and by Dr. Qing He and Dr. Daichun Yi from the Survey and Research Center of China Household Finance.

The data analyzed for Section 3 was also from the China Household Finance Survey. The analyses depicted in this section were conducted in part by Dr. Qing He from the Survey and Research Center of China Household Finance.

The data analyzed for Section 4 was from the Beijing College Student Panel Survey conducted by National Survey Research Center of Renmin University of China.

Funding Sources

Graduate study was supported by an assistantship from the Department of Economics of Texas A&M University.

NOMENCLATURE

BCSPS	Beijing College Student Panel Survey
CEE	College Entrance Examination
CHFS	China Household Finance Survey
HRS	Household Responsibility System
NBS	National Bureau of Statistics of China

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

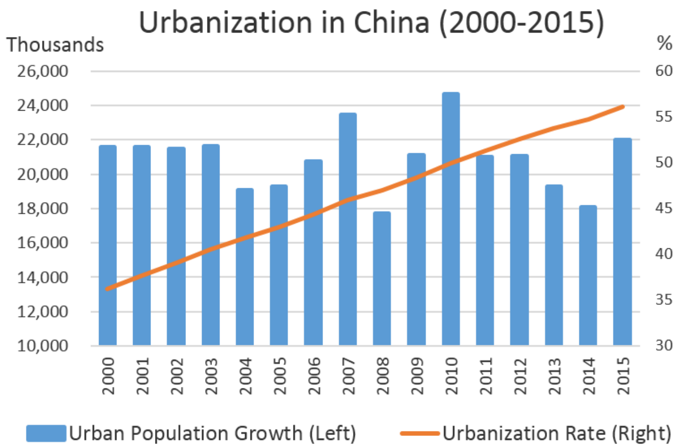
1. INTRODUCTION

1.1 Development in China's Urbanization Process

Urbanization has played a key role in the global development in the past century, and also, in the current century. According to the World Urbanization Prospects from the United Nations, the urban population of the world has grown rapidly from 751 million in 1950 to 4.2 billion in 2018, accounting for 55% of the world population. This ratio is expected to increase to 68% in 2050. Yet, the urbanization of the current century is different, most obviously because of the accelerated growth of urban population and cities in developing countries ([23]). China, as the most populous country, has experienced remarkable economic achievements since the reform in the 1970s, a transformation from a agricultural dominated planned economy to an export, consumer-good-oriented economy with a highly developed industry sector and an emerging and fast growing service sector ([25]). As most manufacturing and service production is most efficient in urban areas such as industrial parks ([59]) and special economic zones ([1]) that benefit from agglomeration, the urbanization process in China has also experienced rapid growth since the reform ([19]). The urbanization rate increased from less than 20% in the late 1970s to almost 60% in the 2010s. This is the largest and perhaps one of the fastest urbanization processes in human history ([12]). Figure 1.1 shows that only in the first 15 years of this new millennium, China's urban population increased from 459 million to 771 million.

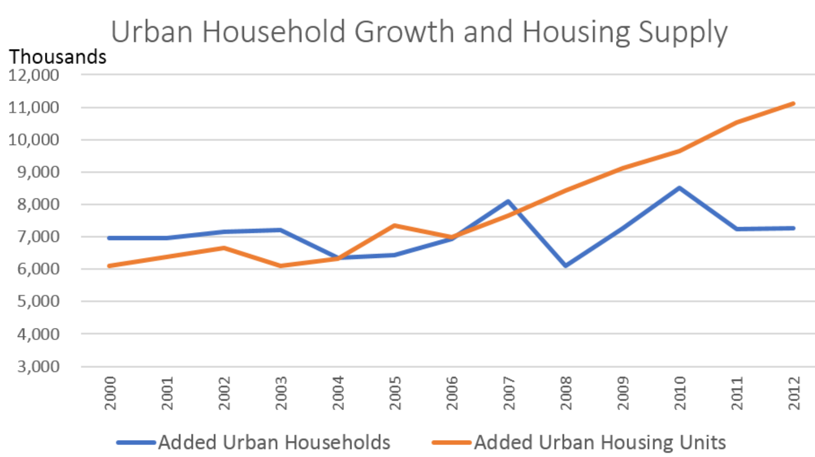
Accompanied by this large wave of urbanization is the booming in the urban housing market. Before 1990s, households in urban areas who worked for the state-owned enterprises were assigned housing units by their work units with low rents. In 1993, approximately 40 percent of urban households in China were residing in state-owned housing ([46]). In a housing reform in 1994, state-owned housing units were no longer provided and private housing market quickly developed. For example, in 2005, the quantiy of newly completed residential construction reached 28.4 million square meteres in Beijing, which accounted for 13.1% of the city's existing housing stock ([58]).

Figure 1.1: Urban Population Growth and the Rate of Urbanization in China



At the national level, Figure 1.2 graphically illustrates the increments of the housing construction in China's urban areas[1]. Between 2000 and 2012, accounting for the demolished housing units, the added housing units in urban areas was about 80.96 million. During the same period, the number of added urban households was around 92.46 million.

Figure 1.2: Urban Household Growth and Housing Supply



[1]Data about the added housing unit size in urban areas is from the 2014 China Statistical Yearbook. The last year in which this variable is available is 2012. There is only information on the total size of newly built residential houses on a yearly basis from 1985 to 2012. I assume the average size per housing unit is 90 square meters. Data about the annual demolished urban housing units are estimated based on the 2013 wave of China Household Finance Survey.

In contrast, according to a recent study by China Household Finance Survey (CHFS) in 2015, there are approximately 50 million vacant housing units located in urban areas all over China. These vacant units consist of houses and apartments which are not occupied by their owners for a variety of reasons, such as temporary migration or that the homeowners own multiple properties that are not occupied by themselves or anyone else. This number is high not only in absolute values but also in percentages; these units accounts for more than 22% of all urban housing units. Table 1.1 shows the housing vacancy rate in other economies. Though there are some discrepancies in accounting for the vacant housing units, urban China's housing vacancy rate still ranks the highest among all listed countries or regions.

Table 1.1: International Comparisons for the Housing Vacancy Rate

	Year	Vacancy Rate	Source
Australia	2017	2.5%[a]	SQM Research
Brazil	2012	11.3%	Ministério das Cidades
China (Urban)	2014	22.4%	China Household Finance Survey
Germany	2011	4.5%[b]	Statistisches Bundesamt
Hong Kong	2013	4.1%[c]	Hong Kong Property Review
Japan	2013	13.5%	Statistic Bureau of Japan
Mexico	2010	14.2%[d]	INEGI 2010 Population and Housing Census
Singapore	2014	7.8%[e]	Yearbook of Statistics Singapore
United States	2014	13.4%[f]	United States Census Bureau

[a] This number refers to the rental vacancy rate, conducted monthly by a private Australian research company.
[b] According to the micro census, a vacant property is defined as a housing unit in which no interview partner was found even after multiple visits.
[c] This number refers to Vacant housing units at the end of the year as a percentage of total housing stock.
[d] A vacant home is defined as dwelling that is offered for sale or rent, rented or sold awaiting occupancy, or held off market for other reasons. This excludes housing for temporal use.
[e] This number refers to the percentage of the existing stock that is vacant. Vacant Houses are defined as vacant units/space that are/is not physically occupied.
[f] This number refers to the share of vacant housing units as of all housing units. Vacant housing units include both year-round vacant units and seasonal vacant units.

This is the first trial to use large-sampled household survey data to estimate the number of vacant housing units nationwide. After the result was released, it raised a nationwide debate on whether the estimate of 50 million vacant housing units is reasonable given that there would be an approximate equality between the added urban housing units and the added urban households if I regard them as the supply and demand sides.

Motivated by these two seemingly contradictory facts, in Section 2.1.2, I aim to address this question from a new perspective on Chinese urbanization. That is, I solve this high-vancacy-rate puzzle by fully understanding composition of those newly added members of the urban population while providing a new statistical framework for China's urbanization.

Besides natural growth from birth and death, urban population growth is a dynamic process that consists of two main sources: migration from rural to urban areas (relocating migrants) and urbanization by expansion of urban areas (in situ migrants) ([44]; [7]). However, rural-urban migrants have been generally believed to be a dominant driving force of the overall urbanization process. For example, [55] compute the amount of annual rural-urban migrants by breaking down urban growth into natural growth and rural-urban migration; they find that the latter made dominant contributions to urban population growth from 1978 to 1999. However, I find that quite a large share of newly added urban residents in the first half of the 2010s resulted from the National Bureau of Statistics' (NBS) reclassifying communities, that is, changing their designations from rural to urban. By encoding all 700,000 communities on the NBS website from 2009 to 2017 and using their respective rural-urban division codes to identify rural communities that were reclassified as urban, I estimate that in situ migrants accounted for nearly 35% of all urban population growth during this period.

This work contributes to the vast literature on urbanization in China, both spatially and in terms of population. Spatially, [19] employ a unique three-period panel data set of high-resolution satellite imagery data to study the extent of and the factors driving urban expansion in China from the late 1980s to 2000. They find that the growth of income and population in a city play powerful roles in China's urban exansion. [37] theoretically demonstrate how fiscal and governance reforms give rise to land conversion decisions and long run urban spatial sizes. These papers mainly talk about how urban area expands due to economic activities, for example, land converted from agricultural use into industrial or residential uses. My work documents and measures the spatial expansion of urban areas from a statistical perspective on which urbanization rate is based. In terms of population, I complement the work of [44] and [7] by providing the estimated composition of urban population growth in the 2010s. [44] uses census data from 1990 and 2000 and estimate different sources of urban population growth in 1990s while [7] use their methods to account for China's urbanization in 2000s[2]. However, both of these two papers are only able to measure at the national level. Different from theirs, I could measure different sources at a more disaggregated level on a yearly basis.

Furthermore, I also employ the survey data set to document the characteristics of those urbanized population due to urban expansions. Employing survey data collected by CHFS at both the household and community levels, I empirically show that these communities, though statistically reclassified as urban, retain their basic rural characteristics, and residents in these communities continue to share similar living conditions with rural villagers. Moreover they have limited affordability to participate in local housing markets. I thus show that urbanized population should not be treated equally since a large share of urban population growth comes from the urban land expansion who are not living in a truly urbanized environment. As a result, previous studies on determinants or consequences of urbanization in China might be misleading. For example, [34] study how housing prices increase in Chinese cities might have impact on household saving behavior. Given that housing prices could be affected by some macro-level factors, they use the province-level changes of urbanization rate as an excluded instrument, which implies that increases in urban population leads to more demand in the urban housing markets, thus increase the price. However, based on my analysis, this does not hold for urban population who comes from urban land expansion.

In addition, I apply this new urbanization framework to land and housing markets at the prefecture level. At the prefecture level, I show that not only the overall urban population growth is a strong predictor of the future residential land supply, but also the scale of in situ migrants is significantly related to the local residential land supply. This implies that a mismeasurement of urbanization might explain the surplus of housing units and result in a high vacant rate in China's

[2]Their method and results will be discussed in Section 2.1.2.

urban areas, as reported by CHFS. Thus, my work also contributes to the literature on the primary land market in China. For example, [53] show that government intervention enlarges the impact of positive productivity shocks on housing price appreciation, through mainly the government control over residential land supply. [33] shows that at the prefecture level, land supply could also be determined by the local leaders' career concern.

1.2 Development in China's Rural Land Tenure Security

Recent literature postulates that institutions have a central role in facilitating economic development ([41]; [45]). Broadly accessible and secure property rights that allow holders to enjoy the benefits from investment without being challenged by outsiders are viewed as an important element of an environment conducive to growth ([17]). In terms of growth in agricultural sector, although the fact from the process of economic development is that labor will leave the agricultural sector in favor of secondary and tertiary industries ([15]), resulting in the decline of the share of agriculture in GDP, it still remains as an important and essential component that supports human well-being. This is particularly true for less-developed countries in which reducing poverty is still one primary target of the government. Rural land, as the most basic inputs of agricultural production, is a key asset for households participating in agricultural activities. In this sense, well-defined and secure property rights in ownership or use over land can be recognized as essential for well-functioning institutions and economic growth in these countries ([18]; [14]).

Despite the emphasis on the importance of private property rights, collectively owned or managed land remains a widespread phenomenon in the developing world ([32]). China, which has one of the largest number of population in the agricultural sector, is among these countries where collective land structure is predominant. The establishment of Household Responsibility System (HRS) starting in early 1980s in China's countryside allows each rural household to be assigned a piece of land for management, the size of which is mainly on an egalitarian basis. On one hand, the reform greatly incentivizes enthusiasm of rural households, leading to huge increase in agricultural production. On the other hand, the Household Responsibility System and its associated tenure arrangement remain almost identical with how it looked like at the beginning of the reform, especially given the tremendous social and economic change over the last forty years. One recent theoretical paper by [10] studies why ration and contract land are a reasonable mechanism design in China's villages. Relying on the assumption that rural peasants migrating to the city for work have a probability of going underground when economic condition is bad and thus cause negative externalities that cannot be internalized by peasants themselves, they argue such ration land could serve as a role to make sure that the peasant migrant worker has incentives to return to the village during economic hard times.

However, the separation of land use rights from land ownership rights renders land tenure security being challenged even by state agents ([28]). In China's rural villages, plots are periodically subject to redistributions among all or a part of households within village by local officials, in order to fit the demographical change over time. Moreover, the rapid urbanization process makes land a demanding input for industrial, commercial and residential use, leading to large scale of acquisitions for rural land by local governments, with much lower compensations than it is sold then, which contributes to another source of systematic insecurity of land tenure.

In Section 3, I empirically test whether a land certification program launched by the central government improves agricultural land utilization. In order to secure land contractual and managerial rights for rural households, in addition to the land contract signed between local collectives and rural households, land certificates stating the size and the boundaries of each parcel as well as the contractual relationship were issued to rural households, since the mid-1990s and widely implemented from 2013. In particular, I estimate the effects of this program on reducing agriculture land abandonment phenomenon.

Cultivated land abandonment is a worldwide phenomenon that is driven by interlinked both environmental and economic forces. Its effects are also related to both economic and ecological fields. Ecologically, its negative impacts include soil erosion ([21]), reduction of landscape heterogeneity ([26]) etc. The most direct economic effect in the context of China is the implication to resource misallocation and food security ([36]). China, as the most populous country in the world, has one of the highest ratio of population to land. Having sufficient agricultural land to support such a large number of population independently is always a tough task. Keeping the bottom line of total agricultural land in rural China is one primary policy object in China's central government[3]. Therefore, increasing land utilization and reducing the size of idle land for agriculture use become fundamental in land use policies.

In contrast, the scale of cultivated land abandonment in rural China is non-negligible. [43] find that the abandonment ratio was 37.5% in 2009 in southern *Ningxia Hui Autonomous Region* in northwestern China. [48] study 2 counties in *Chongqing*, a southwest provincial-level municipality, and show that more than 30% plots have been left abandoned since 1992.

In Section 3, I use both household-level and village-level panel datasets from China Household Finance Survey (CHFS) which consists of more than 1,000 villages all across the country. The village-level dataset includes detailed agricultural land information of the most two recent waves of community survey, including timing and patterns of the land certification program. The panel datasets from CHFS allow me to adopt a generalized difference-in-differences framework to test the causal effect of the land certification program on agricultural land utilization. The empirical results show that the land certification program is effective in improving the utilization of agricultural land. I then try to disentangle possible mechanisms using the household-level panel data. I fail to find the effects of land certification program on labor migration, agricultural investments or overall land transfer activities. However, I find evidence showing that the land certification program alters households' renting behaviors; that is, households with cultivated land certificate are more willing to rent their agricultural land to new agricultural business entities rather than individual farming households, which might help increase the scale of economy and thus the land utilization rate.

This work mainly contributes to the literature that focuses on the effects of land tenure security on agricultural activities and related outcomes in developing countries, for example, [4] and [16] for African countries, [14] and [31] in the context of Latin America. [16] use data from Ethiopia and argue that the impact of land rights on investment incentives varies by type of investment. [14] study the issuance of ownership certificates in Mexico and show that removing the link between

[3]In 2006, the central government raised the policy that the bottom line of the agricultural land size is 1.8 billion Mu. This bottom line was reinforced recently in the 2013 Central Conference on Rural Work.

land use and land rights can result in large-scale adjustments to labor and land allocations, especially the increasing probability of migrating from households with ownership certificates. In the context of China, much of the literature focuses on its effect on developing local land transfer markets. For example, [45] study the effects of having land contract or certificate on household renting behaviors. They find that possession of such documents encourages households to engage in land renting to non-family members. [9] use data collected from more than 5,000 rural households in 2012 and find an increase in land renting activities and land rents from certification.

In addition, this work contributes to a growing literature studying the determinants of cultivated land abandonment in China empirically. For example, [35] employ a multivariate linear regression model to link livelihood strategy of household to land abandonment using a survey dataset conducted in 2011 in 12 villages in *Chongqing* and find positive relationship between off-farm labor choice and land abandonment. [48] use a plot-level dataset collected in 2012 in *Chongqing* and employ a Logit model to test the relationship between the different sources of income and the probability of rural land abandonment. They find that wage incomes are positively related to the probability of land abandonment while the agriculture income has negative effects. However, both these two papers only focus on regional cross-sectional data, which are neither nationwide representative nor well identified.

1.3 Development in China's Human Capital Investments

Education plays a fundamental role in formulating and developing human capital. Household investment on children's education is among one of the most important components of human capital investment, which helps improve children's ability to achieve higher educational outcomes as well as success in the labor market. It is also credential in the economic growth of a country ([29]). There is a large body of literature studying patterns and also determinants of household education expenditures in developed countries ([24]; [38]). However, such patterns in developing countries such as China might differ from those in developed countries mainly in the following two perspectives. First, the current Compulsory Education Law in China requires a minimum of 9 years' compulsory education in contrast to at least 12 years in the United States ([11]). Thus households might prepare extra expenditures for further education beyond this nine-year compulsory education. Second, the supply of high-quality public education resources is strictly limited, compared with an increasingly high demand, leading households to substitute these inaccessible public resources with affordable private investments. [57] show that household educational expenditure accounts for almost 30% of total education spending in 2013. As a result, it would be worth studying factors that determine levels of household educational investments in China.

In Section 4, I study how household educational investments are affected by different levels of opportunity costs in terms of re-taking exams of students in preparation for further education caused by an exogenous policy change. In particular, I rely on curriculum reform of high schools in mainland China launched by the Ministry of Education to study household investment decisions in their children who are in senior high school. In the context of education institution in China, students graduating from senior high school (12th grade) who apply for college are required to take the College Entrance Examination (CEE) which is held annually from the 7th to the 9th of June, organized by educational administration of each province[4]. Beside, it is common for students who fail the exam[5] for the first time to repeat the 12th grade (third year in senior high school), prepare for one more year and re-take the CEE one year later. Since settings and guidelines of CEE are determined by curriculum upon the current cohort in the third year of senior high school, in this sense, students from the last cohort under old curriculum are faced with a totally different CEE framework, in contrast to both cohorts under new curriculum or others under old curriculum. This provides a framework in which I can identify household educational investment in senior high school students given this exogenous curriculum shock leading to their different levels of opportunity costs to re-take CEE. Put differently, students among the last cohort under old curriculum are faced with a different situation upon re-taking CEE. If they decide to re-take CEE, they would spend extra time getting familiar with the new curriculum under which the forthcoming CEE would be guided. Thus, the difficulty for them to succeed in their second trial is higher

[4]In some provinces, the College Entrance Exam lasts for two days.

[5]There is no actual passing line of such exam. Students who fail the exam simply mean that they are either not enrolled or not enrolled into their expected university or major.

compared with other cohorts who may re-take the CEE with consistent curriculum.

I first employ province-level administrative data from the Chinese Education Statistical Yearbooks to establish the pattern that the number of CEE re-takers in years when provinces introduce new curriculum is significantly lower. Using individual-level survey data from the first wave of Beijing College Student Panel Survey, results show that households with students from the last cohort under old curriculum are significantly more likely to have reference books and electronic devices that improve their learning efficiency. My identification strategy lies in the sharp province by cohort variation in school curricula, allowing me to rule out confounding factors such as province specific or cohort specific differences. In addition, the result is robust after I consider the non-randomness of assignments of the introduction year of the new curriculum or after I include the level of public investments on senior high school students. However, due to the limitation of the survey data, I cannot measure investments such as asking for private tutors or attending out-of-school classes, which account for larger portions of household educational expenditures. Moreover, I do not find any long-run effect of these extra investments on students' performance after they are enrolled in college.

This work contributes directly to a growing empirical literature focusing on the effects of curriculum reform. For example, [5] study the causal effect of school curricula on students political attitudes and find that the new curriculum was often successful in changing their attitudes on important issues, in the direction intended by the Chinese government. [50] investigate effects of high school curriculum on various student outcomes including academic performance at the university, happiness, physical and mental health etc. Different from both papers that focus on direct effects of this curriculum reform, this work shows that there exists an indirect effect on students' decisions of re-taking CEE, which affects their education outcome and even their capability subsequently in the labor market.

More broadly, I also contribute to the literature that studies determinants of household educational investments. On one hand, levels of household educational expenditures are primarily affected by household-level characteristics such as disposable income, parents' education levels and their occupations. For example, [11] use Urban Household Survey data in 2007 and 2011 and find that out-of-school education expenditure increased rapidly during this period while household income and parents' education levels are two key factors influencing intergenerational education expenditure. On the other hand, changes in public policies would also affect household educational investment. [52] document robust evidence that increases in public spending on basic education are associated with significant reductions in household private tutoring spending in urban China. [39] find that expansions of pension coverage in urban China cause significant increase in total education investment in children. Similarly, this work analyzes how household educational investments are affected using curriculum reform as an exogenous policy change.

Furthermore, this work also contributes to the literature on how education decisions are affected by different levels of opportunity costs. In particular, [13] study how the reduction of changes in the cost of migration affected the decision of middle school graduates to attend high school in rural China and find a negative relationship between migrant opportunity and high school enrollment. In addition, this work studies how changes in the cost of re-taking a high-stake examination is related to the decision of household educational investment.

2. RELOCATING MIGRANTS AND IN SITU MIGRANTS: A NEW PERSPECTIVE ON URBANIZATION IN CHINA

2.1 Background

2.1.1 Rural-Urban Classification and the Urbanization Rate

The urbanization rate is the share of the population that resides in urban areas. It is also justifies studying how rural and urban areas are classified in China. There are actually no universal standards regarding the rural-urban dichotomy, and one country's specific standard may also vary over time. One traditional distinction between urban and rural areas is based on the assumption that urban areas, no matter how they are defined, offer a different way of life and usually provide a higher standard of living than rural areas (United Nations Statistics Division, 2017). According to a 2012 UNICEF report, the dichotomy could be based on one or a combination of the following standards: administrative criteria or political boundaries; a threshold population size[1]; population density; and local economic industries/sectors or the presence of urban characteristics (e.g., paved streets, electronic systems)[2].

In China's case, the official standard of rural-urban dichotomy changed with different waves of population censuses. For example, in the fourth census wave in 1990, the rural-urban definitions are purely administrative rather than indicative ([44]). For example, geographically, all "districts" under province and prefecture level cities were classified as urban ([6]). The fifth wave of census in 2000 used a combination of standards for administrative criteria (in conjunction with three new elements) to define urban areas: (a) whether or not an area has a population density of 1,500/squared kilometer; (b) whether or not the local government is located in the area; (c) whether or not the area is contiguous to an area where the local government is located[3] ([6]). Compared with the

[1] In United States, for example, the Census Bureau identifies two types of urban areas: Urbanized Areas have 5,000 or more people while Urban Clusters have at least 2,500 and less than 5,000 people.

[2] The summary of these standards comes from *The State of the World's Children*.

[3] The standard is *Guidelines for Statistically Classifying Rural and Urban Areas (for Trial Implementation)* which was issued by the National Bureau of Statistics in 1999.

previous standard, some towns or township areas that are in districts with less than 1,500 people/ square kilometer and that satisfy certain area contiguity criteria are classified as rural areas. Some areas under townships that were previously defined as rural, however, would be considered urban under this new standard if they meet certain conditions. [51] compares these statistical rules and finds that a different method for calculating the total urban population using the aforementioned two standards yields a very small change, only 0.16% of the 1990 and 2000 population censuses.

In 2008, the National Bureau of Statistics issued the current standard and announced that it would also be used for the sixth census in 2010. The standard is based solely on land contiguity, regardless of population densities, economic activities, or residential community infrastructures. Hence, there is a possibility that a community is officially reclassified from rural to urban only because its attribute of land contiguity has been changed. One result of expanding urban areas is that many rural residents are being redefined as urban. Different from migrants who actively relocate themselves from rural to urban areas, these urbanized population never move. This contributes to another part of urban population growth, in situ migrants.

2.1.2 In Situ Migrants

The main sources of urban population growth are natural growth, relocating migrants, and in situ migrants.

The current literature (albeit limited) offers a couple of papers that focus on how each of those sources contributes to China's urban population growth. For example, using population census data from the ten-year period of 1990 to 2000, [44] estimates that natural growth accounted for 17% of the total urban population increase and relocating migrants accounted for 31%. The remaining 52% came from the establishment of new cities or towns or the expansion of current city or town boundaries; this is contextually similar to the "in situ migrants".

Another related paper is [7], who employ census data and find that between 2000 and 2010, urbanization accounted for more than 80% of total urban population growth, more than half of which was attributed to rural-urban migration. In the 2000s, around 40% of the total urban population growth resulted from the land reclassification. Their approach of decomposing urban population growth can be summarized as follows. After net of natural population change, using census data on total urban population in both waves, they first estimate the overall urbanized population in the period, which accounts for both rural-urban migrants who have physically relocated (relocating migrants) and those whose place of residence was reclassified as urban (in situ migrants) ([7]). Second, they back out the flow of relocating migrants from census data which reflects population stocks. More specifically, their accounting framework defines four states based on one's residential status (rural and urban) and Hukou status (agricultural and non-agricultural)[4]. By assuming that those who live in rural areas always have agricultural Hukou in addition to some other assumptions, they break down rural-urban migrants into two components and calculate them separately: changes in urban residents originally with non-local agricultural Hukou (net of natural growth) and migrants who not only changed their Hukou status (agricultural to non-agricultural) but also change their Hukou registration place, called *YiDiNongZhuanFei* in Chinese. After estimating the number of relocating migrants, they back out the scale of urbanized population due to land reclassification by subtracting relocating migrants from the overall urbanized population.

These papers shed light on the sources of urban population growth. However, they seldom receive attention, perhaps for the following reasons. First, they treat "in situ migrants" as a residual term, not a direct measure. Put differently, they calculate the scale of in situ migrants after estimating the scale of relocating migrants and natural growth population; as a result, measurement errors may arise if one wants to study the scale of in situ migrants. Second, their methods can only be performed using two consecutive census waves to estimate the number for a nationwide 10-year period. In addition, their approach often relies on strong assumptions. For example, in [7], they assume those who live in urban areas and have agricultural Hukou in 2000 continue to live in urban areas in 2010, which fails to consider the relatively large scale of return migrants.

In this current work, I propose an approach to directly and annually measure the scale of in situ migrants at the prefecture level. The key to my method is using administrative codes and rural-urban classification codes published by the National Bureau of Statistics.

[4]They further define urban residents with agricultural Hukou into those with non-local agricultural Hukou and local agricultural Hukou.

2.2 Decomposing Urban Population Growth

Based on the new rural-urban division standard, in 2009, the National Bureau of Statistics (NBS) started annually publicizing the ID code for all communities and their corresponding rural-urban division codes[5]. Figure 2.1 includes two screenshots from the NBS website: the upper panel shows the links to information for all available years while the lower panel shows the content of each observation unit. Each row represents a community. The first two columns contain community ID and rural-urban division codes, respectively while the last column shows the community's name (in Chinese). I encode this information, track each community by its ID code year by year, and identify those that have been reclassified from rural to urban during the period. I am then able to measure the scale of in situ migrants resulting from community reclassification.

2.2.1 Community ID and Rural-Urban Division Codes

The twelve-digit community ID and the five-digit rural-urban division code have a combined 17 digits. The community ID consists of five administrative levels: the first two digits represent province-level administrative units (provinces, autonomous regions and province-level municipalities[6]), the next two digits for the prefecture and the fifth and sixth digits represent the county (counties, county-level cities and districts)[7]. The seventh, eighth and ninth digits represent the town while the final three digits represent the community[8]. They are codified according to the NBS' *Coding Rules for Administrative Units at the Town Level and Lower*[9].

Regarding the rural-urban division code, the first two digits represent the village attributes and the last three represent the rural-urban dichotomy. Only the last three digits for each community are available (Note that there are only three digits in the second column of Figure 2.1's lower panel.); they are translated from the twelve-digit administrative code and the first two digits of the

[5]On average, number of communities have been declining in recent years. The number varies by year between 667,909 (in 2016) and 699,089 (in 2009).

[6]Province-level municipalities include *Beijing, Tianjin, Shanghai* and *Chongqing*.

[7]These three levels are codified by the Ministry of Civil Affairs, not by the National Bureau of Statistics.

[8]Towns and communities that have been confirmed by local governments would be codified from 001 to 399. Those which have not been confirmed would be assigned codes from 400 to 599, indicating that they are virtual administrative units (for example, industrial zones, farms and forestry stations).

[9]*XianJi YiXia XingZheng QuHua Daima BianZhi Guize* in Chinese.

Figure 2.1: Screenshots of Community ID and Rural-Urban Division Codes

rural-urban division code. In this three-digit code, the first digit represents the basic rural-urban division[10] while the remaining two provide a detailed sub-classification. Since my main focus is the basic rural-urban dichotomy, I essentially rely on the first digit to distinguish urban communities from rural villages.

2.2.2 Results from Tracking Community ID Over Time

One problem in identifying communities with rural-urban dichotomy changes is that the same community's community ID might vary across years. Due to the coding rule discussed above, any change to the type of upper-level administrative units[11] could result in changes to the community ID. Moreover, it can also change when its (upper-level) administrative unit is reorganized[12]. Communities with ID codes change account more than 5% every year. As a result, before I identify those communities with rural-urban dichotomy change, I first need to track each community over years. Otherwise, my results will be inaccurate if these communities are different in possibilities of changing their rural-urban dichotomy.

I track community ID codes for all nearly 700,000 communities from 2009 to 2017. As shown in Figure 2.1, for each community, I know its community ID and Chinese name. From the code, I can extract information regarding all upper-level administrative units (e.g., province, prefecture, county and town-level governments) it belongs to, beyond which I have no further information. Since, I cannot obtain geographic locations (i.e., longitude and latitude) for each community, it would be impossible for us to track communities due to merges and splits using my algorithm. For example, Communities *A* and *B* merged as a new community. If the merged community in the latter year shares the same community ID and name with one of these two communities, (e.g., *A*), then my tracking methods would track Community *A* in both years and leave *B* in the former year unmatched. If the merged community in the latter year has a new code and name, then they

[10] *"1"* represents urban communities and *"2"* represents rural villages.

[11] For example, a county can be changed into a county-level city or district; a township can be changed into a town.

[12] For example, a county in one prefecture can be relegated to another prefecture. In 2011, *Chaohu* prefecture in the province of *Anhui* had all its counties revoked. All of *Chaohu*'s counties were reassigned to three other prefectures in the same province.

would be left unmatched in both years[13]. Apart from that, my tracking algorithm works pretty well in one-to-one community tracking in any two consecutive years. Put differently, almost all communities that are left unmatched are outcomes due to merges and splits.

Another fact that is worth mentioning is the variable I choose as the identifier. First, community name itself cannot work since there are multiple villages sharing the same Chinese name. Though each name is unique within a town, there might be several villages with the same name across towns in the same county. Second, community ID itself is not ideal as a single identifier, at least in the first few steps. There is a possibility that Community *A* with an ID in the former year is reassigned a new ID in the latter year while this new ID belonged to Community *B* in the same town in the former year. For example, in 2009, under *Pingzhuangcheng* Street, *Yuanbaoshan* District in *Chifeng* City, *Inner Mongolia*, the ID *150403002005* was assigned to *Qianjin Road No.2 Community*. In 2010, this ID was assigned to *Fangzhi Road No.1 Community* while that community used to have another ID *150403002014* in 2009. Thus, in this case, if I use only the community ID as the key variable, I could mistakenly track two different communities. As a result, using only the twelve-digit community ID code as the key identifier would cause a mismatch problem, even within the same town-level units, though this case is very rare in my dataset. In consequence, in order to prevent such mismatch, I use both community ID and its name (in Chinese) as key variables in my tracking algorithm.

Specifically, steps of my tracking algorithm are as follows and summarized in Table 2.1. After each step, tracked communities are dropped and I would deal only with those left unmatched in the next step, which would avoid the double counting problem.

Step 0: In this step, I track communities without code or name change using both variables as the identifier, which account for the majority of all tracked communities.

Starting from Step 1, I begin dealing with communities whose upper-level administrative units

[13]Since I could not distinguish both cases, I am not aware of the share of each case. Nevertheless, I can take the county-level administrative units merge as a reference. If the new county shares the same name with one of the previous counties, then it will keep the corresponding ID (For example, in 2015, *Jing'an* District and *Zhabei* District in *Shanghai* merged as a new district named *Jing'an*, then the new district kept the original *Jing'an* District ID code). Otherwise, it would have a new code (for example, in 2016, *Qinghe* District and *Qingpu* District in *Huai'an, Jiangsu* merged as a new district named *Qingjiangpu*, then the new district was assigned a new county-level ID code).

Table 2.1: Description of Tracking Steps

Description of Algorithm	Step	Key Variables Used in Matching
Tracking with unchanged code and name	0	Community Code, Name
Tracking with changed code at the prefecture level	1	Manually
Tracking with changed code at the county level	2	Manually
Tracking with changed code at the town level		
with: administrative type adjusted within county	3	Community Code[a], Name
with: reassignment across counties within prefecture	4	Community Code[a], Name
Tracking with changed code at the community level		
with: no name change within town	5	Town Code, Name
with: no short name change within town	6	Town Code, Short Name[b]
with: no name change across town within county	7	County Code, Name
with: no short name change across town within county	8	County Code, Short Name
with: first 2 characters of community name within town	9	Town Code, First 2 Characters
with: community code only	10	Community Code

[a] This is revised community code with substituting the 7$^{\text{th}}$ to 9$^{\text{th}}$ digits representing the town-level administrative unit with the corresponding digits of the matched town-level units after matching with short names as identifiers.
[b] "Short Name" refers to community name dropping the characters indicating the type of the community. For example, short name for *Guanghua Village* is *Guanghua*.

have been adjusted. As has been mentioned, since digits in the code represent a corresponding level of administrative units, I will proceed from the highest level (province) to the lowest level (town). For administrative changes at the provincial, city and county-level governments, I find formal documentation on adjustment details from the website of the Ministry of Civil Affairs[14]. First, I manually check changes at the province level. It turns out to be that in any two consecutive years between 2009 and 2017, there are no administrative unit adjustment across provinces. Put differently, no administrative units are reassigned to a different province. Thus, I start from the prefecture level.

Step 1: In this step, I track communities with prefecture-level administrative units change, which corresponds to the third and forth digits of the community ID. There are three reasons for prefecture-level units adjustment: establish of new prefectures, abolishment of current prefectures

[14] The website is http://xzqh.mca.gov.cn/description?dcpid=1

and changes of type of units[15]. For example, in 2011, *Chaohu* City in *Anhui* Province was revoked and its counties were assigned to three cities in the province. Examples of establishment of new prefectures include *Sansha* and *Danzhou* City in *Hainan* province. Though, the majority of administrative units change at the prefecture level takes the form of readjustment, with many prefecture-level regions in Southwestern China areas or autonomous regions reclassified as cities during this period. For all these three cases, I check communities under those changed prefectures and manually do the one-to-one matching.

Step 2: In this step, I track communities with county-level administrative units change. Such changes are shown from the fifth and sixth digits of the code. Similarly, most adjustments come from changes of type of county-level administrative units[16]. In particular, many counties or county-level cities are reclassified as districts. In addition to such reclassification, county-level units merges and splits are also a key component for adjustments. I also check communities under those changed counties and manually do the one-to-one matching.

The next two steps are associated with town-level ID code change. Since there are no detailed formal documents on adjustments at the town level and lower, I mainly rely on adjustments due to type changes of town-level administrative units.

Step 3: I first separate the characters indicating the type of the town-level units[17] from its full name and define what is left as its "short name"[18]. For town-level units that cannot be matched using their full names within a county in two consecutive years, I match them using their short names. Those matched samples may represent type adjustments[19]. Then I am able to track communities under those towns using both the new code by changing the corresponding digits (seventh

[15]Types of prefecture-level administrative units include prefecture-level cities (*Shi* in Chinese), prefecture-level regions (*Diqu*), autonomous prefectures (*Zizhizhou*) and leagues (*Meng*).

[16]Types of county-level administrative units include districts (*Qu* in Chinese), counties (*Xian*), county-level cities (*Xianjishi*) and autonomous counties (*Zizhixian*), banners (*Qi*) and autonomous banners (*Zizhiqi*). The latter two are particular for minority-grouped areas in *Inner Mongolia*.

[17]Types of town-level administrative units include towns (*Zhen* in Chinese), townships (*Xiang*), Streets (*Jiedao*) and autonomous townships (*Zizhixiang*) and types indicating a particular use (e.g., prisons, farms, free trade zones, industrial zones).

[18]For example, if the full name of a town-level administrative unit is *Guanghua Township*, I just keep *Guanghua* as the short name.

[19]For example, if the *Guanghua Township* is reclassified as *Guanghua Street* in the next year, though I cannot match these two units with their full name, I may achieve that result using the short-name matching.

to ninth) in the community ID and community name as key variables.

Step 4: I follow similar procedures in Step 3, only allowing matching town-level units with short name as identifier across county within the same prefecture in two consecutive years.

Beginning from Step 5, I track communities with code changes at the community level.

Step 5: Since the community name is uniquely identifiable within a town-level unit. In this step, I track communities by using town-level ID code and community name as key variables, which accounts for re-orderings for community lists within a town-level administrative unit.

Step 6: In this step, I use similar methods as I did for type adjustments at the town level. For all unmatched communities, I construct their short names by dropping the characters indicating their types[20]. I then match communities within town using their short names as key variables.

Step 7: This step follows Step 5, but using county-level ID code and community name as key variables, allowing communities reassignment to another town-level unit within the same county.

Step 8: Similar to Step 7, this step follows Step 6, using county-level ID code and community short name as key variables.

Step 9: In this step, I keep the first two characters of a community's full name and use this as the key identifier to track communities within town in two consecutive years. Though not as sharp as Step 0-8, in most cases they are accurate.

Step 10: The final step uses community ID code only as the identifier for those unmatched communities[21]. The reason that I would adopt this final step is because I observe that after previous steps, most unmatched communities result from character changes in the name of the community. Especially for communities in the minority-grouped regions, there exist changes in translating community names from their own language to Mandarin Chinese.

My tracking algorithm allows me to track more than 99% of communities for two consecutive years. Table 2.2 shows my tracking results using community administrative codes from 2009 to

[20]Types of community-level administrative units include communities (*Shequ* in Chinese), villages (*Cun*) and types indicating a particular use.

[21]Before the last step, I would match communities with the same name within a town across year, so the aforementioned mismatch problem would be alleviated.

2017[22]. Overall, around 90% of communities could be tracked across all years.

2.2.3 Measuring the Overall Scale of in Situ Migrants

In Table 2.2, I have already shown my tracking results starting from 2009, the first year that the National Bureau of Statistics publicized the rural-urban dichotomy information. As my study period is 2010-2015, Table 2.11 summarizes the tracking results during this sub-period. I tracked around 650,000 communities between 2010 and 2015. By analyzing each year's rural-urban division code for those tracked communities, I am able to obtain detailed information on whether each community has had a rural-urban dichotomy change and, if so, the year it was reclassified. Table 2.11 summarizes this information. Of the 649,182 communities that were tracked during this period, 27,795 were reclassified from rural to urban while 9,983 were reclassified from urban to rural. The percentage of reclassified communities from rural to urban net of those reclassified from urban to rural account for more than 3.7% of total rural communities in 2010[23].

One of the main focuses of this work is to measure the scale of in situ migrants between 2011 and 2015, which requires me to have information on the urban and rural population during this period. The lower the level of administrative units from which I have population data, the more accurate my measurements would be. Currently, I managed to gather data regarding urban and rural population at the prefecture level. This data set comes from different sources. First, for the year 2010, I obtain this information for every prefecture-level city from the 2010 population census. For the rest years (2011-2015), I look up urban population data for each prefecture either from its provincial-level statistical yearbooks or from the prefecture's statistical yearbooks or bulletins[24].

In order to translate the number of reclassified communities into the scale of in situ migrants,

[22] Tables 2.3 to 2.10 show results for the mannual tracking steps for each two consecutive years.

[23] Here I assume that there are no systemic differences between communities reclassified from rural to urban and those from urban to rural as the rural-urban dichotomy mainly relies on the contiguity rule discussed in Section 2.1.1, which can also be supported by descriptive evidence later shown in Section 2.3. In fact, many of communities reclassified from urban to rural were either original rural villages that had been reclassified to urban before another reclassification or urban communities that subsequently reclassified to urban again.

[24] While most provinces document the urban and rural population information each year in their statistical yearbooks, some prefectures in provinces that do not have such information would publish this data in the prefecture-level statistical yearbooks or announce it in the statistical bulletins. However, I fail to find this information in most prefectures in provinces like *Jilin*, *Heilongjiang*, *Tibet*, *Qinghai* and *Xinjiang*.

Table 2.2: Results from Tracking Communities across Years

	2009	2010	2011	2012	2013	2014	2015	2016	2017
Number of Communities	699,089	695,924	694,479	694,663	694,664	671,348	668,389	667,909	673,804
Tracked with Previous Year	N/A	692,925	688,739	691,522	690,941	666,600	664,252	642,728	663,001
Tracked with Previous Year (Percentage)	**N/A**	**99.57%**	**99.17%**	**99.55%**	**99.46%**	**99.29%**	**99.38%**	**96.23%**	**98.40%**
Tracked with Post Year	692,925	688,739	691,522	690,941	666,600	664,252	642,728	663,001	N/A
Tracked with Post Year (Percentage)	**99.12%**	**98.97%**	**99.57%**	**99.46%**	**95.96%**	**98.94%**	**96.16%**	**99.27%**	**N/A**
Tracked over All Years	617,770	617,770	617,770	617,770	617,770	617,770	617,770	617,770	617,770
Tracked over All Years (Percentage)	**88.37%**	**88.77%**	**88.95%**	**88.93%**	**88.93%**	**92.02%**	**92.43%**	**92.49%**	**91.68%**
Tracked over All Years: Urban	156,751	162,753	169,415	173,653	176,229	177,750	179,062	180,724	182,151
Tracked over All Years: Rural	461,019	455,017	448,355	444,117	441,541	440,020	438,708	437,046	435,619
Tracked over All Years: Urban to Rural	N/A	2,013	3,138	2,272	1,968	1,007	1,315	1,610	730
Tracked over All Years: Rural to Urban	N/A	8,015	9,800	6,510	4,544	2,528	2,627	3,272	2,157

Note: Data on number of communities is from the website of the National Bureau of Statistics. See appendix for detailed description for our tracking strategy. Communities tracked with previous (post) years are communities in year t that can be tracked with year t-1 (t+1), independent of other years. Communities tracked over all years refer to those that can be tracked for all years from 2009 through 2017.

Table 2.3: Tracking Results for Communities between 2009 and 2010

Step	Communities	Reason for Adjustment	Related Code (2009)	Related Code (2010)
0	666,778			
1	-			
2	1,051			
	90	merge	*110103*	*110101*
	107	merge	*110104*	*110102*
	336	merge & establishment	*120107, 120108, 120109*	*120116*
	52	abolishment	*320304*	*320302, 320311, 320312*
	326	reclassification	*320323*	*320312*
	39	establishment	*360425, 360426, 360427*	*360482*
	101	reclassification	*532522*	*532503*
3	9,275			
4	453			
5	971			
6	4,820			
7	4,666			
8	556			
9	1,900			
10	2,455			
Total	692,925			

Table 2.4: Tracking Results for Communities between 2010 and 2011

Step	Communities	Reason for Adjustment	Related Code (2010)	Related Code (2011)
0	648,805			
1	6,463			
	1,018	abolishment	*3414*	*3401, 3402, 3405*
	1,718	reclassification	*5222*	*5206*
	3,727	reclassification	*5224*	*5205*
2	1,929			
	72	merge	*310103*	*310101*
	65	reassignment	*321003*	*321002*
	113	merge	*321011*	*321003*
	333	reclassification	*321088*	*321012*
	156	reclassification	*430122*	*430112*
	439	merge & reclassification	*500110, 500222*	*500110*
	312	merge & reclassification	*500111, 500225*	*500111*
	237	reclassification	*511522*	*511503*
	65	reclassification	*530121*	*530114*
	137	reclassification	*532621*	*532601*
3	8,986			
4	580			
5	1,318			
6	3,291			
7	12,209			
8	644			
9	1,646			
10	2,868			
Total	688,739			

Table 2.5: Tracking Results for Communities between 2011 and 2012

Step	Communities	Reason for Adjustment	Related Code (2011)	Related Code (2012)
0	675,115			
1	3			
	3	establishment	*4690*	*4603*
2	981			
	131	reclassification	*130207, 130230*	*130209*
	269	merge & establishment	*320502, 320503, 320504*	*320508*
	310	reclassification	*320584*	*320509*
	26	merge	*340502*	*340503*
	43	establishment	*340521*	*340506*
	202	reclassification	*511821*	*511803*
3	4,327			
4	199			
5	1,029			
6	2,026			
7	4,640			
8	861			
9	929			
10	1,412			
Total	691,522			

I need to obtain average population for those reclassified communities. This data comes from the community survey in the China Household Finance Survey (CHFS), conducted by the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics.

To date, there have been four CHFS waves: 2011, 2013, 2015 and 2017. The first wave, from summer 2011, covered 80 counties in 25 provinces and collected data from 8,438 households across 320 community-level administrative units. The second wave in 2013 increased the sample size to 28,143 households distributed in 1,048 communities across 29 provincial-level regions[25]. The 2015 and 2017 waves consisted of 37,341 households from 1,373 communities and 40,011 households from 1,428 communities, respectively.

In CHFS data, communities are sampled using a stratified method with probability proportionate to size sampling. All counties are divided into groups based on their respective GDP per capita rankings; sampled counties are drawn from these groups. Sampled communities are then

[25]There are 31 provincial-level regions in mainland China. *Tibet* and *Xinjiang* are the only two that are not included in this survey. (These two are Minority Autonomous Regions in Western China.)

Table 2.6: Tracking Results for Communities between 2012 and 2013

Step	Communities	Reason for Adjustment	Related Code (2012)	Related Code (2013)
0	661,722			
1	1,648			
	1,648	reclassification	*6321*	*6302*
2	5,619			
	21	establishment	*150781*	*150703*
	407	reclassification	*220724*	*220781*
	58	merge	*320103*	*320104*
	53	merge	*320107*	*320106*
	120	reclassification	*320124*	*320117*
	148	reclassification	*320125*	*320118*
	319	reclassification	*321284*	*321204*
	93	merge	*370205*	*370203*
	1,023	merge	*370284*	*370211*
	206	reclassification	*441827*	*441803*
	526	reclassification	*445121*	*445102, 445103*
	265	establishment	*445202, 445221*	*445202, 445203*
	168	reclassification	*450322*	*450312*
	57	merge	*450404*	*450403*
	12	reassignment	*450403*	*450421*
	79	establishment	*450421*	*450406*
	116	establishment	*450902*	*450903*
	282	establishment	*511602*	*511603*
	849	reclassification	*511721*	*511703*
	435	establishment	*511902*	*511903*
	59	merge	*520114, 520181*	*520111*
	75	establishment	*520112*	*520115*
	140	reclassification	*532526*	*532504*
	31	establishment	*542430*	*542431*
	75	reclassification	*632721*	*632701*
	2	establishment	*652701*	*652702*
3	6,257			
4	246			
5	733			
6	3,621			
7	6,161			
8	275			
9	1,864			
10	2,795			
Total	690,941			

Table 2.7: Tracking Results for Communities between 2013 and 2014

Step	Communities	Reason for Adjustment	Related Code (2013)	Related Code (2014)
0	639,153			
1	2,514			
	492	reassignment	*1306*	*1390*
	354	reassignment	*1301*	*1390*
	1,668	reclassification	*5423*	*5402*
2	8,261			
	69	abolishment	*130103*	*130102, 130104*
	173	reclassification	*130124*	*130111*
	224	reclassification	*130182*	*130109*
	215	reclassification	*130185*	*130110*
	277	reclassification	*220181*	*220113*
	196	merge	*320705, 320706*	*320706*
	468	reclassification	*320721*	*320707*
	401	reclassification	*330621*	*330602, 330603*
	439	reclassification	*330682*	*330604*
	330	reclassification	*360782*	*360702, 360703*
	281	merge	*370802*	*370811*
	437	reclassification	*370882*	*370812*
	733	reclassification	*371081*	*371002, 371003*
	990	reclassification	*371421*	*371403*
	443	reclassification	*371624*	*371603*
	375	reclassification	*420321*	*420304*
	337	reclassification	*440183*	*440118*
	265	reclassification	*440184*	*440117*
	426	merge & reclassification	*440903, 440923*	*440904*
	388	reclassification	*441421*	*441403*
	142	reclassification	*445302, 445323*	*445302, 445303*
	140	establishment	*460201*	*460202 - 460205*
	326	reclassification	*500224*	*500151*
	170	reclassification	*500227*	*500120*
	16	establishment	*632802*	*632857, 632858, 632859*
3	6,964			
4	487			
5	525			
6	1,875			
7	4,486			
8	343			
9	839			
10	1,153			
Total	666,600			

Table 2.8: Tracking Results for Communities between 2014 and 2015

Step	Communities	Reason for Adjustment	Related Code (2014)	Related Code (2015)
0	638,107			
1	1,860			
	1,142	reclassification	*5421*	*5403*
	496	reclassification	*5426*	*5404*
	222	reclassification	*6521*	*6504*
2	7,425			
	311	reclassification	*120221*	*120117*
	418	reclassification	*120223*	*120118*
	573	reclassification	*130323*	*130302, 130304, 130306*
	221	merge & establishment	*130603, 130604*	*130606*
	193	reclassification	*130621*	*130607*
	272	reclassification	*130622*	*130608*
	304	reclassification	*130625*	*130609*
	330	reclassification	*230182*	*230113*
	80	merge	*320405, 320412*	*320402, 320404, 320411, 320412*
	109	reclassification	*320482*	*320413*
	245	reclassification	*320982*	*320904*
	305	reclassification	*330183*	*330111*
	103	reclassification	*330303, 330322*	*330305*
	217	reclassification	*350784*	*350703*
	279	reclassification	*350822*	*350803*
	338	reclassification	*360122*	*360112*
	223	reclassification	*361122*	*361103*
	345	reclassification	*410224*	*410212*
	60	merge	*440116*	*440112*
	352	reclassification	*441283*	*441204*
	171	reclassification	*441723*	*441704*
	218	reclassification	*450122*	*450110*
	291	reclassification	*451025*	*451081*
	303	reclassification	*500223*	*500152*
	168	reclassification	*500226*	*500153*
	108	reclassification	*511422*	*511403*
	244	reclassification	*513321*	*513301*
	146	reclassification	*520421*	*520403*
	220	reclassification	*530522*	*530581*
	64	reclassification	*533421*	*533401*
	96	reclassification	*610126*	*610117*
	118	reclassification	*630221*	*630203*
3	7,467			
4	70			
5	327			
6	2,140			
7	5,082			
8	256			
9	395			
10	1,123			
Total	664,252			

Table 2.9: Tracking Results for Communities between 2015 and 2016

Step	Communities	Reason for Adjustment	Related Code (2015)	Related Code (2016)
0	604,947			
1	1,083			
	285	establishment	*4690*	*4603*
	554	reclassification	*5422*	*5405*
	244	reclassification	*6522*	*6505*
2	9,901			
	426	reclassification	*110228*	*110118*
	422	reclassification	*110229*	*110119*
	979	reclassification	*120225*	*120119*
	300	merge	*130721*	*130702, 130703, 130705*
	182	reclassification	*130729*	*130708*
	214	reclassification	*130733*	*130709*
	382	reclassification	*131181*	*131103*
	232	reclassification	*210122*	*210115*
	254	reclassification	*210282*	*210214*
	167	reclassification	*211121*	*211104*
	55	reclassification	*230833*	*230883*
	122	reclassification	*231024*	*231086*
	205	merge	*310108*	*310106*
	340	reclassification	*310230*	*310151*
	154	merge & establishment	*320202, 320203, 320204*	*320213*
	118	merge & establishment	*320802, 320811*	*320812*
	122	reclassification	*320829*	*320813*
	40	merge & establishment	*340702, 340703*	*340705*
	115	reclassification	*340721*	*340706*
	257	reassignment	*340823*	*340722*
	278	reassignment	*341521*	*340422*
	82	reclassification	*360402, 360427*	*360483*
	330	reclassification	*370521*	*370505*
	367	reclassification	*371727*	*371703*
	272	reclassification	*411222*	*411203*
	147	reclassification	*450221*	*450206*
	522	reclassification	*500234*	*500154*
	282	reclassification	*510122*	*510116*
	261	reclassification	*510724*	*510705*
	853	reassignment	*512081*	*510185*
	108	reclassification	*513229*	*513201*
	267	reclassification	*520303, 520321*	*520302, 520303, 520304*
	130	reclassification	*530328*	*530303*
	72	reclassification	*530421*	*530403*
	76	reclassification	*533321*	*533301*
	34	reclassification	*540125*	*540103*
	150	reclassification	*610521*	*610503*
	217	reclassification	*610624*	*610603*
	367	reclassification	*610823*	*610803*
3	6,066			
4	32			
5	447			
6	8,682			
7	4,306			
8	2,876			
9	2,162			
10	2,226			
Total	642,728			

Table 2.10: Tracking Results for Communities between 2016 and 2017

Step	Communities	Reason for Adjustment	Related Code (2016)	Related Code (2017)
0	637,228			
1	890			
	890	reclassification	*1390*	*1301, 1306*
2	9,020			
	265	reclassification	*130428*	*130407*
	410	reclassification	*130429*	*130408*
	158	abolishment	*130421*	*130402, 130471*
	109	reassignment	*130427*	*130473*
	250	reclassification	*130823*	*130881*
	316	reclassification	*330185*	*330112*
	75	abolishment	*330204*	*330212*
	398	reclassification	*330283*	*330213*
	194	reassignment	*330212*	*330203*
	319	reclassification	*331021*	*331083*
	148	reclassification	*360421*	*360404*
	294	reclassification	*360721*	*360704*
	181	reclassification	*361029*	*361003*
	910	reclassification	*370181*	*370114*
	1,097	reclassification	*370282*	*370215*
	70	abolishment	*410211*	*410202*
	438	reclassification	*411023*	*411003*
	271	reclassification	*430124*	*430182*
	100	establishment	*440306*	*440309*
	24	establishment	*440307*	*440310*
	210	reclassification	*451281*	*451203*
	343	reclassification	*500228*	*500155*
	210	reclassification	*500232*	*500156*
	207	reclassification	*510124*	*510117*
	126	reclassification	*510626*	*510604*
	413	reclassification	*511028*	*511083*
	504	reclassification	*520222*	*520281*
	136	reclassification	*530122*	*530115*
	190	reclassification	*610125*	*610118*
	313	reclassification	*610721*	*610703*
	341	reclassification	*610821*	*610882*
3	6,722			
4	302			
5	477			
6	2,436			
7	2,722			
8	227			
9	921			
10	2,056			
Total	663,001			

Table 2.11: Results from Tracking Communities: 2010-2015

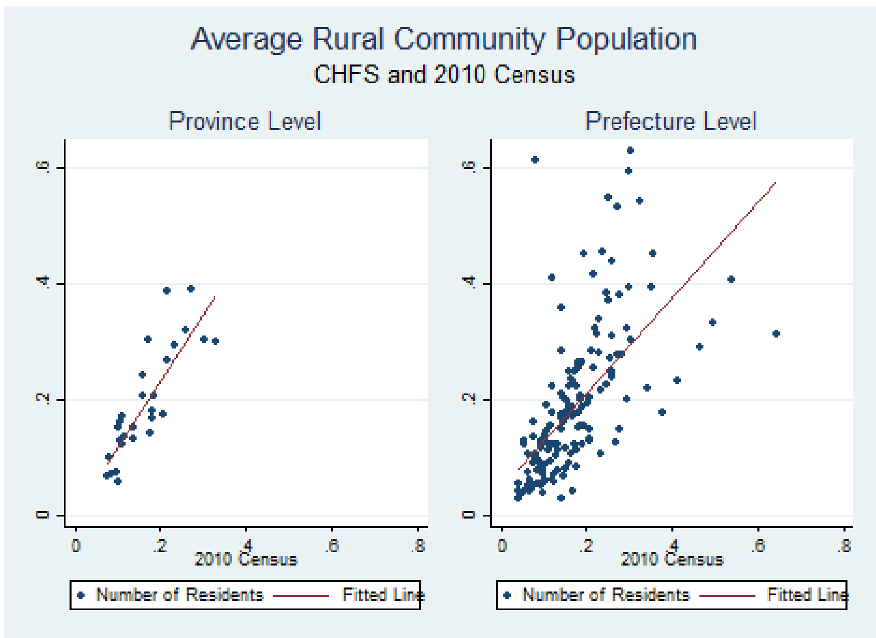
	2010	2015
Number of Communities	695,924	668,389
Tracked Communities	649,182	649,182
Tracked Communities: Urban	168,576	186,388
Tracked Communities: Rural	480,606	462,794
Tracked Communities: Rural to Urban	27,795	
Tracked Communities: Urban to Rural	9,983	
Tracked Communities: Percentage of 2010 Rural	**3.71%**	

Note: Data on number of communities is from the website of the National Bureau of Statistics. See appendix for detailed description for our tracking strategy. Tracked communities as a percentage of 2010 rural villages are calculated as (the number of tracked communities from rural to urban - the number of tracked communities from urban to rural) / the number of tracked rural communities in year 2010.

randomly drawn from those counties. The CHFS dataset is nationally representative panel data which contains rich micro-level information on household demographics, balance sheets, income and expenditures. Starting from the second wave in 2013, CHFS added a community-level survey for all sampled communities. This questionnaire included basic community demographic and geographic information, public infrastructures, economy and local governance.

To calculate the average population for reclassified communities, I pool all three waves of community survey and aggregate at both prefecture and province levels. I first show that my community samples are both representative at these two levels. In Figure 2.2, I show that information regarding average population from the CHFS community survey is quite comparable with the census data. The left panel shows this comparison at the province level. The horizontal axis shows the average population of a rural community within a province calculated as the ratio between the total rural population from 2010 Census and the total number of rural communities in 2010. The vertical axis illustrates the average population from the pooled community survey aggregated at each province. The correlation coefficient between these two variables is 0.84 while the estimated coefficient from a regression framework is 1.14, with the slope very close to a 45-degree line. The right panel show this relationship at the prefecture level. I can calculate the average population for all 339 prefectures from 2010 Census and 155 prefectures from pooled CHFS community survey. The coefficient of correlation is 0.62. In addition, the estimated coefficient is 1.04, which is significantly different from 0 but insignificantly different from 1, the slope of the 45-degree line. As a result, I am confident that the survey data is representative, at least for the rural population information at the prefecture level.

Figure 2.2: Population between CHFS Community Samples and 2010 Census



I then measure the average population for communities reclassified from rural to urban for each prefecture between 2010 and 2015. First, I assume it is equal to the average population from reclassified communities in the pooled CHFS community survey data (I have such information for 36 prefectures.). Since I do not have a large number of reclassified community samples in the CHFS data set, for those prefectures without such information, the average population is then assumed to be equal to the average rural population from the pooled CHFS community survey (I match additional 121 prefectures.)[26]. Recall that I only have prefecture-level aggregated information for

[26] Later in Section 2.3, I would provide some descriptive evidence that the average number of population from reclassified communities (both from rural to urban and from urban to rural) would be higher than the average population in a rural community by less than around 10%. Therefore, I am likely to underestimate the average population from these communities which lead to a lower bound of the measurement of the overall in situ migrants at the national level.

less than half prefectures. For the rest, I set the average population for reclassified communities equal to the province average[27]. Lastly, for the two province-level autonomous regions (*Tibet* and *Xinjiang*) from which I have no observations, the average population from reclassified communities in a prefecture is assumed to be the average number of rural population in 2010 Census in that prefecture also considering the natural growth from 2010 throughout 2015[28].

Then I am able to measure the scale of in situ migrants for each prefecture during 2010-2015, which is just the number of communities reclassified from rural to urban net of those reclassified from urban to rural multiplies the corresponding average population for these reclassified prefectures rescaled by the ratio between average number of total communities to the number of tracked communities during the period. Figure 2.3 illustrates the distributions of prefectures for both total urban population growth and the share of in situ migrants between 2010 and 2015. Every prefecture experienced positive total urban population growth. In terms of in situ migrants, however, I find negative results in a few prefectures, which is a reason that number of communities reclassified from urban to rural is greater than that from rural to urban. In terms of the distribution, I find that most of the prefectures experience an in situ migration scale with around 200,000. I estimate the overall scale of in situ migrants at the national level by summing this number across all prefectures. Results show that from 2010 to 2015, there were approximately 33.65 million in situ migrants.

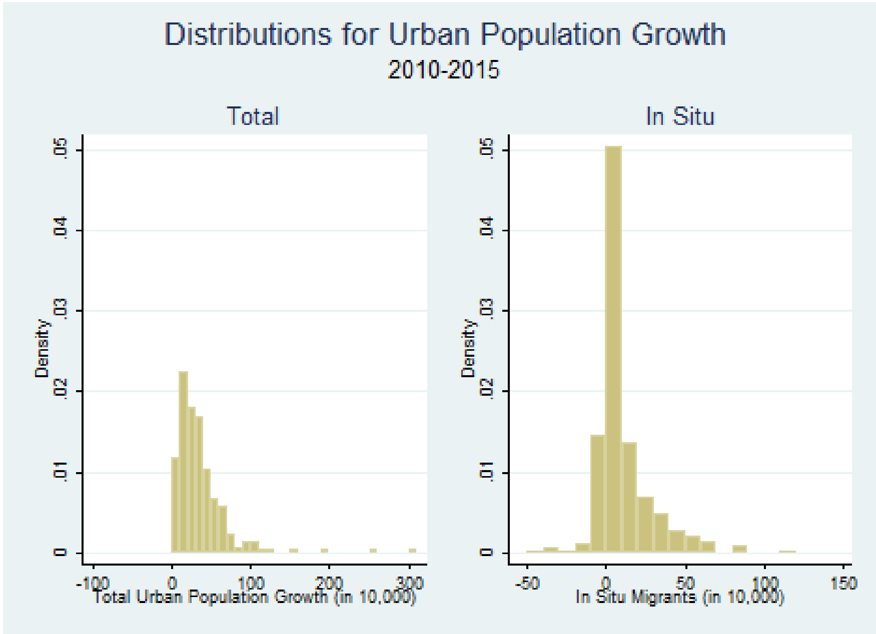
Next, I estimate the prefecture-level urban population growth resulting from natural birth and death. For each prefecture, this number is equal to the difference between the 2010 rural population and the overall prefecture-level in situ migrants times the overall natural growth rate in the corresponding province during 2011-2015. I conclude that the natural growth of the urban population was approximately 16.60 million.

Finally, Table 2.12 shows that the urban population increased by 101.38 million between 2010

[27] I use similar step-wise assumptions as I did at the prefecture level. For those unmatched prefectures, I first assume the average population for reclassified communities is equal to the province average for reclassified communities (I match 124 prefectures.) or rural communities (I match 36 prefectures.).

[28] Data about the natural growth rate is at the province-year level from the National Bureau of Statistics. I am not able to find such information of rural population at the prefecture level. As a result, I need to make assumptions that the variations of natural growth rate during 2010-2015 are not significantly different between rural and urban areas across prefectures within the same province.

Figure 2.3: Distributions for Urban Population Growth across Prefectures



and 2015. This total growth consists of natural growth, relocating migrants and the in situ migrants. By subtracting the amount of natural growth and in situ migrants from the total, I estimate that relocating migrants accounted for approximately 51.13 million, around 50.43% of total urban population growth; in situ migrants accounting for around 33.19%, shown in Figure 2.4.

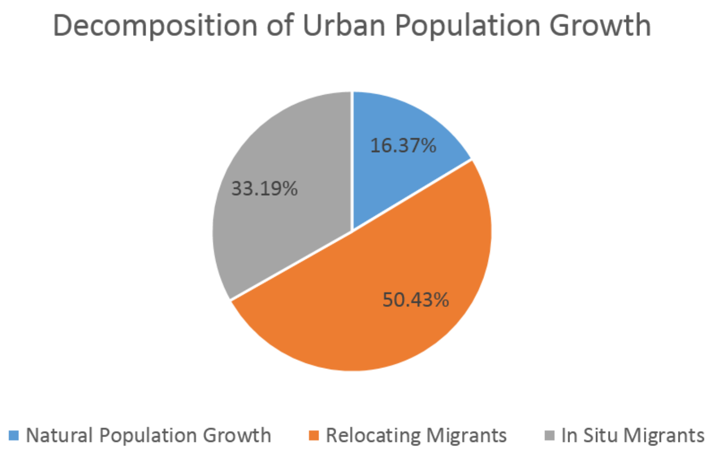
Table 2.12: Total Population and Urbanization Rate

Year	Total Population	Total Population: Urban	Urbanization Rate
2011	1,347,350	690,790	51.27%
2012	1,354,040	711,820	52.57%
2013	1,360,720	731,110	53.73%
2014	1,367,820	749,160	54.77%
2015	1,374,620	771,160	56.10%

Note: Data is from the National Bureau of Statistics. The population number is measured in thousands.

Comparing my result with the aforementioned two papers in this literature, it follows quite a

Figure 2.4: Sources of Urban Population Growth (2010-2015)



similar trend since 1990[29]. Though declining, this relatively large share indicates that urban area expansions are still one of the main factors driving China's increasing urbanization rate.

2.2.4 Comparing Results Using an Alternative Approach

Unlike [44] and [7] who back out the scale of rural-urban migration first and then take the residual as urbanization by urban land extension or reclassification, my approach of decomposing urban population growth takes the opposite direction. I first estimate the scale of in situ migrants and then back out the relocating migrants as the residual term. In order to compare my approach with theirs, especially [7] result that employs the most recent census waves, I adopt their method to back out the scale of rural-urban relocating migrants as a robustness check using 2010 population census and 2015 1% population mini-census. Recall that [7] break down rural-urban migrants into two components: net changes in urban residents originally with non-local agricultural Hukou and *YiDiNongZhuanFei*, summarized in Equation (10) from [7]:

$$\underbrace{P_{12}^{*}+P_{14}^{*}}_{\text{rural-urban migration}} = P_{2n}^{2010} - P_{2n}^{2000} - \underbrace{(b_{2_n} - d_{2_n})}_{\substack{\text{natural population change of} \\ \text{urban residents with non-local Hukou}}} + \underbrace{P_{2_n4} + P_{14}^{*}}_{\textit{YiDiNongZhuanFei}} \tag{2.1}$$

[29]Recall that the share of in situ migrants to the overall urban population growth is 52% in 1990s [44]) and 40% in 2000s ([7]).

where $P_{2n}^{2010} - P_{2n}^{2000} - (b_{2_n} - d_{2_n})$ is change of urban residents with non-local agricultural Hukou net of natural growth.

I try to calculate these two components from the census data. First, I follow their approach to find $P_{2_n}^t$, the total number of urban residents with non-local agricultural Hukou. Since it is not directly available from census data, they back it out using the following Equation:

$$P_{2_n}^t = P_n^t \times \frac{P_{2_n}^t}{P_n^t} \tag{2.2}$$

where P_n^t represents the total number of urban residents with non-local Hukou, which is observable in census while the latter term can be found in the *Long Table*, a subsample of census data. However, in 2015 mini-census, I do not have *Long Table* questionnaire, so such information is not directly accessible. I infer this ratio using 2000 and 2010 census data by linear interpolation, which is approximately 54.7%[30]. In the 2015 mini-census, total number of urban residents with non-local Hukou (P_n^{2015}) is 292.96 million. Thus, changes amount to

$$P_{2n}^{2015} - P_{2n}^{2010} = P_n^{2015} \times \frac{P_{2_n}^{2015}}{P_n^{2015}} - P_n^{2010} \times \frac{P_{2_n}^{2010}}{P_n^{2010}} = 292.96 \times 54.7\% - 261.38 \times 52.7\% = 22.50 \tag{2.3}$$

Moreover, I assume the natural population change as of total change is proportionately identical between 2000-2010 and 2011-2015. By this assumption, the estimation of $b_{2_n} - d_{2_n}$ is approximately 2.84 million. Hence, I can infer a net change of urban residents with non-local agricultural Hukou during 2011-2015 of 19.66 million.

Second, I estimate the scale of *YiDiNongZhuanFei* from 2010 to 2015. Since the agricultural/non-agricultural dual Hukou system was abolished in 2014, there was no information regarding number of people with non-agricultural Hukou in the 2015 mini-census. Similarly, I approximate this ratio in 2015 by linear interpolation, which is around 31.4%[31].Thus, the population with non-agricultural Hukou grew from 384.34 million in 2010 to 430.99 million in 2015. The estimated change of popu-

[30]This ratio is 48.8% and 52.7% in 2000 and 2010 respectively.
[31]This ratio is 24.7% and 29.1% in 2000 and 2010 respectively.

lation with non-agricultural Hukou net of natural growth (*NongZhuanFei*) was 35.26 million, 45% of which were *YiDiNongZhuanFei*, approximately 15.87 million.

By adding up these two components, the scale of rural-urban migrants using the approach in [7] during 2010-2015 was 35.53 million. Subtracting these 35.53 million relocating migrants and 18.16 million natural growth from the 101.38 million total urban population increase between 2010 and 2015, the estimated scale of in situ migrants was around 47.69 million, 42% larger than the result from Section 2.2.3. As has been mentioned, my approach tends to be conservative and is likely to underestimate the scale of in situ migrants. Also, in this alternative method, since some information from the 2015 1% population mini-census is not available, I have to rely on some linear estimation from its own tread, which is also likely to cause inaccurate measurements. Nevertheless, since both approaches show that there are a large share of urban population growth from urban area expansion (from 33% to 45%), I believe that exploring the characteristics of in situ migrants living on those expanded urban areas are of importance to understand the overall urbanization process for the recent decades in China.

2.3 Characteristics of in Situ Migrants

In Section 2.2, I estimate that in situ migrants accounted for more than 30% of total urban population growth from 2010 to 2015. In this subsection, I use micro-level survey data from CHFS to explore some characteristics of in situ migrants and the communities they live in.

2.3.1 Community- and Household-level Descriptive Evidence

I obtained three waves of the community survey and four waves of household survey from CHFS; then I categorized these communities into four groups: rural, urban and those reclassified from rural to urban from 2010 to the survey year[32], those reclassified from urban to rural[33].

Table 2.13 compares community characteristics for the 2013 wave. Of all 1,048 communities surveyed in 2013, I can track 928 using their administrative code as the identifier; of these 505

[32]For example, if a community was reclassified from rural to urban in 2014, it was categorized as rural in the 2013 wave but as reclassified in the 2015 wave.

[33]Some communities experienced the rural-urban dichotomy switch more than twice during the study period. I do not include those observations in the comparison as well as later in the empirical analysis.

were urban, 388 were rural, 22 were communities reclassified from rural to urban (I refer to those communities as reclassified communities for simplicity.), 11 were communities reclassified in the opposite way and 1 community was reclassified more than once. In the community questionnaire, I ask village leaders what type of community they consider their community to be. Results show that the leadership of all 22 reclassified communities regarded themselves as rural, indicating that they were unaware of their own community's rural-urban dichotomy change. What's more, in most of the variables listed in the table, reclassified communities have more characteristics in common with rural villages than their urban counterparts. For example, reclassified communities have an average of 2,277 residents; this is close to the rural population average, and far below the urban community average. Similarly, the average number of registered households in reclassified communities is close to that of rural villages. The questionnaire also asked about basic local facilities, such as number of banks. Data shows that both reclassified and rural communities have less than one bank in their neighborhoods. Conversely, urban communities have more than two banks on average; this reflects that, in terms of the economic development and infrastructure, these reclassified communities are far less developed than urban communities. Interestingly, all the characteristics for communities reclassified from urban to rural are also very close to traditionally rural villages and reclassified communities. This is straightforward in the sense that the current rural-urban dichotomy adopted by the National Bureau of Statistics is simply based on land contiguity, making both directions of switches independent of the social and economic development of the community.

Beginning in 2015, more questions were added to the survey questionnaire regarding infrastructure and economic activities within a community. These new questions will help me make comparisons in more aspects. Of the 1,257 samples from the 2015 wave shown in Table 2.14, 37 were reclassified from 2010 to 2015; of these, more than 70% regard themselves as rural rather than urban. Demographically, these reclassified communities share similar characteristics with rural villages, such as the number of residents and also registered residents. Furthermore, 90% of the population in reclassified communities has an agricultural Hukou. The ratio of communities that own agricultural land is approximately 9% in urban communities but over 70% in reclassified

Table 2.13: Community Comparisons for CHFS 2013

	Urban	Rural	Rural to Urban	Urban to Rural
Self-reported Rural	0.16	0.97	1.00	0.83
	(0.37)	(0.18)	(0.00)	(0.39)
Number of Residents	7,595.99	2,060.70	2,276.59	2,268.42
	(7,061.90)	(1,445.41)	(1,080.74)	(1,787.38)
Registered Households	2,375.17	511.07	567.32	637.83
	(1,900.59)	(333.15)	(270.58)	(480.16)
Number of Banks	2.28	0.51	0.55	0.58
	(2.37)	(0.98)	(0.86)	(1.08)
Poor Households	0.07	0.13	0.11	0.10
	(0.12)	(0.14)	(0.08)	(0.14)
Observations	505	388	22	12

Means and standard errors (in parentheses) are presented.

communities; this indicates that the majority of households in these reclassified communities still rely on agricultural activities. Additionally, these reclassified villages have significantly sparser financial services, job training, and childcare and eldercare options than urban areas. Moreover, the disposable per capital income for reclassified communities was less than 10,000 *RMB*, only half of that of urban communities.

This phenomenon remained unchanged in the 2017 survey. Of all 1,310 communities in 2017, 42 were reclassified. For those communities reclassified between 2010 and 2017, they continued to share common characteristics with traditional rural villages in most of my measures. For example, in Table 2.15, more than 60% of these villages still self-reported as rural, almost 95% of residents held agricultural Hukou and more than 80% had farmland. The average disposable income was still far less than that of urban communities.

To compare the difference between reclassified communities from both their rural and urban counterparts, I pool all three waves to increase the sample size. Table 2.16 presents the results. Variables related to values are measured in 2015 *CNY*. In Column 5 I show the difference between

Table 2.14: Community Comparisons for CHFS 2015

	Urban	Rural	Rural to Urban	Urban to Rural
Self-reported Rural	0.19	0.98	0.76	1.00
	(0.39)	(0.15)	(0.43)	(0.00)
Number of Residents	7,214.69	1,925.94	2,110.22	2,002.50
	(7,151.95)	(1,835.21)	(1,489.81)	(1,671.91)
Registered Residents	5,374.95	2,125.37	2,180.14	1,929.71
	(5,226.85)	(1,826.65)	(1,338.65)	(1,500.85)
Ag-Hukou Ratio	0.26	0.96	0.94	0.86
	(0.41)	(0.15)	(0.15)	(0.31)
Registered Households	2,250.38	557.68	672.58	638.50
	(3,725.63)	(433.81)	(559.74)	(426.91)
Agricultural Land	0.09	0.94	0.77	0.73
	(0.28)	(0.25)	(0.44)	(0.47)
Kindergarten	0.70	0.36	0.49	0.43
	(0.46)	(0.48)	(0.51)	(0.51)
Number of Banks	1.98	0.31	0.35	0.43
	(2.35)	(0.87)	(0.75)	(0.76)
Old-Care Service	0.48	0.16	0.22	0.43
	(0.50)	(0.37)	(0.42)	(0.51)
Job Training Service	0.24	0.07	0.14	0.21
	(0.43)	(0.25)	(0.35)	(0.43)
Disposable Income pc	17,730.16	6,967.72	8,230.83	20,298.14
	(17,221.26)	(6,576.80)	(5,683.49)	(29,994.51)
Poor Residents	0.03	0.05	0.09	0.03
	(0.08)	(0.08)	(0.17)	(0.05)
Observations	676	524	37	14

Means and standard errors (in parentheses) are presented.

Table 2.15: Community Comparisons for CHFS 2017

	Urban	Rural	Rural to Urban	Urban to Rural
Self-reported Rural	0.15	0.96	0.62	0.93
	(0.36)	(0.19)	(0.49)	(0.26)
Number of Residents	7,186.88	1,911.69	2,070.29	1,836.60
	(6,307.84)	(2,026.00)	(1,571.71)	(1,439.96)
Registered Residents	5,189.84	2,040.96	2,044.90	1,847.00
	(4,071.50)	(1,543.53)	(1,474.35)	(1,430.95)
Ag-Hukou Ratio	0.25	0.95	0.93	0.86
	(0.40)	(0.16)	(0.22)	(0.35)
Registered Households	2,072.33	556.69	609.17	618.80
	(1,838.89)	(406.66)	(528.80)	(437.60)
Agricultural Land	0.15	0.92	0.82	0.77
	(0.36)	(0.28)	(0.39)	(0.44)
Kindergarten	0.70	0.38	0.50	0.40
	(0.46)	(0.49)	(0.51)	(0.51)
Number of Banks	1.80	0.33	0.57	0.73
	(2.13)	(1.28)	(0.97)	(1.62)
Old-Care Service	0.45	0.20	0.33	0.27
	(0.50)	(0.40)	(0.48)	(0.46)
Job Training Service	0.22	0.09	0.12	0.13
	(0.41)	(0.29)	(0.33)	(0.35)
Disposable Income pc	23,036.04	7,731.74	9,745.69	10,145.13
	(45,279.23)	(6,706.46)	(7,895.18)	(6,363.47)
Poor Residents	0.03	0.05	0.05	0.02
	(0.07)	(0.08)	(0.09)	(0.02)
Observations	694	550	42	15

Means and standard errors (in parentheses) are presented.

Table 2.16: Community Comparisons for Pooled Sample

	Urban	Rural	Rural to Urban	Urban to Rural	(1)-(3)	(2)-(3)
Self-reported Rural	0.17	0.97	0.75	0.93	-0.59***	0.22***
	(0.37)	(0.17)	(0.43)	(0.26)	(0.04)	(0.02)
Number of Residents	7,307.78	1,956.30	2,129.85	2,019.63	5,177.93***	-173.55
	(6,824.02)	(1,818.16)	(1,436.49)	(1,594.98)	(679.99)	(184.80)
Registered Residents	5,280.86	2,082.14	2,107.32	1,886.93	3,173.54***	-25.18
	(4,674.56)	(1,687.33)	(1,405.91)	(1,439.24)	(530.85)	(195.84)
Ag-Hukou Ratio	0.26	0.95	0.94	0.86	-0.68***	0.02
	(0.40)	(0.16)	(0.19)	(0.33)	(0.05)	(0.02)
Registered Households	2,218.62	544.92	622.79	631.10	1,595.83***	-77.87*
	(2,686.86)	(399.12)	(493.73)	(435.60)	(268.99)	(41.95)
Agricultural Land	0.17	0.94	0.87	0.82	-0.70***	0.08***
	(0.38)	(0.23)	(0.34)	(0.39)	(0.05)	(0.03)
Kindergarten	0.70	0.37	0.49	0.41	0.21***	-0.12**
	(0.46)	(0.48)	(0.50)	(0.50)	(0.05)	(0.06)
Number of Banks	1.99	0.37	0.49	0.59	1.51***	-0.11
	(2.28)	(1.07)	(0.87)	(1.20)	(0.23)	(0.11)
Old-Care Service	0.47	0.18	0.28	0.34	0.19***	-0.10**
	(0.50)	(0.39)	(0.45)	(0.48)	(0.06)	(0.05)
Job Training Service	0.23	0.08	0.13	0.17	0.10**	-0.05
	(0.42)	(0.27)	(0.33)	(0.38)	(0.05)	(0.03)
Disposable Income pc	19,234.17	6,645.82	8,104.43	12,887.85	11,129.74***	-1,458.61**
	(32,388.26)	(6,068.18)	(6,270.48)	(18,886.43)	(3,244.04)	(628.80)
Poor Residents	0.04	0.07	0.08	0.05	-0.04***	-0.01
	(0.09)	(0.10)	(0.13)	(0.09)	(0.01)	(0.01)
Observations	1875	1462	101	41		

Means and standard errors (in parentheses) are presented.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

urban and reclassified communities. In all characteristics, the differences are not only large in magnitude but also highly significant different from zero, all at 1% level. In contrast, Column 6 shows that between rural villages and reclassified communities. While the differences are smaller, most of them are indistinguishable from zero. From such simple comparison, it is clear to see that although reclassified as urban, those commmunities are more aligned with their rural counterparts in terms of demographic and economic development.

In addition to community-level comparisons, I also use household survey data to explore

Table 2.17: Household Comparisons for CHFS 2011

	Urban	Rural	Rural to Urban	Urban to Rural
Family Size	3.21	3.98	3.95	2.83
	(1.39)	(1.72)	(1.62)	(1.26)
Number of Children	0.53	0.77	0.61	0.30
	(0.69)	(0.93)	(0.84)	(0.56)
Head Schooling	10.56	6.76	6.60	8.51
	(3.90)	(3.46)	(3.52)	(4.34)
Head Urban Hukou	0.69	0.03	0.02	0.85
	(0.46)	(0.16)	(0.14)	(0.36)
Ag Activity	0.13	0.82	0.88	0.00
	(0.34)	(0.39)	(0.33)	(0.00)
Agricultural Land	0.22	0.89	0.96	0.07
	(0.42)	(0.32)	(0.20)	(0.27)
Total Income	65,174.46	26,701.83	27,423.48	29,278.56
	(140,145.31)	(134,420.56)	(27,167.06)	(25,316.41)
Total Consumption	49,783.60	24,138.83	19,732.86	22,851.19
	(108,396.96)	(45,063.22)	(21,338.86)	(15,136.80)
Observations	4958	2459	100	40

Means and standard errors (in parentheses) are presented.

whether significant demographic differences exist at the household level. Table 2.17 shows the demographic and economic information of households in different categories of communities from the 2011 wave. Of the total 7,557 households from which I have detailed information, 100 are from reclassified communities. Similar to what I find with community-level data, households living in reclassified communities share more common characteristics with rural households. For example, compared with urban households, they have larger household sizes, more children, and less years of schooling. They are also more likely to hold a rural Hukou while working in the agricultural sector. With respect to economic activity, their incomes and consumption levels are much lower.

The following waves have larger samples. I identified 463, 825 and 969 households from

Table 2.18: Household Comparisons for CHFS 2013

	Urban	Rural	Rural to Urban	Urban to Rural
Family Size	3.19	4.05	4.02	3.64
	(1.42)	(1.88)	(1.87)	(1.53)
Number of Children	0.49	0.76	0.75	0.55
	(0.70)	(1.00)	(0.90)	(0.67)
Head Schooling	10.55	6.84	7.25	7.76
	(4.04)	(3.50)	(3.50)	(3.08)
Head Urban Hukou	0.68	0.03	0.04	0.13
	(0.47)	(0.17)	(0.19)	(0.34)
Ag Activity	0.10	0.73	0.69	0.56
	(0.30)	(0.45)	(0.46)	(0.50)
Agricultural Land	0.21	0.80	0.81	0.66
	(0.40)	(0.40)	(0.39)	(0.48)
Total Income	80,292.60	35,927.02	45,131.39	39,209.15
	(212,872.84)	(104,699.64)	(97,593.95)	(49,702.95)
Total Consumption	57,874.10	35,064.97	39,381.68	36,619.71
	(66,336.46)	(41,180.66)	(44,921.47)	(32,050.64)
Observations	16999	8123	463	293

Means and standard errors (in parentheses) are presented.

communities reclassified from rural to urban in 2013, 2015 and 2017 wave, respectively, and the results were in line with what I found using the 2011 wave, both in terms of demographic and economic conditions shown in Table 2.18, 2.19 and 2.20.

In addition, Table 2.21 compares the difference for household characteristics pooling samples from all four waves. Similar to what I find from the community-level observations, in the household data, I still find large and significant difference between urban households and those from reclassified communities while this difference is much larger than that between rural and reclassified communities.

Using descriptive evidence from both community and household-level comparisons, I conclude

Table 2.19: Household Comparisons for CHFS 2015

	Urban	Rural	Rural to Urban	Urban to Rural
Family Size	3.14	3.78	3.88	3.52
	(1.43)	(1.84)	(1.80)	(1.57)
Number of Children	0.47	0.67	0.69	0.50
	(0.69)	(0.95)	(0.87)	(0.68)
Head Schooling	10.37	6.88	7.18	7.59
	(4.07)	(3.46)	(3.57)	(2.87)
Head Urban Hukou	0.65	0.03	0.06	0.04
	(0.48)	(0.17)	(0.24)	(0.19)
Ag Activity	0.12	0.74	0.59	0.67
	(0.33)	(0.44)	(0.49)	(0.47)
Agricultural Land	0.24	0.89	0.77	0.81
	(0.43)	(0.31)	(0.42)	(0.40)
Total Income	98,828.93	44,847.91	52,322.53	49,098.03
	(734,014.00)	(267,614.44)	(132,083.67)	(132,610.70)
Total Consumption	66,831.60	37,144.16	49,047.86	49,714.57
	(79,987.78)	(49,554.05)	(86,940.28)	(70,898.66)
Observations	22442	10513	825	349

Means and standard errors (in parentheses) are presented.

Table 2.20: Household Comparisons for CHFS 2017

	Urban	Rural	Rural to Urban	Urban to Rural
Family Size	2.97	3.55	3.59	3.35
	(1.39)	(1.77)	(1.80)	(1.64)
Number of Children	0.43	0.62	0.65	0.51
	(0.69)	(0.92)	(0.93)	(0.74)
Head Schooling	10.46	6.96	7.12	7.08
	(3.97)	(3.46)	(3.69)	(3.36)
Head Urban Hukou	0.60	0.03	0.07	0.06
	(0.49)	(0.18)	(0.25)	(0.24)
Ag Activity	0.12	0.75	0.56	0.70
	(0.32)	(0.43)	(0.50)	(0.46)
Agricultural Land	0.18	0.80	0.66	0.71
	(0.39)	(0.40)	(0.47)	(0.46)
Total Income	98,468.22	45,227.48	50,811.92	56,815.12
	(121,807.79)	(73,563.44)	(70,419.81)	(86,970.08)
Total Consumption	67,704.29	37,091.12	44,290.51	41,214.59
	(58,118.76)	(38,198.78)	(43,928.98)	(42,938.89)
Observations	23912	11425	969	394

Means and standard errors (in parentheses) are presented.

Table 2.21: Household Comparisons for Pooled Sample

	Urban	Rural	Rural to Urban	Urban to Rural	(1)-(3)	(2)-(3)
Family Size	3.10	3.78	3.79	3.46	-0.69***	-0.01
	(1.41)	(1.83)	(1.81)	(1.58)	(0.03)	(0.04)
Number of Children	0.46	0.68	0.68	0.51	-0.22***	-0.00
	(0.69)	(0.95)	(0.90)	(0.69)	(0.01)	(0.02)
Head Schooling	10.46	6.89	7.15	7.48	3.32***	-0.26***
	(4.02)	(3.47)	(3.60)	(3.19)	(0.08)	(0.07)
Head Urban Hukou	0.64	0.03	0.06	0.10	0.58***	-0.03***
	(0.48)	(0.17)	(0.23)	(0.30)	(0.01)	(0.00)
Ag Activity	0.11	0.75	0.61	0.63	-0.50***	0.14***
	(0.32)	(0.44)	(0.49)	(0.48)	(0.01)	(0.01)
Agricultural Land	0.21	0.84	0.74	0.70	-0.53***	0.10***
	(0.41)	(0.37)	(0.44)	(0.46)	(0.01)	(0.01)
Total Income	90,373.77	40,786.01	48,339.45	47,617.83	42,034.32***	-7,553.44**
	(441,480.46)	(170,892.02)	(99,504.37)	(94,564.54)	(9,101.54)	(3,563.01)
Total Consumption	62,821.37	35,157.36	43,161.16	41,425.77	19,660.21***	-8,003.80***
	(72,486.90)	(43,216.13)	(61,805.14)	(50,695.26)	(1,511.69)	(953.84)
Observations	68287	32520	2357	1076		

Means and standard errors (in parentheses) are presented.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

that the modernization process is still less developed for these reclassified communities, regardless of the community's infrastructure, economic activities, or income levels. Although statistically treated as urban, reclassified communities are much more similar to rural villages.

2.3.2 Community Reclassification and Housing Behaviors

Aimed to solve the urban high-vacancy-rate puzzle, it is of both interest and importance to test empirically whether reclassification from rural to urban had impacts on homeownership and other housing measures for households in those reclassified communities. In order to do this, I pool all four waves of household-level survey data and adopt a fixed-effects model, with the following regression equation:

$$y_{ijt} = \alpha_j + \rho_t + \beta Reclassification_{jt} + X'_{ijt}\gamma + \epsilon_{ijt} \tag{2.4}$$

where i denotes household, j denotes community and t denotes the survey year. The outcome variable is a dummy variable indicating whether the household is a homeowner. I also explore the effects of reclassification on multiple homeownership, measured by the number of housing units each household owns. α_j and ρ_t are community and year (survey wave) fixed effects. $Reclassification_{jt}$ is the key variable of interest, which is a dummy variable equal to one if Community j has been reclassified from rural to urban by year t. X_{ijt} is a vector a household-level control variables including household size, age of household head, whether the household head is married, number of children and elderly within the household and whether household has agricultural land. Summary statistics is presented in Table 2.22. Since I am comparing households in traditional rural villages and reclassified communities, I drop observations from urban communities as well as communities reclassified from urban to rural. Standard errors are clustered at the community level.

Since I am exploit the within-community variation in reclassification to distinguish its effect from confounding factors, I am in a difference-in-differences framework. Thus, my identifying assumption would be that in the absence of the reclassification, reclassified communities would have experienced changes in homeownership and other behaviors regarding living conditions and housing demand similar to non-reclassified villages.

One may wonder whether those reclassified communities shared different attributes from traditional rural villages even before they were reclassified, which could cause concerns in my identification strategy. My data allows me to test and relax this identifying assumption in several ways. First, I offer a formal statistical test by including an indicator for the two years prior to the reclassification in Equation 2.4. That is, I ask whether reclassified villages diverged in terms of household homeownership even before reclassification. If they do, it suggests that the identifying assumption of my research design is violated. Second, as has been discussed in previous subsections, the current standard of rural-urban dichotomy is simply based on land contiguity, independent of the social and economic development of the community. I can empirically test this by taking advantage of my community-level survey panel data. To be more specific, I use community-level demographic

Table 2.22: Housing Comparisons for Pooled Sample

	Urban	Rural	Rural to Urban	Urban to Rural
Willing to Buy	0.16	0.08	0.09	0.09
	(0.36)	(0.27)	(0.28)	(0.28)
Willing to Build	0.04	0.11	0.10	0.08
	(0.18)	(0.31)	(0.30)	(0.27)
Live in Own House	0.78	0.95	0.91	0.90
	(0.41)	(0.23)	(0.28)	(0.31)
Have Own House	0.87	0.96	0.94	0.93
	(0.33)	(0.19)	(0.23)	(0.25)
Housing Usable Size	96.94	126.86	138.72	121.19
	(81.13)	(94.91)	(101.17)	(93.09)
Self-Built House	0.30	0.87	0.86	0.71
	(0.46)	(0.34)	(0.35)	(0.45)
Collective-Owned Land	0.15	0.48	0.58	0.49
	(0.36)	(0.50)	(0.49)	(0.50)
Housing Cost	189,740.68	63,376.50	82,744.14	65,611.57
	(380,473.69)	(120,495.65)	(153,465.42)	(113,515.44)
Housing Fair Value	778,791.16	146,127.94	290,012.42	175,446.81
	(1,342,186.87)	(833,538.40)	(2,139,777.83)	(331,484.54)
Observations	68282	32520	2357	1076

Means and standard errors (in parentheses) are presented.

and economic variables to predict whether this village would be reclassified within the next survey period (i.e., two years). The dependent variable is an indicator that equals one if a village that had never been reclassified was reclassified within the next survey period[34]. Results are shown in Table 2.23. Column 1 only includes geographic, demographic variables and those proxy for economic development, which could have explanatory power for household homeownership and housing behaviors. The overall R-squared is low (only 0.008) and I fail to reject the hypothesis of the overall insignificance using the joint F-test, though some variables are statistically significant. This significance disappears in specifications which I gradually include county-level fixed effects, year fixed effects and their interaction terms. Across all specifications, the F-statistics are insignificant. As a result, I am assured that observable community-level characteristics that ought to reflect the social and economic development of a community are not correlated with its reclassification status.

Results from Equation 2.4 using homeownership dummy and number of housing units are shown in Table 2.24 and 2.25. Column 1 only includes community fixed effects and year fixed effects while Column 2 includes additionally household-level covariates to increase precision. Results remain unchanged with the coefficients or reclassification dummy both economically and statistically insignificant, indicating that the community reclassification from rural to urban does not have any impacts on household-level homeownership and multiple homeownership. The third column shows that the estimated coefficients remain small in magnitude and insignificance to replacing community fixed effects with household fixed effects. To further mitigate the concern for my identification strategy due to the possibility of differential time trends that would be correlated with the timing of reclassification, in columns 4 and 5, I show that the results are robust to controlling for specific time trends more flexibly. In Column 4, I allow communities in different provinces to have different time trends by including province-year fixed effects while Column 5 adopts interact all household-level characteristics with each year fixed effect. My main results for both outcome measures remain economically small and statistically insignificant after introducing

[34]Since the first community-level survey was conducted in 2013, I drop communities that had been reclassified before the year 2013 since I do not have any pre-period information on them. I also drop observations from the 2017 survey since I do not know their rural-urban dichotomy after 2017.

Table 2.23: Balanced Test

	(1)	(2)	(3)	(4)
Ln(Area)	-0.001	-0.002	-0.002	-0.002
	(0.001)	(0.002)	(0.002)	(0.004)
Number of Roads	-0.004	-0.003	-0.004	-0.006
	(0.006)	(0.008)	(0.008)	(0.014)
Agricultural Land	0.012**	-0.001	0.001	0.000
	(0.005)	(0.007)	(0.007)	(0.005)
Ln(Residents)	0.016**	0.011	0.014	0.020
	(0.007)	(0.010)	(0.012)	(0.017)
Ln(Households)	-0.019**	-0.005	-0.008	-0.010
	(0.010)	(0.014)	(0.016)	(0.021)
Ln(Migrants)	-0.002	-0.001	-0.001	0.001
	(0.002)	(0.003)	(0.003)	(0.004)
Poor Residents	0.008	-0.006	0.010	0.040
	(0.039)	(0.057)	(0.057)	(0.076)
Ln(Household Income)	0.005	-0.004	-0.005	-0.009
	(0.006)	(0.005)	(0.005)	(0.009)
County FE	No	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
County*Year FE	No	No	No	Yes
R^2	0.008	0.519	0.520	0.612
Observations	813	813	813	813

Robust standard errors are clustered at the county level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

several additional controls for differential time trends.

In Table 2.26, I show results including the dummy variable for the two years prior to the reclassification. Adding this indicator makes the omitted category to three or more years prior to treatment, which alters the interpretation of the reclassification dummy. However, the fact that both indicators are small and insignificant reflects that these two reclassified community follow similar trend with rural villages in terms of homeownership before and even after reclassification.

One may worry that the control group in my empirical model might not approximate the treatment group well if they fall into different regions of the country. I address this concern in two way. At the macro level, Figure 2.5 shows the regional distribution of the overall ratio of reclassified communities across provinces during the period 2009 throughout 2017. Though provinces along the coastal areas show somehow larger reclassification rate, provinces with the highest rates are in the central and western areas. Furthermore, neighbouring provinces with similar levels of economic development may show different reclassification behaviors. For example, *Shaanxi* Province is among the highest reclassification rate, while *Shanxi* Province, which is to the east of *Shaanxi* shows the lowest rate of reclassification, although both provinces share similar tradition and culture with similar economic development. Second, at the micro level, in Table 2.27, I show the baseline results are robust if I drop the observations from provinces that do not have any reclassified community.

Another potential issue of concern is attrition from the CHFS survey. In Table 2.28, I run the basic regression used to identify the role of reclassification on homeownership, equation 2.4, on attrition. The effect of reclassification on attrition is both small and statistically insignificant. There is therefore no evidence that reclassification would cause different behaviors in terms of attrition from the sample or household moving out of the community.

The rich information on housing from CHFS household survey questionnaire also allows me to empirically test the reclassification impacts on other housing measures. Table 2.29 shows corresponding results. Samples are restricted to households owning at least one housing unit. In Column 1, I regress number of rooms for the current housing unit they live in on the reclassifi-

Table 2.24: Effects of Village Reclassification: Homeownership

	(1)	(2)	(3)	(4)	(5)
Reclassified	-0.015	-0.015	-0.034	-0.017	-0.015
	(0.060)	(0.060)	(0.063)	(0.057)	(0.059)
Family Size		0.018***		0.018***	0.017***
		(0.001)		(0.001)	(0.003)
Female Head		0.004		0.003	0.006
		(0.005)		(0.005)	(0.014)
Head Age		-0.000***		-0.000***	-0.000
		(0.000)		(0.000)	(0.000)
Head Schooling		0.001		0.001	0.002
		(0.000)		(0.000)	(0.001)
Number of Children		-0.016***		-0.015***	-0.018***
		(0.002)		(0.002)	(0.004)
Number of Elderly		-0.016***		-0.016***	-0.019***
		(0.002)		(0.002)	(0.006)
Married		0.036***		0.035***	0.038***
		(0.006)		(0.006)	(0.010)
Agricutural Land		0.039***		0.039***	0.045***
		(0.004)		(0.004)	(0.008)
Community FE	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Household FE	No	No	Yes	No	No
Province*Year FE	No	No	No	Yes	No
Household Characteristics*Year Effects	No	No	No	No	Yes
R^2	0.058	0.093	0.003	0.097	0.094
Observations	34876	34789	34876	34789	34789

Robust standard errors are clustered at the community level.

R-squred reported in the specification with household fixed effects is the within R-suqred.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.25: Effects of Village Reclassification: Multiple Homeownership

	(1)	(2)	(3)	(4)	(5)
Reclassified	-0.042	-0.043	-0.040	-0.038	-0.043
	(0.086)	(0.086)	(0.097)	(0.086)	(0.085)
Family Size		0.061^{***}		0.061^{***}	0.050^{***}
		(0.003)		(0.003)	(0.005)
Female Head		0.017^{*}		0.017^{*}	0.023
		(0.010)		(0.010)	(0.019)
Head Age		0.001^{**}		0.001^{**}	0.002^{**}
		(0.000)		(0.000)	(0.001)
Head Schooling		0.008^{***}		0.008^{***}	0.010^{***}
		(0.001)		(0.001)	(0.002)
Number of Children		-0.039^{***}		-0.039^{***}	-0.025^{**}
		(0.005)		(0.005)	(0.010)
Number of Elderly		-0.039^{***}		-0.039^{***}	-0.037^{***}
		(0.005)		(0.005)	(0.012)
Married		0.034^{***}		0.034^{***}	0.038^{**}
		(0.010)		(0.010)	(0.019)
Agricutural Land		0.050^{***}		0.049^{***}	0.057^{***}
		(0.008)		(0.008)	(0.014)
Community FE	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Household FE	No	No	Yes	No	No
Province*Year FE	No	No	No	Yes	No
Household Characteristics*Year Effects	No	No	No	No	Yes
R^2	0.064	0.109	0.001	0.113	0.111
Observations	32706	32622	32706	32622	32622

Robust standard errors are clustered at the community level.

R-squred reported in the specification with household fixed effects is the within R-suqred.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.26: Effects of Village Reclassification: Add Pre-Treatment Indicator

	Homeownership		Multiple Homeownership	
	(1)	(2)	(3)	(4)
Reclassified	-0.010	-0.006	0.029	0.039
	(0.037)	(0.036)	(0.053)	(0.053)
1-2 Years Before	0.007	0.012	0.097	0.112
	(0.044)	(0.043)	(0.080)	(0.079)
Community FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Household Characteristics	No	Yes	No	Yes
R^2	0.058	0.093	0.064	0.109
Observations	34876	34789	32706	32622

Robust standard errors are clustered at the community level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

cation dummy variable, controlling for household-level characteristics, community fixed effects and year fixed effects. The coefficient of reclassification indicator is around 0.4 and insignificant. Next, I use logarithm of housing size as dependent variable in Column 2, the coefficient estimate is negative and also insignificant, indicating that community reclassification does not to improve the quantity of housing consumption. One character of housing behavior in rural China is that households often build their houses on assigned land (*ZhaiJiDi*) rather than purchase from the housing markets. In Column 3, I find that the coefficient is around 0 in magnitude.

Starting with the third wave in 2015, the questionnaire also included questions regarding housing demand. From both waves in 2015 and 2017, I distinguished their demand as either self-built or purchasing from housing markets. In 2015, the ratio of households that are willing to improve their living conditions by having more houses was similar in all three categories (rural, urban and reclassified), around 20%. While most urban residents would prefer to purchase apartments, more than half of reclassified villages would like to build housing units themselves. This trend was still obvious in the 2017 wave; this indicates that although noticeable housing demands exist, the major-

Table 2.27: Effects of Village Reclassification: Drop Provinces w/o Reclassified Villages

	Homeownership		Multiple Homeownership	
	(1)	(2)	(3)	(4)
Reclassified	-0.016	-0.016	-0.048	-0.049
	(0.060)	(0.060)	(0.086)	(0.087)
Family Size		0.019***		0.066***
		(0.001)		(0.003)
Female Head		0.008		0.023**
		(0.005)		(0.011)
Head Age		-0.000***		0.001**
		(0.000)		(0.000)
Head Schooling		0.001		0.008***
		(0.000)		(0.001)
Number of Children		-0.017***		-0.043***
		(0.002)		(0.005)
Number of Elderly		-0.018***		-0.044***
		(0.002)		(0.005)
Married		0.034***		0.025**
		(0.007)		(0.011)
Agricutural Land		0.037***		0.039***
		(0.005)		(0.009)
Community FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.061	0.096	0.061	0.111
Observations	27294	27223	25424	25355

Robust standard errors are clustered at the community level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.28: Effects of Village Reclassification: Attrition from Survey

	(1)	(2)	(3)
Reclassified	-0.048	-0.049	0.008
	(0.232)	(0.234)	(0.007)
Family Size		-0.005***	
		(0.002)	
Female Head		0.010	
		(0.007)	
Head Age		-0.001***	
		(0.000)	
Head Schooling		-0.001*	
		(0.001)	
Number of Children		0.001	
		(0.003)	
Number of Elderly		0.002	
		(0.003)	
Married		-0.049***	
		(0.007)	
Agricutural Land		-0.016***	
		(0.006)	
Community FE	Yes	Yes	No
Year FE	Yes	Yes	Yes
Household FE	No	No	Yes
R^2	0.360	0.363	0.080
Observations	21393	21364	21393

Robust standard errors are clustered at the community level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

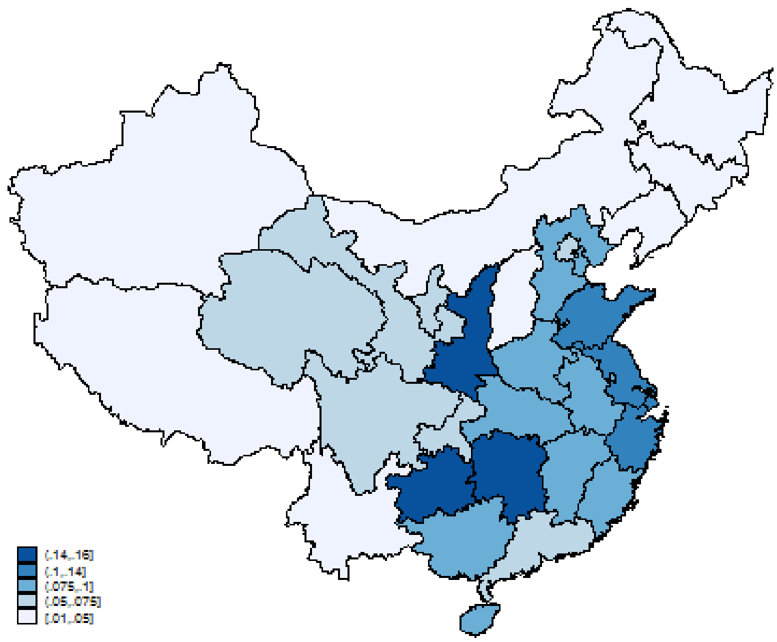
Table 2.29: Effects of Village Reclassification: Alternative Housing Measures

	(1)	(2)	(3)	(4)
	No.Room	Ln(Size)	Self-Built	Willing to Buy
Reclassified	0.400	-0.030	-0.000	-0.054
	(0.556)	(0.030)	(0.037)	(0.033)
Family Size	0.466*	0.055***	0.009***	0.016***
	(0.242)	(0.004)	(0.002)	(0.002)
Female Head	-0.303	0.022	-0.004	0.006
	(0.377)	(0.015)	(0.008)	(0.007)
Head Age	0.093	-0.001**	0.002***	-0.002***
	(0.093)	(0.001)	(0.000)	(0.000)
Head Schooling	-0.122	0.009***	0.001	0.004***
	(0.167)	(0.002)	(0.001)	(0.001)
Number of Children	0.188	-0.032***	-0.012***	-0.012***
	(0.310)	(0.007)	(0.004)	(0.004)
Number of Elderly	-0.198	-0.038***	0.000	-0.012***
	(0.128)	(0.008)	(0.003)	(0.003)
Married	2.748	0.138***	0.017**	-0.002
	(2.513)	(0.017)	(0.008)	(0.006)
Agricutural Land	0.057	0.037***	0.026***	0.005
	(0.090)	(0.011)	(0.006)	(0.007)
Community FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.027	0.225	0.154	0.089
Observations	31823	30632	31875	17244

Robust standard errors are clustered at the community level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.5: Geographic Distribution of the Reclassified Communities



ity of households from reclassified communities still prefer to build houses rather than participate in the local housing market. In Column 4 of Table 2.29, I show that households in reclassified communities are no more willing to participate in more formal housing markets; i.e., they are no more involved in purchasing a unit compared with households in rural villages.

2.3.3 Community Reclassification and Household Welfare

In the previous subsection, I fail to find any effects of community reclassification from rural to urban on any improvement on household housing behavior. Nevertheless, households living in those communities are never regarded as rural citizens in any statistical way. I am interested in the question that in what aspects the reclassification affects households within the community. Did it affect household-level income and consumption that might not be revealed from housing behavior? If this urbanization lead to more local job opportunities, then this could translate to more welfare gain. To investigate this, I use the household-level survey in all four waves of CHFS. Each survey has a detailed income and consumption module that allows for calculation of different sources of income and consumption. Regression results using them as dependent variables are shown in Table

2.30.

In Column 1, I regress the logarithm of household income per capita on the reclassification indicator, household-level covariates as well as the community fixed effects and year fixed effects. Though the coefficient estimates of the reclassification dummy is positive with 0.109 in magnitude, it is not statistically different from zero. In addition, I do not observe a statistically significant effect on overall consumption as well as consumption of non-food items. All results combined, I fail to detect any effect of the community reclassification on improving welfare for the local households.

2.3.4 Long-Run Reclassification Effects

One may think that after being reclassified as urban, those communities would follow a differential development trajectory than their rural counterparts. For example, reclassified communities might experience more investments in the infrastructures from upper-level governments, which might take years. Thus, it is also important to discover whether there exists any long-run reclassification effects for the reclassified villages.

In Table 2.31, we show household characteristics in both rural villages and communities reclassified from rural to urban across years. As the first year we can identify reclassified communities is 2010 and our last wave of survey measures household behaviors in 2017, it has been at least six years for the first cohort of communities since reclassification. However, we fail to discover any noticable differences in all household-level characteristics.

In terms of housing behavior, Figure 2.6 shows the dynamic reclassification effects on household homeownership and multiple homeownership from a regression framework. The outcome variable is either the homeownership dummy or number of housing units. The key explanatory variables are a series of indicators for 1 or 2 years before reclassification, the year of reclassification, 1 or 2 years after reclassification and 3 or more years after reclassification. The omitted category is 3 or more years before reclassification. We also include community and year fixed effects. The red dots indicate the coefficients of estimate while the 95 percent confidence intervals are indicated by lines. In line with the descriptive evidence, we fail to find any long-run effects of reclassification on household housing behaviors.

Table 2.30: Effects of Village Reclassification: Welfare

	(1) Ln(Income pc)	(2) Ln(Consumption pc)	(3) Ln(Non-Food Consumption pc)
Reclassified	0.109	0.009	-0.046
	(0.103)	(0.075)	(0.102)
Female Head	-0.003	0.041**	0.032
	(0.026)	(0.016)	(0.020)
Head Age	-0.006***	-0.007***	-0.009***
	(0.001)	(0.001)	(0.001)
Head Schooling	0.047***	0.024***	0.027***
	(0.002)	(0.002)	(0.002)
Number of Children	-0.134***	-0.140***	-0.137***
	(0.009)	(0.006)	(0.006)
Number of Elderly	-0.164***	-0.097***	-0.106***
	(0.011)	(0.007)	(0.008)
Married	0.065**	-0.007	0.099***
	(0.026)	(0.017)	(0.020)
Agricutural Land	0.143***	-0.004	0.057***
	(0.024)	(0.013)	(0.017)
Community FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R^2	0.184	0.219	0.185
Observations	33627	34786	34682

Robust standard errors are clustered at the community level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.31: Household Comparisons for CHFS 2017

	Rural	2010	2011	2012	2013	2014	2015	2016	2017
Family Size	3.55	3.92	3.53	3.84	3.80	4.11	3.29	3.23	2.90
	(1.77)	(1.92)	(1.86)	(1.77)	(1.76)	(1.88)	(1.51)	(1.70)	(1.37)
Number of Children	0.62	0.85	0.64	0.76	0.78	0.89	0.41	0.48	0.30
	(0.92)	(0.97)	(0.94)	(0.97)	(0.98)	(1.16)	(0.64)	(0.76)	(0.69)
Head Schooling	6.96	7.50	7.09	7.36	7.62	4.88	7.78	7.40	6.75
	(3.46)	(3.46)	(3.66)	(3.25)	(3.40)	(4.33)	(3.68)	(3.55)	(4.11)
Head Urban Hukou	0.03	0.08	0.04	0.04	0.03	0.05	0.00	0.03	0.31
	(0.18)	(0.27)	(0.19)	(0.21)	(0.17)	(0.21)	(0.00)	(0.18)	(0.46)
Ag Activity	0.75	0.58	0.70	0.55	0.65	0.13	0.44	0.39	0.54
	(0.43)	(0.49)	(0.46)	(0.50)	(0.48)	(0.33)	(0.50)	(0.49)	(0.50)
Agricultural Land	0.80	0.59	0.79	0.64	0.69	0.41	0.65	0.58	0.56
	(0.40)	(0.49)	(0.41)	(0.48)	(0.47)	(0.50)	(0.48)	(0.50)	(0.50)
Total Income	45,227.48	49,469.86	41,368.92	63,153.77	45,783.76	35,410.72	78,264.57	45,560.84	60,679.98
	(73,563.44)	(59,060.08)	(56,717.22)	(93,987.32)	(90,011.87)	(33,782.32)	(93,935.51)	(61,929.20)	(64,604.45)
Total Consumption	37,091.12	44,165.32	32,322.48	58,397.08	61,332.89	58,984.09	54,761.42	37,124.31	38,896.41
	(38,198.78)	(42,957.02)	(31,324.77)	(50,308.48)	(71,967.49)	(37,473.43)	(51,325.93)	(41,200.22)	(30,678.66)
Observations	11425	142	325	135	65	64	85	62	91

Means and standard errors (in parentheses) are presented.

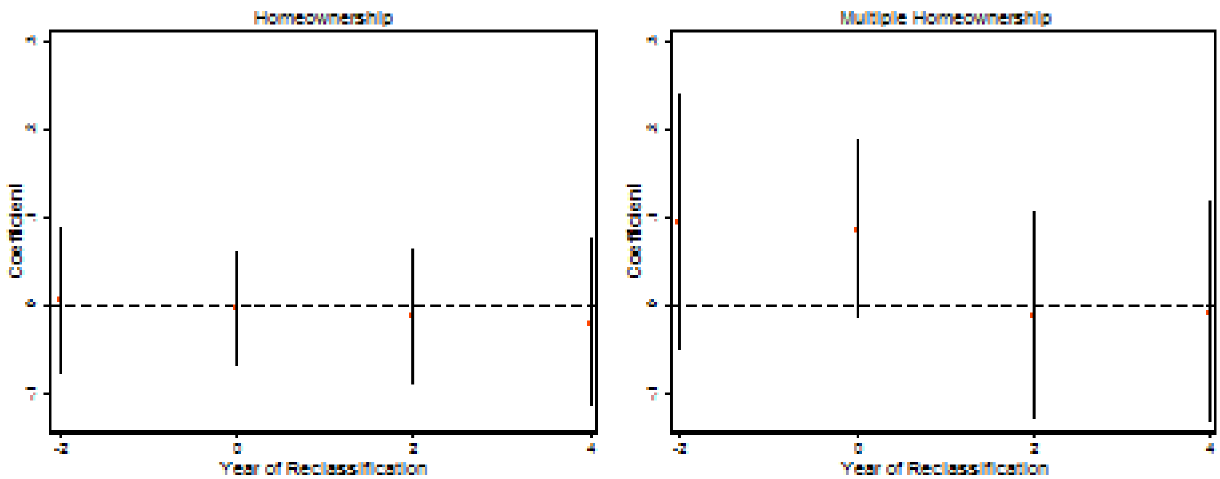
To summarize, I first descriptively demonstrate that reclassified communities are different from urban communities not only in community characteristics but also in household housing behaviors and demand. In all of these aspects, reclassified areas are identical to rural villages. In addition to this cross-sectional comparison, I empirically show that being reclassified does not have impacts on the development trajectory for the corresponding communities; they still followed the trend that resembles rural villages. Both horizontally and vertically, I fail to detect any difference between reclassified and rural villages.

2.4 In Situ Migrants and Housing Markets

2.4.1 Urban Population as a Determinant of Land Supply

In China, according to the Constitution under the “Two-Tier Land Tenure System”, urban land is owned by the state, and rural land is owned by local collective communes ([33]). In fact, local governments have de facto control over city planning and land use. As a result, the fundamental restriction on China’s housing supply is that the land supply is controlled by government ([20]). The amount of supplied land is based on city planning, land quota, economic activity, population, income and expenditures, etc. Though there are policies that restrict the ratio of each type (e.g.,

Figure 2.6: Dynamic Reclassification Effects

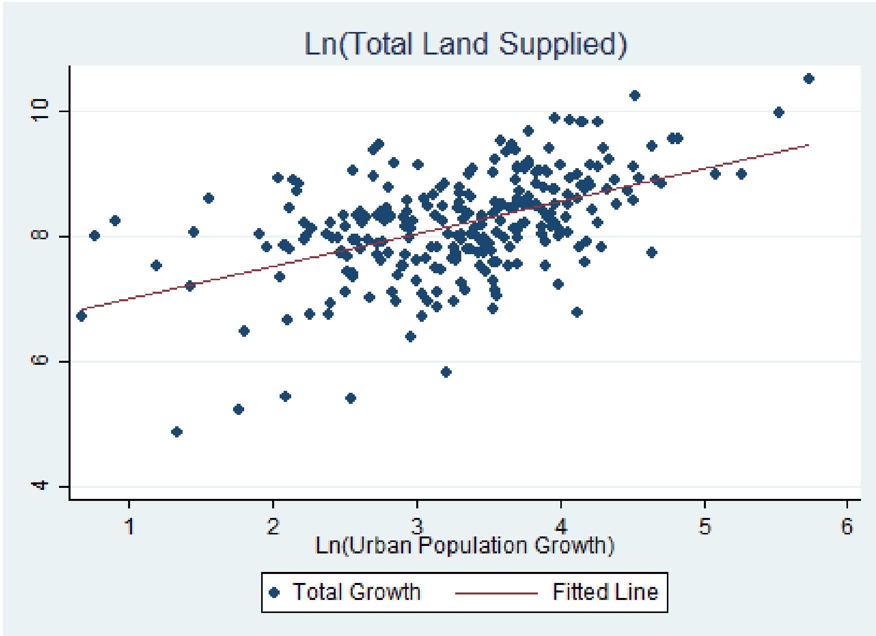


industrial, commercial, residential) of land supplied, the local government has discretion to some extent on allocating shares of land on different types of use. According to [37] and [20], after government acquired all land from rural farms and established a land reserve center for management, approximately 20%-30% percent of land reserves were transferred for residential use and offered to the markets, mostly through public auctions. Additionally, around 50% was transferred to investors for industrial use, mostly through negotiation. The price for industrial use was much lower than for residential use. In order to develop the local economy and balance their budgets, local governments would attract outside investments by increasing the supply of industrial land with lower price. At the same time, they tend to supply less residential land in order to inflate the average unit price to finance government expenditures.

In order to show how in situ migrants lead to oversupply in the local land and housing markets, I aggregate my community reclassification information at the prefecture level and link it to the land and population data sets. Data on land comes from the *China Urban Construction Statistical Yearbooks*. Figure 2.7 graphically shows a strong positive relationship between land supply and urban population growth.

In terms of the land supply for residential use, to empirically show whether local governments

Figure 2.7: Urban Population Growth and Total Land Supply



consider urban population growth as one of the key determinants on the amount of residential land supplied in the primary land market, I adopt the following OLS fixed-effects panel data model using prefecture-year observations for all prefectures which I can obtain urban population data from 2010-2015. Specifically, the regression equation is:

$$Residential\_Land_{j,t} = \alpha_j + \rho_t + \beta Urban\_Population_{j,t-1} + X'_{j,t-1}\gamma + \epsilon_{jt} \tag{2.5}$$

where $Residential\_Land_{j,t}$ is the total land for residential use in prefecture j as of year t while $Urban\_Population_{j,t-1}$ the one-year lagged measure for the urban population in prefecture j. Both variables are logged so the coefficient β can be interpreted as the elasticity of residential land with respect to urban population. I control for the prefecture fixed effects and year fixed effects as well as a list of one-year lagged prefecture-level time-varying covariates including the prefecture area, per capita GDP, per capita fiscal income, per capita size of paved roads and green spaces. Standard errors are clustered at the prefecture level.

Regression results are shown in Table 2.32. Column 1 presents the results only controlling for the prefecture and year fixed effects. The elasticity of residential land with respect to urban population is large in magnitude and statistically significant. A one percent increase in urban population is associated with 0.43 percent increase in residential land in the next year, indicating that total urban population growth is a strong predictor of residential land supply. In Column 2, I add the lagged prefecture-level covariates as in my baseline specification in Equation 2.5, the coefficient increases to 0.52 and still highly significant. In the next column, I further include the region-by-year fixed effects[35] to account for potential differential changes in different regions due to the central government policies in favor of less developed central and western provinces, the coefficient is almost unchanged. Lastly, in Column 4, I allow each prefecture to have its own linear trend; again, the elasticity is high both economically and statistically. As expected and in line with the urban planning literature in China, the growth in urban population is a key determinant on the amount of residential land supplied within the prefecture.

In addition, it is interesting to explore whether this elasticity is heterogeneous across prefectures with different attributes. While it is plausible to categorize different prefectures by different dimensions, I would divide prefectures into three tiers, which has been widely used officially and unofficially. To be more specific, I am dealing with this potential heterogeneity between the second-tier and third-tier cities[36]. Second-tier cities consist of most provincial capitals and municipalities that are local economic and political centers while the remaining prefectures are considered the third-tier. Table 2.33 shows the results for prefectures from the two tiers. Unlike Tier 3 cities which are relatively less developed, for tier 2 prefectures, residential land supply is inelastic to urban population growth. The different sensitivities of land supply to population growth might reflect the different weights of urban population in the urban planning process. Since the second-tier prefectures are usually provincial capitals or local economic centers, they tend to attract population from nearby less developed third-tier prefectures on a fixed bases. As a result, urban population is less predictive of land supply in these prefectures.

[35] I divide the country into three regions: East, Central and West.

[36] The first-tier cities include *Beijing*, *Shanghai*, Guangzhou and *Shenzhen*, two provincial-level municipalities and two prefecture-level cities in the coastal *Guangdong* province. These four cities have a large urban population and enjoy a large wave of immigrants, with very different policies in the land and housing markets.

Table 2.32: Urban Population and Residential Land

	(1)	(2)	(3)	(4)
Ln(Urban Population)	0.430***	0.521***	0.562***	0.625**
	(0.147)	(0.171)	(0.154)	(0.281)
Ln(City Size)		0.125***	0.125***	0.008
		(0.044)	(0.042)	(0.037)
Ln(Road Size pc)		0.282	0.284	0.434
		(0.236)	(0.220)	(0.287)
Ln(Greenland Size pc)		-0.040	-0.047	-0.026
		(0.065)	(0.070)	(0.041)
Ln(GDP pc)		-0.117	-0.126	-0.156
		(0.101)	(0.101)	(0.165)
Ln(Fiscal Income pc)		-0.011	-0.005	-0.004
		(0.012)	(0.011)	(0.007)
Prefecture FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Region by Year FE	No	No	Yes	No
Prefecture Linear Trend	No	No	No	Yes
R^2	0.093	0.126	0.142	0.631
Observations	1560	1437	1437	1437

Standard errors are clustered at the prefecture level.

Reported R-squres are the within R-squares from prefecture fixed effects regressions

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.33: Urban Population and Residential Land: by Prefecture Tiers

	Tier 2-3	Tier 2	Tier 3
Ln(Urban Population)	0.465***	0.195	0.509***
	(0.167)	(0.256)	(0.174)
Ln(City Size)	0.117***	0.038	0.122**
	(0.044)	(0.072)	(0.049)
Ln(Road Size pc)	0.267	0.067	0.295
	(0.236)	(0.084)	(0.266)
Ln(Greenland Size pc)	-0.043	-0.042	-0.045
	(0.065)	(0.077)	(0.072)
Ln(GDP pc)	-0.047	-0.037	-0.050
	(0.062)	(0.066)	(0.085)
Ln(Fiscal Income pc)	-0.014	0.203	-0.015
	(0.012)	(0.128)	(0.013)
Prefecture FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R^2	0.134	0.456	0.124
Observations	1419	225	1194

Standard errors are clustered at the prefecture level.

Reported R-squres are the within R-squares from prefecture fixed effects regressions

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.4.2 In Situ Migrants and Residential Land Supply

The fact that urban population growth is a strong determinant of local land supply for residential use is expected. However, different prefectures might have different compositions of urban population growth. For example, in metropolises, migration both from rural areas within the prefecture and from outside the prefecture might account for the majority of urban population growth while in some small and remote prefectures in situ migrants might dominate. As a result, it is of importance to find whether the elasticity of land supply with respect to urban population differs across different sources. I shed light on this question by regressing residential land supply on different sources of urban population growth.

I define nominal urban population growth as the difference in the total urban population as of 2015 and 2010 within a prefecture. I estimate the prefecture-level scale of in situ migrants by using the same approach when I estimate the overall scale nationwide. By assuming that the mean of residents among all rural villages in a prefecture is similar to that of reclassified communities in the same prefecture (which has been supported by descriptive evidence using community-level data in Section2.3[37]), I am able to estimate the scale of in situ migrants by multiplying the prefecture's rural population in 2010 by the share of reclassified communities. I then define real urban population growth by subtracting the scale of in situ migrants from nominal urban population growth.

The regression equation is as follows:

$$y_{c,p,\Delta t} = \beta_0 + \beta_1 * Real\_UPG_{c,p,\Delta t} + \beta_2 * In\_Situ_{c,p,\Delta t} + X'_{c,p,t_0}\gamma + \phi_p + \epsilon_{c,p,\Delta t} \tag{2.6}$$

where $y_{c,p,\Delta t}$ measures the land supply of prefecture c in province p during time period Δt; $Real\_UPG_{c,p,\Delta t}$ and $In\_Situ\_Migrant_{c,p,\Delta t}$ measure the real urban population growth and the scale of in situ migrants respectively. X_{c,p,t_0} is a list of control variables similar to those in Equation 2.5 but they are measured as of the base year (2010). I also add the level of residential land

[37] Actually, in summary statistics from all waves of community survey, the number of residents in reclassified communities is slightly larger than that in rural villages.

as a control variable. ϕ_p controls for province fixed effects[38]. Standard errors are clustered at the province level.

Table 2.34 shows the results of the relationship between different sources of urban population growth and residential land supply. In Column 1, I only include the total urban population growth, an overall measure. Similar to my results using the prefecture-level panel data set, the coefficient which can be interpreted as elasticity is 0.58, large and highly significantly different from zero, implying a strong relationship between urban population growth and residential land supply. In the next two columns, I add my measure of real urban population growth and in situ migrants respectively. Both measures are of similar size and significant. In the last column, I include both measures in one regression. Both coefficients are qualitatively and quantitatively unchanged. While this result is suggestive, it shows that the elasticity of residential land supply with respect urban population from in siut migrants is almost as large as predictive as that with respect to real urban population growth, the group of which is thought to have real demand for the urban housing markets.

In addition, I divide the cities into different tiers and estimate Equation 2.6 separately. While residential land supply is inelastic to urban population growth in Tier-2 cities, it is very sensitive to both real urban population growth and in situ migrants in smaller Tier-3 cities, consistent with what I find from the panel data regressions.

[38]For the four province-level municipalities, *Beijing, Tianjin, Shanghai* and *Chongqing*, the results are virtually unchanged if these four observations are dropped.

Table 2.34: Urban Population Growth and Annual Residential Land Supply

	(1)	(2)	(3)	(4)
Ln(Nominal UPG)	0.580***			
	(0.103)			
Ln(Real UPG)		0.311***		0.301***
		(0.048)		(0.047)
Ln(In Situ Migrant)			0.218***	0.249***
			(0.040)	(0.049)
Ln(Residential Size pc)	0.149	0.015	-0.011	0.038
	(0.094)	(0.155)	(0.124)	(0.128)
Ln(GDP pc)	0.464***	0.426*	0.502**	0.510*
	(0.164)	(0.221)	(0.219)	(0.263)
Ln(Fiscal Income pc)	-0.192	-0.153	-0.098	-0.236
	(0.148)	(0.163)	(0.159)	(0.207)
Ln(Road Size pc)	0.137	0.237*	0.220	0.265
	(0.135)	(0.130)	(0.166)	(0.176)
Ln(Green Size pc)	-0.170**	-0.218*	-0.227*	-0.192
	(0.081)	(0.121)	(0.127)	(0.125)
Province FE	Yes	Yes	Yes	Yes
R^2	0.688	0.610	0.575	0.678
Observations	248	226	239	194

Standard errors are clustered at the province level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.35: Urban Population Growth and Annual Residential Land Supply

	Tier 2-3	Tier 2	Tier 3
Ln(Real UPG)	0.307***	0.667	0.218***
	(0.054)	(1.617)	(0.061)
Ln(In Situ Migrant)	0.247***	-0.150	0.226***
	(0.048)	(0.931)	(0.054)
Ln(Residential Size pc)	0.032	3.953***	-0.064
	(0.130)	(0.624)	(0.122)
Ln(GDP pc)	0.491*	-0.391	0.326
	(0.260)	(2.986)	(0.214)
Ln(Fiscal Income pc)	-0.208	0.404	-0.172
	(0.210)	(2.634)	(0.200)
Ln(Road Size pc)	0.266	-0.969	0.291
	(0.176)	(0.857)	(0.186)
Ln(Green Size pc)	-0.189	-1.768**	-0.202
	(0.135)	(0.706)	(0.150)
Province FE	Yes	Yes	Yes
R^2	0.680	0.960	0.632
Observations	192	28	164

Standard errors are clustered at the province level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3. TENURE SECURITY AND AGRICULTURAL LAND UTILIZATION: EVIDENCE FROM CHINA

3.1 Background

3.1.1 Land Tenure System in China

After the foundation of the People's Republic of China in 1949, all means of production were gradually transferred to public ownership. Land, which serves as the most important input in agricultural production, was no exception. In most villages at that time, land was mainly classified into two types in terms of managerial rights: private plots and collectively-owned plots. During the People's Commune period from early 1960s to late 1970s, although all land was collectively owned, commune members were allowed to manage their own private land for some extra agricultural activities, such as raising livestocks. Private plots only accounted for very small proportions of the total arable land and they were not allowed for transfer. The Regulations on the Work of the Rural People's Commune implemented in 1962 by the Central Committee of Communist Party of China (CCCPC) stated that the size of all private plots managed by commune members should be around 5% to 7% of the size of total arable land under the control of the commune.

With the introduction of the Household Responsibility System (HRS) in the early 1980s, the property rights of land were divided into ownership right and managerial right. On one hand, land is still collectively owned; on the other hand, households are allocated an amount of land, mainly on the egalitarian basis and are entitled the managerial rights on the land with residual claimant to outputs. They were subsequently provided a 15-year land use right. This reform greatly encouraged farmers' enthusiasm, resulting in tremendous increase in both outputs and productivity. This success prompted a recommendation to renew contracts for additional 30 years upon expiration of the original 15-year lease in the late 1990s ([17])[1].

Even though the lease contract has been expanded, there still exist institutional risks threatening

[1]This renew is generally called second-round contract

the land tenure security. Rural land in contract is still subject to administrative redistribution by village leaders or compulsory expropriation by local governments. Due to demographic changes over time, total or partial contractual land would be redistributed across households within the village. This reallocation tends to have considerable administrative support.

Another threat to land tenure security is through expropriation. Agricultural land is requisite by local government and then sold for non-agricultural use. There are compensation standards for land being expropriated, but implementations vary. Although such phenomenon is much alleviated by the introduction of the new *Rural Land Contracting Law* in 2003 ([17]), it still remains one major risk for land tenure security.

Land rental activities were prohibited in the early phases of HRS, ([17]), nor were households allowed to use land as collaterals. Although such restriction was loosened later, the insecure land tenure still restricts development of the land rental market, as households who would like to be involved in land rental activities are worried about their land being expropriated or required for readjustment. For example, renting out land by a household could be interpreted as a signal that land is not needed any more, which might lead to a range of land readjustment. This is argued to create disincentives for free migration of villagers from agricultural sector to non-farm activities, both locally and to the city ([49]). [22] show that in response to an expected land reallocation in the following year, the probability that a rural resident migrates out of the county declines by 2.8 percentage points, almost 17.5 percent of the annual share of village residents who worked as migrants. However, as wage difference between agricultural and non-agricultural sector is large and still increasing, many rural laborers migrate to city for work. In the absence of a secure land tenure and an well-functioning land transfer market, households with insufficient agricultural labors might only leave their land abandoned, leading to the underutilization of agricultural land.

3.1.2 The Rural Land Certification Program

The establishment of land transferring system was first raised in a document about stabilizing and improving the rural land contract by the State Council in 1995. It stated that under the condition that the ownership and the use of the land are unchanged, the contractor may rent, swap and use land as stock shares, with the consent of the rural collectivity. The first time that the rural land certificate, formally the certificates of the right to contracted management of rural land, was mentioned was from another State Council document about further stabilizing and improving the rural land contract in 1997. This document formally emphasized that the contract was expanded for another 30 years; more importantly, it stated that for households that expanded their leasing period, they should be issued certificates of the right to contracted management of rural land by the local government at the county level or above in a timely manner. These reforms were finally legislated into the new *Rural Land Contracting Law*, which required responsibility of the local government for the issuance of the rural land certificate and transferring of land managerial rights. The Ministry of Agriculture then detailed all information that are required in the certificate, including land contract date and the size and boundaries of each parcel. However, the implementation varied across villages and the process was slow.

It was not until the year 2009 that the No.1 policy document of the Central Committee of the Communist Party and the State Council reinforced the stability of rural land contract as a basis of stabilizing agricultural institutions in rural areas and proposed the development of land certificate program in some experimental pilots all over the country. It also prohibited reallocating land during the process of certification. The Ministry of Agriculture chose villages from 8 provinces as pilots to restart the land certification program. After several years' trial, the 2013 No.1 policy document developed this program nationwide and set the plan to finish certifying rural land in all villages within five years. In 2014, three provinces (*Anhui, Shandong* and *Sichuan*) were chosen to develop the program in all villages. As of March 2016, around 20% of rural arable land had been finished certifying.

This round of land certification program can be seen as the extension of the policy back in mid-1990s, since their targets are the same as to provide land contract households with a certificate that helps secure land tenure. The differences lie in that the land certifying program in recent years were implemented by the upper-level governments with more administrative power. In addition, measurement of specific parcel information is more accurate by using electronic tools while all information collected would be stored electronically in a public dataset platform supported by the State Archives Administration and will be linked to the registration of the real estate property.

3.2 Data

3.2.1 Data Source

Data used in this paper are from the China Household Finance Survey (CHFS), conducted by the Survey and Research Center for China Household Finance at the Southwestern University of Finance and Economics. So far the project has collected data in 4 waves: 2011, 2013, 2015 and 2017. The first wave, in the summer of 2011, covered 80 counties in 25 provinces and collected data from 8,438 households all over 320 communities. The second wave in 2013 enlarged its sample to 28,143 households in 1,048 communities across 29 provincial-level regions [2]. The 2015 wave and 2017 wave consist of 37,341 households from 1,373 communities and 40,011 households from 1,428 communities respectively. The samples from the most recent two waves are representative at both province and deputy-provincial city levels.

The sampling of communities in CHFS data follows a stratified method with probability proportionate to size. In the first step, all counties are divided into ten categories with respect to their rankings in GDP per capita, from which sample counties are drawn. Then sample communities are randomly drawn from these sample counties.

With all waves combined, the dataset is a nationally representative panel containing rich micro-level information on households' balance sheet, income, consumptions, and saving behaviors. Besides, starting from the second wave, it also included a community-level survey for all sampled communities, the questionnaire of which consists of questions regarding the basic demographic and geographic information, public infrastructure, community economy and local governance at the community level. For rural villages, it also asks detailed information on agricultural activities. Moreover, the community survey of CHFS 2017 asked each community with rural land whether agricultural land in the village had been certified and if so, the specific year that the certification

[2] There are 31 provincial-level regions in mainland China. The only two that are not included in the sample are *Xinjiang* and *Tibet*, two Minority Autonomous Regions in western China. The agriculture sector in these two regions is mainly animal husbandry, where arable land is limited.

was finished and the method of certification, from which I can back out the information whether cultivated land in the village had been certified in previous waves.

As a result, this paper exploits data from the 2015 and 2017 surveys, both the household and community levels, as these two waves have a larger sample size and I could construct a panel with detailed information with respect to agricultural activities. The land abandonment variables are available in both waves of community survey but only available in the 2017 household survey. In consequence, I employ the community-level panel data for our main results while use the 2017 household-level cross-sectional data as a check for robustness. Additionally, the two-period household-level panel dataset provides plenty of information that allows me to test possible mechanisms through which the land tenure affects land utilization. Such explanations include agricultural investments, land transfer activities and labor migration, illustrated by the related literature.

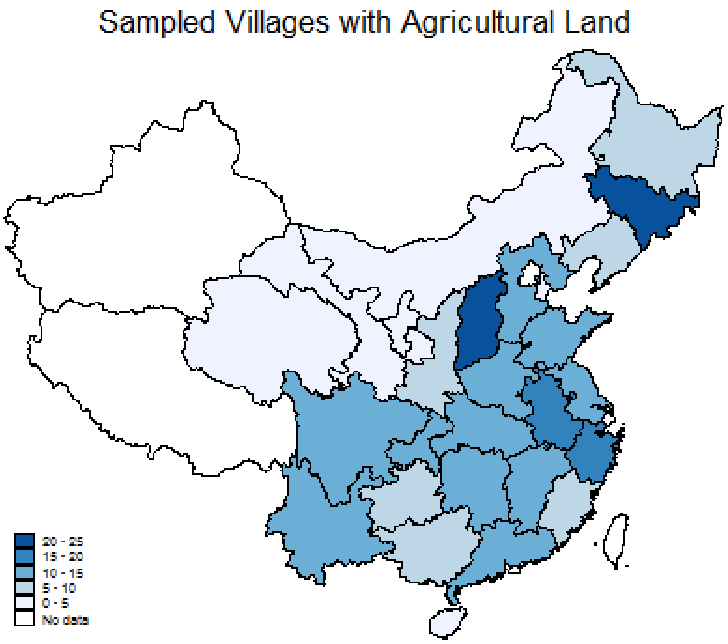
3.2.2 Descriptive Statistics

There are 1,362 observations in the CHFS 2015 community survey while the CHFS 2017 survey consists of 1,417 communities. I combine rural villages with agricultural land that are in both waves and construct a panel dataset. Figure 3.1 shows the distribution of villages across provinces. As can be seen, I have data of sampled villages from almost every province while in most provinces, I have at least ten villages, indicating the representativeness of the dataset. After appending the two waves' community survey data into a panel, I find for some communities, there were large changes in terms of the size of total agricultural land, which is abnormal given the fixed size of total village area. I thus drop these extreme observations in the empirical analysis[3]. In the analysis using community level panel data, my final sample size is 198 villages every year.

Table 3.1 and 3.2 present demographic and geographic information for rural communities in my balanced panel dataset in CHFS 2015 and 2017, respectively. Geographically, of those with agricultural land, they are on average over 70 kilometers away from the center of the city with less than 2 roads toward. For community-level demographics, a typical village has more than 600

[3]Specifically, I calculate the ratio of agricultural land size across the two waves and drop villages with the ratio less than 0.75 or greater than 1.25. Results are qualitatively similar for alternative thresholds.

Figure 3.1: Rural Communities with Agricultural Land



households; number of residents[4] are around 2,106 in CHFS 2015 and 2,066 in CHFS 2017. Number of registered residents are also quite stable across years, from 2,431 in 2015 to around 2,336 in 2017. Over 95% of registered residents hold rural Hukou, with very little variation between the two survey periods. The difference between number of residents and Hukou residents indicate the net emigration from rural villages to city. Moreover migrant workers from these communities account for 23% of total registered residents. Percentage of communities with Internet connection increased from 86% in 2015 to 92% in 2017. With respect to income, data suggest that the average disposable income per capita across those communities increased by more than 1,000 yuan during the two-year period. Table 3.1 and 3.2 also show the different characteristics for villages that have land certified and that have not, in both waves. Most variables are quite identical across these two types of villages and stable over years. However, in terms of demographics, certified villages tend to have less households and residents while the disposable income is on average higher.

In terms of information regarding agricultural land, Table 3.3 summarizes community-level

[4]Residents include those who hold a local Hukou and live in the local community and those who do not have a local Hukou but live in the local community for more than 6 months.

Table 3.1: Descriptive Statistics for Communities in CHFS 2015

	Overall	Certi=0	Certi=1
Distance to Center	78.72	78.34	81.11
	(64.94)	(66.52)	(55.07)
Roads to Center	1.42	1.42	1.43
	(0.63)	(0.63)	(0.63)
Households	618.03	618.39	615.86
	(418.95)	(420.39)	(417.70)
Residents	2,106.05	2,165.04	1,747.89
	(1,844.32)	(1,895.79)	(1,472.34)
Hukou Residents	2,431.04	2,495.01	2,044.93
	(2,113.48)	(2,206.33)	(1,398.74)
Ag-Hukou Ratio	0.96	0.96	0.95
	(0.14)	(0.14)	(0.14)
Migrant Ratio	0.23	0.23	0.23
	(0.17)	(0.17)	(0.15)
Internet	0.86	0.85	0.93
	(0.35)	(0.36)	(0.26)
Featured Industry	0.44	0.42	0.57
	(0.50)	(0.50)	(0.50)
Disposable Income	6,991.61	6,785.11	8,245.36
	(6,249.67)	(5,974.37)	(7,724.68)
Dibao Ratio	0.05	0.04	0.06
	(0.08)	(0.05)	(0.19)
Observations	198	170	28

Means and standard errors (in parentheses) are presented.

Table 3.2: Descriptive Statistics for Communities in CHFS 2017

	Overall	Certi=0	Certi=1
Distance to Center	83.98	83.06	85.14
	(65.30)	(67.84)	(62.34)
Roads to Center	1.68	1.65	1.72
	(0.75)	(0.75)	(0.76)
Households	636.28	664.77	600.66
	(414.43)	(436.98)	(383.89)
Residents	2,065.93	2,181.25	1,921.80
	(1,701.72)	(1,864.49)	(1,471.08)
Hukou Residents	2,336.53	2,469.14	2,170.76
	(1,750.39)	(1,932.05)	(1,486.93)
Ag-Hukou Ratio	0.95	0.96	0.94
	(0.15)	(0.14)	(0.17)
Migrant Ratio	0.24	0.22	0.26
	(0.17)	(0.15)	(0.20)
Internet	0.92	0.91	0.94
	(0.27)	(0.29)	(0.23)
Featured Industry	0.54	0.51	0.57
	(0.50)	(0.50)	(0.50)
Disposable Income	8,119.12	7,515.50	8,873.64
	(5,306.92)	(5,306.35)	(5,240.05)
Dibao Ratio	0.05	0.04	0.06
	(0.08)	(0.05)	(0.11)
Observations	198	110	88

Means and standard errors (in parentheses) are presented.

Table 3.3: Descriptive Statistics for Agricultural Land

	2015	2017
Ag Land Size	3,538.55	3,562.78
	(3,490.25)	(3,538.40)
Transfer Activity	0.77	1.00
	(0.42)	(0.00)
Land Certification	0.14	0.44
	(0.35)	(0.50)
Abandoned Land	0.31	0.33
	(0.46)	(0.47)
Mechanization Ratio	0.71	0.78
	(0.34)	(0.32)
Economic Crop Ratio	0.32	0.36
	(0.32)	(0.34)
Land Acquisition	0.25	0.25
	(0.44)	(0.43)
Polluted Land	0.06	0.07
	(0.23)	(0.26)
Observations	198	198

Means and standard errors (in parentheses) are presented.

characteristics across the two waves. Total agricultural land size increases slightly from 3,539 Mu per village to around 3,563 Mu within two years. While only 77% percent communities have land transferring activities[5] in 2015, in 2017 land transfer activities are common in all sampled communities. Around 30% of villages have abandoned land. 70% of agricultural land is cultivated with machines while around one third of agricultural land is used for economic crops.

As for the land certification program, these two years witnessed wider spread of the program. Since the two waves all took place in the summer, questions are all about the previous year[6]. Thus,

[5]In the survey questionnaire, land transfer activity does not restrict to arable land; it consists of all types of land for rural use such as forestry and aquaculture.

[6]For example, the CHFS 2015 survey asked detailed questions regarding the year 2014.

for the CHFS 2015 communities, I define villages that finished certifying by the year 2014 as certified villages. In CHFS 2015, only 14% percent of villages with agricultural land had finished certifying while this increased considerably to more than 44% in CHFS 2017 survey. Figure 3.2 shows the distribution of years with land being certified. The frequencies closely follow the pattern in which the land certification program evolved. After the policy was first raised in 1997, there were several villages in my sample implementing the policy. However, the program stagnated in the 2000s. Beginning in 2009 when the Ministry of Agriculture started the experiment in 8 provinces for this program, the number of communities with land certification started to grow again, especially after the year 2013 when this program was quickly expanded all across the country. Out of all communities that finished this program by the end of 2016, the majority were done during the period 2014-2016 while almost half were done in the single year 2016. Figure 3.3 and 3.4 show the share of certified community across provinces. In Figure 3.3, only *Chongqing* municipality has over 75% certified villages. However, in CHFS 2017, *Shandong, Jiangxi, Shaanxi* provinces and *Ningxia Hui Autonomous Region* all had more than 75% communities with certification and these provinces are widely located all over the country. Despite the trend that the number of certified villages has been rapidly growing, the program is still less developed and more than half rural villages so far still lack sufficient security for their land tenure.

Additionally, Table 3.4 and 3.5 show some household-level demographic and economic information in CHFS 2015 and 2017 respectively. In CHFS 2015, on average, there are less than four members in a household while this number decreases to 3.62 in CHFS 2017. Across two waves, more than 90% of households have all members with agricultural Hukou. The ratio of households with migrant worker increases from 10% in CHFS 2015 to around 18% in CHFS 2017, while the total value for assets also increases by around 10%. Additionally, almost all households have their own housing units and this number is quite stable across years. In terms of agricultural characteristics, more than 80% of households participated in agricultural activities with average land size increasing from 7.57 Mu to 8.32 Mu. In CHFS 2015, 13% of households transfer their land out and 20% transfer land in; these ratios are 21% and 16% respectively in CHFS 2017.

Figure 3.2: Years of Rural Land Certification Program

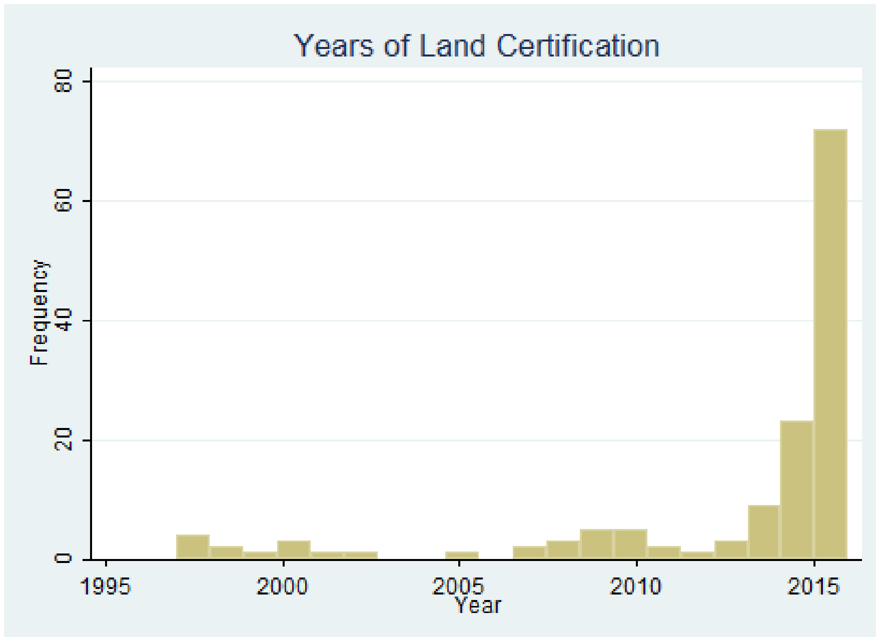


Figure 3.3: Rural Communities with Land Certification from CHFS2015

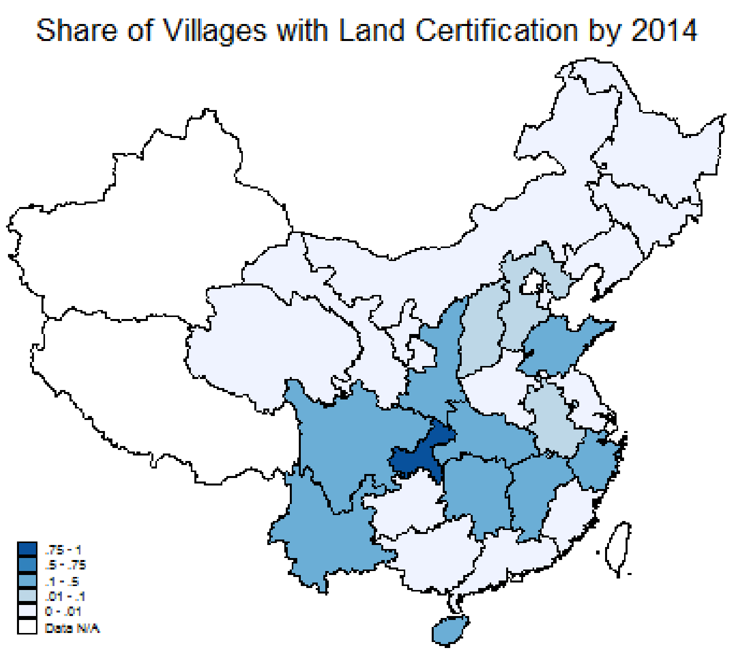


Table 3.4: Descriptive Statistics for Households in CHFS 2015

	Overall	Certi=0	Certi=1
Household Size	3.86	3.92	3.55
	(1.79)	(1.81)	(1.65)
Head Age	56.30	56.14	57.23
	(11.12)	(11.10)	(11.20)
Head Schooling	6.97	7.03	6.63
	(3.22)	(3.19)	(3.36)
All with Ag Hukou	0.92	0.92	0.93
	(0.26)	(0.26)	(0.26)
Head Non-ag Hukou	0.03	0.02	0.04
	(0.16)	(0.15)	(0.19)
Number of Laborers	2.45	2.50	2.15
	(1.38)	(1.39)	(1.28)
Have Migrant	0.10	0.10	0.06
	(0.30)	(0.30)	(0.24)
Number of Migrants	0.15	0.16	0.08
	(0.52)	(0.54)	(0.36)
Asset Value	293,746.79	293,847.53	293,164.31
	(512,979.37)	(512,556.07)	(516,080.72)
Have House	0.98	0.98	0.97
	(0.14)	(0.13)	(0.18)
Have Vehicle	0.13	0.12	0.15
	(0.33)	(0.33)	(0.35)
Observations	2645	2255	390

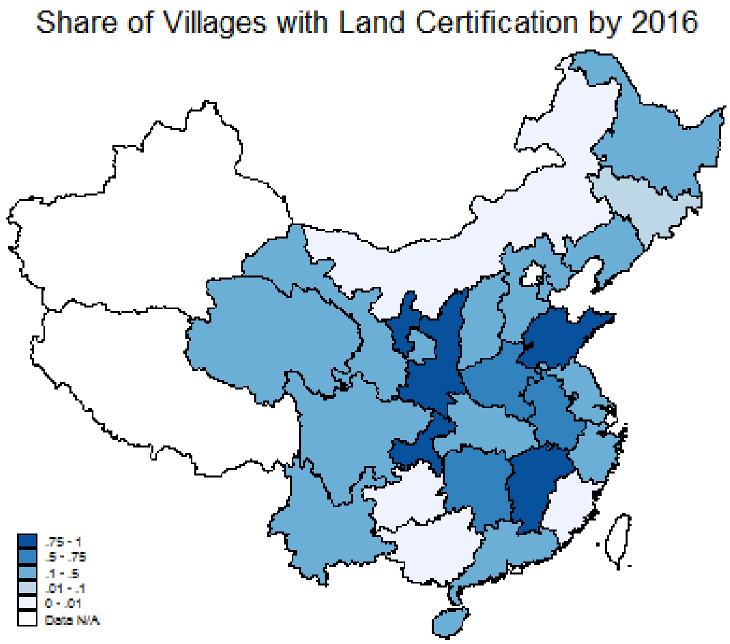
Means and standard errors (in parentheses) are presented.

Table 3.5: Descriptive Statistics for Households in CHFS 2017

	Overall	Certi=0	Certi=1
Household Size	3.62	3.66	3.57
	(1.73)	(1.71)	(1.76)
Head Age	57.25	56.73	57.87
	(11.24)	(11.34)	(11.11)
Head Schooling	7.07	7.04	7.10
	(3.26)	(3.22)	(3.32)
All with Ag Hukou	0.93	0.93	0.93
	(0.26)	(0.25)	(0.26)
Head Non-ag Hukou	0.03	0.03	0.02
	(0.16)	(0.17)	(0.15)
Number of Laborers	2.17	2.24	2.08
	(1.34)	(1.33)	(1.35)
Have Migrant	0.18	0.18	0.18
	(0.38)	(0.38)	(0.38)
Number of Migrants	0.25	0.25	0.26
	(0.62)	(0.60)	(0.63)
Asset Value	320,730.24	330,236.52	309,317.96
	(684,408.51)	(600,217.06)	(773,578.39)
Have House	0.96	0.97	0.96
	(0.18)	(0.17)	(0.20)
Have Vehicle	0.17	0.17	0.18
	(0.38)	(0.37)	(0.38)
Observations	2645	1443	1202

Means and standard errors (in parentheses) are presented.

Figure 3.4: Rural Communities with Land Certification from CHFS2017



Overall, in both my community- and household-level panel data I construct using two waves of CHFS datasets, I find quite identical results across certified and uncertified villages in both years. These balanced descriptive statistics may partly alleviate the concern that villages that are selected into certification in my sample are due to some demographic or agricultural characteristics.

3.3 Empirical Results

3.3.1 Estimation Strategy

In order to distinguish the effect of agricultural land certification program from confounding factors, I exploit the within-community variations caused by timing of the program. Specifically, I use a generalized difference-in-differences research design and compares changes in agricultural land utilization in certified communities with uncertified communities.

Formally, I take advantage of my two-period community-level dataset to estimate a fixed-effects ordinary least squares panel data model, where my outcome variable is the ratio of abandoned land at the community level. I also use a dummy indicating whether there exists abandoned agricultural land in a community as another outcome variable for robustness check. The OLS regression equation is as follows,

$$y_{jt} = \beta Certi_{jt} + X'_{jt}\delta + \alpha_j + \gamma_t + \epsilon_{jt} \quad (3.1)$$

where $Certi_{jt}$ is a dummy variable taking the value of 1 if community j in year t has finished land certification program, and 0 otherwise. X_{jt} is a vector of control variables. Specifically, I include some community-level demographic and economic variables, for example, number of households, number of residents, number of residents with agricultural Hukou, all in log terms, Internet connection, the ratio of residents relying on subsistence allowance, village income and a dummy variable indicating whether there are some special featured industry within the village. As mentioned before, land acquisition by upper-level governments is a major threats to land tenure security and thus affects land utilization. In this sense, I include a dummy variable of whether rural land in the community has been acquired since 2000. α_j and γ_t control for community and year fixed effects. In this way, my estimation equation is a generalized difference-in-differences model estimating the pure effects of land certification program on agricultural land abandonment. In all specifications, standard errors are clustered at the county level.

3.3.2 Results

Regression results from Equation 3.1 using the ratio of abandoned agricultural land as outcome variable are show in Table 3.6. In Column 1, I include only control variables regarding demographic information. I add the land acquisition variables in Column 2. In Column 3, I further add three variables indicating the economic development of the village. In all specifications, results are similar, both qualitatively and quantitatively. For example, in Column 1, the land certification program reduced the ratio of idle land in the village by 2.6%, and is significantly different from zero at the 10% level, which implies a positive effect in improving land utilization for agricultural use.

For estimators of the control variables, some results are also of interest. In terms of village demographics, larger number of residents suggests more labor supply in the village and thus are

Table 3.6: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Land Certification	-0.026*	-0.027*	-0.029*
	(0.016)	(0.016)	(0.016)
Ln(Households)	-0.065	-0.065	-0.070
	(0.048)	(0.049)	(0.052)
Ln(Residents)	-0.011	-0.010	-0.016
	(0.032)	(0.032)	(0.029)
Ln(Ag-Hukou Residents)	-0.057**	-0.058**	-0.061***
	(0.023)	(0.023)	(0.023)
Dibao Ratio	0.072	0.070	0.067
	(0.064)	(0.063)	(0.062)
Land Acquisition		-0.023	-0.022
		(0.023)	(0.024)
Featured Industry			-0.030
			(0.021)
Internet			0.030
			(0.042)
Village Income			-0.003
			(0.003)
R^2	0.157	0.166	0.195
Observations	388	388	382

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

associated with less abandoned land. Additionally, having some featured industries in the village is associated with less abandoned land. One possible explanation might be that with the featured industry, more labors would stay in the village rather than migrate, thus increase labor supply in the agricultural activity during the busy seasons.

My specification using the dummy variable of whether there exists idle land in a village yields similar results. For example, in Column 1 of Table 3.7, having agricultural land being certified reduces the probability of having abandoned land by more than 12%, a large magnitude compared with the fact that villages with land abandoned account for 36% in my sample.

3.3.3 Robustness Checks

In this subsection, I conduct multiple robustness checks.

First, instead of a year fixed effects, I estimate a specification that controls for province-year fixed effects. Doing this way, I allow each province to have its own trends in these two waves. Results in Table 3.8 show that results are qualitatively indifferent from my baseline model controlling for year fixed effects.

In addition, villages in my sample have a large gap in terms of years of certification. For example, a few villages have land certified in late 1990s while most villages have it done after 2015. One might worry that land certification in earlier periods might have differential effects compared with that in later periods. Thus, I restrict my sample to villages with land certification in different cutoff years. In Table 3.9, I exclude villages with land certification before 2009 while in Table 3.10 I exclude those before 2013. Results are both quite identical to my baseline full-sample estimates.

In order further to estimate the impact of the most recent wave of land certification, I drop all sampled villages that had finished the program before 2016 and conduct an event study. After dropping those villages, all left are those either uncertified (the control group) or certified at the year 2016 (the treatment group). Thus, I can compare village-level characteristics prior to the treatment assignment to conduct a balance test and adopt the propensity score matching approach to construct a matched uncertified control group to further mitigate the bias from selection into

Table 3.7: Dependent Variable: Land Abandonment Dummy

	(1)	(2)	(3)
Land Certification	-0.122*	-0.123*	-0.127*
	(0.072)	(0.072)	(0.073)
Ln(Ag Land Size)	0.591	0.586	0.598
	(0.428)	(0.427)	(0.439)
Ln(Households)	-0.074	-0.074	-0.083
	(0.123)	(0.124)	(0.130)
Ln(Residents)	-0.129	-0.128	-0.134*
	(0.080)	(0.079)	(0.080)
Ln(Ag-Hukou Residents)	-0.072**	-0.074**	-0.077**
	(0.035)	(0.035)	(0.034)
Dibao Ratio	0.077	0.074	0.075
	(0.176)	(0.175)	(0.178)
Land Acquisition		-0.026	-0.019
		(0.061)	(0.063)
Featured Industry			-0.110*
			(0.063)
Internet			0.070
			(0.082)
Village Income			0.001
			(0.015)
R^2	0.063	0.064	0.083
Observations	388	388	382

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Land Certification	-0.039**	-0.039**	-0.045**
	(0.017)	(0.018)	(0.019)
Ln(Households)	-0.095**	-0.095**	-0.101**
	(0.043)	(0.043)	(0.047)
Ln(Residents)	-0.011	-0.010	-0.021
	(0.027)	(0.027)	(0.026)
Ln(Ag-Hukou Residents)	-0.054***	-0.056***	-0.059***
	(0.018)	(0.018)	(0.018)
Dibao Ratio	0.103*	0.097*	0.087
	(0.057)	(0.057)	(0.055)
Land Acquisition		-0.022	-0.024
		(0.025)	(0.026)
Featured Industry			-0.037*
			(0.020)
Internet			0.027
			(0.032)
Village Income			-0.005
			(0.005)
R^2	0.311	0.317	0.355
Observations	388	388	382

Robust standard errors are clustered at the county level.

Community fixed effects and province-year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Land Certification	-0.034*	-0.034*	-0.034*
	(0.018)	(0.019)	(0.018)
Ln(Households)	-0.070	-0.069	-0.074
	(0.050)	(0.051)	(0.054)
Ln(Residents)	-0.009	-0.008	-0.016
	(0.033)	(0.033)	(0.031)
Ln(Ag-Hukou Residents)	-0.056**	-0.057**	-0.059**
	(0.023)	(0.023)	(0.023)
Dibao Ratio	0.149	0.146	0.138
	(0.122)	(0.120)	(0.117)
Land Acquisition		-0.014	-0.013
		(0.019)	(0.021)
Featured Industry			-0.023
			(0.021)
Internet			0.027
			(0.037)
Village Income			-0.003
			(0.004)
R^2	0.181	0.184	0.207
Observations	369	369	363

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Land Certification	-0.033*	-0.034*	-0.034*
	(0.018)	(0.018)	(0.018)
Ln(Households)	-0.070	-0.070	-0.074
	(0.051)	(0.052)	(0.055)
Ln(Residents)	-0.010	-0.008	-0.015
	(0.034)	(0.034)	(0.032)
Ln(Ag-Hukou Residents)	-0.058**	-0.059**	-0.061**
	(0.025)	(0.025)	(0.024)
Dibao Ratio	0.149	0.145	0.135
	(0.123)	(0.121)	(0.117)
Land Acquisition		-0.015	-0.015
		(0.020)	(0.022)
Featured Industry			-0.018
			(0.021)
Internet			0.027
			(0.039)
Village Income			-0.003
			(0.005)
R^2	0.189	0.193	0.212
Observations	353	353	347

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: Logistic Estimations for Selection into Treatment

Ln(Ag Land Size)	0.045	(0.045)
Ln(Households)	-0.120	(0.103)
Ln(Residents)	0.169*	(0.096)
Ln(Ag-Hukou Residents)	-0.030	(0.032)
Dibao Ratio	0.883	(0.672)
Land Acquisition	-0.042	(0.080)
Distance to Center	-0.000	(0.001)
Featured Industry	0.092	(0.079)
Internet	-0.040	(0.106)
Ln(Disposable Income)	0.032	(0.047)
Observations	149	

Marginal effects are displayed.

Robust standard errors are clustered at the county level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

treatment. Selection into treatment may occur if the decision of timing of certification for villages is dependent on some village-level characteristics, especially the agricultural activities. Therefore, my matching technique is exploited to establish a set of control villages without certification that are similar to villages certified in 2016, so that the analysis has a comparable counterfactual. I then estimate OLS regressions of Equation 3.1 on my constructed sample of matched villages.

The set of covariates used to estimate and predict propensity scores is identical to those controlled for in my baseline specification in Column 2 of Table 3.6. Additionally, I add the log size of agricultural land to control for the endowment of a village. Results from the Probit model using these covariates are shown in Table 3.11. Reassuringly, most of these covariates are not significantly different from zero, implying that the assignment of land certification program is not selective on these village-level observable characteristics. Only the village-level scale of residents marginally predict the assignment of the program.

Results from a matched sample is shown in Table 3.12, with sample size 290. For both my proxies of land utilization measurements, the coefficients of land certification dummy are quantitatively similar to the baseline estimates, significantly at the 10% level. The land certification program is associated with a decrease of the abandoned rural land by 3.6% for villages that finished this program in 2016 compared with uncertified villages. My results indicate that the program does improves land utilization.

So far, I have done a list of robustness checks using village-level panel data. For the household-level survey data, since the question regarding whether a household has abandoned land was only asked in the CHFS 2017 questionnaire, it is not possible to construct a panel and do similar regressions using household-level data. Alternatively, I keep only the CHFS 2017 households with non-missing values of the dependent variable and run a regression using this cross-sectional data with households in the aforementioned matched village samples. This result, though not a causal inference, may provide some statistical evidence at the household level. Estimation results are shown in Table 3.13. In different columns, I control for different sets of variables. Standard errors are clustered at the village level. In all specifications, results are identical: in my matched samples, households living in certified villages are associated with 2% less abandoned land compared with households in uncertified villages.

3.3.4 Heterogeneous Effects

We then report OLS estimates using interactions of the land certification dummy with some village-level economic conditions to show possible heterogeneous effects of land certification program. For all my specifications, I use fixed-effects model and cluster standard errors at the county level.

First, Table 3.14 divides communities into two groups according to their levels of disposable income per capita in the first wave. It is clearly shown that the effect of land certification on reduction of land abandonment is increasing in higher income villages. For those villages with low disposable income, the coefficient is both economically and statistically indifferent from zero. On the other hand, I find a large and significant result for high income villages.

We also explore some geographic variation of the effect of the program. I divide provinces into subgroups according to their rankings in the share of GDP from agricultural sector. I measure the share using the 2014 data published by the National Bureau of Statistics. Nationally, the agriculture

Table 3.12: Estimation using matched control groups from PSM Results

	Land Abandonment Ratio	Land Abandonment Dummy
Land Certification	-0.036*	-0.132*
	(0.019)	(0.076)
Land Acquisition	-0.017	-0.045
	(0.024)	(0.052)
Ln(Households)	-0.071	0.015
	(0.060)	(0.134)
Ln(Residents)	-0.020	-0.196**
	(0.041)	(0.095)
Ln(Ag-Hukou Residents)	-0.062**	-0.084**
	(0.026)	(0.037)
Dibao Ratio	0.575*	1.664
	(0.293)	(1.019)
Featured Industry	-0.018	-0.136**
	(0.022)	(0.063)
Internet	0.022	-0.020
	(0.042)	(0.100)
Village Income	-0.005	-0.012
	(0.006)	(0.019)
R^2	0.266	0.169
Observations	290	290

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in both specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.13: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Land Certi	-0.019	-0.019*	-0.019*
	(0.011)	(0.011)	(0.011)
Household Size	0.000	-0.000	0.000
	(0.003)	(0.003)	(0.003)
Head Age	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Head Schooling	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
All with Ag Hukou	-0.034	-0.034	-0.031
	(0.022)	(0.021)	(0.022)
Number of Laborers	-0.014***	-0.013***	-0.013***
	(0.005)	(0.005)	(0.005)
Asset Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	-0.004	-0.001	-0.002
	(0.011)	(0.011)	(0.011)
Ln(Farmland Size)		-0.023***	-0.025***
		(0.005)	(0.005)
Land Acquisition			-0.003
			(0.013)
Households in Village			-0.016*
			(0.008)
Ag-Hukou Residents in Village			0.005
			(0.006)
R^2	0.019	0.030	0.032
Observations	1845	1845	1845

Robust standard errors are clustered at the village level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.14: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Certi*Low Income	-0.015	-0.015	-0.014
	(0.023)	(0.023)	(0.023)
Certi*High Income	-0.040**	-0.042**	-0.049**
	(0.016)	(0.017)	(0.020)
Ln(Households)	-0.067	-0.067	-0.072
	(0.049)	(0.050)	(0.053)
Ln(Residents)	-0.010	-0.009	-0.015
	(0.031)	(0.032)	(0.029)
Ln(Ag-Hukou Residents)	-0.056**	-0.058**	-0.060***
	(0.023)	(0.022)	(0.022)
Dibao Ratio	0.073	0.070	0.068
	(0.065)	(0.063)	(0.063)
Land Acquisition		-0.024	-0.023
		(0.023)	(0.024)
Featured Industry			-0.033
			(0.022)
Internet			0.029
			(0.042)
Village Income			-0.003
			(0.003)
R^2	0.160	0.169	0.200
Observations	388	388	382

Robust standard errors are clustered at the county level.

Community fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

sector consists of 9% of the total GDP and ranges from the highest 23% in *Hainan* Province to less than 1% in *Beijing* and *Shanghai*, two provincial-level municipal cities [7]. Table 3.15 presents the results. It shows that land certificate is more effective in provinces with higher share of GDP contributed from agricultural sector.

3.3.5 Possible Mechanisms

So far I have established that there is a positive effect of the land certification program on reducing idle agricultural land in China's rural villages. The CHFS household-level panel data also allow me to exploit its effects on other outcomes from agriculture and social welfare concepts, which could help me better understand through which channel this effect works.

One possible explanation is that through land certification, farmers would have incentives to increase investments in the certified land. In Table 3.16, I use values of machine for agricultural use as a proxy to see whether the land certification improves investment related to production inputs. The coefficients is positive but insignificant, which does not support the idea that land certification program improves agricultural investments in rural areas.

Moreover, in some of the rural-urban migration literature, the effect of land certification on migration is mixed. On the one hand, with land tenure being secured, rural labors would be more likely to migrate to urban areas for work since they are no longer worried about their land being expropriated. On the other hand, [40] uses surveys from China's villages and found the effect of tenure security on increasing migration is more significant for forestry land. In Table 3.17, I do not find any significant results showing the effect of land tenure security on migration choice, which is in line with [40] in their finding on agricultural land.

Moreover, much of the literature focuses on its effects on land rental market. From the household-level survey, I am able to observe the land rental decision for each household. Thus, I use whether a rural household rents in and out land to as dependent variables. Results are shown in Table 3.18 and 3.19. The coefficients are mixed and statistically indifferent from zero, showing that land certification overall does not improve the land rental market, at least in the short run.

[7] Geographically, Beijing, Shanghai and Hainan are all eastern provinces.

Table 3.15: Dependent Variable: Land Abandonment Ratio

	(1)	(2)	(3)
Certi*Low Ag Share	-0.003	-0.004	-0.008
	(0.029)	(0.029)	(0.028)
Certi*High Ag Share	-0.043***	-0.044***	-0.043**
	(0.014)	(0.015)	(0.017)
Ln(Households)	-0.066	-0.066	-0.071
	(0.049)	(0.050)	(0.053)
Ln(Residents)	-0.012	-0.010	-0.016
	(0.031)	(0.032)	(0.029)
Ln(Ag-Hukou Residents)	-0.059**	-0.060**	-0.062***
	(0.024)	(0.024)	(0.023)
Dibao Ratio	0.072	0.070	0.066
	(0.065)	(0.063)	(0.062)
Land Acquisition		-0.023	-0.022
		(0.023)	(0.024)
Featured Industry			-0.027
			(0.022)
Internet			0.032
			(0.041)
Village Income			-0.003
			(0.003)
R^2	0.164	0.173	0.200
Observations	388	388	382

Robust standard errors are clustered at the county level.
Community fixed effects and year fixed effects are included in both specifications.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.16: Dependent Variable: Ln(Machine Value)

	(1)	(2)	(3)
Land Certi	0.195	0.241	0.199
	(0.209)	(0.207)	(0.204)
Household Size	-0.121	-0.136	-0.134
	(0.085)	(0.085)	(0.085)
Head Age	-0.011	-0.011	-0.011
	(0.011)	(0.011)	(0.011)
Head Schooling	0.058**	0.054*	0.053*
	(0.028)	(0.028)	(0.028)
All with Ag Hukou	0.338	0.268	0.291
	(0.284)	(0.282)	(0.285)
Number of Laborers	0.262**	0.244**	0.236**
	(0.105)	(0.105)	(0.105)
Asset Value	0.000***	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
Have Vehicle	0.003	0.002	0.018
	(0.266)	(0.266)	(0.267)
Ln(Farmland Size)		0.601***	0.591***
		(0.141)	(0.142)
Land Acquisition			-0.080
			(0.222)
Households in Village			-0.604
			(0.369)
Ag-Hukou Residents in Village			0.010
			(0.092)
R^2	0.018	0.025	0.026
Observations	4985	4955	4941

Robust standard errors are clustered at the village level.

Household fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.17: Dependent Variable: Migrant Dummy

	(1)	(2)	(3)
Land Certi	0.022	0.022	0.019
	(0.022)	(0.022)	(0.022)
Household Size	0.032***	0.031***	0.031***
	(0.010)	(0.010)	(0.010)
Head Age	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Head Schooling	-0.002	-0.001	-0.001
	(0.003)	(0.003)	(0.003)
All with Ag Hukou	-0.057	-0.056	-0.054
	(0.047)	(0.047)	(0.048)
Number of Laborers	0.071***	0.074***	0.073***
	(0.012)	(0.011)	(0.011)
Asset Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	0.008	0.007	0.008
	(0.029)	(0.030)	(0.030)
Ln(Farmland Size)		-0.023	-0.024
		(0.015)	(0.015)
Land Acquisition			0.000
			(0.021)
Households in Village			-0.048
			(0.032)
Ag-Hukou Residents in Village			0.002
			(0.012)
R^2	0.120	0.120	0.121
Observations	4986	4956	4942

Robust standard errors are clustered at the village level.

Household fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.18: Dependent Variable: Rent in Land

	(1)	(2)	(3)
Land Certi	0.029	0.030	0.027
	(0.022)	(0.022)	(0.023)
Household Size	0.008	0.009	0.009
	(0.009)	(0.009)	(0.009)
Head Age	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Head Schooling	0.001	0.002	0.001
	(0.003)	(0.003)	(0.003)
All with Ag Hukou	0.022	0.023	0.024
	(0.042)	(0.042)	(0.042)
Number of Laborers	0.004	0.004	0.003
	(0.011)	(0.011)	(0.011)
Asset Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	-0.013	-0.013	-0.012
	(0.028)	(0.027)	(0.028)
Ln(Farmland Size)		-0.006	-0.007
		(0.019)	(0.019)
Land Acquisition			-0.003
			(0.018)
Households in Village			-0.035
			(0.043)
Ag-Hukou Residents in Village			0.007
			(0.006)
R^2	0.013	0.013	0.014
Observations	4986	4956	4942

Robust standard errors are clustered at the village level.

Household fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.19: Dependent Variable: Rent out Land

	(1)	(2)	(3)
Land Certi	-0.020	-0.022	-0.023
	(0.026)	(0.025)	(0.024)
Household Size	0.020**	0.020**	0.019**
	(0.009)	(0.009)	(0.009)
Head Age	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Head Schooling	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)
All with Ag Hukou	0.005	0.002	0.011
	(0.047)	(0.048)	(0.047)
Number of Laborers	-0.023**	-0.023**	-0.023**
	(0.011)	(0.011)	(0.011)
Asset Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	0.015	0.018	0.015
	(0.023)	(0.023)	(0.023)
Ln(Farmland Size)		0.029**	0.029**
		(0.014)	(0.014)
Land Acquisition			-0.002
			(0.031)
Households in Village			-0.018
			(0.054)
Ag-Hukou Residents in Village			-0.009
			(0.009)
R^2	0.035	0.037	0.037
Observations	4979	4951	4937

Robust standard errors are clustered at the village level.

Household fixed effects and year fixed effects are included in all specifications.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

However, though land certification program might not be able to improve the total land rental activities, it might affect the land rental behavior through different entities who rent agricultural land in. In order to further exploit the rental activities, I divide entities that transfer land into two groups: individual households and agricultural business entities. There is a large body of literature showing that land transfer between relatives and neighbors within a village is very common in rural countryside, as there is no information asymmetry and little risk between acquaintances. In CHFS 2017 survey, questions are asked for information regarding agricultural land transfer market, for example, different types of entities that transfer land in, the size and rent for those entities. Table 3.20 shows the difference of land rental behavior across villages with certification and those without. Despite the fact that certified villages have more arable land endowment than non-certified villages, the size of land that is transferred to individual households is not with much variation. However, the size of land transferred to institutional agricultural entities such as cooperations or companies related to agricultural sector as well as local farms that specialize in one crop with large scales is much larger for villages with land certificate. Also, in certified villages, shares of land transferred to those entities are much larger than that to individual households. Moreover, the rent per Mu for land to those entities are also greater on average than rents to individual households, suggesting that land transfer market is more formal in certified village.

In order to empirically test this hypothesis, I present OLS regressions using household-level panel data in Table 3.21 and 3.22. Samples are restricted to households that participate in transferring land out. In Table 3.21, the dependent variable is whether the household transfers land to another household while the dependent variable in Table 3.21 is a dummy that takes the value of one if the household transfers land to a new agricultural business entity. From both tables, results show that the land certification program decreases the possibility for a household to transfer land to another household but significantly increases the probability for them to transfer to entities.

Thus, from both household-level regression estimates and village-level descriptive statistics, I find evidence that the effect of land certification program on improving land utilization might be a result that households with land being certified are more likely to transfer their cultivated land to

Table 3.20: Comparison on Land Transfer Activities across Villages in 2017

	Certi=0	Certi=1
Agricultural Business Entities	0.72	0.75
	(0.45)	(0.44)
Size to Households	539.51	461.96
	(1,387.88)	(789.12)
Size to ABE	329.45	644.84
	(851.31)	(1,246.02)
Rent for Housholds	427.40	484.81
	(343.42)	(665.60)
Rent for ABE	505.82	705.95
	(340.28)	(1,322.17)
Observations	110	88

Means and standard errors (in parentheses) are presented.

new agricultural business entities that might possibly increase land utilization.

Table 3.21: Dependent Variable: Rent out Land to Farmers

	(1)	(2)	(3)
Land Certi	-0.078	-0.102	-0.115
	(0.076)	(0.075)	(0.070)
Household Size	0.047*	0.050**	0.050**
	(0.026)	(0.025)	(0.025)
Head Age	-0.003	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)
Head Schooling	-0.009	-0.019	-0.015
	(0.014)	(0.014)	(0.013)
All with Ag Hukou	-0.060	-0.100*	-0.140*
	(0.061)	(0.058)	(0.075)
Number of Laborers	-0.114***	-0.122***	-0.117***
	(0.037)	(0.035)	(0.035)
Asset Value	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	0.109	0.110	0.107
	(0.086)	(0.074)	(0.071)
Ln(Farmland Size)		0.142**	0.157***
		(0.062)	(0.056)
Land Acquisition			-0.144**
			(0.055)
Households in Village			0.171*
			(0.095)
Ag-Hukou Residents in Village			-0.048**
			(0.020)
R^2	0.072	0.109	0.150
Observations	809	802	794

Robust standard errors are clustered at the village level.
Household fixed effects and year fixed effects are included in all specifications.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.22: Dependent Variable: Rent out Land to Companies

	(1)	(2)	(3)
Land Certi	0.145*	0.164*	0.166**
	(0.083)	(0.084)	(0.077)
Household Size	-0.050*	-0.054*	-0.052*
	(0.028)	(0.027)	(0.027)
Head Age	0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)
Head Schooling	0.007	0.014	0.010
	(0.014)	(0.014)	(0.013)
All with Ag Hukou	0.087	0.108*	0.130*
	(0.059)	(0.060)	(0.069)
Number of Laborers	0.113***	0.118***	0.112***
	(0.037)	(0.037)	(0.037)
Asset Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Have Vehicle	-0.104	-0.103	-0.095
	(0.096)	(0.095)	(0.093)
Ln(Farmland Size)		-0.054	-0.077
		(0.063)	(0.058)
Land Acquisition			0.097
			(0.064)
Households in Village			-0.158*
			(0.091)
Ag-Hukou Residents in Village			0.072***
			(0.025)
R^2	0.074	0.087	0.125
Observations	809	802	794

Robust standard errors are clustered at the village level.
Household fixed effects and year fixed effects are included in all specifications.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4. CURRICULUM REFORM, EXAM RE-TAKING OPPORTUNITY COST AND HOUSEHOLD EDUCATIONAL INVESTMENTS

4.1 Background

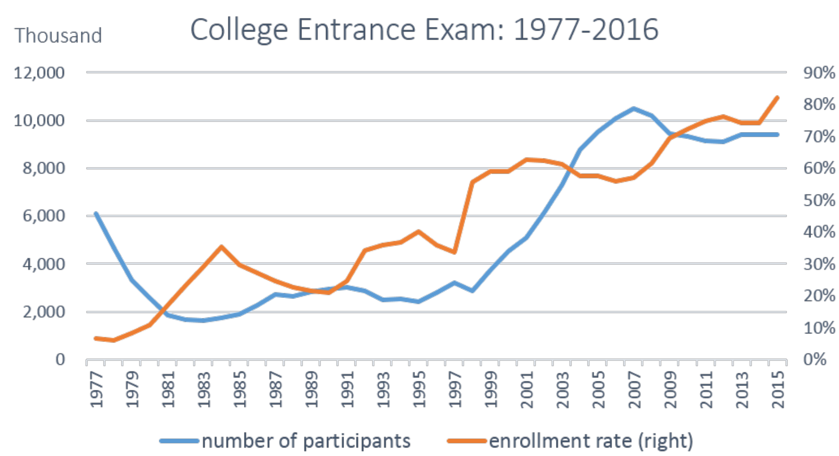
4.1.1 Household Educational Investments in China

Traditionally, the Chinese place a very high value on education ([42]). Under the lasting influence of Confucianism, the officials (*Shi*) in imperial China were held in the highest regard, and civil exam was the only road to officialdom for commoners ([8]). [3] show that the abolition of this civil exam that lasts for more than 1,300 years caused political instability at the early 1900s.

The establishment of the People's Republic of China greatly changed education policies, under which schools at all levels were publicized. The Cultural Revolution starting in 1966 had a catastrophic effect on China's educational system as family background and political acceptability instead of intelligence and exam scores became criteria for entrance into college ([30]). After such a ten-year long interruption, universities in China reopened and resumed to enrolling students in the late 1970s. The first College Entrance Examination (CEE) after Cultural Revolution took place in the winter of 1977, since which it became the only way for most high school graduates to apply for universities and colleges.

Figure 4.1 shows the total number of exam takers and enrollment rate since 1977. During the first two decades after the resumption of CEE, the enrollment rate was comparatively low and never exceeded 40%. At that time, being enrolled in a university is extremely competitive due to a limit capacity of high education resources and high demand for education opportunities. [54] provide estimates of the returns to schooling in urban China over the period 1988-2001, and show that returns to education have changed from only 4% per year in 1988 to 10.2% per year in 2001. They argue that such increasing returns are influenced by institutional reforms in the labor market that raise the demand for skilled labor. In 1998, the central government and the Ministry of Education began to invest more in facilities of universities and colleges all around the country and expand the scale of high education. In the year of 1999, the enrollment rate reached 56%, compared with only 34% in 1998. Since then, the college enrollment never drops below 50% and recently in the 2010s, it bypassed 75%, indicating that 3 out of 4 high school graduates who took CEE would end up being enrolled somewhere to achieve tertiary education.

Figure 4.1: College Entrance Exam: 1977-2016



Even though the enrollment rate increases, the competition has never been less intense, which nowadays is mainly for a high quality of education, rather than the education opportunity ([56]). [27] study the value of elite education in Chinese universities[1] and find that receiving an elite education increases the monthly wages of workers by 30-40% and such premium is more likely to be explained by university-related networks and signaling. Such competition for a high quality of education includes not only students but also parents. Since most households especially those in urban areas have only one child, parents would wish their children to be admitted into an elite university and thus obtain a better employment and social status upon graduation.

Household investments in children's education take a variety of forms at different school levels. For example, recently there has been a trend in some cities for affordable parents to send their chil-

[1]In China, tier-1 universities are publicly recognized as elite universities.

dren into dual-language kindergartens[2] for their children, sometimes even before they were born because of the over demand, in the hope that their kid would be exposed to an environment with English communications in their very early childhood. In addition, more and more households are willing to pay premiums for apartments located in elite primary or junior middle schools' neighborhoods[3] in order for their children being qualified to be enrolled into these schools. Furthermore, wealthy parents prefer to send their children abroad in their early teens, where they could apply for oversea high school and college without being involved in the competitive CEE in mainland China.

Normal educational investments for Chinese families with children in senior high school include asking private tutors for certain subjects, attending out-of-school classes during weekends or vacations ([52]). According to [47], more than half of senior high school students are involved in private tutoring classes in urban China in 2004. Educational investments also take forms such as purchasing electronic devices that help students' learning efficiency. Furthermore, some parents also provide comforting living environments for their children during their last year in high school. For example, they would quit job and rent an apartment to accompany their children for the entire academic year to prepare for CEE.

4.1.2 The Eighth Curriculum Reform in China

China's eighth curriculum reform [4] is a nationwide education reform undertaken by the Chinese central government beginning in 2001 ([5]) and formally started in 2004. It is characterized as three news, i.e. new curriculum, new notion and new College Entrance Examination.

The new curriculum is introduced at the province level. All provinces and provincial-level municipalities in China (except Shanghai) witnessed the introduction of new curriculum between 2004 and 2010. The introduction years for each province are shown in Figure 4.2. Incoming students in the Fall 2004 in *Guangdong, Hainan, Ningxia* and *Shandong* provinces were among the first cohorts to study under new curriculum and took CEE under new guidelines at the year 2007.

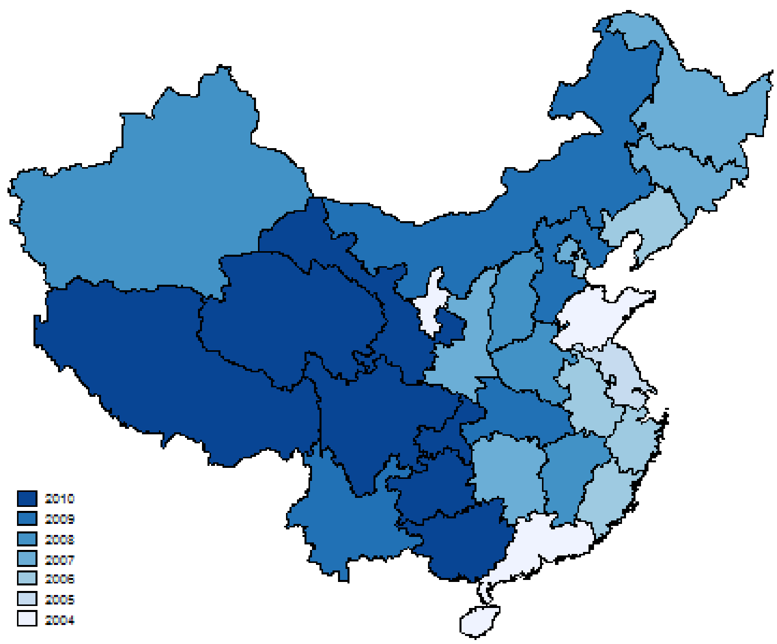
[2]Most dual-language schools refer to those who teach both Chinese and English.

[3]In Chinese, those apartments are called *Xue Qu Fang*.

[4]Formally, it is called *New Curriculum Program for High School*.

With respect to this curriculum reform, [5] document both qualitative and quantitative evidence on the changes of politics textbook content and the CEE guideline. Changes are also evident in other subjects ([50]).

Figure 4.2: Years of Introduction of the New Curriculum



In addition, the curriculum reform is also associated with changes of College Entrance Exam settings and guidelines as one characteristic is called "new CEE". In some provinces the CEE was modified not only in guidelines but also in courses offered as well as their corresponding weights in total scores. For example, in *Shandong* province in the year of 2007 in which the first cohort under new curriculum took CEE, students were required to take a new subject called "Basic Ability Test" which covered physical education, arts, music as well as basic technology questions. This test lasted for seven years and was abolished in 2014. On the other hand, *Jiangsu* province, which introduced new curriculum for the 2005 incoming high school cohort, adopted a more complex CEE reform framework. Before the reform, the CEE was in form *3+2*[5]. Each subject has a full

[5]All students are required to take Chinese, mathematics and English. Students in natural science track can choose

score 150, making the total score 750. After the adoption of new curriculum, the setting of CEE was changed hugely. Students were asked to take academic assessments (mini CEE) at the end of their second year in high school, including the natural science subjects or the social science subjects which were compositions of regular CEE before. They only receive grade (e.g. A, B, C etc.) instead of scores. Students with all A's would receive extra points in addition to their final CEE scores. In CEE, they only need to take Chinese, mathematics and English exams. For students in natural science track, the score is 160 for Chinese, 200 for mathematics and 120 for English. For social science track students, the score is 160 for mathematics, 200 for Chinese 120 for English, making the total score 480.

4.1.3 College Entrance Exam

Normally, students in senior high school are in their 10th to 12th years while most of their time is spent to prepare for the College Entrance Examination. In their first year of senior high school, they normally take some required courses in Chinese, English, mathematics as well as the basic natural science courses (physics, chemistry and biology) and social science courses (politics, geography, history). At the end of their first year, students can choose either the natural science track or the social science track to specialize in as their field of the College Entrance Examination. The timing of choosing field varies across schools. For examples, some high schools require their students to choose their fields upon entry and some high schools allow students to do so at the end of the second year. Regardless of the timing, most students would spend at least one whole year totally focused on reviewing and preparing for the CEE without learning any new knowledge. More importantly and related to my empirical work, if students do not perform well or are not enrolled into the college or major that meets their expectations, households might wish to let them repeat the 12th year (the last year in senior high school) and take one more chance next year in the CEE.

This mode of high school time schedule provides a promising context in which to study house-

two out of physics, chemistry and biology and those in the social science track may choose two out of history, politics and geography.

hold educational investments caused by exogenous shocks. To be more specific, students entering high school at the introduction year of new curriculum would have a totally different three-year curriculum and CEE guideline from those entering just the year before. In terms of costs of re-taking CEE, the last cohort under the old curriculum would be in a different situation compared with other cohorts. For other cohorts, if they fail in their first trial at the end of their third year in senior high school, they would choose to prepare for one more year and retake CEE one year later, without spending additional time and effort getting familiar with new knowledge and methods under new curriculum. However, for the last cohort under the old curriculum, their opportunity costs are much higher since if they fail in their first trial in CEE and choose to retake one year later, they have to face a different curriculum and CEE guideline. As a result, students in the last cohort under old curriculum need to face a higher opportunity cost with respects to re-taking CEE. In my empirical work, I utilize this exogenous shock faced by households with students in the last cohort under old curriculum to study their educational investments.

Moreover, the process of introduction of the new curriculum makes this context a good one in which to study the causal effect of different opportunity costs faced by cohorts under different curricula. Between the year 2004 and 2010, the new curriculum as well as high school textbooks and CEE guidelines were adopted in different provinces at different years. Students in old pre-reform cohort would not be “partially treated" since the College Entrance Exam is based either on the old curriculum or on the new one ([5]). By examining the sharp province by cohort variation in school curricula, I can plausibly identify the causal effect of curriculum effect on household educational investments.

However, the introduction year of the new curriculum was not randomly assigned to provinces. Provinces introduced the new curriculum when they had successfully trained teachers and developed supplemental materials based on new textbooks ([5]). If this nonrandom assignment coincides with different patterns in household educational investments, my identification strategy would be invalid. I would address this concern later in my empirical analysis.

4.2 Curriculum Reform and CEE Re-taking

This subsection shows evidence that the introduction of new curriculum causes a reduction in the number of the CEE re-takers at the year when provinces switch curricula.

4.2.1 Data

Since both curriculum reform and CEE are organized and implemented by the provincial department of education, I then use panel data documenting the number of examinees across years at the province level to analyze the relationship between the adoption of new curriculum and CEE re-taking behavior.

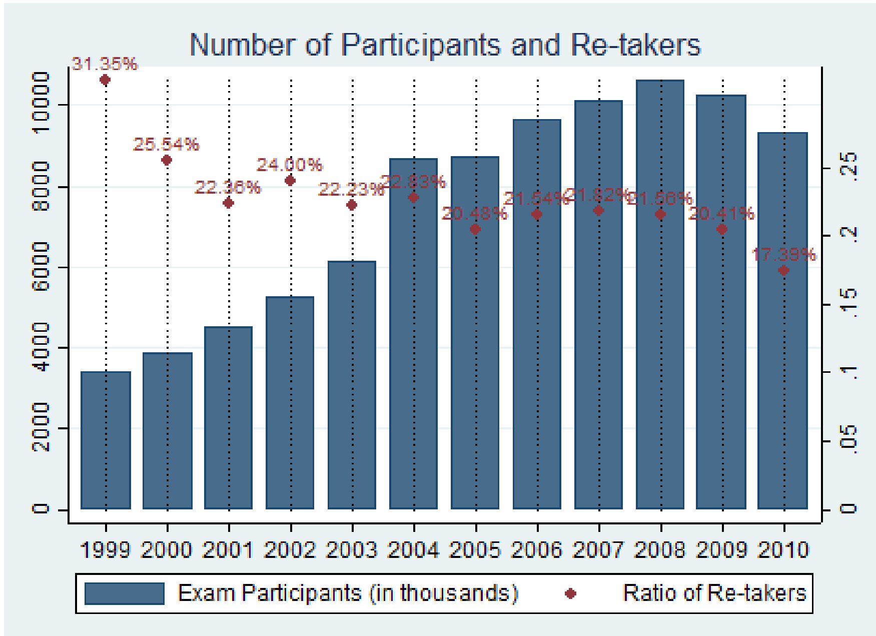
The data used in this subsection are from *Educational Statistics Yearbook of China*, edited and published by the Department of Development and Planning, Ministry of Education. Since information of the number of participants of College Entrance Exam are only available in yearbooks before 2010, I restrict my panel dataset to years between 1999 and 2010[6]. One limitation of such time range is that the period only witnessed the change in curriculum in 15 provinces, only half of total provinces in mainland China. However, it is the only administrative dataset that clearly divides the number of CEE participants into first-triers and re-takers at the provincial level, which is a key feature of my empirical strategy.

The introduction year of the new curriculum at each province is shown in Figure 4.2. The 15 provinces that adopted the new curriculum between 2004 and 2007 vary both geographically and economically, from provinces in comparatively developed coastal areas in the east (e.g. *Guangdong, Jiangsu, Zhejiang*) to those in western inland areas that are comparatively less developed, from populous provinces in which the CEE is quite competitive (e.g. *Anhui, Shandong*) to provincial-level municipalities in which resource of high-quality education is comparatively abundant.

Figure 4.3 shows the total number of participants in CEE from 1999 to 2010. It steadily increased until it reached the peak in the year 2008, when more than 10 million participants took CEE. With respect to CEE re-takers, the ratio is quite stable, around 20% to 25%. Overall, for

[6]Information of the number of participants of CEE is first available in 1999.

Figure 4.3: Number of CEE participants and Share of Re-takers



every 5 students taking CEE, one would be the re-taker. Besides, Figure 4.4 and 4.5 show that the trend of re-taking behavior does not differ much between rural and urban students during this period, however, the ratio of CEE re-takers for students with rural registration is consistently higher.

Moreover, enrollment rate did not change much during the period 1999-2010, shown in figure 4.1. As has been mentioned, in 1998 Chinese government began expanding the scale of high education. Consequently, there was a large jump of college enrollment rate between 1998 and 1999. Since then, it stayed quite stably until then end of 2000s, which coincides with the period of my panel data, indicating that there were no huge changes on the supply side during that period[7].

In addition, the probability that a student migrates to other provinces to take CEE is rare. In China, the school roll status is tightly linked to the students' Hukou status as well as their parents'. For most migrant workers in cities without local residential status, their children's school roll status is still in their hometowns[8]. Students have to take the CEE in provinces where their school roll is

[7]Although the way of being enrolled into a college or university is not limited, including university-administered screening tests (*Zizhu Zhaosheng* in Chinese), this admission process accounts for very small proportion and is restricted to top elite universities. Most students still solely rely on the CEE test scores as the only criterion to apply for college.

[8]Though recent policies allow some of migrant children to attend local primary school or junior high school with

Figure 4.4: Number of CEE participants and Share of Re-takers for Rural Students

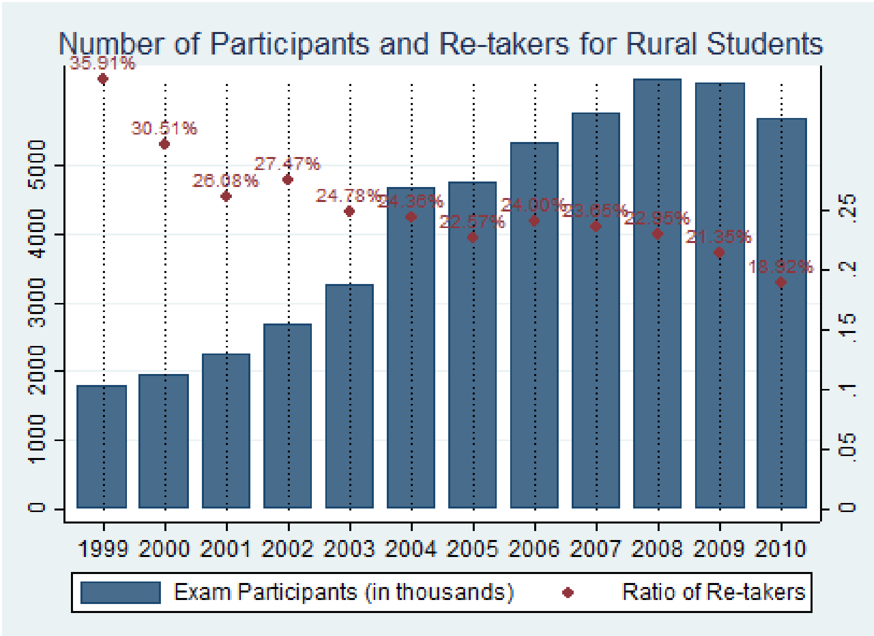
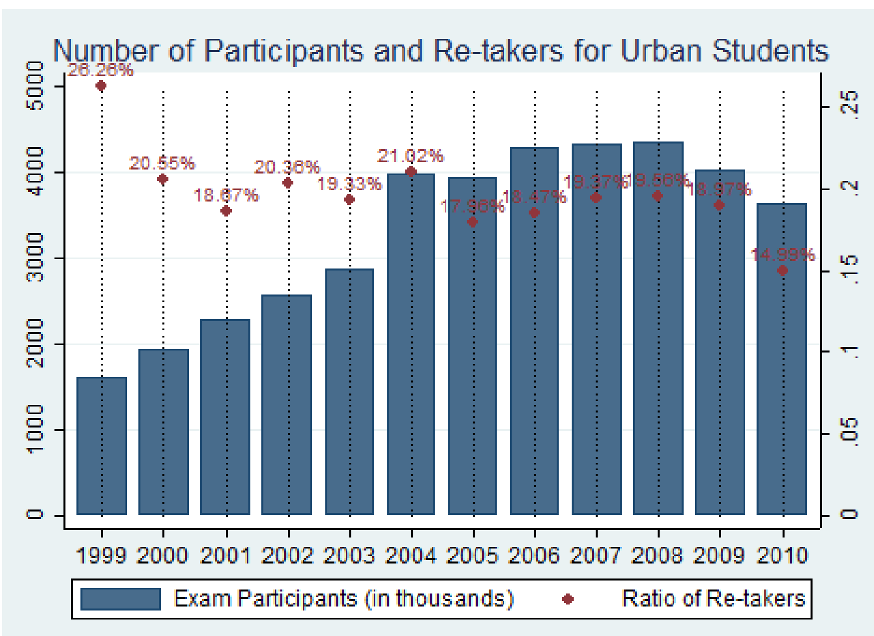


Figure 4.5: Number of CEE participants and Share of Re-takers for Urban Students



registered. As a result, migrants in purpose of CEE is not a concern in this context.

4.2.2 Empirical Strategy

In order to study the casual relationship between the switch of new curriculum and the CEE re-taking patterns at the province level, I employ the following generalized difference-in-differences model:

$$Retaker_{j,t} = \beta * New\_Curriculum_{j,t} + Province_j + Year_t + \epsilon_{j,t} \tag{4.1}$$

where $Retaker_{j,t}$ is different measures of CEE re-takers in province j at year t. $New\_Curriculum_{j,t}$ is a dummy variable taking the value of 1 when province j has its first cohort take CEE under new curriculum in year t and 0 otherwise. In the regression, I include both the province dummies and the year dummies to control for province and year specific characteristics. The coefficient of interest is β, which reflects the effect of the curricula switch on CEE re-taking behavior.

One assumption using the difference-in-differences model is that both the treatment group and control group should follow similar trends prior to treatment. Since not all provinces have their first cohort taking CEE under new curriculum during the period 1999-2010 when data are available, I consider those provinces that adopted new curriculum after 2007 as my control group[9]. Figure 4.6 to 4.9 show patterns of CEE re-takers over time for treatment provinces and control provinces by year of new curriculum introduction. For example, Figure 4.6 shows the log number of exam re-takers in provinces that introduced the new curriculum in 2004, which had their fist cohort take CEE under new curriculum in 2007, relative to the control group with provinces that had not introduced the new curriculum by 2010.

From these figures, several patterns emerge. The first is that for all years, the number of CEE re-takers of adopting provinces in the treatment group exhibit similar trajectories with those non-adopting provinces before the switch of curriculum, which serves as a support for common trends assumption prior to treatment. Second, there is a huge decrease in all four years for number of CEE re-takers in adopting provinces at the year of introduction of new curriculum. For example,

certain qualifications, their eligibility of attending secondary education is still not allowed.

[9]For those provinces, they never had any cohort taking CEE under new curriculum before 2010.

Figure 4.6: Number of Exam Re-takers in 2007

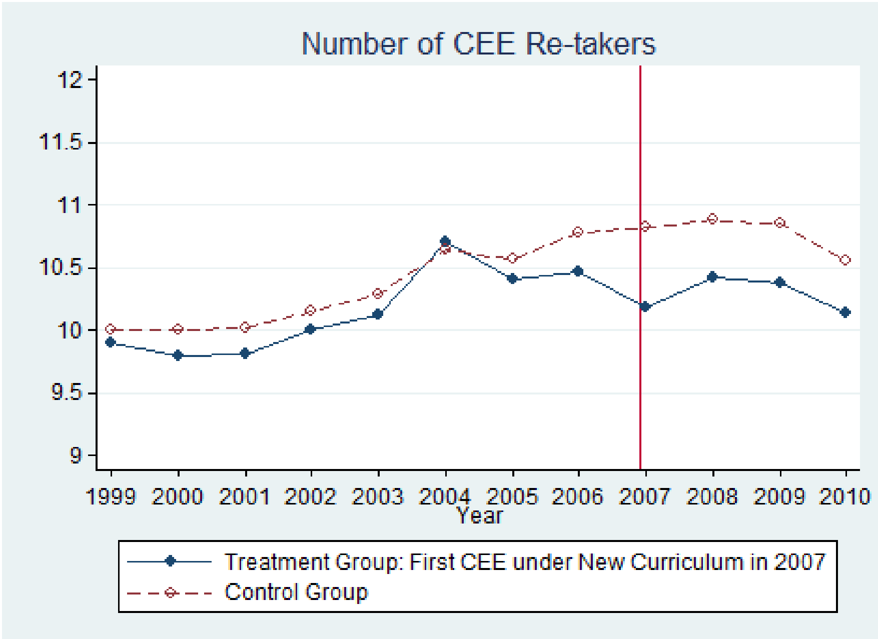


Figure 4.7: Number of Exam Re-takers in 2008

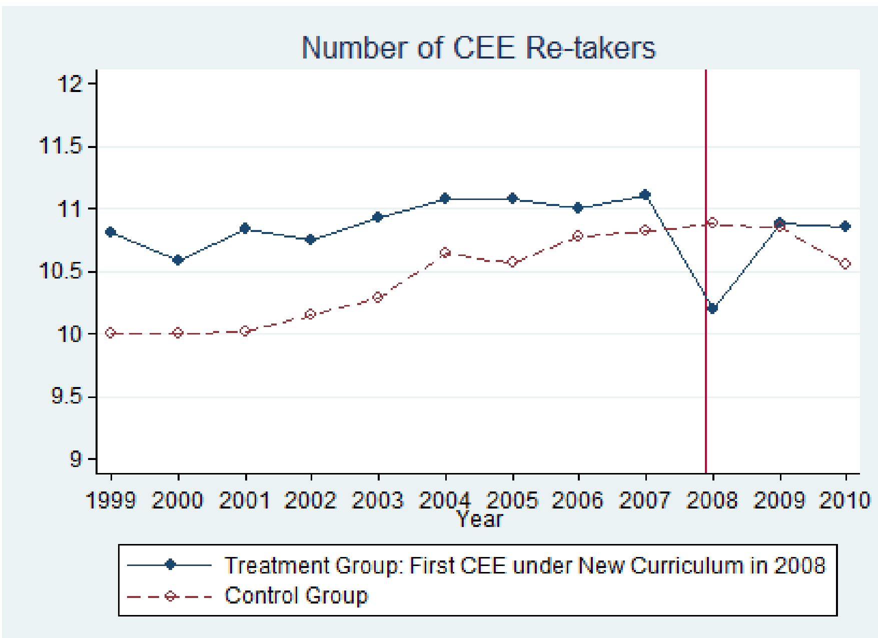


Figure 4.8: Number of Exam Re-takers in 2009

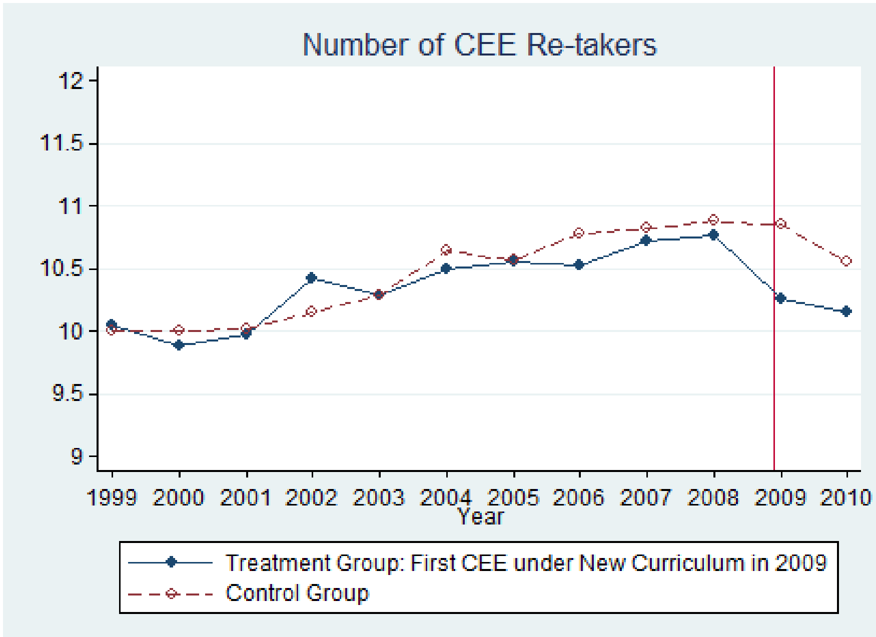


Figure 4.9: Number of Exam Re-takers in 2010

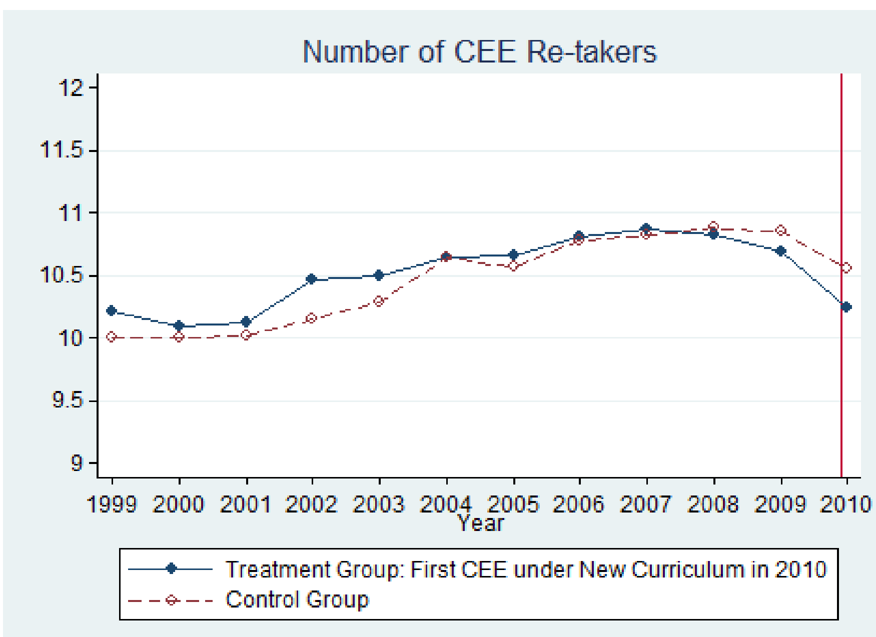


Table 4.1: New Curriculum and CEE Re-takers

	(1) Numbers	(2) Ln(Numbers)	(3) Ratio
New Curriculum	-16764.833***	-0.418***	-0.035***
	(4931.436)	(0.097)	(0.010)
Province FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R^2	0.860	0.952	0.803
Observations	372	372	341

Standard errors are clustered at the province level.

Re-taker ratio is as of the total number of previous-year exam takers.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

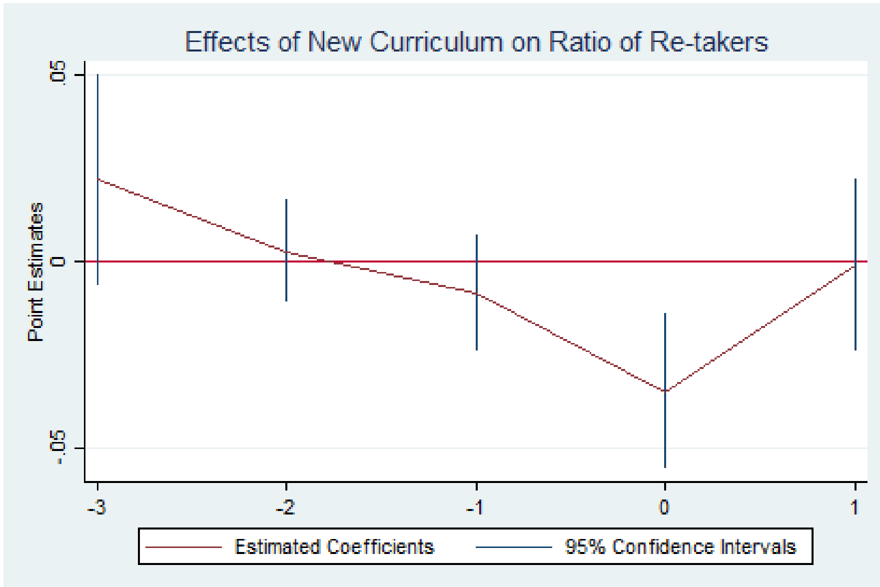
in Figure 4.6, in 2007, the average log number of CEE re-takers in the treatment group dropped from 10.5 to around 10 while there is a little increase for provinces in the control group. Third, the effect of curricula switch on CEE re-takers is only a one-period temporary effect while the number of CEE re-takers would rebound to normal levels after the switch. For example, the decrease of number of CEE re-takers in Figure 4.6 only occurred at the year 2007 when the new curriculum was introduced. After that it went back to the normal trajectory similar with that of provinces in the control group.

4.2.3 Results

The result from Equation 4.1 is shown in Table 4.1. After controlling for province and year fixed effects, according to Column 1, in the province at the year of the first cohort taking CEE under new curriculum, the re-takers of CEE drops by around 16,765 compared with other years, around 3.5% lower in Column 3. In addition, all results from different specifications are significant at the 1% level.

Moreover, a falsification test on estimates between different years relative to the introduction year of new curriculum and the ratio of CEE re-takers is displayed in Figure 4.10. For different cohorts, the coefficient is only significant at the year in which the new curriculum was introduced,

Figure 4.10: Estimated Coefficients on CEE Re-taking Ratio



which also supports the fact the the adoption of the new curriculum significantly reduced the CEE re-taking behavior.

Additionally, the data allow me to look into some heterogeneous effects between rural students and urban students. Anecdotal evidence shows that students from rural areas are more likely to re-take CEE since they value the education opportunity more than their urban counterparts given their comparatively disadvantageous family economic conditions. [2] use data collected from a second-tier university in *Sichuan* Province where rural students account for more than half, almost 40% students took the CEE at least twice. The results for urban and rural subsamples are shown in the last two columns in Table 4.2. The negative effect of the curricula switch on CEE re-taking behavior is larger for rural students.

In sum, results in this subsection establish the fact that the introduction of new curriculum causes the reduction of the CEE re-taking behavior, both for rural and urban students. In the subsection that follows, I will focus on how this changing opportunity costs from re-taking CEE affect household educational investments.

Table 4.2: Last Old Curriculum and Re-taking Ratio

	(1) Whole Sample	(2) Rural Sample	(3) Urban Sample
New Curriculum	-0.035***	-0.041***	-0.027**
	(0.010)	(0.011)	(0.011)
Province FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R^2	0.803	0.709	0.566
Observations	341	339	341

Standard errors are clustered at the province level.

Re-taker ratio is as of the total number of previous-year exam takers.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3 Effects on Household Educational Investments

4.3.1 Data

The data I use to analyze the effects of the changing exam re-taking opportunity costs on household educational investments are from the Beijing College Student Panel Survey (BCSPS), conducted by National Survey Research Center of Renmin University of China. This survey was conducted in several waves to make a panel. However, only the first wave is publicly accessible, which is used in this paper. In the first wave, the sample includes 4,711 college students from 15 universities in Beijing who came from all over China[10]. The questionnaire covers detailed information about students' demographics and family backgrounds, their grades in high school and score in CEE, their current academic, political performance in university, plans upon graduation as well as attitudes towards various aspects. The first wave was conducted in the summer of 2009, and interviewees are mainly freshmen and juniors, the majority of whom took CEE in 2006 or 2008.

Since the identification strategy relies on variations of students among the last cohorts under old curriculum and those who are not, in the sample, students who took the CEE in *Shandong,*

[10]Almost all these 15 universities are ranked top 50 in mainland China. As a result, samples drawn from this population consist of students who are regarded top ranked from CEE.

Guangdong, Hainan, Ningxia in 2006, in *Jiangsu* Province in 2007 as well as in *Beijing, Liaoning, Zhejiang, Anhui* and *Fujian* in 2008 are categorized into the treatment group, the number of which is 524, accounting for 11% of all observations in the sample.

Typical household investments in children's education during senior high school include asking private tutors or attending out-of-school classes, purchasing electronic devices or reference books that facilitate students' learning efficiency, or accommodating a comforting living environment by providing an independent study room or bedroom. Since this survey is mainly focused on the students' life in college while information about their high school is limited, I cannot measure expenditures on the household's investment on asking for private tutors or out-of-school classes. Instead, I mainly focus on investments on the latter two forms. Specifically, the data allow me to study whether students are endowed with a personal computer, Internet, cell phone, learning machine (e.g. electronic dictionary) reference books (e.g. encyclopedia, hard-copy dictionaries) in high school[11].

4.3.2 Empirical Model

Similar to the empirical strategy in Section 4.2, I rely on a generalized difference-in-differences model to estimate the effects of the changing opportunity costs in re-taking CEE caused by the introduction of new curriculum on household educational investments. The regression equation is as follows:

$$y_{i,j,c} = \beta * Last\_Old\_Curriculum_{j,c} + \gamma * X_{i,j,c} + Province_j + Cohort_c + \epsilon_{i,j,c} \tag{4.2}$$

where $y_{i,j,c}$ is a list of dummy variables indicating whether student i in province j at the high school entry cohort c was provided with the following equipments: a personal computer, Internet connection, learning machine and reference books during senior high school. $Last\_Old\_Curriculum_{j,c}$ is a dummy variable taking the value of 1 when province j adopts the new curriculum for high school entry cohort $c+1$ and 0 otherwise. Put differently, $Last\_Old\_Curriculum_{j,c}$ represents

[11] Admittedly, compared with expenditures on asking for private tutors, investments measured in this paper are small in amount and most of them are one-time expenditures.

the last cohort under old curriculum in province j. Therefore, β is the coefficient of interest, capturing the changing-opportunity-cost effect. $Province_j$ and $Cohort_c$ are full sets of province and cohort fixed effects. In addition, a list of control variables in $X_{i,j,c}$ is added to measure individual level characteristic and household level background that would affect the household educational investment decisions, including gender, ethnicity (i.e., whether the student is from a minority people), whether the student held a rural Hukou when in high school, whether the student had siblings, whether the student was from an elite senior high school. Also included are two variables indicating the family economic status and education level of the student's father, which are key determinants of levels of household educational expenditures in China ([11]). In the dataset, family annual income is provided, but it is related to the last year of the survey which corresponds to the year 2008 when more than half of students have already attended college for at least two years, which might reflect some measurement errors after the students graduated from senior high school. Instead, I use the self-reported family economic status variable to index for the family economic condition. My assumption is that the family comparative economic condition would not change since the student was in high school. In all specifications, standard errors are clustered at the province-cohort level.

My empirical model is similar to [5], who use their survey data in Peking University to study the effects of the introduction of the new curriculum on students' political attitudes and ideology. In this paper, I make use of the same framework with respect to the curriculum reform. However, I focus on different levels of CEE re-taking opportunity costs caused by this exogenous curriculum change. Therefore, my key independent variable is the dummy variable that takes the value of 1 if the student is among the last cohort under the old curriculum. Similar to their paper, this framework allows me to address several concerns about the identification method. First, since I have controlled for province fixed effects and exploited cross-cohort variation within provinces, the province-level differences in the valuation of entering into an elite college and the opportunity costs of re-taking CEE would not be a concern. Second, one might be concerned about the evolution of the attitudes towards re-taking CEE across cohorts without curriculum switch. In this framework, I would be able to difference out these cross-cohort changes by including the cohort fixed effects. Moreover, this empirical model would allow me to rule out confounding factors such as some province time-varying shocks that affect neighboring cohorts similarly. I will address any additional concerns later in robustness checks.

4.3.3 Balance of Student Characteristics

Summary statistics of the survey sample are provided in Column 1 from Table 4.3, while Column 3 and 5 show the mean characteristics of students by subgroups. One concern of the identification strategy is the lack of the balance resulted from different selection into treatment and control groups. Thus, I first check for balance of observed characteristics between treatment and control groups in the sample before presenting estimation results.

In Column 7, I provide differences of mean for measures of characteristics between treatment and control groups, while Column 8 presents p-values using the simple t-test with the null hypothesis that the differences of the two groups are indifferent from zero. It is clear to see that there exists significant differences across these two groups. [5] argue that this unconditional imbalance is to be expected due to the fact the new curriculum was introduced later in time and the assignment of introduction years across provinces were nonrandom, resulting in differences across students from different provinces. Thus, I follow their method by showing the differences between students across groups, conditional on province and cohort fixed effects, with the p-values testing for the statistical significance of these conditional differences. In all specifications, the standard errors are clustered at the province-cohort level. It is apparent that after accounting for average characteristics in the province where students spent their senior high school and accounting for average characteristics of a cohort, the differences of the two groups become statistically insignificant on most observable characteristics. The only significant characteristic is students' minority status. Specifically, the proportion of students (6.30%) who are among the last cohort under old curriculum is much smaller than that of students who are not (11.89%). One possible reason is that in the sample period, most of the minority-populous provinces and autonomous regions did not witness the change of curriculum (except *Ningxia Hui Autonomous Region*, which is the smallest

Table 4.3: Summary Statistics

	All		Last Curriculum=0		Last Curriculum=1		Unconditional		Conditional	
variable	Mean	SD	Mean	SD	Mean	SD	Diff	p-Value	Diff	p-Value
Personal										
Age	20.01	1.34	20.05	1.33	19.75	1.33	-0.30	0.00	-0.02	0.76
Male	0.53	0.50	0.52	0.50	0.57	0.50	0.05	0.03	0.00	0.96
Minority	0.11	0.32	0.12	0.32	0.06	0.24	-0.06	0.00	-0.05	0.00
Height	169.13	8.06	169.01	8.07	170.11	7.96	1.10	0.00	-0.21	0.51
Weight	59.74	11.30	59.70	11.41	60.07	10.32	0.38	0.47	-0.41	0.40
Health	0.70	0.46	0.70	0.46	0.71	0.45	0.01	0.63	0.02	0.43
Parents										
Father is CCP	0.43	0.50	0.43	0.50	0.43	0.50	0.00	0.94	-0.01	0.61
Father urban	0.28	0.45	0.28	0.45	0.30	0.46	0.02	0.39	0.02	0.47
Father Education Level	5.55	2.29	5.55	2.29	5.53	2.30	-0.03	0.80	-0.07	0.69
Family Rural Hukou	0.28	0.45	0.28	0.45	0.31	0.46	0.03	0.13	0.03	0.24
Number of Siblings	0.49	0.79	0.49	0.78	0.46	0.83	-0.03	0.42	-0.04	0.36
Ln (annual family income)	10.59	1.14	10.58	1.14	10.67	1.09	0.09	0.09	0.03	0.53
Family social status=high	0.20	0.40	0.20	0.40	0.25	0.43	0.05	0.01	0.04	0.17
Education										
CCP member in high school	0.06	0.24	0.05	0.23	0.10	0.30	0.05	0.00	0.01	0.53
High school is elite	0.77	0.42	0.76	0.43	0.87	0.34	0.11	0.00	0.02	0.25
Without CEE	0.05	0.23	0.06	0.24	0.01	0.10	-0.05	0.00	0.00	0.49

Notes: The last two columns report differences conditional on cohort and province fixed effects.

among the five autonomous regions). Thus, it is reasonable to expect not many minority students among the last cohort under old curriculum. I would also include the minority dummy in the baseline model to control for any systematic differences between Han people and those minorities. In addition, I will restrict my sample with only Han students in the robustness check.

Moreover, Table 4.4 shows the comparison of means for a list of outcome variables. From the simple comparison, it is difficult to tell any differences between the last cohort under old curriculum and other cohorts. Thus it is interesting to see whether such differences in household educational investment exist after controlling for other characteristics including both province and year fixed effects.

4.3.4 Regression Results

The estimation results using the generalized difference-in-differences model in Equation 4.2 are provided in Table 4.5. I estimate the curriculum reform on household educational investment such as equipments and devices that can facilitate students' learning efficiency. Most of the coefficients of the last old curriculum variable are significant at least at the 5% level. Households with students from the last cohort under old curriculum would increase their investment in laptops, reference books, learning machines, PCs as well as Internet. Compared with the mean of the cohorts that are not under the last old curriculum, the possibility that the household has laptops is 13.20% higher while it is only 3.25% higher for a household that has a student under the last old curriculum to be equipped with reference books.

Another type of household educational investments is by providing comforting living environments for their children in high school, such as independent bedroom or study room. However, I did not find evidence of such investments in the sample. In Table 4.6, none of the coefficients with respect to the last cohort under old curriculum dummy are significant. One interpretation of the insignificant result is that unlike the electronics, living conditions are less likely to be accommodated, as household would be less likely to buy a larger apartment simply because their children are among the cohort under the last curriculum and their opportunity costs of re-taking CEE is higher. Moreover, in some high schools in China, students who live far from campus may choose

Table 4.4: Summary Statititics by Last Old Curriculum Cohort

	Last Old=0	Last Old=1
Laptop	0.095	0.109
	(0.293)	(0.311)
Reference Books	0.768	0.799
	(0.422)	(0.401)
Learning Machines	0.605	0.605
	(0.489)	(0.489)
Internet	0.616	0.644
	(0.486)	(0.479)
PC	0.677	0.708
	(0.468)	(0.455)
Study Room	0.541	0.596
	(0.498)	(0.491)
Own Room	0.832	0.874
	(0.374)	(0.333)
Cell Phone	0.334	0.298
	(0.472)	(0.458)
Leisure Books	0.626	0.660
	(0.484)	(0.474)
Video Games	0.320	0.246
	(0.467)	(0.431)
Observations	4207	562

Means and standard errors (in parentheses) are presented.

Table 4.5: Last Old Curriculum and Household Investment

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
Last Old Curri	0.022***	0.055**	0.023	0.038**	0.034**
	(0.007)	(0.021)	(0.016)	(0.017)	(0.016)
Male	0.010	-0.024	-0.066***	-0.032***	-0.024**
	(0.008)	(0.019)	(0.014)	(0.010)	(0.010)
Minority	-0.009	-0.003	0.012	0.028*	0.001
	(0.013)	(0.023)	(0.021)	(0.016)	(0.020)
Rural	-0.038**	-0.253***	-0.267***	-0.368***	-0.345***
	(0.016)	(0.022)	(0.022)	(0.025)	(0.022)
Elite HS	-0.005	-0.004	0.002	-0.031	-0.023
	(0.010)	(0.017)	(0.013)	(0.019)	(0.020)
Siblings	-0.066***	-0.074***	-0.103***	-0.142***	-0.131***
	(0.016)	(0.015)	(0.018)	(0.020)	(0.016)
Family Status	0.018	0.016	0.088***	0.088***	0.064***
	(0.016)	(0.019)	(0.026)	(0.026)	(0.023)
Father Edu	-0.002	0.038***	0.037***	0.048***	0.048***
	(0.006)	(0.005)	(0.005)	(0.007)	(0.008)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.044	0.263	0.272	0.458	0.457
Observations	4450	4452	4452	4451	4452

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to live in dormitories inside schools[12].

Besides the key independent variable, the relationship between other control variables included in the estimation equation and household educational investments are also shown in both tables. Most of these variables are in line with expectation. For example, male students are less likely to own learning devices. Being from a rural household significantly reduces the possibility being provided with learning devices and good living environment, which reflect huge inequalities with respect to educational expenditures between rural and urban households.

In addition, students from wealthier families or with more educated father are likely to be accommodated with good living conditions. This result is in line with [11], who also find empirically that household income and parents' levels of education are key determinants of household educational investments.

As a measure of placebo test, in Table 4.7, I show the result using the same framework only with the outcome variable being replaced with a list of devices that are mostly used for leisure, such as leisure books, cell phones and PSP. As expected, these results are negative and insignificantly different from zero.

In addition, I also conduct a series of falsification tests, I regress the outcome variables on the dummy variable taking the value 1 if the cohort is among the second last cohort under the old curriculum. Results in Table 4.8 show that most of the coefficients are not significantly different from zero and some are even negative, implying that this difference is not likely to exist before the switch of curricula.

4.3.5 Robustness Checks and Additional Discussions

I then explore the robustness of the baseline results from the previous subsection.

When comparing the balances of the sample divided by whether students are among the last cohort under old curriculum or not, I find the proportion of minorities is significantly smaller for the treatment group. In order to address the concern that households from minority groups might have

[12]Such living information is not available in the dataset, so I cannot directly test this hypothesis by simply restricting the sample to those who commute to school from home everyday.

Table 4.6: Last Old Curriculum and Household Investment

	(1) Study Room	(2) Independent Bedroom
Last Old Curri	-0.008	0.028
	(0.019)	(0.025)
Male	-0.014	0.011
	(0.013)	(0.012)
Minority	0.002	-0.017
	(0.027)	(0.018)
Rural	-0.205***	-0.156***
	(0.039)	(0.029)
Elite HS	0.014	0.025*
	(0.021)	(0.014)
Siblings	-0.089***	-0.058***
	(0.021)	(0.019)
Family Status	0.159***	0.068***
	(0.024)	(0.009)
Father Edu	0.051***	0.024***
	(0.004)	(0.004)
Province FE	Yes	Yes
Cohort FE	Yes	Yes
R^2	0.222	0.159
Observations	4449	4452

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Last Old Curriculum and Household Investment

	(1) Cell Phone	(2) Leisure Books	(3) PSP
Last Old Curri	-0.014	-0.011	-0.025
	(0.025)	(0.024)	(0.026)
Male	-0.010	0.004	0.164***
	(0.015)	(0.012)	(0.018)
Minority	-0.005	-0.024	0.064***
	(0.020)	(0.019)	(0.019)
Rural	-0.071*	-0.257***	-0.112***
	(0.039)	(0.020)	(0.022)
Elite HS	-0.008	-0.010	-0.039**
	(0.021)	(0.015)	(0.018)
Siblings	-0.059**	-0.058***	-0.015
	(0.024)	(0.019)	(0.018)
Family Status	-0.004	0.055***	0.148***
	(0.039)	(0.021)	(0.019)
Father Edu	0.011**	0.069***	0.011*
	(0.004)	(0.004)	(0.006)
Province FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
R^2	0.072	0.298	0.096
Observations	4449	4452	4452

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: Second Last Old Curriculum and Household Investment

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
2nd Last Old	-0.016	-0.045**	0.019	0.021	-0.009
	(0.017)	(0.020)	(0.024)	(0.026)	(0.027)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.044	0.262	0.272	0.458	0.456
Observations	4450	4452	4452	4451	4452

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.9: Robustness: Han Only

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
Last Old Curri	0.020**	0.056***	0.010	0.042**	0.041**
	(0.008)	(0.021)	(0.016)	(0.017)	(0.019)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.048	0.256	0.266	0.466	0.463
Observations	3960	3962	3962	3961	3962

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

different value on education, especially when students of minority peoples often enjoy extra points in CEE, I restrict the sample to only students from Han People, which account for 88.45% for all observations. The result from Table 4.9 shows that the coefficients of the last cohort under old curriculum do not differ much from the baseline result, both in magnitude and level of significance.

Another issue in my baseline sample is that all students are included regardless of whether they are under the new curriculum or old one. One might be concerned that some topics from the new curriculum might require students to be equipped with more advanced learning devices or

Table 4.10: Robustness: Old Curriculum Only

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
Last Old Curri	0.022***	0.056***	0.022	0.037**	0.033**
	(0.007)	(0.021)	(0.016)	(0.017)	(0.016)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.044	0.262	0.272	0.459	0.456
Observations	4370	4372	4372	4371	4372

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

teachers may encourage students to use such devices more and search for information, either online or from reference books, leading students to behave differently when they are under different curricula. First, I use similar strategy as the placebo test and replace my key independent variable with the dummy variable taking on the value of 1 for the first cohort in the province under new curriculum, results show insignificant coefficients. Second, I limit the sample to cohorts only under old curriculum. The results from Table 4.10 imply that this is not a big issue.

Another question about the empirical analysis is whether the adoption of the new curriculum coincided with some provincial variations that might affect household level input at the province-cohort level. One might argue that government spending on secondary education is a good substitute for household educational expenditures on children in senior high school. For example, [52] provide such evidence from student in primary and junior high school. If senior high schools have already provided plenty of computers and Internet connection, or students can easily obtain reference books from school library, then some of the effects may be driven by school spending change. I thus follow [5] to control for provincial spending on secondary education at the province-cohort level that is calculated as a province's average level of spending during the 3 years of senior high school for each cohort. The data are from the *Educational Finance Statistical Yearbook of China* (2002-2008) published by the Department of Finance in the Ministry of Education. As can be seen

Table 4.11: Robustness: Province-Cohort level Control

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
Last Old Curri	0.018**	0.055***	0.017	0.030*	0.030**
	(0.007)	(0.021)	(0.013)	(0.017)	(0.014)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.045	0.263	0.273	0.460	0.457
Observations	4450	4452	4452	4451	4452

Standard errors are clustered at the province-cohort level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

from Table 4.11, the inclusion of the province-cohort level control variable does not qualitatively change the estimated effects of changing opportunity costs on household educational investments.

Moreover, as has been mentioned before, the introduction year of the new curriculum is not randomly assigned across provinces. One might be interested in the determinants of a province adopting the new curriculum in a particular year and be concerned that these factors might also affect household investment. According to the [5], their most robust finding is that greater province income in 2003 is quite predictive of earlier introduction of the new curriculum while educational variables are generally less predictive. To empirically dealing with this concern, I follow their strategy to control for the interaction between a province's 2003 Gross Regional Product per capita and cohort fixed effects. From the results shown in Table 4.12, one can see that the findings are not affected after these controls are included.

Furthermore, it is interesting to see whether there exist some long-run effects of this extra investment from household when students have been enrolled in college. Since students in the sample are from different majors in different universities, their scores in classes are not comparable. Instead, I use their College English Test-4 (CET-4) scores as a proxy for some long-run effects. CET-4 is a nationwide English test for all undergraduates and passing this test is one of the most important requirements for graduation. In addition, across years, the test is in similar levels of

Table 4.12: Robustness: Controlling for 2003 Provincial GRP*Cohort FE

	(1) Laptop	(2) Reference Books	(3) Learning Machine	(4) Internet	(5) Computer
Last Old Curri	0.022***	0.055**	0.023*	0.039***	0.035***
	(0.007)	(0.021)	(0.014)	(0.013)	(0.013)
Province FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
R^2	0.046	0.263	0.274	0.460	0.457
Observations	4388	4389	4389	4388	4389

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

difficulty and has the same passing grade, making it a reasonable proxy for learning outcomes for comparison of students across universities and years. The result from Table 4.13 indicates that such long-run effect is minimal, implying there might not be a significant difference for students due to such extra investment after they enter college.

Table 4.13: Last Old Curriculum and Long-run Effects

	(1)
	CET-4 Score
Last Old Curri	-3.198
	(12.749)
Male	-22.001***
	(2.126)
Minority	-13.768***
	(4.341)
Rural	-4.618
	(3.851)
Elite HS	8.320
	(5.747)
Siblings	-4.247
	(3.367)
Family Status	16.025***
	(4.910)
Father Edu	3.867***
	(1.334)
Province FE	Yes
Cohort FE	Yes
R^2	0.202
Observations	2167

Standard errors are clustered at the province-cohort level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5. SUMMARY AND CONCLUSIONS

In this dissertation, I study the development of China in different perspectives.

First, a large wave of newly added households in urban areas exists simply because the areas they live in have been reclassified from rural to urban. These in situ migrants comprised almost 35% of total urban population growth from 2010 to 2015; they had little demand in urban housing markets since they "bring" their own housing units into urban areas. To quantify its scale, I simply combine our estimated scale of in situ migrants with information regarding their living conditions from descriptive evidence found in CHFS datasets. Results show that between 2010 to 2015, in situ migrants account for 33.65 million, which is approximately 8.88 million households from reclassified communities. Around 94% of these households have their own housing units. As a result, they brought around 8.35 million housing units into urban areas. Combining results from [7] by employing similar assumptions, from 2000 to 2010, these in situ migrants brought around 30.4 million housing units into urban areas. In total, during the first 15 years of this new millennium, the scale of newly added urban households from in situ migrants almost 40% and they brought around 38.75 million housing units into urban areas, accounting for nearly 80% of total number of vacant housing units estimated by CHFS in 2014. This result provides evidence in favor of the high vacancy rate found by CHFS. In fact, when comparing the added urban housing units and urban households, the scale of in situ migrants has been taken into account in calculating the increment of newly added urban households but the housing units brought by these in situ migrants are not taken into account in considering the increments of newly added urban housing units. This mismatch due to neglecting the group of in situ migrants causes excess supply which in turn leads to large numbers of vacant housing units in urban China.

Accompanied by rapid economic development, in 2016 China's urbanization rate reached 58.4%. A large wave of the population migrated and settled in urban areas, causing a consistent demand for housing. As a result, urbanization is considered to be an important factor for the enlarged urban housing demand. In this paper, however, I find that a large share of all new urban residents during the first half of the 2010s actually came from communities that were reclassified from rural to urban. Relying on the rural-urban division codes which are issued by the National Bureau of Statistics every year, I estimate that around 35% of the total urban population growth from 2010 to 2015 resulted from rural-urban dichotomy reclassification.

Further exploration of the demographic and economic characteristics of reclassified communities and in situ migrants finds descriptive evidence that these communities are far less urbanized than those in urban areas; most households still live in houses that they built on collectively-owned land, and have little demand to participate in local housing markets. I then use prefecture-level land supply data and our measures of urban population growth to empirically estimate the relationship between local land supply and different types of urban population growth. Our results show that the elasticity of the scale of in situ migrants is both larger in magnitude and more significant than that of real urban population growth. If this is the case, I can anticipate large amounts of surplus housing units, especially in cities with large scales of in situ migrants. All of this evidence supports a recent finding from the China Household Finance Survey that there exist around 50 million vacant housing units in urban China.

Secondly, motivated by the theory that property rights and well-functioning institutions are important for development and given the fact that findings of effects of securing agricultural land tenure are mixed in China, I focus on the rural land certification program in China's countryside and empirically test its effects on agricultural land utilization at the community level by adopting a panel dataset from the community questionnaires of China Household Finance Survey in 2015 and 2017. Using panel data fixed-effects model, I have three main findings.

First, my results show that land certification has a positive effect on reducing abandoned agricultural land. Moreover, the effects are more significant in high-income villages or those located in provinces with higher share of GDP from agricultural sector. Second, I do not find significant relationships between secure land tenure and other outcomes such as the improvement in productivity or off-farm labor participation. In addition, though I do not find a significant effect on overall land transfer activity, there is descriptive evidence from the most recent round of survey showing that in certified villages, the size of land transferred to agricultural business entities is much larger than that in non-certified villages and the rent is also higher. I also find supporting results using my household-level panel data.

Land is a key asset for rural household. In China, given the fact that the property right is separated from its ownership right, keeping tenure secure is of great importance for stabilizing relationship between rural households and collectives. In this paper, findings from this paper only suggest a improvement in land utilization. As a result, endowing rural households with full property security is still essential for rural development.

Finally, I document how the curriculum change guided by the Chinese Ministry of Education has caused different levels of opportunity costs with respect to re-taking the College Entrance Exam in mainland China. Furthermore, I show how this changing levels of opportunity costs result in different household educational investments on senior high school students. I first employ provincial-level panel data and show that the CEE re-taking is significantly lower at the year when provinces switched senior high school curricula. Due to the change of the curriculum and the CEE guideline, the opportunity costs of a student to re-take CEE are much higher for the last cohort under old curriculum. I then adopt the Beijing College Student Panel Survey to empirically test whether this exogenous policy change leads to different household investment behavior on students' education. Due to the limitation of the dataset, it is inaccessible to test expenditures such as asking for private tutors and attending out-of-school classes. Instead, I test this change on the household investments in electronic devices as well as comforting living environments.

Results show that effects of the changing opportunity costs by curricula switch on the household educational investment are noticeable. Households with children in the last cohort under old curriculum would invest more on electronic devices that improve their learning efficiency. However, I do not find such investment has any long-run effect when their CET-4 test scores are used for proxy, implying that these inputs might be only helpful when students are in senior high school.

REFERENCES

[1] Simon Alder, Lin Shao, and Fabrizio Zilibotti. Economic reforms and industrial policy in a panel of Chinese cities. *Journal of Economic Growth*, 21(4):305–349, 2016.

[2] Chong-en Bai, Wei Chi, and Xiaoye Qian. Do college entrance examination scores predict undergraduate GPAs? a tale of two universities. *China Economic Review*, 30:632–647, 2014.

[3] Ying Bai and Ruixue Jia. Elite recruitment and political stability: the impact of the abolition of China's civil service exam. *Econometrica*, 84(2):677–733, 2016.

[4] Timothy Besley. Property rights and investment incentives: Theory and evidence from Ghana. *Journal of Political Economy*, 103(5):903–937, 1995.

[5] Davide Cantoni, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. Curriculum and ideology. *Journal of Political Economy*, 125(2):338–392, 2017.

[6] Kam Wing Chan and Ying Hu. Urbanization in China in the 1990s: New definition, different series, and revised trends. *China Review*, pages 49–71, 2003.

[7] Qin Chen and Zheng Song. Accounting for China's urbanization. *China Economic Review*, 30:485–494, 2014.

[8] Ting Chen, James Kai-sing Kung, and Chicheng Ma. Long live Keju! the persistent effects of China's imperial examination system. *The Persistent Effects of China's Imperial Examination System (June 26, 2017)*, 2017.

[9] Lingguo Cheng, Ye Zhang, and Zhibiao Liu. Does land titling improve land rental markets in rural China. *Management World*, 1:88–98, 2016.

[10] Yuk-Shing Cheng and Kim-Sau Chung. Designing property rights over land in rural China. *The Economic Journal*, 128(615):2676–2710, 2017.

[11] Wei Chi and Xiaoye Qian. Human capital investment in children: An empirical study of household child education expenditure in China, 2007 and 2011. *China Economic Review*, 37:52–65, 2016.

[12] Peter Christensen and Gordon C McCord. Geographic determinants of China's urbanization. *Regional Science and Urban Economics*, 59:90–102, 2016.

[13] Alan De Brauw and John Giles. Migrant opportunity and the educational attainment of youth in rural China. *Journal of Human Resources*, 52(1):272–311, 2017.

[14] Alain De Janvry, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. Delinking land rights from land use: Certification and migration in Mexico. *American Economic Review*, 105(10):3125–49, 2015.

[15] Klaus Deininger and Songqing Jin. The potential of land rental markets in the process of economic development: Evidence from China. *Journal of Development Economics*, 78(1):241–270, 2005.

[16] Klaus Deininger and Songqing Jin. Tenure security and land-related investment: Evidence from Ethiopia. *European Economic Review*, 50(5):1245–1277, 2006.

[17] Klaus Deininger and Songqing Jin. Securing property rights in transition: Lessons from implementation of China's rural land contracting law. *Journal of Economic Behavior & Organization*, 70(1-2):22–38, 2009.

[18] Harold Demsetz. Toward a theory of property rights. In *Classic papers in natural resource economics*, pages 163–177. Springer, 1974.

[19] Xiangzheng Deng, Jikun Huang, Scott Rozelle, and Emi Uchida. Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics*, 63(1):96–115, 2008.

[20] Jianyong Fan, Jiawei Mo, and Jipeng Zhang. Housing models and urbanization in China - empirical research from the perspective of land supply. *Social Sciences in China*, 4, 2015.

[21] Francesc Gallart and Pilar Llorens. Catchment management under environmental change: impact of land cover change on water resources. *Water International*, 28(3):334–340, 2003.

[22] John Giles and Ren Mu. Village political economy, land tenure insecurity, and the rural to urban migration decision: evidence from China. *American Journal of Agricultural Economics*, 100(2):521–544, 2017.

[23] Edward Glaeser and J Vernon Henderson. Urban economics for the developing world: An introduction. *Journal of Urban Economics*, 98:1–5, 2017.

[24] Robert Haveman and B Wolfe. A review of methods and findings. *Journal of Economic Literature*, 33(4):1829–1878, 1995.

[25] J Vernon Henderson. Urbanization in China. *The Oxford Companion to the Economics of China*, page 225, 2014.

[26] Franz Höchtl, Susanne Lehringer, and Werner Konold. "Wilderness": what it means when it becomes a reality a case study from the southwestern Alps. *Landscape and Urban Planning*, 70(1-2):85–95, 2005.

[27] Ruixue Jia and Hongbin Li. Access to elite education, wage premium, and social mobility: the truth and illusion of China's college entrance exam. *Stanford Center for International Development, Working Paper*, (577), 2016.

[28] Songqing Jin and Klaus Deininger. Land rental markets in the process of rural structural transformation: Productivity and equity impacts from China. *Journal of Comparative Economics*, 37(4):629–646, 2009.

[29] Michael Kaganovich and Itzhak Zilcha. Education, social security, and growth. *Journal of Public Economics*, 71(2):289–309, 1999.

[30] John Knight and Li Shi. Educational attainment and the rural-urban divide in China's. *Oxford Bulletin of Economics and Statistics*, 58(1):83–117, 1996.

[31] Jean O Lanjouw and Philip I Levy. Untitled: A study of formal and informal property rights in urban Ecuador. *The Economic Journal*, 112(482):986–1019, 2002.

[32] Jessica Leight. Reallocating wealth? insecure property rights and agricultural investment in rural China. *China Economic Review*, 40:207–227, 2016.

[33] Xuan Li. The political cycle in China's primary land market. 2018.

[34] Xuesong Li and Yanyan Huang. Rising housing price, multi-suite decision and household savings rate in urban China. *Economic Research Journal*, 9:100–113, 2015.

[35] Zanhong Li, Jianzhong Yan, Xiaobo Hua, Liangjie Xin, and Xiubin Li. Factors influencing the cultivated land abandonment of households of different types: A case study of 12 typical villages in Chongqing municipality. *Geographical Research*, 33(4):721–734, 2014.

[36] Erik Lichtenberg and Chengri Ding. Assessing farmland protection policy in China. *Land Use Policy*, 25(1):59–68, 2008.

[37] Erik Lichtenberg and Chengri Ding. Local officials as land developers: Urban spatial expansion in China. *Journal of Urban Economics*, 66(1):57–64, 2009.

[38] Teresa Mauldin, Yoko Mimura, and Mark Lino. Parental expenditures on children's education. *Journal of Family and Economic Issues*, 22(3):221–241, 2001.

[39] Ren Mu and Yang Du. Pension coverage for parents and educational investment in children: Evidence from urban China's. *The World Bank Economic Review*, 31(2):483–503, 2015.

[40] Katrina Mullan, Pauline Grosjean, and Andreas Kontoleon. Land tenure arrangements and rural–urban migration in China. *World Development*, 39(1):123–133, 2011.

[41] Douglass C North. Institutions, institutional change and economic performance. *New York*, 1990.

[42] Joanne Xiaolei Qian and Russell Smyth. Educational expenditure in urban China's: income effects, family characteristics and the demand for domestic and overseas education. *Applied Economics*, 43(24):3379–3394, 2011.

[43] YJ Tian, XB Li, LJ Xin, GX Ma, and ZM Li. Impacts of the rise of labor opportunity cost on agricultural land use changes: A case study of Ningxia Hui Autonomous Region. *Journal of Natural Resources*, 24(3):369–377, 2009.

[44] Fang Wang. The composition of urban population growth between 4th census and 5th census. *Population Research*, 28(3):60–67, 2004.

[45] Hui Wang, Jeffrey Riedinger, and Songqing Jin. Land documents, tenure security and land rental development: Panel evidence from China. *China Economic Review*, 36:220–235, 2015.

[46] Shing-Yi Wang. State misallocation and housing prices: theory and evidence from China. *American Economic Review*, 101(5):2081–2107, 2011.

[47] Haiping Xue and Xiaohao Ding. A study on additional instruction for students in cities and towns in China's. *Educational Research*, 2009.

[48] Jianzhong Yan, Ziyan Yang, Zanhong Li, Xiubin Li, Liangjie Xin, and Laixiang Sun. Drivers of cropland abandonment in mountainous areas: A household decision model on farming scale in Southwest China. *Land Use Policy*, 57:459–469, 2016.

[49] Dennis Tao Yang. China's land arrangements and rural labor mobility. *China Economic Review*, 8(2):101–115, 1997.

[50] Han Yu and Naci Mocan. The impact of high school curriculum on confidence, academic success, and mental and physical well-being of university students. Technical report, National Bureau of Economic Research, 2018.

[51] Hongwen Yu. Traits of population situation seen through 2000 census. *Population Research*, 25(4):12–18, 2001.

[52] Cheng Yuan and Lei Zhang. Public education spending and private substitution in urban China's. *Journal of Development Economics*, 115:124–139, 2015.

[53] Jipeng Zhang, Jianyong Fan, and Jiawei Mo. Government intervention, land market, and urban development: Evidence from Chinese cities. *Economic Inquiry*, 55(1):115–136, 2017.

[54] Junsen Zhang, Yaohui Zhao, Albert Park, and Xiaoqing Song. Economic returns to schooling in urban China's, 1988 to 2001. *Journal of Comparative Economics*, 33(4):730–752, 2005.

[55] Kevin Honglin Zhang and Shunfeng Song. Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Economic Review*, 14(4):386–400, 2003.

[56] Yu Zhang. Does private tutoring improve students' national college entrance exam performance? a case study from Jinan, China's. *Economics of Education Review*, 32:1–28, 2013.

[57] Yu Zhang and Xuehan Zhou. Can higher household education expenditure improve the national college entrance exam performance? empirical evidence from Jinan, China's. *Current Issues in Comparative Education*, 19(2):8–32, 2017.

[58] Siqi Zheng and Matthew E Kahn. Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics*, 63(2):743–757, 2008.

[59] Siqi Zheng, Weizeng Sun, Jianfeng Wu, and Matthew E Kahn. The birth of edge cities in China: Measuring the effects of industrial parks policy. *Journal of Urban Economics*, 100:80–103, 2017.