Assignment name: MA thesis
Name: Zechen Li
Handed in: 2019-05-29 22:45
Generated at: 2019-06-24 19:58

LUND UNIVERSITY
School of Economics and Management

Master's Programme in Economic Growth, Population and Development (Economic Demography)

A Study of the Relationship Between Crude Birth Rate and Population Density in Xinjiang Uyghur Autonomous Region in China

by

Zechen Li

ze8315li-s@student.lu.se

Abstract: Recently, the urbanization in China caused much social problem in China, and one of the distinctive features of urbanization is the increasing population density. This thesis aims to discover the relationship between population density and fertility in China by choosing Xinjiang region as an example, other determinants of fertility in Xinjiang region were also examined, such as average income, employment rate, ethnic minority rate and urban population rate. In this thesis, the data was collected from Xinjiang statistical yearbooks the period was from 1996 to 2013. In this thesis, the quantitative method was used, and the models were formed based on the fixed-effect model and random effect model, there was also the OLS model added to compare. In results, it is discovered that there is no significant relationship between population density and crude birth rate, whereas the economic factors GDP per capita has a positive impact on the crude birth rate. The one-child policy generally has no significant impact on the crude birth rate, and there is no notable difference between urban and rural fertility. The employment rate has a positive impact on fertility level, which contradicts the mainstream theory. The general conclusion is that Chinese fertility has a unique feature compare with other countries, but further research on other regions in China is needed to support this thesis.

Keywords: population density; fertility; one-child policy; China; Xinjiang region

EKHM01

Master's Thesis (15 credits ECTS)

June 2019

Supervisor: Luciana Quaranta

Examiner:

Word Count: 13623

Acknowledgements

I would first like to thank my thesis supervisor Luciana Quaranta of the department of Economy History at Lund Unversity. It was always easy to get in touch with her whenever I ran into a trouble or had any questions about my research. I would also like to thank for Phd. Finn Hedefalk for the initial help and my work will not be accomplish without his suggestion of the literatures in the beginning of this semester. I also would like to thank for my parents for the financial and mental support during the whole year, and my friends, Sebstian Maric, who suggested me the data source, Wooseong Kim who gave me a lot of help on Stata work.

Table of Contents

List of Tables

List of Figures

1 Introduction

In the past few decades, urbanization has accelerated across the world, especially in developing countries such as India, China (Wang Q. Su M. Li R., 2018), Vietnam (Fan, P. Park, H. Ouyang, Z. Nguyen, D.D Nguyen, T.T.H. Chen, J., 2018) and so on. Moreover, the population grows very fast in large cities. According to researches done by Tong Y. & Wu Y. (2012), the urban population in China has multiplied four-fold, and the rate of urbanization has doubled ever since 1978.

The rapid population growth of the urbanization also causes significant problems in several regions, some of the social problems (Li, 2013) and the migration from rural areas to the cities also caused the rises of population density in cities (Malpezzi, 2013). According to the research done by Li (2013), this phenomena in China has caused a lot of problems, some of which are related to the rural population are newly converted into the urban population. It takes time and effort for them to adapt to the society of cities (Li, 2013). As the migration reshapes city life, the cities are also facing the pressure of increasing migration. The migration has posed some new challenges, such as the social structure changes and institution changes in the city and even in the state level (Zhang & Song, 2003).

As the urbanization rate is increasing, the population density in Chinese cities is also increasing quickly. Some Chinese regions are being too crowded to maintain the living standard, and some of the residences reduce the number of children in their family (Tsui, M., 1989). Also, the research done by Lutz & Qiang (2002) suggested that there is a stable relationship between fertility rate and population density, the higher the population density is, the lower the fertility rate. What is more, the research also suggested that the population density has an impact on other human behaviours, such as fertility behaviour (Lutz, W., M. Testa, and D. Penn. 2007). In order to check if population density can be a contributing factor of fertility rate in China, the research based on Chinese data will be the theme of this research. In this thesis I will choose Xinjiang Uyghur Autonomous region (hereafter referred to as Xinjiang region) as the sample, the reason why this province has been chosen is that this province was least affected by the one-child policy. Moreover, the reason why it is least affected will be explained in the theory section.

Generally speaking, the one-child policy was looser for the rural population and ethnic minority population (Xinjiang regions government, 2010). Xinjiang region has more than 50% of the ethnic minority population. Furthermore, in 2015, the Xinjiang region had the highest fertility rate among all the provinces in China (Chinese national statistical bureau, 2016). This information will be addressed again in the theory part in detail.

Figure 1.1 the location of Xinjiang region

Source: Google map

For all the above reasons, I picked the Xinjiang region as the sample of the thesis. By choosing the Xinjiang region, it will be more evident that how the different social and economic factors impact the fertility level without any policy disturbance. Because the impact of one child policy was the weakest in the Xinjiang region, also, in this thesis, the county and city will be chosen as the thesis objects. County in China is the administrative division under the prefecture, and the prefecture is the administrative division under the province (World Bank, 2012). One thing to notice is there are three levels of cities in China, here I choose the county-level cities because in this way the city and county are more comparable.

In this thesis, I also investigate the difference between the cities and the counties. As the national statistical bureau addressed, the city is a region that contains mostly urban area, whereas the county is where contains both urban and rural area (Chinese national statistical bureau, 2019). Which means the socioeconomic feature of these two types of regions is different. Therefore there will be introducing the different model to the counties and cities area later. The reason why I choose county/city level data as the object of this thesis is that this is the smallest unit that is available for these data, such as GDP per capita, crude birth rate and so on. The data is only available for four administrative division levels, which are the county, prefecture, province and country level, by chosen the smallest unit the more detailed information I am going to use in this thesis.

1.1 Aim and Scope

This thesis will focus on the impacts of population density on fertility in cities and counties of China. The density changes can affect human fertility behaviours. Research shows that the higher population densities can result in a lower fertility rate and effect other demographic features (Croix & Gobbi, 2017; Lutz et al., 2007). This thesis will explore more about the relationship between the two elements above.

As the last section addressed, there is some research about the negative correlation between population density and fertility level. Furthermore, China as the most populated country is suffering from overpopulation and at the same time, cities are facing low crude birth rate and ageing society. By researching the relationship between the population density and fertility level of the large cities, it can provide a new angle of how to look at the fertility problem in the cities of China and by this new angle, we can have a broader vision of the current low fertility problem across the world. Moreover, in recent years the Chinese government keep changing their population policy; it has a different level of impact on the population so it might be interesting if Chinese data will be collected and observed. By doing that, more information about China instead of the European and American countries will be found out, by knowing about what happened in the Chinese fertility rate.

1.2 Research Problem

This thesis aims to explore the relationship between population density and the fertility level in China. Furthermore, it aims to identify the difference in fertility rates in

Chinese rural area and urban area, and the thesis mainly focuses on counties and cities in Xinjiang region, because as previously described it would be more suitable for this thesis, as it suffers the least from the one-child policy. Also, the thesis period will be from 1990 to 2012, because this is the period when family control policy in China was not changed, in other word, more stable. So the **first research question** is, does population density affect fertility in Xinjiang region from 1990 to 2012? The **second research question** is, what are the other contributing factors which are correlated with fertility in Xinjiang region from 1990 to 2012?

1.3 Outline of the Thesis

The remainder of the thesis is organized as follows. They are chapter 2 theory, chapter 3 data, chapter 4 method, chapter 5 empirical analysis and chapter 6 conclusion. Each chapter contains many sub-chapters to express the chapter in a more specific way.

The theory chapter contains previous research and theoretical framework. Previous research reviews the researches that are linked with this topic, which contains the relationship between population density and fertility, the determinants of fertility and one-child policy in China and the Chinese fertility situation. As for the theoretical framework, it takes the theories from previous researches and constructs the basis of the models.

The data chapter contains the structure of the data, descriptive statistics and quality of the data. The structure of the data mainly describes the source of the data and every variable that is going to be in the model. The descriptive statistics are about the values of the variables that are going to be in the model. The quality of the data evaluates the reliability, representability and validity of the data set and comment on the data set.

The method chapter contains the model that is going to be used in the thesis. The chapter explained the mechanism of the fixed-effect model. It also listed the models that are going to be used in the analysis and separate them based on the different proposes.

The empirical analysis chapter contains results and discussion. The results part describe the results of the regression of different models, followed by a sensitivity test. The discussion part discusses the possible explanation of the results. Then the conclusion summarizes the entire thesis, give the practical implications and suggest future researches.

2 Theory

The thesis aims to find out the significant interaction between population density and fertility rate so it will include different fields of knowledge including determinants of fertility, relationship between fertility and population density, the Chinese fertility feature and impcact of one-child policy. To know about these topics well, there is a need to break down this topic into several sub-topic to investigate the literature behind the sub-topics. To do research related to population density and fertility, the effect of population density and the determinants of fertility must be reviewed.

2.1 Previous Research

2.1.1 The population density and fertility

There is already research about population density and fertility, and there are also a few types of other researches related to human population density: the research of Malpezzi (2013), suggested that there is a close relationship between high population density and urbanization, the urbanization is above-average population density, which means there is a linkage between high population density and urbanization. Furthermore, the study done by Griffit & Veitch (1971) suggested that the different human population density and weather can affect human behaviour in different ways, especially the different population density and weather can affect their psychology. The effects are surprisingly significant. The research provided a micro point of view on how density affects human behaviour, a higher population density will lead to a worse situation of mental health. There have also been several pieces of research suggested the significant negative relationship between population density and fertility rate (Easterlin, 1976; Lutz & Qiang, 2002; Vanlandingham & Hirschmann, 2001).

Lutz and Qiang (2002) pointed out that there is a significant negative relationship between human population density and human fertility level. The research is about 147 countries across the world. They also discovered on the individual level, the fertility preferences also decline as human population density grows. Another study done by Croix (2017) pointed out the negative convergence of population density and fertility rate in 147 developing countries.

Furthermore, there are more studies regarding population density in Asia. A research done by Firebaugh (1982), suggested a significant relationship between population density and fertility in Indian rural area. In the research, the author used the data of 22 villages in Punjabi and discovered the population density in the village has a moderate inhibiting effect on fertility in the villages. This thesis provides the theory for the significant relationship between population density and fertility rate in the rural area. Moreover, the research done by Kato (2018) investigated the determinants of fertility rate in Japan. This research examined the relationship between population density and fertility rate of 47 municipalities in Japan and discovered a significant relation. This research also suggested the rationality of using population density as one of the variables. The explanation is that as the population density increases, the living cost increases too, so it will affect the willingness of giving birth, it is called a proxy variable by the author. However, the relationship of population density and fertility level regarding China were poorly studied, which is a country that has a significant population, that is one of the reasons for this study.

2.1.2 The determinants of fertility

The control variables are necessary to be added in this thesis because there are many other factors that can effect fertility other than population density. In order to add more variables, the factors which will affect the fertility level will be the variables. Even if there are some researches claimed that the social, economic factors were not the main reason for the fertility decline. A research was done by Bongaarts & Watkins (1996) point out that According to the theory developed by Notestein, the social and economic conditions can weakly affect the fertility declines. However, in this thesis, the authors admitted that the social, economic changes do change the cost and benefit of having children. Another research done by Kurkin, Branislav, Luděk, and Jiřina (2018) concluded that the achieved education level of woman and women's income are essential factors for fertility level. The research was about the factor of fertility in the Czech Republic based on the census in 2011. Then they discovered higher the women's income is, lower the fertility level is, and the higher their education level is, lower the fertility level is. Furthermore,-the research also suggested two different types of country, in terms of the effect of the female unemployment rate. In traditional male breadwinner society, the higher unemployment increases the level of fertility, whereas the more equal counties such as Nordic countries, a higher female unemployment rate has a negative effect on the fertility level (Yu, 2014). Then another study done by Bongaarts & Watkins (1996), enhanced the importance of the impact of education level on fertility. He said women with a primary education level tend to have more children than the women with a secondary degree, and woman without degree trend to have more children than the woman with a primary degree. Based on these articles, the literacy rate, the unemployment rate, and average income will be considered as the possible control variable in this study. However, as for the literacy rate, there are also some researches indicate that the literacy rate in China is rather high, according to Chinese statistical bureau, in 2015, Chinese average literacy rate was 96%, and it has been stable for a long time (National statistical bureau, 2015).

The determinants that can cause fertility change must be investigated because many other factors are affecting the fertility level other than population density. The population of cities is different from the rural area in terms of their characteristics of fertility. Generally speaking, the fertility change has been a hot topic for several years, so there are large amount of study regarding fertility determinants. A study done by Groat & Neal (1967) pointed out a correlation between social or psychological variables, such as education and religion, and fertility level. Then the research was done by Singh & Casterline (1985) suggested the decisions of the fertility of individuals are affected by what region they belong to, whether they live in rural areas or cites, and different literacy rate. In a high literacy rate country, there will be a more significant difference in fertility rate in rural area and cities, and the population density can be one of the social, economic problems as well.

2.1.3 Chinese fertility situation and the one-child policy

To research China, the historical events and its unique political system of China should be investigated. Recently, Many Chinese regions are suffering the low fertility problem and the migration has been keeping entering the cities (Beijing Municipal People's Government Research Office Social Affairs Office, 2015), the cities around the country are also suffering the ageing society problems (Beijing Municipal People's Government Research Office Social Affairs Office, 2015). As the study was done by Ho (2011), the amount of desired children per family is decreasing as the average income increases, and the cities in China have a lower fertility rate. At the same time, the policy of the Chinese government is also very noticeable, especially family control policy (Ho, 2011). There are also some studies on Chinese family control policy. Study of Tsui (1989) described although the nuclear family increased, and the traditional family still remain powerful, due to the fact that the traditional family concept is still active. The effect of the one-child policy will be discussed in the last paragraph of this section.

The history of the one-child policy is as follows: in 1981, the one-child policy start to be validated, before one child policy there were some restrictions on childbirth already, but they are not compulsory (Ma & Sun, 2010). Moreover, after 1981, the one-child policy law has been settled (Ma & Sun, 2010), which was not merely a law that every family could only have one child, it was more complicated. The original law was as follows "The state stabilizes the current policy of birth and childbirth, encourages citizens to marry late, and encourages a couple to have a child late. If they meet the requirements of laws and regulations, they can ask for the second child." the real regulations are as follows: first of all, All the urban area and the rural area of Beijing, Tianjin, Shanghai, Chongqing, Jiangsu, Sichuan provinces, has the basic implementation of Han residents which is a couple can only have one child (Ma & Sun, 2010). , there is very loose regulation for some ethnic minority and no limitation for the number of children per family for some ethnic minorities (Ma & Sun, 2010). Xinjiang region has more than 50% of the population of the ethnic minority, most of the ethnic minority are Uyghur people, Xinjiang region also has roughly 50% of the rural population, which is higher than the average rural population rate in China (Ma & Sun, 2010).

Moreover, there are five provinces have the regulations of the couples can have two children (Ma et al., 2010). There were 19 provinces had the regulations of, if the first child is a girl, the couple can have another child after an interval of a few years (4 to 5 years) (Ma & Sun, 2010). Besides that, people that belong to the categories of the ethnic minorities and remarriage and people who came back from overseas countries can have a second child, even third if needed (Ma & Sun, 2010). There are also many other rules in one-child policy, but they will not be listed in this thesis because they are not related to the research objective. For Xingjiang region, there are three regulations for couples that can have more than one child: Rural couples in Xinjiang region are allowed to have a second child (Xinjiang regions government, 2010). Ethnic minority couples in the Xinjiang region are allowed to have a second child (Xinjiang regions government, 2010). Minority farmers and herdsmen in Xinjiang can have three children (Xinjiang regions government, 2010). Also one thing that has to be noticed is the majority of the ethnic minority in Xinjiang region are Muslim (Luqiu, Luwei, Fan, 2018), which is possibility linked with very recent Islamophobia in China, this will be examined in the result part.

The national population policy had changed at the end of 2013 when the government decided to apply the "partially two-child policy", which means one of the couples was the only child of their family, the couple can have the second child (Zhang, Huang, Zhang, Chen, 2016). Then in 2016, the national population policy has changed again, this time it has changed into "two-child policy", which as the name implies, every couple can have two children (Zhang et al., 2016).

The period of the thesis is in the period of one child policy of China, (family planning policy). Which was from 1981 to 2014, and the family planning of China is still an ongoing policy, but the content has changed into two-child policy since 2016. So that becomes a topic of whether the one-child policy affects fertility level. A research was done by Goodkind (2017) point out the one-child policy has been the main driver of Chinese fertility decline, he proved that by comparing with different countries with a similar feature, and those countries has significantly higher fertility, so the birth restriction is the driving force. However, the study done by Zhao & Wang (2018) contradicted Goodkind (2017) in their research. They suggested that the socioeconomic factor has been the main driver for fertility level since the 1990s, the historical event was also the key factors. So according to their arguments, we can see the importance of the socioeconomic factors for Chinese fertility level, and the effect of one child policy also has to beware in the thesis. All in all, to know about how the Population density can impact on the fertility rate in Chinese cities, we need to look through those theories separately and get our linkage from the data proof and we have to look after the policies and the Chinese reality since Chinese society is relatively separated from the rest of the world.

2.2 Theoretical Approach

As the previous research addressed, there is a negative correlation between human population density and the fertility (Easterlin, 1976; Lutz & Qiang, 2002; Van Landingham &

Hirschmann, 2001), and the relationship between these two elements are significant, this phenomenon happens not only in the city but also in the rural area (Firebaugh, 1982). The most of the research about human population density and fertility are done in developing countries, China was missing from these researches, but China is a developing country too, so density in this thesis it is going to be the leading theory that is going to be tested in Chinese situations.

Also As the previous researches indicated, many other determinants can contribute to fertility change. The research done by Kurkin et al. (2018) pointed out that the female income and education level can affect the fertility rate. Also in the research done by Croix (2017), there are some of the determinants for the fertility which were significant, they are the female literacy rate, the female unemployment rate, and average income. However, the female literacy rate in China was rather high, and the data was not available on the city and county level. For the unemployment rate, a good substitution is the employment rate, since unemployment was not available in the database that was chosen.

There are also several specific determinants for the Chinese case; all of them are related to the one-child policy. As the previous study addressed, the one-child policy was different towards a different group of people (Zhang et al., 2016). The main difference results in two categories: The first one is the difference between ethnic minority and ethnic majority, the one-child policy was mainly targeting the ethnic majority group in China which was Han nation (it has more than 90% of Chinese population) (Zhang et al., 2016), and rest of other ethnic groups can give birth with a better policy. Another is whether the household registered as a rural household or urban household. The policies on urban households are stricter than rural households. In this thesis, there will be a rate of the ethnic minority population in every county and city, and the rate of the urban population in every county and city, thus this two variable can reflect whether the policy affects the fertility level. There is the difference between city and rural fertility level, as previous research addressed, the fertility in the urban area and rural area are different so that I will have both city and county as the smallest unit in this thesis, and they will be analyzed separately later.

Also as the previous research wrote, one child policy will not profoundly affect the result, but in order to keep the consistency of the policy, I decided to focus on the period and the place when the one-child policy was most stable and where the one-child policy was least strict. Since the one-child policy was not changed from 1990 to 2013, and since the policy was not strict on ethnic minority, the province which has the most proportion of ethnic minority will be chosen as the thesis sample. Xinjiang region has the second largest proportion of ethnic minority, and first largest fertility rate in 2015 (Wang P., 2010), so it is feasible to study. Also I choose county and city as the object to study, because county and city is the smallest unit for the statistical data, and it can also distinguish the rural area and urban area, since majority area of a city in China is urban area, and majority of the area in county in China is rural area (National statistical bureau, 2019).

3 Data

Generally speaking, the data of this thesis is secondary data. One source of the data is the National Statistics Bureau in Xinjiang region, and I am using the quantitative data from the yearbooks published by the statistics bureau in Xinjiang region. Moreover, some of the data are from the website called: China Data Online, which was a private company who has the data from the Chinese national statistics bureau and many other reliable sources. The time range of the thesis was from 1996 to 2013, which were the years that the one-child policy was stable, and the data covered most of the county/city in Xinjiang region.

In this thesis, I collected the data from Xinjiang region, but as the most cases, not all the years and areas are available for those variables. There is 101 countries/cities level of administrative areas in Xinjiang region in 2013, but there are only 55 counties/cities available in the database for the variable I have chosen, rest of the county/city was not available. The total population of the 55 counties/cities that are available is 12,73 million population in 2013, which represented 56,02% population of Xinjiang region in 2013. The missing regions were not selected, they are randomly missing from the database. For the years I have chosen, the data from the yearbooks are available from 1997-1998, and 2000-2009, the rest of the yearbooks are missing from the database that has been chosen. Moreover, for the data directly exported from county-level data of China Data Online, the situation was similar, despite one category called: Annual Per Capita Net Income of Rural Households (yuan) (constant prices), it only has the data from 2004 to 2013, it is missing the data before 2004.

3.1 The structure of the data

The data that is going to be used in this thesis, as is description above, were collected throughout two channels, the first channel is the database of China county statistics from China Data Online. The variables that were collected through this channel are as follows: The name of districts, the years, area of administrative region (10000 square kilometers), population at the year-end (10000 persons), GDP (100 million yuan) (constant prices), number of employed persons at year-end (10000 persons), number of rural labours (10000 persons) and annual per capita net income of rural households (yuan). Another channel is the statistical yearbook of Xinjiang region. The variables that were collected through this channel are as follows: average wage of the staff on the job (yuan) (constant prices), crude birth rate (‰), the population of minority, urban population (10000 people). Then based on these variables, more important variables are constructed. There are totally for 55 counties/cities and 17 years in the period for this thesis. There is also the category for every country/city whether it is a county or a city, 1 represents county and 0 represents the city.

The valuables that are going to be used in the model are as follows:

The crude birth rate variable. This variable was directly taken from the statistic yearbooks of Xinjiang region as we previously addressed. The crude birth rate has been chosen instead of the fertility rate, because the fertility rate of the areas is not available. In fact, the fertility rate did not exist in any level of statistical data in Chinese official statistics in county and city level.

The population density variable. This variable was constructed by using the area of the administrative region (10000sq, km) (hereafter referred to as *area*) and the population at the year-end (10000 persons) (hereafter referred to as *population*) of every county and cities in this thesis. For area all the year from 1997 to 2013 were available, but the majority of the counties and cities are lack of the data from 1996, the Microsoft excel command predicts them: *forecast*, which predicted the missing value by a linear method, the same situation for the population as well.

The ethnic minority rate variable. It is a constructed variable consisting of the population of an ethnic minority from the statistical yearbooks of Xinjiang region and total population at the year-end from China Data Online. The function is the population of ethnic minority divided by the total population at the year-end. All the 55 counties/cities are available in the years of 1997, 1998, 2000 to 2009, the rest of the value are missing because the population of an ethnic minority from the statistical yearbooks of Xinjiang region are missing in those years.

The employment rate variable. It is a constructed variable consist of the number of rural Labourers (10000 persons) from China Data Online, amount of population employed [sic] from China Data Online, and the population at the year-end from China Data Online. The function is the sum of the available rural labour force and amount of worker on job divide by the population at the year-end. One obvious thing is that the worker on the job represents the works that work in the urban area. For the workers who work in the rural area, they are categorized as the available rural labour force. The years of 1996-1999 and 2013 were missing, the rest of the years are available.

The urban population rate variable. It is constructed by urban population (10000 persons) from statistic yearbooks of Xinjiang region and the population at the year-end (10000 persons) from. The function is urban population divided by population at the year-end (10000 persons). The year of 1996-1999, 2010-2013 were missing from the majority of the counties and cities, because many yearbooks were missing from the database China Data Online and they are not available from many other sources as well.

The GDP per capita variable. This variable was constructed by GDP (100 million yuan) from China Data Online and the population at the year-end (10000 persons) from China Data Online. The function is GDP divided by the population at the year-end (10000 persons), and the majority of the years were available, except 1996. The GDP per capita was corrected by using the inflation rate of China by year from the World Bank.

The average income variable. This variable was constructed by average wage of the employment population (yuan) from China Data Online, annual income per capita of rural households (yuan) from China Data Online, amount of staff on job (10000 persons) from China Data Online and number of rural labourers (10000 persons) from China Data Online. It is a complicated function, which is:

$$\text{The average income} = \frac{\text{average wage of the empolyment population (yuan)} * \text{employed population (10000 persons)}}{employed\ population + number\ of\ rural\ labours} + \\ + \\ + \frac{\text{annual income per capita of rural households (yuan)} * \text{ number of rural Labors (10000 persons)}}{employed\ population + number\ of\ rural\ labours}$$

And divided by inflation rate from World Bank (2019) to get the real average income. Only the year of 2004 to 2009 was available.

In sum, the variable above are the variables I am going to use in the analysis. Also, the GDP per capita and average income variable were the variables which are has a similar meaning. The average income was constructed in a complicated way, and many of them were missing.

Therefore the average income will be abandoned in one model to test which model is the better model.

3.2 Descriptive Statistics

Table 3-1 descriptive statistics of total, counties and cities

Total	count	mean	sd	min	max
Birth Rate	714	15.34	5.47	5.15	54.89
Population Density	989	63.93	286.03	0.00	3619.35
The employment rate	721	0.35	0.12	0.06	1.00
Rate of ethnic minority	604	0.70	0.28	0.18	1.00
Urban population rate	542	0.47	0.30	0.06	1.00
Real GDP per capita	944	12387.31	17411.89	1038.48	204157.03
Average income	361	7600.60	6045.21	1339.81	32163.07
N	990				

Counties	count	mean	sd	min	max
Birth Rate	584	15.74	5.50	5.15	54.89
Population Density	809	24.82	38.63	0.00	233.01
The employment rate	581	0.37	0.12	0.06	1.00
Rate of ethnic minority	494	0.74	0.27	0.18	1.00
Urban population rate	450	0.36	0.20	0.06	0.97
Real GDP per capita	765	11010.84	15970.27	1038.48	204157.03
Average income	285	5157.42	2935.95	1339.81	19621.86
N	810				

Cities	count	mean	sd	min	max
Birth Rate	130	13.55	4.99	7.47	33.44

Population Density	180	239.71	637.84	0.00	3619.35
The employment rate	140	0.29	0.09	0.13	0.86
Rate of ethnic minority	110	0.57	0.29	0.22	1.00
Urban population rate	100	0.98	0.13	0.10	1.00
Real GDP per capita	179	18269.98	21644.61	1298.45	135628.01
Average income	76	16762.56	5918.60	6915.97	32163.07
N	180				

Data source: Xinjiang statistical yearbook (2000-2009); China data online (2019)

Table 1 shows the descriptive statistics, separately for the total sample, the county sample and the city sample. The total number of the sample differs from different variables. The highest is 989 and the lowest is 367 in total, and the data was not statistically balanced, and the county and city all together were 55, there are ten cities and 45 counties in this sample, so there are generally more observations in the county data than the city data. The mean of population density in the city is slightly larger than that in the counties (239.7068 in the cities and 24.81862 in the counties). Also, the range of real GDP per capita in the city is smaller than that in the counties, which indicates the gap between the rich and the poor is smaller in the city than it is in the county.

There is also a difference in the crude birth rate between city and county. Counties have a 2% higher crude birth rate than cities on average, but it is not confirmed to be significant. Also, the county has a higher employment rate on average, mainly because there are more labour force in the rural household considered as employed, due to the fact that basic economic unit in the rural area is family in China (Wang & Ding, 2005), and counties have the more rural population. Therefore it is natural that counties have a higher employment rate compared to cities. Not surprisingly, the urban population rate in the cities was almost double the county's urban population rate, but the maximum number of the urban population rate in counties is over 90%, which was abnormal. As for the average income, there is no significant difference between counties and cities, but the difference between max and min in counties were slightly higher than cities, whereas cities have equal income distribution.

3.3 Quality of the data

3.3.1 Reliability

The data was directly exported from China Data Online and yearbooks, which have comprehensive data from many reliable sources. They are mainly from the China statistical bureau. Moreover, there are many published articles which have been cited by this database (China data online, 2019). It is a good indication that the database has high reliability.

The data from the Chinese national statistical bureau was not trustworthy, according to many scholars. Some researchers such as Wang and Meng (2000) suggested that not all of the Chinese government based source are available. Due to some technical problems, many of the Chinese data were different from the reality. So the data from Chinese databases always needed to be examined. Then the Chinese statistics bureau's yearbook *Xinjiang region statistics yearbooks* are going to be used as one of the sources. They are a series of yearbook about socioeconomic indicators including population and crude birth rate of different Xinjiang region's counties and cities. And most of the yearbooks are available online, so the data from the local government is very comprehensive for this thesis.

There are also two types of data to use in this thesis: predicted data and the dropped data. The predicted data will be partially used as it is described in the variable section since nearly half of the data is missing from the data set. Moreover, the reason why they are missing is that many yearbooks were missing from the database and some of the city did exist until a particular year, and some of the data was not recorded in some individual years. So in order to analyse, the empty data must be dropped, or be filled with predictions by excel function *forecast*. This is a function that was based on linear prediction. The variables that were applied this methods are population density and average wage since there are little changes in these two variables they were predicted in the missing years as previously addressed. However, other variables were not easy to predict by using this function since there are too many changes in the variable. However, the prediction was based on the secondary data, so the trustworthiness of the data need to be taken into account too.

3.3.2 Representativity

All of the data were not census data, but there is a need to use secondary statistical data to keep consistency because the thesis is on the county and city level. The government data is more comprehensive and more trustworthy, compare with the survey done by the private institute. Because of their comprehensive power-on information and statistics. Moreover, many of the data took average value, such as GDP per capita, annual income per capita of rural households. The and the population data, including was directly from the local police office according to the description of the data sheets, which means they are real data and it is not predicted data.

The data that is going to be used in this thesis, however, was not totally representative. Because not all of the data planned to be used were available, 53% of the county and city were available for the analysis, but as aforementioned information addressed, the missing cities and counties were randomly missed (see the missing regions in appendix B), so the result will not be affected severely. Furthermore, nearly one-third of the year is missing, therefore the representativity of the area in the data should be aware.

3.3.3 Validity

The data were on the same level of this thesis, which was both on the county and city level, and most of the designed variable was available. The variables that are planned to use in this thesis were mostly available in the database I chose. The names of the variables in the database need to be changed for the researchers' purpose, but generally, the variables were valid. The data of Chinese statistics, the unit was not unified in different years and different regions in the original data either, but it was unified in this thesis.

However, some variables I was planning to use are not available, such as the female literacy rate and the female unemployment rate, but as it addressed before, the literacy rate in China was on average 95% in 2015 so it will not have much effect on the regression. Also, the fertility variable was not available in the database, so it is represented by the crude birth rate, which reflected how many percentages of the newborn baby the year divided by population of the area. Furthermore, instead of the female unemployment rate, the only variable that was close in the database was the employment rate, that is also a problem that is needed to be noticed in the later analysis.

4 Methods

The analysis of this thesis is based on panel data, so a simple OLS will not fulfil the need for the analysis. OLS model was not designed to cope with time variant and invariant variables on the same individual at the same time. In order to avoid the false, the fixed-effect model will be introduced in this thesis, and the two different types of the model will be compared and decided which one is the proper one to use. By comparing the OLS model and fixed-effect model. Especially the variables population density and crude birth rate need fixed-effect since there is a possibility of having a time-variant in those variables. According to the research done by Lutz et al. (2006) in this paper, the researcher also used the fixed-effect model to analyse the regression of population density and fertility. As for the necessity of using a fixed-effect model instead of the random effect model, The Hausman test will be performed to testify which model fits the data more.

The basic logic of the fixed effects model used in the analysis comes from Angrist and Pischke (2008). Assume Yit equals to the crude fertility rate of year t, Oit indicates category the region (county or city). The observed Yit is Y1it or Y0it, depending on the region attribute. Then, further assume that

$$E(Y0tE(Y0t|Ci,Xit,t,Oit)=E(Y0t|Ci,Xit,t) \qquad (1)$$

Equation (1) implies that crude birth rate variable random assigned variable, provided that other variables associated with the area, Ci and other variables included in the model, such as employment rate of the region, ethnic minority population rate, The urban population rate, etc. The main point of fixed-effect estimation is to assume that the unobserved Ci is time-invariant in the linear model of (Y0it|Ci, Xit, t):

$$E(Y0it|Ci,Xit,t)=\alpha i + \lambda t + \theta Xit\beta \qquad (2)$$

In equation (2), λt is the annual effect that should be estimated in the final model. Next, assume that the causal effects of regional attributes are additive and constant:

$$E(Y1it|Ci,Xit,t)=E(Y0it|Ci,Xit,t)\ \delta \qquad (3)$$

Therefore, this equation means

$$E(Yit|Ci,Xit,t,Oit)=\alpha i + \lambda t + \beta\delta + Oit + \varepsilon it \qquad (4)$$

Where δ is the causal effect of interest. Equation (4) also implies

$$Yit = \alpha i + \lambda t + \beta\delta + Oit + \varepsilon it \qquad (5)$$

where

$$\alpha i \equiv \alpha C i \theta$$

Equation (5) explains how to repeat a personal observation based on panel data. The causal effect of population density on crude fertility can be estimated by considering the fixed effect as a time-invariant variable to estimate. The regression result calculates the mean of the individual, which can be expressed as

$$\bar{Yt} = \alpha i + \bar{\lambda} + \bar{Xi}\beta + \delta \bar{Oi} + \bar{\varepsilon i} \quad (6)$$

Subtract (6) from (5)

$$Yit - YtYt = (\lambda t - \bar{\lambda})\ (Xit - XiXi)\beta\delta(Oit - OiOi)\ (\varepsilon it - \bar{\varepsilon i})$$

Thus, deviations from the mean eliminate unobserved individual effects (Angrist and Pischke, 2008). The Hausman test is used to determine the preferred specifications of the model (Angrist and Pischke, 2008). The random effects model assumes that time-invariant individual effects are strictly independent of all X variables contained in the regression model. Due to this assumption, the model does not automatically exclude time-invariant variables like fixed-effect models. Therefore, the model will be decided after the Hausman test in this thesis. By using individual fixed-effect models, the changes in the attributes of the observations in the data set, only a portion of the total observations in the dataset are used.

In this thesis, all of the variables will be logged in the models. That will make it easier to observe the percentage change in the models. The reason why the dependent variable "crude birth rate" is because the crude birth rate is a percentage variable, to observe the change in percentage is more convenient for understanding. By logging the control variables and independent variables, the changes in control variables and the independent variable are unified with the changes independent variables, so it is easier to analyse.

Also in order to avoid the problem of heteroscedasticity in the models, the white test is going to be done for every model, to test whether the model need robust standard errors on the model. The results of the white tests will be present in appendix C, and the result for the white test is none of the models has a significant of heteroscedasticity. Therefore, a fixed-effect model without robust standard error is going to be used as the model.

The county and city model will also be developed in this thesis. As previous research addressed, there is a big difference between the fertility rate of the urban area and rural fertility rate, and also the one-child policy was different towards rural population and urban population as previous research pointed out Ma et al. (2010). As the description of Xinjiang region, statistical yearbooks addressed, county represent the more rural area, whereas the city represents mostly urban area, and the residents of the city are mostly urban populations. So the counties will be separate from cities, and there will be separate models for both cities and counties to observe the difference.

I am also going to set up two different way of modelling, whether the model is with the income variable or without the income variable. In the data set, both average income variable and GDP per capita variable are economic indicators, so in order to keep the reliability of the model, the two variable will be put in different models. Furthermore, the income variable was not so reliable since several other variables constructed it, and many of the years were missing from this variable, the sample size will increase a lot if this variable is abandoned. So the model will also be separated based on whether if the model contains the average income variable or GDP per capita variable.

As the researcher Malpezzi (2013) suggested, the population density has a relationship with urbanizations, the higher the urbanization is, and the higher the population density an area will have. In order to check if urban population rate in the region will have any effect on other variables especially the population density variable. The models without urban population rates will be present after the normal models.

The models that are going to be presented in this thesis are as follows:

For the model with average income variable:

$$\begin{aligned}\log(\mathit{Crude\ Birth\ Rate}) &= \beta_0 + \beta_1 \log(\mathit{Population\ Density}) \\ &+ \beta_2 \log(\mathit{Ethnic\ minority\ rate}) + \beta_3 \log(\mathit{Employment\ rate}) \\ &+ \beta_4 \log(\mathit{Urban\ popualtion\ rate}) + \beta_5 \log(\mathit{Average\ income})\end{aligned}$$

For the model with GDP per capita variable

$$\begin{aligned}\log(\mathit{Crude\ Birth\ Rate}) &= \beta_0 + \beta_1 \log(\mathit{Population\ Density}) \\ &+ \beta_2 \log(\mathit{Ethnic\ minorty\ rate}) + \beta_3 \log(\mathit{Employment\ rate}) \\ &+ \beta_4 \log(\mathit{Urban\ popualtion\ rate}) + \beta_5 (\mathit{GDPper\ capita})\end{aligned}$$

For the models with average income variable but without urban population rate variable:

$$\begin{aligned}\log(\mathit{Crude\ Birth\ Rate}) &= \beta_0 + \beta_1 \log(\mathit{Population\ Density}) \\ &+ \beta_2 \log(\mathit{Ethnic\ minorty\ rate}) + \beta_3 \log(\mathit{Employment\ rate}) \\ &+ \beta_4 \log(\mathit{Average\ income})\end{aligned}$$

For the models with GDP per capita variable but without:

$$\log(Crude\ Birth\ Rate) = \beta_0 + \beta_1 \log(Population\ Density) + \beta_2 \log(Ethnic\ minorty\ rate) + \beta_3 \log(Employment\ rate) + \beta_4 (GDPper\ capita)$$

All those four models will be developed based on software Stata14 IC. Every model will followed by an OLS regression to compare which variable is time variant related. Also by comparing the OLS model and fixed-effect model, the advantage of fixed-effect model will be seen. All of the four models will be tested in different data set, they are the data set in total, the data set of counties and the data set of cities.

5 Empirical Analysis

This chapter is going to be a chapter of the results and discussion. In the result part, the regression results will be present. The models will be separated based on the data. The first group of models will be the complete data set; the second group of models will be based on the counties data and the last group of the models will be based on the cities data. A robustness test section will follow the regression result. In this section, the models that fit the most will be tested, by using the robustness test. By doing that, it will be clearer if the model I choose is reliable or not, if not what is the reason behind it.

The discussion section will be consist of three parts. The first part is the possible explanation about the results and what is the models suggest, the second part is about theories linked with the results and the third part is the strength and weakness of the models.

5.1Results

Table 5-1 the name in the regression tables and the actual meaning of the variables

log density	log ethnic	log urban	log GDP	log income	log employ
Logged population density	Logged ethnic population rate	Logged urban population rate	Logged GDP per capita	Longed average income	Logged employment rate

5.1.1 Regression results

Table 5-2 the regression models based on all regions

The model in total With urban population rate	(1) OLS with GDP	(2) Fixed-effect with GDP	(3) OLS with income	(4) Fixed-effect with income
log density	0.00617 (1.50)	0.0255 (1.12)	0.00875 (1.75)	0.0277 (1.07)
log ethnic	0.375*** (13.60)	-0.00622 (-0.13)	0.339*** (10.24)	0.0146 (0.28)
log employ	0.00833 (0.20)	0.302** (3.20)	-0.0487 (-0.97)	0.299 (1.97)
log urban	-0.0591** (-3.30)	-0.0562* (-2.58)	-0.148*** (-3.53)	-0.0511 (-0.55)
log GDP	0.0477* (2.32)	0.146 (1.85)		
log income			0.0935* (2.31)	-0.0151 (-0.09)
_cons	2.353*** (11.69)	1.239 (1.71)	1.849*** (4.74)	2.338 (1.58)
N	485	485	344	344
R^2	0.366	0.338	0.365	0.298

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As table above shows above, there are four models presented and all of the four models has urban population rate variable. The first model is OLS model with GDP per capita variable. The second model is a fixed-effect model with GDP per capita variable, the third model is the OLS model with average income variables and the fourth model is a fixed-effect model without income variables. The reason why there are no random effect models is that, the Hausman test (see Appendix A) suggested that the fixed effect model should be used for these models. Moreover, the first two models have more samples because the income variable has too many empty data.

In the first model, the ethnic minority rate and urban population rate were significant, which means, without time variant control, the ethnic minority rate and urban population rate have effects on the crude birth rate, it reflects the decision of family decision of having children. This proved that the policy does affect the fertility decision. However, the second model suggested that in Xinjiang region the employment rate has a positive effect on crude birth rate and urban population rate has a negative effect on crude birth rate, every 1 percentage employment changes in the model, every 0.302 the crude birth rate will change in the region, and every 1 percentage the employment rate rises, the crude birth rate will be decreasing 0.05 percent. The model also suggested the effect of ethnic minority rate was correlated with time variant. The population density variable was not significant in this model.

For the third model, after the income variable replaced the GDP per capita variable as the economic indicator. There are no significant differences with the previous OLS model which has GDP per capita as the economic indicator. However, as for the fixed effect model (the fourth model), there are no significant variables. Compare the second model with fourth model. After the average income variable was used as economic indicator, all the significant variables became insignificant. And compare the model 4 with model 3, the ethnic minority rate variable and urban population rate variable changed from significant into insignificant, which suggests they are correlated with time variant.

The model fit the data the most is probably the second model. Because the error term is under control, for a panel data it is essential of avoiding the error terms. Also without the income variable, the model contains more samples and the R square are better.

Table 5-3 the regression models based on all regions without urban population rate variable

The model in total Without urban population rate	(5) OLS with GDP	(6) Fixed-effect with GDP	(7) OLS with income	(8) Fixed-effect with income
logged density	0.00878^{*} (2.16)	0.0219 (0.95)	0.00906 (1.78)	0.0267 (1.03)
log ethnic	0.386^{***} (13.93)	-0.0125 (-0.26)	0.361^{***} (10.88)	0.00945 (0.19)
log employ	0.0352 (0.85)	0.301^{**} (3.17)	-0.0231 (-0.46)	0.298 (1.96)
log urban				
Log GDP	0.0218 (1.14)	0.118 (1.50)		
log income			-0.0181 (-0.70)	-0.0385 (-0.24)
_cons	2.667^{***} (14.88)	1.633^{*} (2.29)	2.979^{***} (13.11)	2.677^{*} (1.99)
N	485	485	344	344
R^2	0.351	0.327	0.341	0.298

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From table above it is clear to see that after the urban population rate was removed from the model, the models did not affect too much by the change. Especially the main model fixed-effect model with GDP per capita factor did not affect too much by the change, which means in the total model, the other variables did not affect by the urban population rate model. There is only a small change in the OLS with GDP factor model, the population density changed its significance level, but this significance level was only in OLS, model, which means it is time variants affected. Furthermore, the little change suggested that the urban population rate had an effect on population density in the model, but it was not strong enough to change it. Therefore, the model 2 in the table 5-1 which has the urban population rate variable is chose as the model for this data set.

Table 5-4 the regression models based on counties with urban population rate variable

The model of county data With urban population rate	(1) OLS with GDP	(2) Fixed-effect with GDP	(3) OLS with income	(4) Fixed-effect with income
log density	0.0353^{***}	0.0399	0.0380^{**}	0.249
	(3.67)	(0.36)	(3.11)	(1.26)
log ethnic	0.385^{***}	0.0292	0.355^{***}	0.0473
	(12.47)	(0.52)	(8.67)	(0.78)
log employ	-0.00482	0.281^{**}	-0.101	0.353^{*}
	(-0.10)	(2.91)	(-1.64)	(2.14)
log urban	-0.0737^{***}	-0.0607^{**}	-0.145^{**}	-0.0760
	(-3.58)	(-2.64)	(-3.32)	(-0.77)
log GDP	0.0847^{***}	0.195^{*}		
	(3.76)	(2.53)		
log income			0.163^{**}	0.0318
			(3.00)	(0.15)
_cons	1.934^{***}	0.763	1.160^{*}	1.456
	(8.45)	(1.00)	(2.24)	(0.79)
N	403	403	280	280
R^2	0.367	0.373	0.346	0.335

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table above showed the models of all the counties in the data set. There are four models presented. The Hausman test showed that the fixed-effect model needed to be used in these models (see Appendix A). The first model is OLS model included GDP per capita. The second model is a fixed-effect model with GDP per capita variable, the third model is the OLS model with income variables and the fourth model is a fixed-effect model with income variables.

Surprisingly, after getting rid of the city samples in these models, the population density variable became significant in the OLS models, every one percentage the population density increases, 0.03 percentage of the crude birth rate will increase, the ethnic minority rate and urban population rate were significant as well. However, after included the time variants by using the fixed-effect model, the population density variable was insignificant, which means all of the variables were significant in the first model were probably affected by time variant. The fixed-effect shows that employment rate and urban population rate were significant and the GDP per capita was remotely significant. Compare with the fixed-effect model with GDP per capita variable in the total data set, the GDP became significant, which means in the county the economic factor can have more effect on the crude birth rate.

However, the regression was better in the fixed-effect model with GDP per capita. There is still no significant difference with the previous OLS model. There are three significant variables in the fixed-effect model with GDP per capita. They are employment rate, GDP per capita and urban population rate, every 1 percentage employment changes in the model, every 0.28 the crude birth rate will change in the region, and every 1 percentage the employment rate rises, the crude birth rate will be decreasing 0.06 percentage. The GDP per capita also has a positive effect on the crude birth rate. The tendency was quite similar to the total model, both the number and significant level.

The model fit in this data set the most is the fixed-effect model without GDP per capita variable. Because as the table shows above, the second model has the higher R squared and contains samples, it is more representative. More importantly, the fixed-effect model controls the time variant. It can eliminate the error terms, which fits the panel data more.

Table 5-5 the regression models based on counties without urban population rate variable

The model of county data Without urban population rate	(5) OLS with GDP	(6) Fixed-effect with GDP	(7) OLS with income	(8) Fixed-effect with income
log density	0.0339^{***}	0.00873	0.0353^{**}	0.243
	(3.47)	(0.08)	(2.85)	(1.24)
log ethnic	0.402^{***}	0.0189	0.373^{***}	0.0374
	(13.00)	(0.34)	(9.02)	(0.63)
log employ	0.0105	0.275^{**}	-0.0638	0.348^{*}
	(0.22)	(2.82)	(-1.04)	(2.11)
log urban				
log GDP	0.0526^{*}	0.172^{*}		
	(2.51)	(2.23)		
log income			0.0400	-0.00502
			(0.99)	(-0.02)
_cons	2.324^{***}	1.192	2.401^{***}	1.994
	(11.36)	(1.58)	(6.61)	(1.17)
N	403	403	280	280
R^2	0.347	0.360	0.320	0.333

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From the table above it is clear to see that after the urban population rate has been removed, these models did get affected too much by the change. Especially the main model fixed-effect model with GDP factor, did not affect too much by the change. There is an only small change in the OLS with GDP per capita factor model, the population density changed its significance level, but this significance level was only in OLS, model, which means it is time variants effected. Furthermore, the little change suggested that the urban population rate had an effect on population density in the model, but it was not strong enough to change it. Therefore, I will still choose the model with urban population rate as the chosen model for this data set.

Table 5-6 the regression models based on cities with urban population rate variable

The model of city data With urban population rate	(1) OLS with GDP	(2) Random-effect model with GDP	(3) OLS with income	(4) Random-effect model with income
log density	0.00353	0.00601	-0.000815	-0.00202
	(0.71)	(1.21)	(-0.13)	(-0.31)
log ethnic	0.256^{***}	0.134^{*}	0.244^{***}	0.236^{***}
	(4.43)	(2.02)	(4.25)	(3.44)
log employ	0.191	0.223^{*}	0.166	0.215
	(1.73)	(2.04)	(1.16)	(1.27)
log urban	-0.0239	-0.0817	0.517	2.959
	(-0.31)	(-1.06)	(0.47)	(1.65)
log GDP	-0.0467	-0.172^{***}		
	(-1.29)	(-3.64)		
log income			0.274^{**}	-0.231
			(2.84)	(-0.81)
_cons	3.387^{***}	4.371^{***}	0.254	4.952
	(9.02)	(9.99)	(0.27)	(1.85)
N	90	90	72	72
R^2	0.357	0.524	0.399	0.461

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table shows the four models of cities data set, the order of the model are the same as the previous two analysis. However, instead of using the fixed-effect model, the random-effect models are used. Because according to the Hausman test (see Appendix A), the fixed-effect model of these two models has been rejected, therefore the random-effect models are used in this model.

Quite different from the previous two data sets, the regressions in this data set shows a different trend. However, the population density was not significant in all of the regressions. The first OLS regression shows that only ethnic minority population rate was highly significant, and has a positive impact on crude birth rate, but in the random-effect, the significance level of ethnic minority population rate was declined. At the same time, employment rate became significant, and the GDP per capita became highly significant, the GDP per capita has a negative effect on the crude birth rate in this regression. Then the regressions with average income showed a different picture. The average income was significant in the OLS with average income, but in the random-effect model with average income, the average income variable was not significant, which means the income variable might be time variant related in this model. Other than that difference, there is no significant change between random effect models with OLS regression with average income.

Table 5-7 the regression models based on counties without urban population rate variable

The model of city data Without urban population rate	(5) OLS with GDP	(6) Random-effect with GDP	(7) OLS with income	(8) Fixed-effect with income
log density	0.00370	0.0116	-0.000414	0.0246
	(0.75)	(0.47)	(-0.07)	(0.83)
log ethnic	0.253***	-0.0381	0.250***	-0.0504
	(4.46)	(-0.42)	(4.49)	(-0.51)
log employ	0.192	0.283	0.161	0.00781
	(1.74)	(1.15)	(1.13)	(0.02)
log urban				
log GDP	-0.0474	-0.239		
	(-1.32)	(-1.29)		
log income			0.279**	-0.145
			(2.94)	(-0.46)
_cons	3.394***	4.898**	0.195	3.708
	(9.11)	(2.88)	(0.21)	(1.28)
N	90	90	72	72
R^2	0.356	0.239	0.397	0.214

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table above shows the regressions without urban population rate. According to the Hausman test result (see Appendix A), the random-effect model is decided to be used in the model with GDP per capita whereas the fixed-effect model will be used for the regression with average income. As the table showed, after the urban population rate has been taken away, both the random-effect model of the regression with GDP per capita and fixed effect model of the regression with average income shows no significant variables, which means in this dataset, the urban population rate might have interaction with other variables, so the pervious one was not trustworthy.

Based on the analysis above, the fixed-effect model without average income variable include all the region and the fixed-effect model without average income based on the counties in the sample will be chosen as the most suitable model for the two datasets. As for the city sample, since not any of the variables was significant in the chosen regression, the city model will be abandoned. The explanation will be present later in the discussion and conclusion.

5.1.2 Robustness test

In order to examine the robustness of the model, a several sensitivity tests will be done by separating the average income into 2 groups. The sensitivity test is aiming for test if the model is changing in different section, for this thesis, the different section is based on the average income, since different income region might have different significance level and coefficient for independent variable and control variables. The intervals for the sensitivity test are the average income from 1339.81 to 16751.44 (group 1 in the table 5-8) and 16751.44 to 32163.07 (group 2 in the table 5-8). Generally speaking, the chosen models have several variables significant, which are employment rate variable and urban population rate variable, the GDP per capita was not highly significant, but it is significant in 95% confidence level in the total model. Population density variable was not significant in the final model, and these elements will be compared with the result of sensitivity test, by comparing the difference, it is going to be clearer that which part in the regression is trustworthy and which part is not.

The distribution of average income

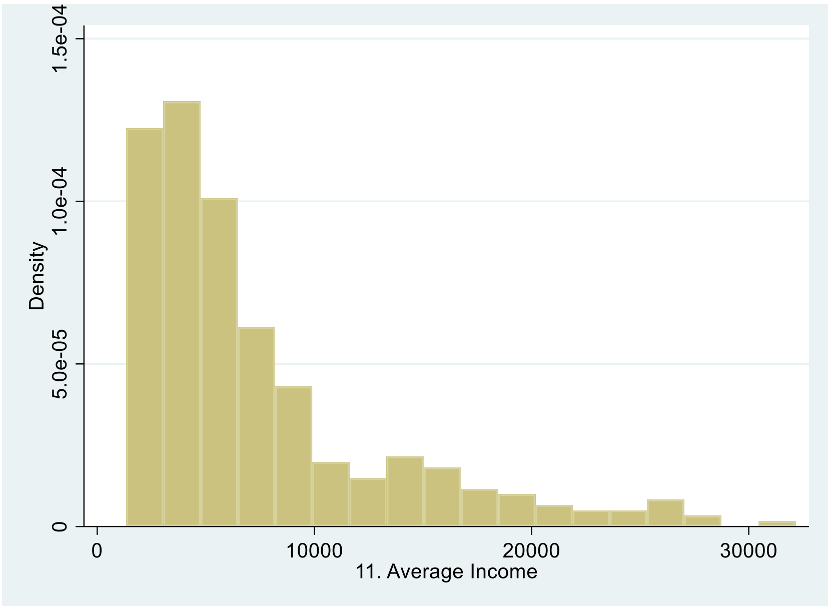

Figure 5.1 The histogram of average income in total

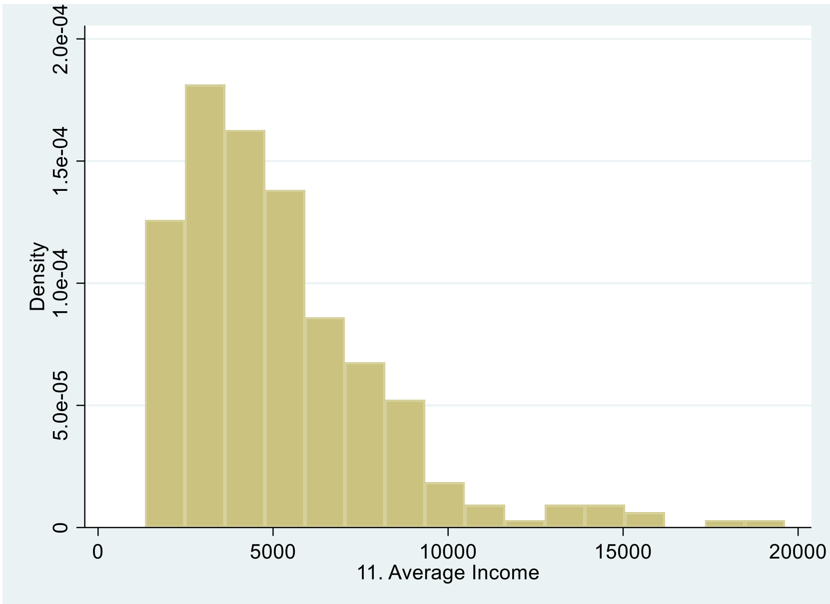

Figure 5.2 The histogram of average income of the counties

As the two figures shows above, the average income distributed not very evenly. The average income in total shows that more samples concentrated in the lower average income, the same situation for the counties, but quite differently for counties, there are less high income sample than the total data set. Furthermore, the average number has been chosen as the threshold instead of the median number, because it is easier to unify the interval size for both model in total and model for counties. However, the downside was quite obvious as well, because the sample size for the two income groups were not equal.

Sensitivity test

One thing that has to be bearded in mind is, the sample of average income is smaller than the rest of the variables. After the model has separated by using average income, the total sample shirked and the number of the sample can be seen in the tables below:

Table 5-8 Sensitivity test for the model in total and county

	(1) In total group 1	(2) In total group 2	(3) County group 1	(4) County group 2
log density	0.0247	-0.100	0.297	0.0762
	(0.98)	(-1.16)	(1.50)	(0.64)
log ethnic	0.0385	0.139	0.0574	-1.548
	(0.69)	(1.22)	(0.94)	(-1.62)
log employ	0.310^{*}	0.237	0.287	0.301^{*}
	(2.03)	(1.70)	(1.90)	(2.39)
log urban	-0.0484	-0.00826	-0.0891	-0.00562
	(-0.54)	(-0.40)	(-0.93)	(-0.33)
log GDP	0.248^{*}	-0.163	0.203^{*}	-0.0790
	(2.19)	(-1.15)	(1.98)	(-0.51)
_cons	0.222	4.787^{***}	-0.184	2.667
	(0.21)	(3.57)	(-0.16)	(1.86)
N	313	172	278	125
R^2	0.323	0.236	0.349	0.168

T statistics in parentheses
* $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As information mentioned earlier described, the sensitivity test also tested the different data sets, but as the city sample are too small, and none of the city models was valid, there will only be the model in total and the model of county be in the sensitivity test. The reason to do it is as follows: there are two final models, which are the model in total and model for counties. By having this kind of sensitivity test, the models will be examined more in detail. In total, all of the significant variables were not as significant as it is in the model, but some of the insignificant variables show a different significance under different conditions.

The very noticeable feature in the table is the GDP per capita. In the total model, the GDP per capita was insignificant, but in this sensitivity test, the first group of the total and the county are significant in 99% of confidence level, which means in the low-income group, the rise of the economic condition does affect the crude birth rate. The population density did not have a remarkable difference between different models.

The employment rate in the model was significant, but after it broke down into different sections, the significant part was only restricted in a certain area but for different models, there are different groups which are significant. For the total model, the lower income group has signed on employment rate, which means the crude birth rate in lower income region in total will be impacted by the employment rate more than the richer region. For the counties, it is quite the opposite, the richer group's employment rate has a significant impact on the crude birth rate. Which can indicate that the county has lower income on average, whereas the cities have more income than counties. The poor region has a higher need for employment, so that also explained the positive impact by the employment rate.

The urban population rate also needs to be noticed in this sensitivity test. The urban population rate was significant in the final model of total and county data set, but in this sensitivity test, it is not significant in all four regressions. It also means in these data sets, it is doubt that if there is a rate significant relationship of urban population rate with crude birth rate.

The sample size of every group is significantly different. The one in the beginning and the one at the end are the same sizes, whereas the two in the middle are the same. Also, there is two tests which have a very small sample size or preferably 0 samples, to be aware of. As for the R square, all of the tests has a higher R-square than the normal one, the revised version of the model are more reliable than the original model, and this will be investigated later.

5.2 Discussion

5.2.1 The interpretation of the results

The models in the regressions and the sensitivity test suggested many interesting results, and they can be combined to explain. The model suggested that population density has no significant effect on the crude birth rate; the ethnic minority population rate was not significant as well. However, the urban population rate has a significant effect, and it is every one percentage change in urban population rate in the region can bring 0.29 percentage change in the crude birth rate. For the urban population rate, every one percentage of change can bring -0.06 percentage change in crude birth rate, and it is significant, but the sensitivity test suggested that it is correlated with average income. The GDP per capita also brings a specific effect on the crude birth rate.

The independent variable of this thesis is population density, but the result suggested there is no significant relationship between crude birth rate and population density in the Xinjiang region regardless of the category of the region. The possible explanations for the insignificant population density can be, the population density as a proxy variable for representing the living cost (Kato, 2018), if the region were not developed enough to suffer from the living cost, or because in that ten years (2000 to 2009) s of the economy in China was booming (Wu, 2007). The theory of Kato (2018) suggested the cities situation, but as a matter of fact, most of the sample are counties, and the unique geographical feature of the region (Han et al. 2010) decided that if the population density was higher in some condition, the living cost might be lower. Also the result contradicted most of the theories of the relationships between population density and fertility (Croix, 2017; Lutz et al., 2007), it can imply that in China, the population density can have a small or no impact on crude birth rate in certain regions, which also means the determinants of fertility in China also differs from other counties.

The ethnic minority also was not significant in any of the models. The reasonable assumption is, even if the general policy suggested the ethnic minority could give more birth, the people on the same economic condition could not afford to give new birth, and still, economic factor has more impact than the ethnic minority rate. Recently in China, Islamophobia is very popular because of the unequal ethnic policy (such as the looser one-child policy towards ethnic minority in China) (Luqiu, Luwei and Fan, 2018). One of the hate speech targets toward the unequal one-child policy, the very loose one-child policy towards ethnic minorities who are also Muslims (Luqiu et al., 2018). As previously mentioned, the ethnic minority in Xinjiang region are mostly Muslim population (Luqiu et al., 2018). This result suggests that the essence of Islamophobia in China was not based on reality, and it might reduce Islamophobia. Furthermore, the insignificant ethnic minority rate and urban rate suggested the one-child policy did not have a severe impact on the cured birth rate in Xinjiang province in the one-child policy years. Zhao & Wang (2018) did the research pointed out that the one-child policy did not impact the fertility in China because the socioeconomic factor impacts Chinese fertility more.

Employment rate, however, shows a difference with the mainstream theories. The mainstream theories suggested that the increase in unemployment can bring an increase in the fertility rate (Kurkin et al., 2018). However, the thesis results suggest that quite reversely, more employment rate bring the fertility rate up, this indicates that an employment rate means a better economic condition, instead of having a negative impact on fertility level (Kurkin et al., 2018), so if the economic situation is getting better, the crude birth rate is getting better too. GDP per capita in the total model was only significant in the confidence of 95%, and the coefficient is 0.195, which also indicated that economic factor affects crude birth rate.

The results regarding the difference between county and cities as follows. First, it is possible that there is not enough sample, because the amount of county was substantially more significant than the cities. Therefore the cities model may get more errors than the counties. Second is because the crude birth rate in the cities are not as it is in the counties, it is more stable in the cities than it is in the counties, because most of the population in cities are urban populations.

The sensitivity test did not only suggested the weak robustness of the models, but it also suggested other links between the variables. GDP per capita was significant in the low-income group, both in the total model and in the county model. Moreover, the urban population rate variable was not significant any more in the sensitivity test.

First of all, the significant GDP per capita in the low average income group. It indicated that among the low-income regions, the better economic conditions they have, and the higher crude birth rate it is. It is the same as one theory they presented in the research of Kurkin et al. (2018). In this research, they suggested that the economic situation matters for the crude birth rate.

The impressive results happened in employment rates insensitivity test. In the total model, the lower income group has a significant employment rate variable, whereas the higher income group in the county model has a significant employment rate variable. It can be explained as follows: in the low-income region, people tend to work more in order to have enough income to raise the children, whereas, in the higher income region, the income was enough for the family to raise the children, so the employment rate will not matter anymore.

The insignificant urban population rate in the sensitivity test, it is probably because of the urban population rate is correlated with an average income level that makes the significance level of urban population rate dissolved in the sensitivity test. That also means the average income has some relationship with the urban population rate, and as the data suggested, the population in the city has a higher average income than the population in the counties in Xinjiang region.

The insignificant urban population rate in the sensitivity test, it is probably because the urban population rate in correlated with average income level that makes the significance level of urban population rate dissolved in the sensitivity test. That also means the average income has some relationship with urban population rate, and as the data suggested, population in city has a higher average income than the population in the counties in Xinjiang region.

5.2.2 The strength and weakness of the study

Strength

The model has been separate into county and city, which is a good indication of urban area and rural area. It solved the problem of how to separate urban area and rural area in a statistically possible way. Furthermore, by doing this, the difference between Chinese county and city will be presented.

In order to consider the one-child policy, the ethnic minority population rate and urban population rate has been taken into account, which is an innovation of the research. By taking those variables into account, the effect of the one-child policy will be better controlled, it can not only apply to this province, but it can also apply to other provinces in China as well, and since the One- Child policy is universal in China.

The general design of the model by using the fixed-effect model, the fixed effect model solved the time variant error term and common error term at the same time. Moreover, by comparing with the OLS model, it is clearer which variable has the time variant error, and it is possible to observe which model is more suitable in this situation.

Weakness

First of all, there is too small amount of sample. As the data section suggested, the sample size was quite small especially for the city sample, and it is shown in the model that the regression of the city was not a successful regression. It can lead to a problem of uncertainty and a very low representative level. Also the considerable amount of years and the regions that were missing from the data set can also causes problems, it can make the results less reliable. Furthermore, taking data from Chinese government related resources can been seen as not trustable, that is also needed to be considered.

The lack of single female variables. Majority of the literature shows the advantage of using a female unique such as female unemployment rate, female labour participate rate, female literacy rate and so on, but because of there is lack of this data in the database, it is impossible to acquire those data for this thesis. However, the problem for using employment rate or income without separating gender is there will be many hidden problems which cannot be discovered or explained.

There are many variables in this thesis constructed, such as average income, it is not fully reliable as previously mentioned. There are some disadvantages of using these constructed variables in the analysis. If there data that constructed this variable were from a different database, that means there will be a risk if they are not using the same statistical method, which gives the data more uncertainty.

6 Conclusion

This thesis mainly investigated the interactions of population density and crude birth rate in Xinjiang region as the example of China. The thesis also examined the impact of the one-child policy on crude birth rate by using ethnic minority population rate and urban population rate of each region. Study shows an insignificant relationship of the interaction between population density and crude birth rate in Xinjiang region, but the study also discovered other facts from the data which has specific implications.

6.1 Research Aims and Objectives

This thesis aimed to discover if the negative relationship between population density and the crude birth rate. The study also examined if the one-child policy affects the crude birth rate by using ethnic minority population rate and urban population rate. Also, the study is aimed to examine if socioeconomic factors can affect the crude birth rate. The argument as the previous research addressed that the socioeconomic factors can affect fertility level, but some scholars (Bongaarts & Watkins, 1996) pointed out the social, economic status can have a minimal effect on fertility rate. This thesis proved that the economic factor affects the crude birth rate. More specifically, the employment rate and GDP per capita, were significant in the model.

The objectives were to prove the existence of negative relationships between population density and crude birth rate. In the theory part, many researchers addressed that there is a negative relationship between population density and fertility level, but all of them are proved in other counties other than China. This thesis tried to prove that there is a specific relationship between population density and crude birth rate, and the results suggested an insignificant relationship between population density and crude birth rate.

To prove the one-child policy has minimal effect on the crude birth rate in China, by using an example of Xinjiang region. The one-child policy is believed to have an impact on fertility rate or crude birth rate, but some researchers suggested that there is no significant impact on fertility rate, and this thesis would aim to prove there are minimal impact from the one-child policy on fertility level. And the results showed the one-child policy had a weak impact on crude birth rate in Xinjiang region.

6.2 Practical Implications

The weak relationships between population density and the crude birth rate were the result of the regressions. In this thesis, the result showed that there is a weak correlation between population density and crude birth rate in Xinjiang region, which also implement that the unique feature of determinants of fertility in Xinjiang province, which can also indicate the unique feature of determinants of Chinese fertility.

The relationship of the socioeconomic factor with the crude birth rate the low-income group. The conclusion is that among the low-income regions in Xinjiang region, the better economic conditions they have, and the higher crude birth rate it is. It proved the theory in the research done by Bongaarts & Watkins (1996) works in the Xinjiang region, which was the theory suggested economic situation matters for the crude birth rate. Therefore, there might be a possibility to improve the fertility rate in areas with this kind of condition.

The one-child policies have week impacts the crude birth rate. The result indicated that with the same economic condition but were permitted to have a second child will not be able to give another birth since the family cannot afford the cost. Therefore the economic factor has more impact than the ethnic minority rate. This results also suggest the base of Chinese Islamophobia and hate speech towards Chinese Muslim group (Luqiu et al., 2018) has no real support, since the Muslim ethnic minority was not benefited from the looser one-child policy.

6.3 Future Research

There is a need for more research on the relationships between population density and fertility rate in China. The fertility of this thesis is based on the crude birth rate of regions, but as many studies show that (Bongaarts & Watkins, 1996; Kurkin et al. 2018) total Fertility rate is more widely used than the crude birth rate. It is more accurate to use specific measurement instead of average ratio, such as category if the individual is an ethnic minority or not instead of and the ethnic minority. It also needs researches in other regions in China to examine the results of this thesis and compare in order to know the situation in China better.

There is a need for more research on the impact of population density in other regions in China. This thesis aiming to discover the relationship between population density and crude birth rate in China, but only chose one province as an example because the thesis was trying to avoid the effect of the one-child policy. In the future, the research could include more regions and cover more population so that there will be more information about China. Moreover, other researches can be compared with this thesis, to study the difference in fertility of different Chinese region.

It also needs more researches on the variation of the one-child policy in China, the different reality in different places. In this thesis, the difference of one-child policy was one of the theories. Moreover, as the literature addressed that difference exists between different provinces and even regions, but there are very few studies about how different the one-child policy implemented in different regions and the effect on the fertility in different regions. There can be some research comprehensively cover the differences in the one-child policy effect on different regions in China.

References

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. *Princeton university press*, Princeton, New Jersey

Beijing Municipal People's Government Research Office Social Affairs Office (2015) Comprehensive report on the floating population of the eight major cities，available at：http://oa.shxyj.org/UploadFile/20130926008/2015-09-18/Issue/klri4fsb.pdf

Bongaarts, J. & and Watkins, S. C. (1996) Social Interactions and Contemporary Fertility Transitions, *Population and Development Review*, 22, no. 4 (December 1996): 639. Available at: https://doi.org/10.2307/2137804.

Boucekkine, R., Croix, D. and Peeters, D. (2007) Early Literacy Achievements, Population Density, and the Transition to Modern Growth, *Journal of the European Economic Association* 5, no. 1 (March 1, 2007): 183–226. https://doi.org/10.1162/JEEA.2007.5.1.183.

China data online, the citation articles of the databases, last access: 2019.05.29, available at: https://www.china-data-online.com/asp/citations.asp

Chinese National statistical bureau, the regulations of how to define the rural and urban area, last access: 2019.05.29, available at: http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/html/append7.htm

Croix, D. and Gobbi, P. E. (2017) Population Density, Fertility, and Demographic Convergence in Developing Countries, *Journal of Development Economics* 127 (July 2017): 13–24. https://doi.org/10.1016/j.jdeveco.2017.02.003.

Easterlin, R. A. 1976. “Population change and farm settlement in the northern United States”, *Journal of Economic History*. 36(1): 45–75.

Fan, P. Park, H. Ouyang, Z. Nguyen, D.D Nguyen, T.T.H. Chen, J., (2018) Urbanization, economic development, environmental and social changes in transitional economies: Vietnam after Doimoi, *Landscape and Urban Planning,* Available online 3 November 2018, available at: https://www.sciencedirect.com/user/login?returnURL=/getaccess/pii/S0169204618311745

Firebaugh, G. (1982) Population Density and Fertility in 22 Indian Villages, *Demography,* 19, no. 4 (November 1982): 481. https://doi.org/10.2307/2061014.

Groat, H. & Neal, A. (1967). Social Psychological Correlates of Urban Fertility. *American Sociological Review*, 32(6), 945-959. Available at http://www.jstor.org/stable/2092847

Goodkind, D. (2017) The Astonishing Population Averted by China's Birth Restrictions: Estimates, Nightmares, and Reprogrammed Ambitions, *Demography*, 54:1375–1400 DOI 10.1007/s13524-017-0595-x

Google map, The xinjiang region map, last access: 2019.05.29, *available at: https://www.google.com/maps/d/viewer?ie=UTF8&oe=UTF8&msa=0&mid=1oo-ajDsuC-nurC3gYT0h8JJcX-w&ll=42.62264071588134%2C84.94628899999998&z=5*

Han, C. Ma, Y. and Luo, H. (2010), Economic Spatial Evolution and Its Mechanism in Xinjiang *Arid Area Geography*, Volume 33, Issue 3 (2010 / 05 / 30) , P449 - 455

Ho, J. (2011), The change of will of giving birth in China (1978-2011), Central University of Finance and Economics Social Research and Database Construction Project, Central University of Finance and Economics Phase III 211 Project

Kato, H. (2018) The Analysis on Disparities of Fertility Rate of Japanese Municipalities, *Public Policy Review*, 24.

Kurkin, R. Branislav, Š. Luděk, Š. and Jiřina, K. (2018) Fertility Factors in Czechia According to the Results of the 2011 Census.Auc, *geographica,* 53, no. 2 (June 22, 2018): 137–48. https://doi.org/10.14712/23361980.2018.14.

Li, C. Lin, L. Dai, M. Chen, Y. Li, X. Ma, J. and Jing J., (2018) One-Child Policy, Weight Status, Lifestyles and Parental Concerns in Chinese Children: A Nationwide Cross-Sectional Survey. *European Journal of Clinical Nutrition* 72, no. 8 (August 2018): 1150–58. https://doi.org/10.1038/s41430-018-0178-y.

Li, Y. (2013), China's roads and new urbanization, "*Real: Modern Management*" 2013 No. 1, pp 24-25

Luqiu L., Luwei R., and Fan Y., (2018) Islamophobia in China: News Coverage, Stereotypes, and Chinese Muslims' Perceptions of Themselves and Islam. *Asian Journal of Communication* 28, no. 6 (November 2, 2018): 598–619. https://doi.org/10.1080/01292986.2018.1457063.

Lutz, W. & Qiang, R. (2007) Determinants of human population growth, *Philosophical Transactionsm of the Royal Society of London Series B, Biological Sciences* available at://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=12396512&site=eds-live&scope=site.

Lutz, W. Maria R. T. and Dustin J. P. (2007) Population Density Is a Key Factor in Declining Human Fertility, *Population and Environment* 28, no. 2 (April 4, 2007): 69–81. https://doi.org/10.1007/s11111-007-0037-6.

Ma, X. & Sun, C. (2011), Chinese family policy 60 years, Beijing Municipal Party School, *Department of Sociology, Teaching and Research Department* 02-0046-07

Malpezzi, S. (2013), Population Density: Some facts and some predictions, *Journal of Policy Development and Research*, volume 15, number 3 pp 183-201 available at: https://www.huduser.gov/portal/periodicals/cityscpe/vol15num3/ch13.pdf

Singh, S. & Casterline, J. (1985), Reproductive Change in Developing Countries: insights from the world fertility survey, edited by John Cleland and John Hobcraft, in collaboration with Betzy Dinesen. London, England, *Oxford University Press*, pp 199-222. available at:https://www.cambridge.org/core/journals/journal-of-biosocial-science/article/reproductive-change-in-developing-countries-insights-from-the-world-fertility-survey-edited-by-john-cleland-and-john-hobcraft-pp-301-oxford-university-press-oxford-1985-1500/F1DA5EB2CFFF85C3264E3E35FE302EC4

Tong, Y. & Wu, Y. (2012) Population Changes and Characteristics Brought by China's Urbanization Process. A*cademic Frontiers: Scientific Development: Deepening Reform and Improving People's Livelihood (I)*. Beijing Social Science Federation, 2012: 15 pp 210-224, available at: http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CPFD&dbname=CPFD0914&filename=BJSL201212004026&v=MDAwNDFNT1VLcmlmWnU5dkZDbmpVTHpNSTFvV0p5ZllZckc0SDlQTnJZOUZZT3NOQ2hOS3VoZGhuajk4VG5qcXF4ZEVl

Tsui, M. (1989). Changes in Chinese Urban Family Structure. Journal of Marriage and Family, 51(3), 737-747. doi:10.2307/352172

Vanlandingham, M. & Hirschmann, C. (2001), Population pressure and fertility in pre-transition Thailand, *Population Studies*, 55 (2001), 233–248

Wang, J. & Ding, S. (2005) How much agricultural surplus labour is there in China? *Chinese Social Science,* No. 5, 2005

Wang, Q. Su, M. Li, R., (2018) Toward to economic growth without emission growth: The role of urbanization and industrialization in China and India, *Journal of Cleaner Production*, Volume 205, pp 499-511 available at:https://www.sciencedirect.com/science/article/pii/S095965261832746X

World Bank, (2012), Overview of China's small and medium towns, available at: http://documents.worldbank.org/curated/en/652091468219309036/pdf/704950ESW0CHIN0BLIC00CHINESE0SMT0cn.pdf

World Bank, Chinese inflation rate, last access: 2019.05.28 available at: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=CN

Wu, H. W. (2007) The Chinese GDP Growth Rate Puzzle: How Fast Has the Chinese Economy Grown?, *Asian Economic Papers* Volume 6 , No. 1

Xinjiang regions government, 2010: available at: http://law.51labour.com/lawshow-99738-2.html

Zhang, K. & Song, S. (2003), Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses, *China Economic Review*, Volume 14, Issue 4, 2003, Pages 386-400

Zhang, X. Huang, C. Zhang, Q. Chen, S. and Fan, Q. (2016), Partially two children policy and the two children family policy family Comparison of birth intentions and its enlightenment, *Population research*, Vol. 40，No. 1 January 2016

Zhao, Z., and Zhang, G. (2018) Socioeconomic Factors Have Been the Major Driving Force of China’s Fertility Changes Since the Mid-1990s, *Demography* 55, no. 2 (April 2018): 733–42. https://doi.org/10.1007/s13524-018-0662-y.

Appendix A

The Hausman test for the total models:

The total model with urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 53.43

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The total model with urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 40.51

Prob>chi2 = 0.0002

(V_b-V_B is not positive definite)

The total model without urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 53.79

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The total model without urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 45.73

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The Hausman test for the county models:

The county model with urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 46.87

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The county model with urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 54.51

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The county model without urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 50.43

Prob>chi2 = 0.0000

(V_b-V_B is not positive definite)

The county model without urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 38.93

Prob>chi2 = 0.0002

(V_b-V_B is not positive definite)

The Hausman test for the city models:

The city model with urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 9.88

Prob>chi2 = 0.7037

(V_b-V_B is not positive definite)

The city model with urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 21.73

Prob>chi2 = 0.0597

(V_b-V_B is not positive definite)

The city model without urban population rate, with GDP per capita:

Test: Ho: difference in coefficients not systematic

chi2(12) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 10.64

Prob>chi2 = 0.5600

(V_b-V_B is not positive definite)

The city model without urban population rate, with average income:

Test: Ho: difference in coefficients not systematic

chi2(12) = (b-B)'[(V_b-V_B)^(-1)](b-B)

= 24.85

Prob>chi2 = 0.0155

(V_b-V_B is not positive definite)

Appendix B

Missing counties and cities from the data sets:

missing regions	Name	Population in 2013 (10,000 persons)
1	Tianshan District	58.61
2	Shayibak District	55.05
3	Xinshi District	59.69
4	Shui Mogou Distric	28.47
5	Tou Tunhe District	23.06
6	Da Bancheng District	4.21
7	Midong District	27.67
8	Dushanzi District	5.72
9	Karamay District	19
10	Bai Jiantan District	4.07
11	Urhe District	0.23
12	Yining [Gulja] City	53.57
13	Kuytun City	30.15
14	Yining [Gulja] County	44.85
15	Qapqal Xibe Autonomous County	19.79

16	Huocheng [Korgas] County	41.29
17	Gongliu [Tokkuztara] County	20.8
18	Xinyuan [Kunes] County	32.52
19	Zhaosu [Mongolkure] County	19.12
20	Tekes County	17.45
21	Nilka County	18.99
22	Tacheng [Qoqek] City	17.19
23	Usu City	23.14
24	Emin [Dorbiljin] County	22.14
25	Shawan County	21.88
26	Toli County	10.19
27	Yumin [Qagantokay] County	5.98
28	Hoboksar Mongol Autonomous County	5.47
29	Altay City	23.42
30	Burqin County	7.22
31	Fuyun [Koktokay] County	9.67
32	Fuhai [Burultokay] County	7.57
33	Habahe [Kaba] County	8.81
34	Qinghe [Qinggil] County	6.45
35	Jemnay County	3.91
36	Lop County	25.84
37	Qira County	15.96
38	Yutian [Keriya] County	26.23
39	Minfeng [Niya] County	3.83
40	Shihezi City	62.26

41	Aral City	18.61
42	Tumxuk City	16.41
43	Wujiaqu City	9.1

Appendix C

The white tests:

In total:

The model with urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(20) = 62.54
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	62.54	20	0.0000
Skewness	4.91	5	0.4274
Kurtosis	6.37	1	0.0116
Total	73.82	26	0.0000

The model with urban population rate and with average income:

Cameron & Trivedi's decomposition of IM-test
White's test for Ho: homoskedasticity
Against Ha: unrestricted heteroskedasticity

chi2(20) = 45.00
Prob > chi2 = 0.0011

Source	chi2	df	p
Heteroskedasticity	37.09	20	0.0114
Skewness	3.83	5	0.5743
Kurtosis	4.01	1	0.0452
Total	44.93	26	0.0120

The model without urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 57.21
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	57.21	14	0.0000
Skewness	4.28	4	0.3698
Kurtosis	5.49	1	0.0192
Total	66.97	19	0.0000

The model without urban population rate and with average income:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 22.80
Prob > chi2 = 0.0636

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	22.80	14	0.0636
Skewness	4.08	4	0.3947
Kurtosis	3.14	1	0.0763
Total	30.03	19	0.0515

County:

The model with urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 38.72
Prob > chi2 = 0.0004

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	38.72	14	0.0004
Skewness	2.86	4	0.5811
Kurtosis	4.02	1	0.0450
Total	45.60	19	0.0006

The model with urban population rate and with average income:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(20) = 30.46
Prob > chi2 = 0.0627

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	30.46	20	0.0627
Skewness	12.62	5	0.0272
Kurtosis	5.58	1	0.0181
Total	48.67	26	0.0045

The model without urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(20) = 33.03
Prob > chi2 = 0.0335

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	33.03	20	0.0335
Skewness	14.67	5	0.0119
Kurtosis	2.60	1	0.1068
Total	50.30	26	0.0029

The model without urban population rate and with average income:

White's test for Ho: homoskedasticity

against Ha: unrestricted heteroskedasticity

chi2(14) = 28.92
Prob > chi2 = 0.0107

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	28.92	14	0.0107
Skewness	14.76	4	0.0052
Kurtosis	2.63	1	0.1050
Total	46.31	19	0.0004

City:

The model with urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(20) = 30.46
Prob > chi2 = 0.0627

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	30.46	20	0.0627
Skewness	12.62	5	0.0272
Kurtosis	5.58	1	0.0181
Total	48.67	26	0.0045

The model with urban population rate and with average income:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(20) = 33.03
Prob > chi2 = 0.0335

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p

Source	chi2	df	p
Heteroskedasticity	33.03	20	0.0335
Skewness	14.67	5	0.0119
Kurtosis	2.60	1	0.1068
Total	50.30	26	0.0029

The model without urban population rate and with GDP per capita:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 28.92
Prob > chi2 = 0.0107

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	28.92	14	0.0107
Skewness	14.76	4	0.0052
Kurtosis	2.63	1	0.1050
Total	46.31	19	0.0004

The model without urban population rate and with average income:

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 28.47
Prob > chi2 = 0.0123

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	28.47	14	0.0123
Skewness	12.65	4	0.0131
Kurtosis	5.54	1	0.0186
Total	46.65	19	0.0004