Paper draft prepared for the Population Association of America Annual Meeting 2015

A Study of County-level Fertility Rate in China using Geographically Weighted Regression Method

Donghui Wang, Ph.D. Candidate

Department of Rural Sociology, Pennsylvania State University

Geographic Information Analysis Core, Population Research Institute, Pennsylvania State University

Jingjing Wang, Ph.D. Candidate

Department of Applied Economics, University of Minnesota

Abstract

The fertility pattern in China exhibits great heterogeneity across space primarily for two reasons: first, the socioeconomic development displays salient regional disparities; second, the fertility policy-making and enforcement is also made on a localized basis. The heterogeneous nature of the Chinese fertility therefore suggests that research on the determinants of the fertility in China based on a single "global" model may fail to capture the intricacies and complexities of the relationships at the local level. Using the county-level geocoded 2010 population census of China, our study adopts Geographically Weighted Regression (GWR) approaches to identify the local relationships between the county-level fertility rate and a set of socioeconomic and policy-related fertility determinants. Our study highlights some facets of the relationships that are completely hidden in a global model, and also goes beyond the provincial boundaries. Our study also draws attention to the areas that require special attention in the future studies of fertility in China.

.

We thank Stephen A Matthews for important insights and suggestions on the draft of this paper and also thank great support from the Geographic Information Analysis Core at Population Research Institute at the Pennsylvania State University.

Introduction

The overall fertility level in China has experienced a remarkable decline over the past few decades, and has drawn considerable scholarly attention. The vast literatures have confirmed that the overall fertility has reached to a below-replacement level since the 1990s, and this trend is continuing until present (Feeney, 1994). The decline in fertility is generally regarded as a response to the socioeconomic development and the relative coercive fertility policy of China (Poston, 1987; Cai, 2010). Consensus has been reached that the economic development level and the fertility policy vary across space. However, among the studies on fertility in China, only a small share of available research has seriously considered the spatial variation of the fertility decline.

In an effort to incorporate the spatial context into the study of fertility, there are generally two research orientations: one direction of research focuses on the spatial dependency of fertility. The spatial dependence orientation highlights that “the structure of the correlation between random variables is derived from a specific ordering, determined by the relative position of the observations in geographic space” (Anselin, 2010). The other research orientation that is less explored is the spatial nonstationarity of fertility. The spatial nonstationarity is a condition in which “a simple global model cannot explain the relationships between some sets of variables” (Brunsdon, 1996). Studies on the spatial nonstationarity of fertility usually tend to highlight the underlying different processes that could not be included into an all-embracing framework.

Our study takes the second research orientation by focusing on the spatial non-stationary of the fertility in China. The study uses the Geographic Weighted Regression techniques. It hopes to add to the existing literatures by highlighting the spatially uneven process of the fertility transition in China, as well as to explore the different impacts the same factor has on fertility based on different geographic locations (Brunsdon, 1998). The structure of our study is as follows: the next section will provide a review of the background on the fertility transition pattern in China, as well as its underlying driving force. Then the primary Exploratory Spatial Data Analysis will be conducted for the county-level fertility variable as a way to offer the preliminary insights about the spatial variation of the fertility level. The paper will then proceed with a discussion of the global estimates of Ordinary Least Square Regression and the local estimates of the Geographic Weighted Regression. The paper will end with the conclusion section and the discussion of the limitation.

II Background Literature

Economic Development, Fertility Policy and Fertility Decline

The rapid fertility decline is believed to be the response of both the socio-economic structural change and the coercive family planning policy. China embarked on the economic reform in the late 1980s that basically shifted the planned economy to the market-oriented economy. The reform brought about a significant structural change in the society, which provided an overarching driving force of the fertility decline.

Parallel with the economic reform in the late 1970s, the central government set an ambitious population target that asserted the population growth should achieve 0.5 growth by 1985 and zero population growth in the year 2000 (Coale, 1981). And the only effective way to achieve this target was to restrict every couple to have one child. In 1982, the central government revised the constitution, stating that "both husband and wife have the duty to implement fertility planning". This makes China probably the only country that has content of fertility policy in its Constitution (177: Zhou, 1996).

But what makes the fertility pattern in China potentially bear the salient feature of spatial heterogeneity could be attributed to the uneven regional economic development patterns and the fertility policy and implementation differences at local levels. To begin with, economic development does not keep the same pace in China. An inter-provincial comparison revealed that ever since the initiation of the market reform, the eastern coastal provinces keep the highest growth rate, followed by the provinces in the middle regions, and whereas the western inland provinces are the least developed ones (Huang, 2003). Not only does the uneven economic development exist among provinces, but also within provinces, which is often presented as the rural-urban inequalities. (Zhao, 2000).

In terms of fertility policy and enforcement differentials, the national "one child per couple" policy was localized in the 1990s. The coercive population policy created a big tension between the central government and the citizens. To tackle this problem, the policy went through several amendments and eventually left the state policy of population control up to each province. Two most salient policy exemptions occur among agricultural households and the ethnic minority households. By analyzing the provincial-level fertility policy, Baochang's study (2007) indicates that the one-child policy is more strictly applied to the non-agricultural households who usually reside in the urban area, and the Han people, who are the majority ethnic group in China, whereas the non-agricultural households and the ethnic minorities are largely exempted from the one-child policy. Apart from the legal documentation, Attane's research (2002) revealed that the local resistance, as well as the enforcement of the fertility policy also varies, which also adds to a layer of complexity of the fertility policy.

To briefly sum up, the uneven economic development, coupled with the fertility policy differentials at local levels contributes to the spatial heterogeneity of the fertility level in China.

Analytical Strategy

We first calculate the basic non-spatial descriptive statistics and conducts the Exploratory Spatial Data Analysis for the dependent variable of fertility level. Second, we examined the relationships between county level fertility and a set of covariates at the global level. Finally, the GWR results will be presented. To better visualize the results, maps of the local R-squared and for each of the independent variables will be presented. The first step of preliminary ESDA is conducted with *Geoda* and Arc GIS, and the descriptive analysis and the global model were conducted with SAS. Finally the GWR of was conducted with GWR4.0.

III Data, Measurement and the Analytical Strategy

Data

Our initial unit of analysis is the administrative counties. There are 2,872 counties or county equivalent administrative units in total. The 2010 county-level shapefile was linked to the 2010 population census. The dataset and the shapefile were both obtained at the *China GeoExploer II* [1]maintained by the University of Michigan.

Measurement

The Fertility level- The 2010 census does not provide information on the age-specific-fertility rates at the county level. Therefore, for this study, we use the average number of alive births born to a woman who are at their reproductive age (15-45) as our measure of county-level fertility rate.

Ethnic Composition

Aside from the majority *Han* people, there are 55 types of minority groups that officially recognized by the central government in China. They are identified by ethnic, cultural and religious criteria (Attane, 2000). Despite its overall coerciveness, the fertility policy is relatively lenient for the ethnic minorities in China. Upon its promulgation in 1970s, the national legislation has been consistent in stating that minority couples (including one member of the couple belongs to an officially recognized minority population group) are allowed to have the second child (Attane, 2000). Furthermore, the local fertility policy also grants further exemptions to ethnic minorities. A review of the local fertility policy in China further revealed that some provinces allow the minority couples to have the third child (Baochang 2007).

[1] For more information, see http://chinadataonline.org/cge/

In our analysis, we use the percentage of the ethnic minorities to capture the differences of the birth quota allowance between ethnic minorities and the majority Han Chinese.

Non-agricultural employment and population density

Industrialization and urbanization are constantly identified as the driving force of the fertility transition in both western and non-western countries. The process of industrialization and urbanization is accompanied by high levels of labor force demand, which drives people out of the traditional agricultural production activities. Urbanization brings about the change in the norms, and the rising cost of childbearing, which lead to a decline in fertility.

In our analysis, we include the percentage of population that undertaking non-agricultural employment as an indicator of the level of industrialization. In addition, I also include population density as an indicator of the urban-rural differences.

Female Education

Formal education has also been constantly identified as an important determinant for fertility level. As Jejeebhoy has pointed out, the higher level of female education direct affect fertility" by delaying entry into marriage, by reducing breastfeeding duration and intensity, by lack of observance of traditional postpartum taboos. " (1996). We use the percentage of females that have at least a high school education to the total female population as a measurement of female education at the aggregated level.

Family Structure

Family structure is also a salient predictor that has constantly been identified in the fertility literature. Living in an extended family structure may promote fertility, one of the most important economic reasons is that other relatives may share part of the burden of child rearing and caring (Davis, 1956). Strong clan or joint family structure encourages the early entry into sex union and formal marriage and in doing so have a positive effect on fertility. In order to take into account of this factor, we use the percentage of one generational households to the total households as an indicator the aggregated trend of family structure.

Household Living Standard

The impact on household income and living standard of fertility is also frequently explored in the literature. The increased income at the aggregated level is often associated with the decrease of fertility. The explanation often falls to the classic “quality-quantity” tradeoff, meaning “households prefer to quality service to their children rather than increasing the number of children” (Hondroyiannis, 2004).

Because there is no direct measurement of household income in the 2010 population census, we use the percentage of households that do not have access to tap water is used as an indicator of poor household living standard, and more broadly the low level household income at the county level in the subsequent analysis.

IV Results

4.1ESDA and the Global Model

The quantile map of the dependent variable is presented in figure 2. This map indicates that the fertility rate is higher in the southeast and southwest regions. Another core area characterized by high fertility comprises part of the Tibetan and Xinjiang provinces.

Table 1 shows that the average number of live births for a woman aged from 15 to 60 is 1.41 in the year 2010. The county level percent of the ethnic minority population is 16.22 percent, but the statistic varies from zero percent to 99 percent. The county level percentage of nonagricultural population is 29.51.On average, there are only 19.46 percentage of females who have received at least high school education (12 years of schooling), and only 38.45 percent of households have access to tap water. The average population density is 1435 people per square kilometers. Finally, the global Moran's I suggest that all the variables are spatially auto-correlated.

Table 1 Descriptive statistics of Mean, SD, Min, Max, Moran's I (N=2,872)

Variable	Mean	Std Dev	Min	Max	Moran's I[a]
Avg. Live Birth born to a woman	1.41	0.30	0	2.45	0.67***
% Ethnic minority population	16.22	28.99	0	99.78	0.87***
%Non-agricultural population	29.51	23.56	0	99.40	0.54***
% Females have at least high school Ed	19.46	0.13	0.01	68.31	0.56***
%Nuclear Family	31.70	0.10	0.07	80.66	0.70***
Population Density (Person/ km²)	143.5	9486.68	0	33968.25	0.31**
% Households do not have access to tap water	38.45	0.27	0	100.00	0.61***

**p<0.01; ***p<0.001 The significance levels are based on 999 times of permutations. a: Moran's I is calculated based on the first order queen's weight matrix.

To shed light on the patterning and magnitude of local spatial patterns, we explore the local clusters of fertility level. There are two roles for LISA: (1) identify the significant local spatial clusters and (2) a diagnostic of local instability in measures of global spatial patterns (Anselin 1995). Figure 2.b displays the local cluster of fertility level uses Queen's first continuity metrics. The local cluster map reveals that counties with high levels of fertility tend to cluster on the western boundaries of Xinjiang provinces and the southern provinces such as Yunnan. The low-low values mainly reside in northeastern counties that near to the coastal line.

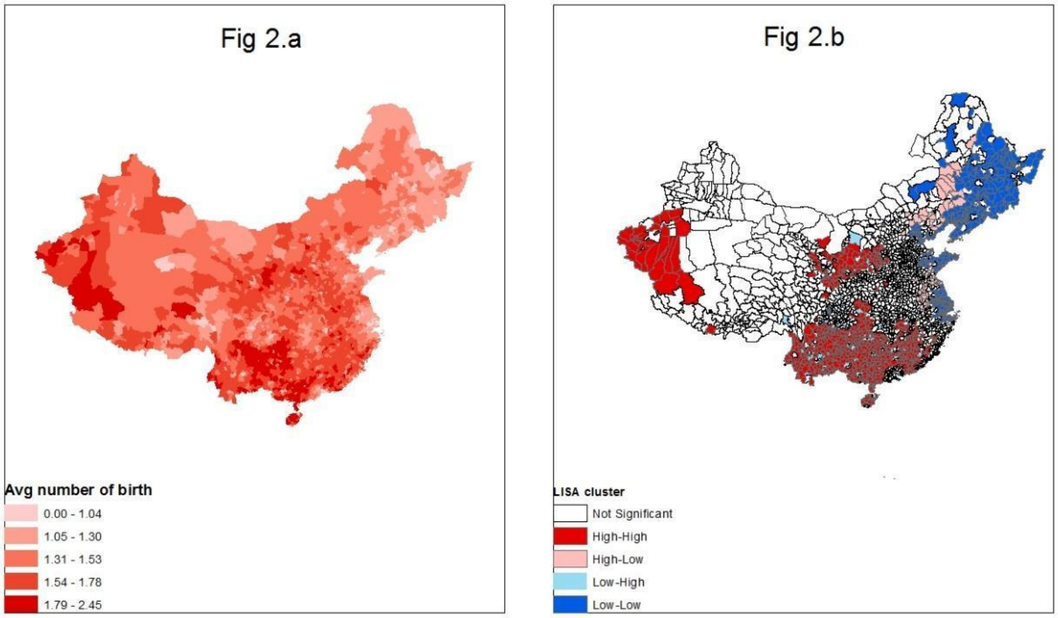


Table 2 presents the standardized estimates of the global model. Model 1 only includes two covariates that closely related with the fertility policy differentiation: one is the percentage of ethnic minority, the other is the percentage of non-agricultural households. As mentioned before, the ethnic minority and the agricultural households are largely exempted from the one child policy, therefore the percentage of ethnic minorities is expected to be positively associated with the county-level fertility while the percentage of nonagricultural households is expected to be negatively associated with the county-level fertility in the global model. Model 2 is the full model that includes all the identified predictors. Model3 adds additional 32 provincial-level dummy variables to take into account of the provincial-level difference. The reference group is the Guizhou province. In terms of the overall fitness, if only taking into account of the ethnic component and the household registration, the model has already explained overall 54.3 percentage variations of the county-level fertility rate (Model1). Adding additional socioeconomic covariates in Model 2 further explains overall 67.6 percentage variations of the fertility level. In Model 3, when adding provincial-level dummies the percentage of ethnic minorities and the percentages of agricultural households become insignificant, this is also expected, as the fertility-related laws and regulations are regulated at the provincial level. In model3, other predictors that capture socioeconomic factors, such as the percentage of females who have at least high school educations, the percentage of one generational families and the percentage of household that have access to tap water, are all significantly associated with county level fertility, and the signs of the relationship are consistent with previous literatures. The only exception is population density, which is statistically insignificant in both model2 and model3. We also conducted the multiconllinarity test for the all the explanatory variables. Turns out the variance inflation factors for the global explanatory variable parameters are all within the reasonable range, suggesting that multiconlinnarity is not a severe issue for the global model.

Table 2 Ordinary Least Squared Estimates of the Global model

	Model1	Model2	Model3
Variable	Beta	Beta	Beta
%Minority	0.141*** (0.000)	0.041*** (0.000)	0.028 (0.000)
%Non-agricultural Population	-0.695*** (0.000)	-0.167*** (0.000)	0.039 (0.000)
% one-generation family		-0.122*** (0.039)	-0.101*** (0.037)
% Females have at least high school Ed		-0.527*** (0.048)	-0.608*** (0.043)
% Households do not have access to tap water		0.096*** (0.015)	1.119*** (0.013)
Population Density		-0.006 (0.00)	-0.000 (0.000)
Provincial-level Dummies[a]	No	No	Yes
Adjusted R^2	0.544	0.676	0.816
AICc	1,021.011	-2,043.474	-3633.773

a: The reference province is Guizhou province

The GWR Model Results

The 5-number summary of the local results is presented in Table 3, as well as the Diff of Criterion that test for the spatial variability that automatically generated in GWR 4.0. The non-stationarity test indicates that all the variables in the model are non-stationary across the space. In terms of the model fitness, compared to the local model, the local model explains 91.1% percent variation of the dependent variable, and has an AICc of -5,157, which suggest that the local model is a significantly better fit compared to the glacial model. The optimized bandwidth suggested by GWR 4.0 is 75. From Table 3 it can be inferred that the local model coefficients differ from global coefficients in two ways. First, the value of the local estimate varies, which again suggests that one global model may not able to reflect the local uneven processes. Second, a comparison between the lower quantile and upper quantile estimates indicates that the sign of the coefficient may change for some of the predictors, such as the percentage of minority, the non-agricultural population and so on, suggest an inverse relationship, which is also something that cannot captured by global model.

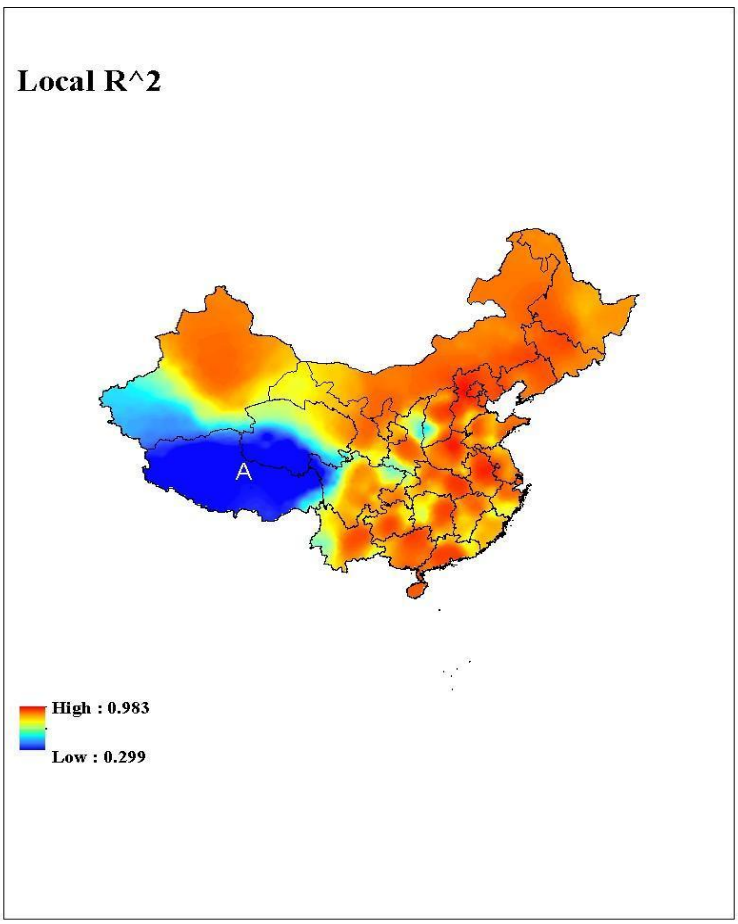
Table 3 GWR Parameter Summary and Test for Spatial Variability of Parameters (N=2, 872)

	Min	Lower Quantile	Median	Upper Quantile	Max	DIFF of criterion[a]
Intercept	-83.467	1.234	1.367	1.526	3.509	
%Minority	-12.976	-.196	.0027	.071	4.069	-739.948
%Non-agricultural population	-.373	-.070	-.022	.229	0.452	-1634.460
% Nuclear family	-.166	.038	.005	.044	0.182	-976.0320
% Females have at least high school Ed	-.627	-.243	-.185	-.135	0.643	-2316.285
Population Density	-562.081	-.067	.002	.056	9.552	-1930.070
%Households have no access to tap water	-.201	-.002	.022	.053	0.466	-961.545
Adjusted R-square	0.911					
AICc	-5,157.152					

a: positive value of Diff-Criterion suggests non spatial variability

Fig 3 is the map of the local R square. As shown in the map, the variance of the local R square ranges from 29.9 to 98.3 percent. The local R-squared map indicates that the model fit quite well in most of the regions in China, which suggest that the identified determinants works for most of areas in China. But there are exceptions. For example, the model is poorly fitted for the greater Tibet area, which include Tibet province and its surrounding areas of Qinghai and Xijiang provinces (Indicated as region A), suggesting that the fertility behavior in Tibetan areas could not be entirely explained by the above-identified covariates. Other slightly poorly fitted areas include the borders between Sichuan and Shaanxi provinces and the mountainous region of Shanxi and Shaanxi provinces. Researchers should at least may pay special attention on the distinctiveness of the fertility pattern of these three highlighted areas. The standard deviation map of the GWR model residual is also provided in Fig A.2 in Appendix. Compared with the OLS model, the overall residual of GWR model is smaller, and more randomly distributed.

Fig.3 Local R-squared



The Map of Local Estimates

To better inform the results of local estimates, Figure 4 displays the local association between the fertility level and six predictors. Noted that the map only presents the significant association at the level of 0.05. Firgure4. a is the local estimates of the percentage of ethnic minority population. Recall that the global model indicates that there is a positive association between the percentage of ethnic minorities and the fertility level, which probably due to the reason that the fertility policy for the ethnic minorities are less cohesive. The GWR results indicate that such positive association does not apply to all the areas. For one thing, only a relatively small portion of areas present statistic significant association between the fertility level and percentage of ethnic minorities. For another, some areas actually present the *negative* relationship between the percentage of ethnic minorities and the fertility level. These counties are most notably located in the eastern coastal provinces such as Guangdong and Shandong provinces. This finding is potentially important, as it suggests that the policy differential between the Han majority and ethnic minority may not be the sole driving force for their fertility differentials.

The negative association between the percentage of non-agricultural population and the fertility level is partly confirmed in Fig2.b. Similar to the local estimates of percentage of ethnic minorities, the percentage of non-agricultural population is also only statistically significant at part of the northeast and northwest regions, leaving the vast majorities of region statistically insignificant. Furthermore, there exists a positive relationship between the percentage of non-agricultural population and the fertility rate at few counties that near to the coastline, which is different from the conventional story of urbanization and industrialization, which suggest that the increase of the non-agricultural employment will disincentives the individual’s willingness to have more children.

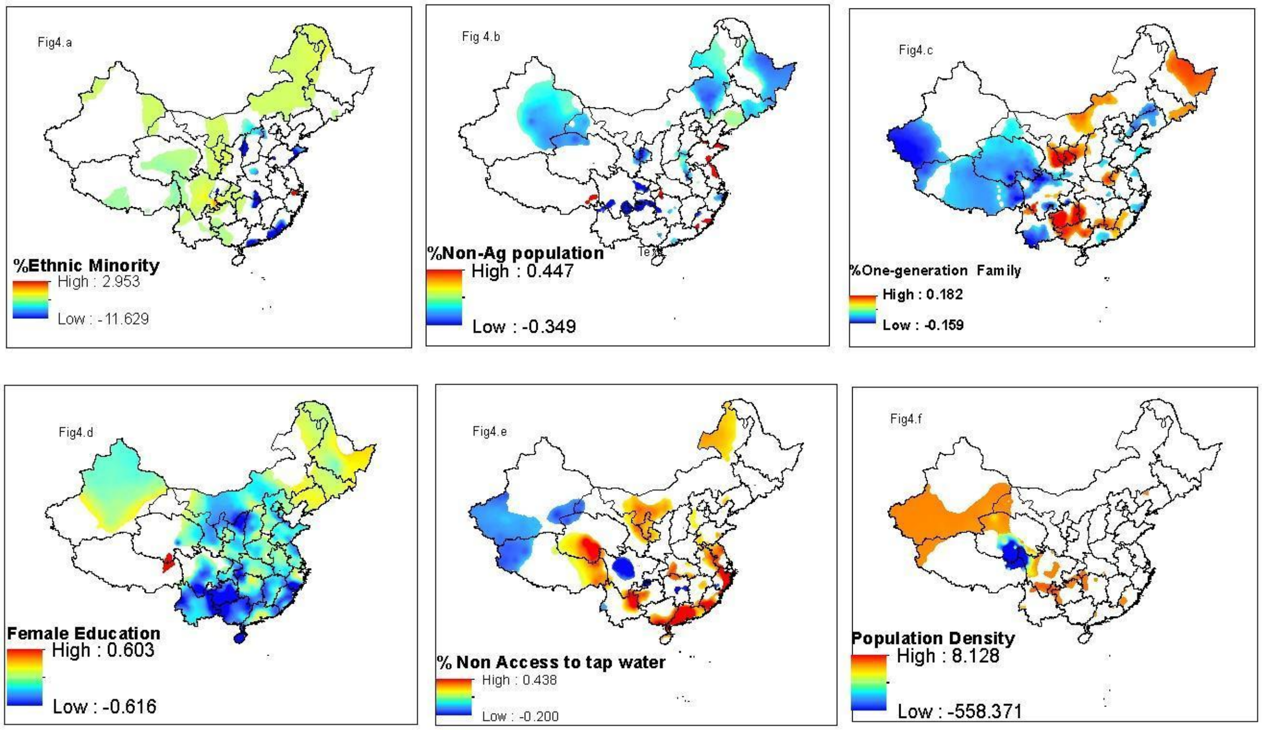
The local coefficient map of the percentage of nuclear families (Fig 4.c) presents a mixed result: The western regions (include Qinghai, Tibetan and part of Xingjiang), and part of east coast area are consistent with the global story to the extent that a higher percentage of nuclear families are associated with lower fertility rate. However, in the central provinces, the association between the percentage of nuclear families and the fertility level is positive. This begs the question of the universal conclusion that the positive association of the nuclear family and the low fertility. As suggested by Back (1973), the nuclear family structure may not always associated with the low level of fertility, and in order to gain further insight, “the relationship between certain aspects of familial interaction and fertility are studied: machismo, female matrimonial fidelity and modesty, husband-wife dominance and conjugal role relationships, marital communication, sexual gratification, and marital adjustment. ”

Female education (Fig 4.d.) offers the most significant explanatory power in the local model, as the negative association between female education and the fertility is significant in most of the areas in China. But there are exceptions of Tibetan area and the northern areas of Inner Mongolia, where the association is insignificant. This finding may have significant policy implication, as these insignificant areas may the targeted improvement areas.

Noted that I use the percentage of the household that do not have access to tap water as an indicator for how poor the living standards are for the local residents, the positive association between poor household living standards and the high level of fertility is also confirmed in Fig.4.e, as among the areas that has significant associations between these two, most of the associations are positive. Exemptions also exist among the counties at western border and inland Sichuan provinces.

Finally, population density is statistically insignificant in predicting fertility at the global level, however the GWR estimates suggest that in certain part of Xingjiang province, the population density is actually modestly positively associated with fertility level, and furthermore, the association becomes negative in the borders of Qinghai and Tibetan areas.

Fig.4 The GWR Estimate Results



V Conclusion

The fertility transition in China is not a homogeneous process. The most important conclusion that can be drawn from this study is that the differences and the complexity of localities cannot be taken into account by a single global model. The wide range of socioeconomic structures, coupled with the complexities of fertility policy, makes the spatial nonstationarity a salient feature of the fertility in China.

The region that deserves special consideration is Tibet and its surrounding areas, which also include the northern part of Xingjiang and Qinghai provinces. The universal determinants have very poor explanatory power of the fertility pattern in this area, since the local R square obtains the lowest value of roughly 0.3 at this region. And this also coincides with the visual inspection of the local estimate map. Of all the six predictors, five of them are statistically insignificant at this region, leaving only one predictor, the percentage of one-generation family, which is consistent with the global story.

Apart from the variations of the overall fitness of the model, other findings based on the observation from the map of local estimates may potentially help inform the policy design. For example, the local map underscores the importance of females' education in reducing fertility, as females' education has a positive and significant association in most of the areas. Although the importance of females' education has been well identified, this study reveals that the positive role of females' education is consistent with local estimates.

While these findings are highly suggestive, they need to be tempered by a discussion of this study's limitations. First, the study is limited by only using the population census, which does not include the direct measure of the economic development level such as GDP per capita or the average household income. Although the study makes an effort to use other indicators to take into account of the aggregated economic development level, it nevertheless brings measurement errors. The 2010 census somehow lacks the statistics on child mortality, which is also one of the predictors that has been confirmed to be highly associated with fertility rate. Further improvement can be made by combining different sources of the dataset so that to have a better measurement of the independent variables. Second, the study only studies the fertility pattern using a single-year census; therefore it cannot be used to make inference on the causal relationships of these determinants and the ongoing fertility transition.

References

Anselin, Luc. 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89 (1): 3-25.

Attane, Isabelle. 2002. China's family planning policy: An overview of its past and future. *Studies in Family Planning* 33 (1): 103-13.

Back, Kurt W., and Paula H. Hass. 1973. Family structure and fertility control. .

Baochang, Gu, Wang Feng, Guo Zhigang, and Zhang Erli. 2007. China's local and national fertility policies at the end of the twentieth century. *Population and Development Review* 33 (1): 129-48.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28 (4): 281-98.

Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton. 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (the Statistician)* 47 (3): 431-43.

Cai, Yong. 2010. China's Below-Replacement fertility: Government policy or socioeconomic development? *Population and Development Review* 36 (3): 419-40.

Davis, Kingsley, and Judith Blake. 1956. Social structure and fertility: An analytic framework. *Economic Development and Cultural Change*: 211-35.

Dorius, Shawn F., and Glenn Firebaugh. 2010. Trends in global gender inequality. *Social Forces* 88 (5): 1941-68.

Guilmoto, Christophe Z., and S. Irudaya Rajan. 2001. Spatial patterns of fertility transition in indian districts. *Population and Development Review* 27 (4): 713-38.

Hondroyiannis, George. 2004. Modeling household fertility decisions in greece. *The Social Science Journal* 41 (3): 477-83.

Huang, JR-Tsung, Chun-Chien Kuo, and An-Pang Kao. 2003. The inequality of regional economic development in china between 1991 and 2001. *Journal of Chinese Economic and Business Studies* 1 (3): 273-85.

Işik, Oğuz, and M. Melih Pinarcioğlu. 2006. Geographies of a silent transition: A geographically weighted regression approach to regional fertility differences in turkey. *European Journal of Population/Revue Europeenne De Demographie* 22 (4): 399-421.

Jejeebhoy, Shireen. 1996. Womens education fertility and the proximate determinants of fertility. .

Loftin, Colin, and Sally K. Ward. 1983. A spatial autocorrelation model of the effects of population density on fertility. *American Sociological Review*: 121-8.

Ping, Tu. 2000. Trends and regional differentials in fertility transition. *The Changing Population of China.Blackwell Publishers*.

Poston, Dudley L. 2000. Social and economic development and the fertility transitions in mainland china and taiwan. *Population and Development Review*: 40-60.

Shrestha, Prasanna Man. 2006. *Comparison of Ordinary Least Square Regression, Spatial Autoregression, and Geographically Weighted Regression for Modeling Forest Structural Attributes using a Geographical Information System (GIS)/remote Sensing (RS) Approach*.

Wang, Jiamin. 2008. China's regional disparity in demographic transition: A spatial analysis. *Review of Regional Studies* 38 (3).

Weeks, John R., Arthur Getis, Allan G. Hill, M. Saad Gadalla, and Tarek Rashed. 2004. The fertility transition in egypt: Intraurban patterns in cairo. *Annals of the Association of American Geographers* 94 (1): 74-93.