A Space-Time Model of Fertility and Development in China, 1982-2000

Katherine King
University of Michigan

Abstract

Although China is extremely regionally diverse and fertility policy is implemented at a local level, yet research often implicitly assumes spatial homogeneity. This paper applies exploratory spatial data analysis (ESDA) and spatial panel regression models to examine county-level variation in fertility rates in China. Spatial statistics reduce bias resulting from spatial and temporal autocorrelation in regression models, while spatial analysis allows detailed examination of relationships between pairs of predictors. Using county-level data from China's 1982, 1990, and 2000 censuses, I model general fertility rates and changes in rates using variables related to social economic development characteristics of the counties. While previous results from earlier studies showed a decline in the association of development factors with fertility, the 2000 census data shows that development factors have re-appeared as important predictors of fertility rates.

Introduction

Although the role of community economic, educational, and social transformation in fertility reduction has been well-documented worldwide, and China's fertility policy has been demonstrated to vary in content and implementation at the county level. Still, most studies of China's fertility decline focus on "micro-demographic" household-level characteristics with perhaps a dummy variable for rural or urban residence. One reason is likely the limitations in available datasets, which seek to represent the country without having large enough datasets for regional analyses. Another reason is that the few studies which did look at regional variation in fertility had found almost no significant relationships between development characteristics and fertility in China in the 1980's and 1990's (Poston 2000). I argue that these factors deserve re-examination. Until recently spatial statistical methods to remove spatial autocorrelation bias from OLS coefficients were not widely used in sociology, (or county boundary mapfiles needed to implement them were unavailable,) which may be partially responsible for the non-significant results. The spatially-coded version of the Chinese census, however, is suitable to test macro-level regional hypotheses about development and preliminary results are promising. Specifically, I ask whether factors traditionally associated with fertility reduction in other countries, such as rising education and education costs, urbanization, industrialization, and the reduction in infant mortality, have also been associated with local birth rates in China, even after considering the extremely restrictive fertility policy?

Background and Significance

Compared to other sociological topics, fertility in China is comparatively well-understood due to extensive research, but there are still considerable gaps. First, researchers have sometimes failed to take into account the complexity of the policy, which is implemented in different ways in each locale (Short and Zhai 1998). In large cities, the policy is to allow couples to have only one child, with enforcement through fines, work unit sanctions against parents, reduced services to additional children, and in some cases more severe sanctions such as forced sterilizations or abortions. In rural areas, some of the same sanctions apply, but various exceptions are made. A widespread practice that takes into account son preference is to allow rural families with an only daughter to have a second child, male or female. Other exemptions exist for recognized minority groups, who may be allowed to have multiple children, and for residents of various locales. Not only the policy, but also the sanctions for violating the policy, may vary over place and time, highlighting the importance of spatial panel methods for modeling this dynamic process.

Second, an extensive literature from other countries explores the relationship between fertility and "development" variables such as education, urbanization, infant mortality levels, etc., but research on China has focused on household decisions when constrained by the one-child policy. Some researchers had found that social and economic development factors, which had been highly predictive of fertility in the early 1970's, weakened as the one-child policy took hold in the late 1970's (Peng 1989, Poston 2000). Possibly the only study using the new spatial analysis technology is Skinner et. al. (2000), which looks at spatial variation in fertility based on distance from urban centers in the Lower Yangzi, but to my knowledge there is no spatial econometric model of fertility rates in China.

Current Investigation

My preliminary results are inconsistent with previous findings that development factors can do little to explain regional variation in fertility rates. Two explanations occur to me. First, regression coefficients may have been downwardly biased or found insignificant because spatial autocorrelation in the data was not controlled for. This possibility could be examined by repeating the OLS models along with tests for spatial error or lag. Indeed, my preliminary spatial and multilevel models and have substantially better AIC values than my OLS model and the tests for spatial lag and error were positive. Still, modeling is probably only responsible for part of the difference, and the original models might well prove to be similar to the spatial versions if someone reran them.

More likely, real changes occurred between 1990 and 2000 in the relationship between development and fertility reduction in China. Enforcement of the one-child policy relaxed after 1990, even as variation in development continued to grow. Overall, then, given the release of a new wave of (2000) census data and the development of more sophisticated statistical and visualization tools, the macro-level hypotheses deserve re-examination.

This paper applies exploratory spatial data analysis (ESDA) and spatial panel regression models to examine inter-county variation in fertility rates in China. Spatial models are important for several reasons. First, when spatial effects are present in a regression model, excluding explicit specification of spatial effects can lead to inaccurate inferences about predictor variables. Second, fertility policy in China is implemented at a local level, so substantial differences are expected between counties. Third, spatial models permit visualization and specific comparison of regional differences. Panel models also allow use of data from different censuses without inducing multicollinearity or temporal autocorrelation (Anselin, Gallo, and Jayet forthcoming). Using county-level spatial panel data from China's 1982, 1990, and 2000 censuses, I model general fertility rates and changes in rates using variables related to economic and demographic characteristics of the counties. Formal tests for spatial autocorrelation reveal significant spatial autocorrelation among county fertility rates, while maps display the substantial regional variation in an easily understandable format.

Data

The Historical China County Population Census Data with GIS Maps was released February 10, 2005 and includes county-level aggregated data from the 1953, 1964, 1982, 1990, and 2000 censuses. The datafiles are linked to digital maps managed with ArcGIS software. The GIS files were copyrighted by All China Marketing Research Co., Ltd., collected by the National Bureau of Statistics of China, and distributed by the University of Michigan China Data Center. Total population as shown in this publication is 1,242.61 million, or 23.22 million less than the population figure of 1,265.83 million (including 2.5 million servicemen) released earlier by the National Bureau of Statistics ,which was based on the advance tabulation. The post-enumeration sample survey indicates an undercount of 1.81% in the census enumeration. Only the 1982-2000 censuses contain fertility data, with the 2000 census the most detailed. For the 2000 data, some variables represent the full sample while some are from the 9.5% long form. Some scholars have questioned the quality of fertility data from China (Zhang and Zhao 2006, Retherford et al. 2005), as I will further discuss in my paper, but census data is often used for fertility research.

Ideally, this paper would model log total fertility rates, but some results reported here model log birth rates while others report general fertility rates, because the analysis is preliminary. The crude birth rate of a population is the number of childbirths per 1000 persons per year. The general fertility rate (GFR) is the number of live births per 1,000 females of childbearing age between the ages of 15-44 years. The total fertility rate (TFR) is the sum of the age-specific birth rates of women ages 15-44 in five-year age groups multiplied by five.

Although I have hopes of finding the total fertility rates, I may only be able to use the general fertility rates. For each census, the birth rate reported is for the year preceding the census. Thus, the "1982 birth rate" I report actually includes births from the last half of 1981 and the first half of 1982.

Proposed Methodology

Spatial Analysis

Using county-level data, I map independent and dependent variables including county-level birth rates, per capita GDP, female labor force participation, education, etc. These maps are useful in at least two ways. First, maps of variables can be examined to find locations where variable values change suddenly, such as city boundaries and mountainous areas. If having urban household registration has an effect independent of the effect of access to city resources, then the change in fertility at the boundaries of a county-level city will show up. If provincial policies are important rather than just local policies, then some provincial boundaries may be discernible from the maps. Although these results may not be conclusive, they can suggest important areas for investigation. If fertility is higher in mountainous regions, then hypotheses can be suggested to explain this, such as low educational expectation and thus lower cost of children, or less stringent policy enforcement in more remote areas. This is the simplest form of spatial analysis.

Second, maps of variables can be systematically analyzed to reveal patterns. Maps 1-3 show the 1982, 1990, and 2000 birth rates, using the same scale for each map. Clearly, fertility reduction has been dramatic nationwide, but birth rates in the west continue to be higher than elsewhere in the country. Comparison of birth rate maps and a map of infant mortality in 1982 (Map 4) suggests that the two may be linked as indicated by theory: families in higher mortality areas had more children in order to ensure their desired family size (see Riley 2004). In particular, I find local indicators of spatial association (LISA) for pairs of variables, looking for locations with usual and unusual relationships between variables (Anselin 1995). Figure 1 is a LISA plot of percent minority versus birth rate in which high minority areas with high birthrates are red and those with low birthrates are pink, while low minority areas with high birthrates are dark blue and those with low birthrates are light blue, although some areas are not colored because the association between the variables is not clear. Parts of the western region are high-high, while the central region is mostly low-high, and values of the variables occur in clusters. This plot can raise questions such as why some high minority areas are low fertility, or why values occur in clusters. If this plot were of education and fertility, I would ask why some places with high education levels have high fertility and why some places with low education levels have low fertility. Outliers can be identifed on a Moran scatterplot (Anselin 1996).

Spatial Regression

Using county-level data from the GIS edition of China's 2000 census, I model 2000 birth rates using variables related to economic and demographic characteristics of the counties. The weights matrix is of first order contiguity, created in Geoda, while the calculations are performed in R. To show that a spatial regression is justified, I first check that fertility levels in counties in China are strongly related to those in neighboring counties. Nearby places seem likely to have similar policies, but this has not been shown (Short and Zhai 1998). The hypothesis is tested by a significance test of the Moran's I value for spatial autocorrelation of total fertility rates (Moran 1950). Once a significant high value is found, a spatial regresssion is justified because spatial autocorrelation can affect regression results (Anselin 1996, Voss et al. 2006). Using R (Anselin 2005), I created a Moran scatterplot of log GFR vs. spatially lagged log GFR (Figure 2) showing that high values are related to high values and low to low values. The very high and statistically significant Moran's I value of .836 indicates that fertility in China is very strongly associated with region, as expected.

I will then use data from 1982, 1990, and 2000 in spatial panel econometric models to predict log GFR, using the Spatial Econometrics Toolbox written for the Matlab programming language by James LeSage and the spatial panel modeling routines created by J. Paul Elhorst. Different alternatives for the weights matrix will be tested. The use of a spatial panel model improves on the models shown here by allowing some insight into the direction and momentum of local trends in the birth rate. Perhaps unlike in other developing countries, the local birth rates may change direction over time in cases where policy enforcement was especially rigorous around 1990 and relaxed thereafter. In addition, trends in birth rates may be expected to be related to trends in economic and educational or infrastructure development. Because educational attainments in a county in 1982, 1990, and 2000 are likely to be correlated, the spatial panel model allows full use of the data without the multicollinearity that would result from the inclusion of time-lagged variables. In particular, I will construct a hierarchical spatial model, nesting counties within provinces, because national policy is implemented by provinces, provincial policy is implemented by counties, etc. That is, if two counties touch but are in different provinces, their fertilities might be less similar than two adjacent counties in the same province. Inclusion of provincial-level variables can help capture this.

Variables may include the percent illiterate, percent with college education, percent employed in industry, infant mortality, death rate, per capita gross domestic product, percent minority, population density, etc. Percent minority is included because minorities often have higher birth allowances, and population density is held to be a significant predictor of birth rates

(Loftin and Ward 1983). The model will be run three times: once with OLS and tests for spatial lag and spatial error, once in a spatial lag model, and once in a spatial error model. Substantively, the spatial lag model makes sense because I consider space to be important rather than a nuisance. The spatial lag and error tests should be positive and the best model can be chosen by the AIC. The model is correctly specified with respect to space when the Moran's I of the residuals is close to zero. The hypotheses will be supported if the coefficients are significant in the regression.

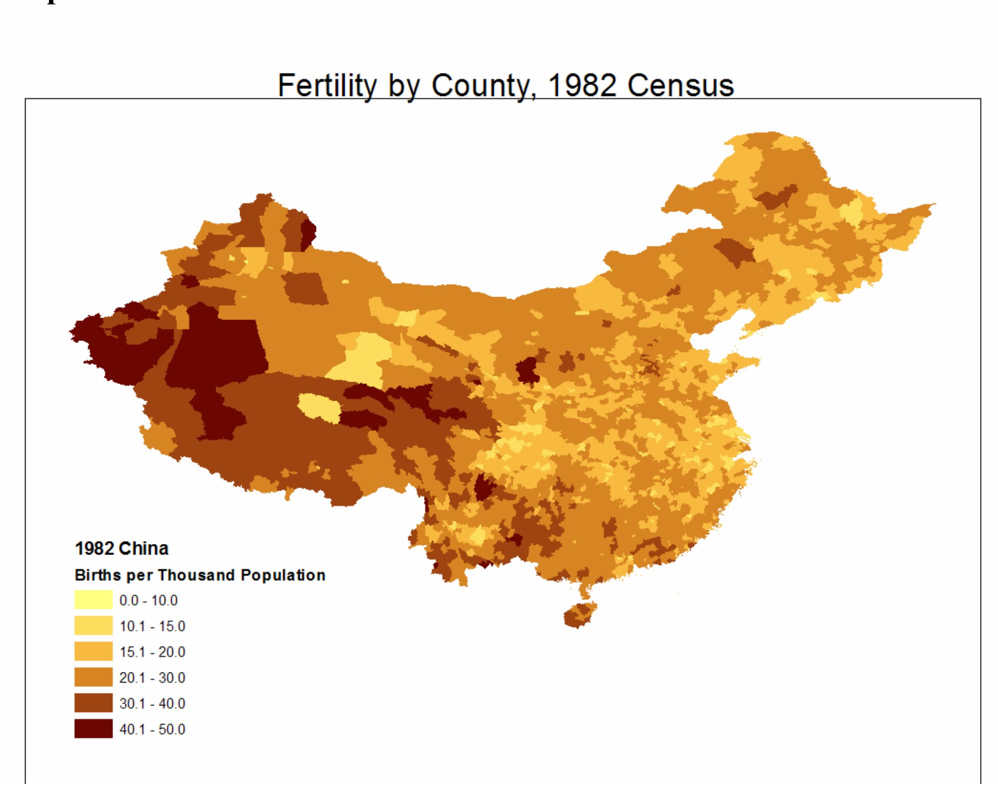
I present preliminary results of OLS and spatial regression models in Table 1. These models are not precisely panel models but include variables from different years. The inclusion of similar variables from different years is likely to introduce multicollinearity into the model and reduce the significance of results, but several variables are still significant. The panel models will remove this defect. Future models will also include additional variables, including those related to education, the labor force, health, and other features of development. Preliminary results show that as expected, the variables are spatially autocorrelated and exhibit both spatial lag and spatial error. Local population age structure, ethnic composition, education, and employment are significant after considering geography. Results are sensitive to variable and model selection, so considerable care must be taken in model creation. Nevertheless, attributes of place retain a causal influence even after controlling for other local attributes and previous birth rates, indicating that consideration of space is important in understanding fertility in China.

Summary

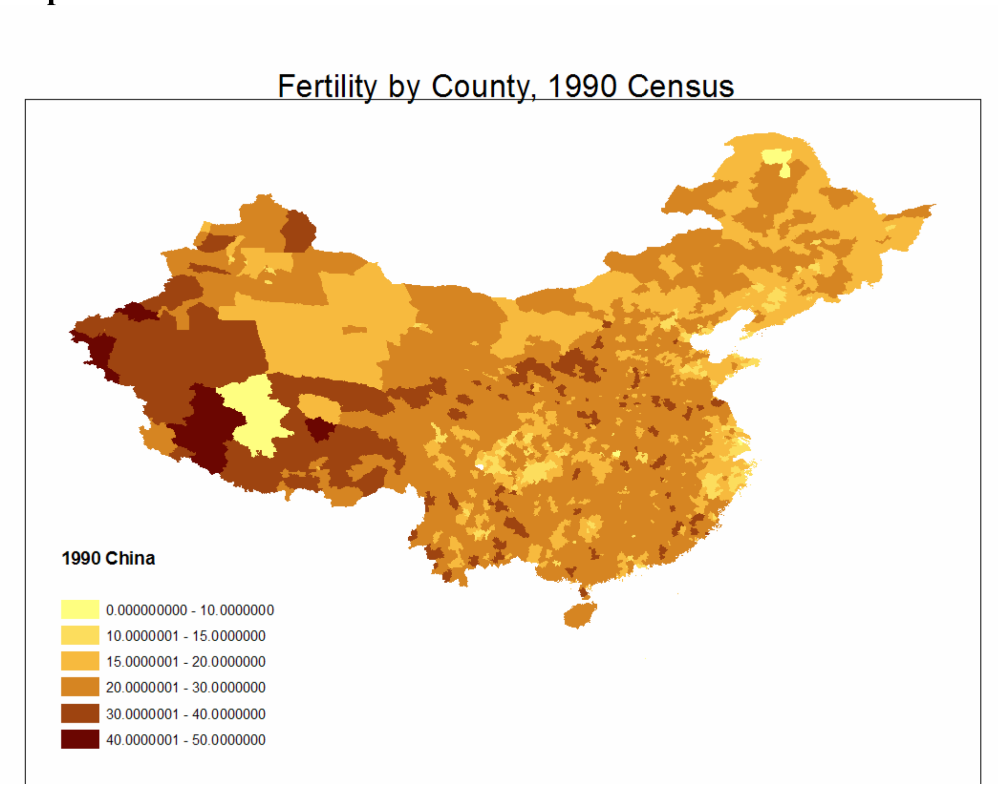
Fertility rates in China are significantly related to place in China and change rapidly over time, so models of fertility in China should account for space and time. Although not suitable for testing micro-level hypotheses about fertility, the spatial version of the census data is the largest and most representative survey available for testing macro-level hypotheses about fertility and development, most of which have never been tested before. Preliminary findings are quite promising, although future modeling should substantially improve the results. This space-time perspective can benefit researchers in many areas, not just fertility, although it is particularly suited to large and regionally diverse, rapidly changing nations.

Figures

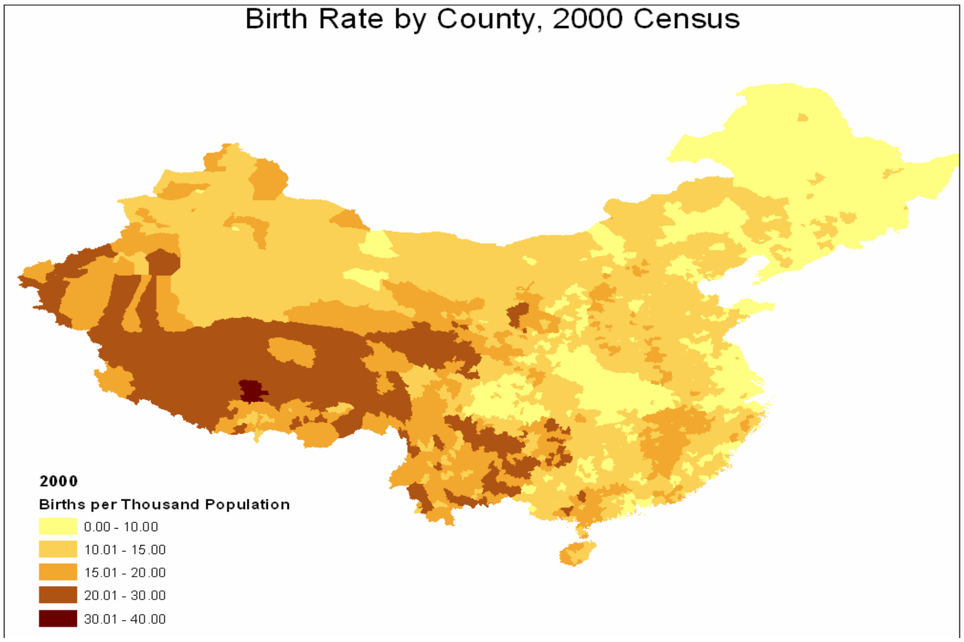
Map 1: 1982 Birthrates



Map 2: 1990 Birthrates



Map 3: 2000 Birthrates



Map 4: Infant Mortality

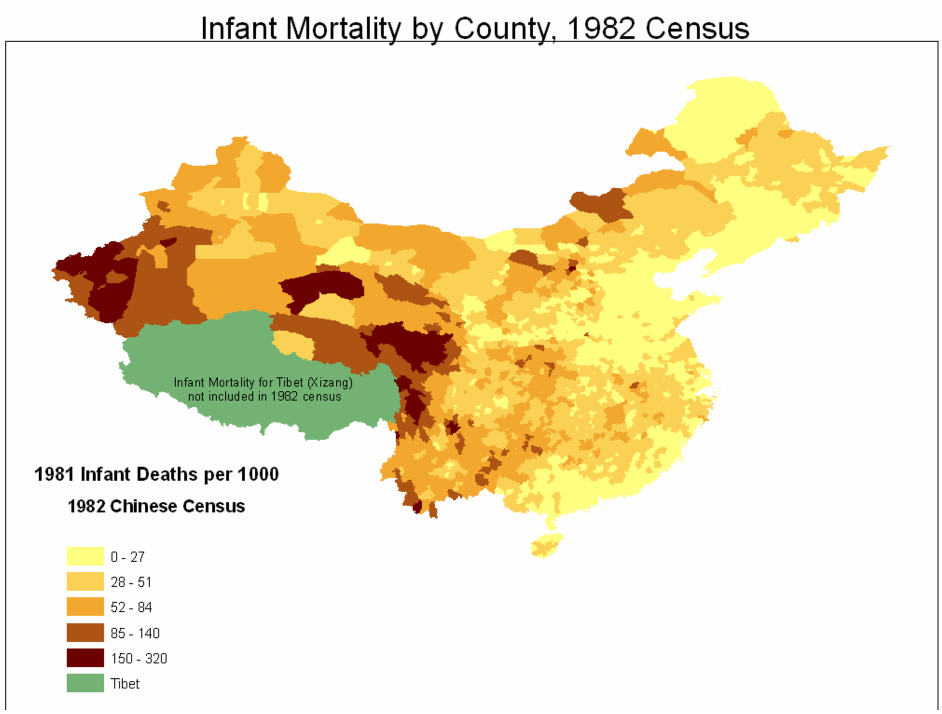
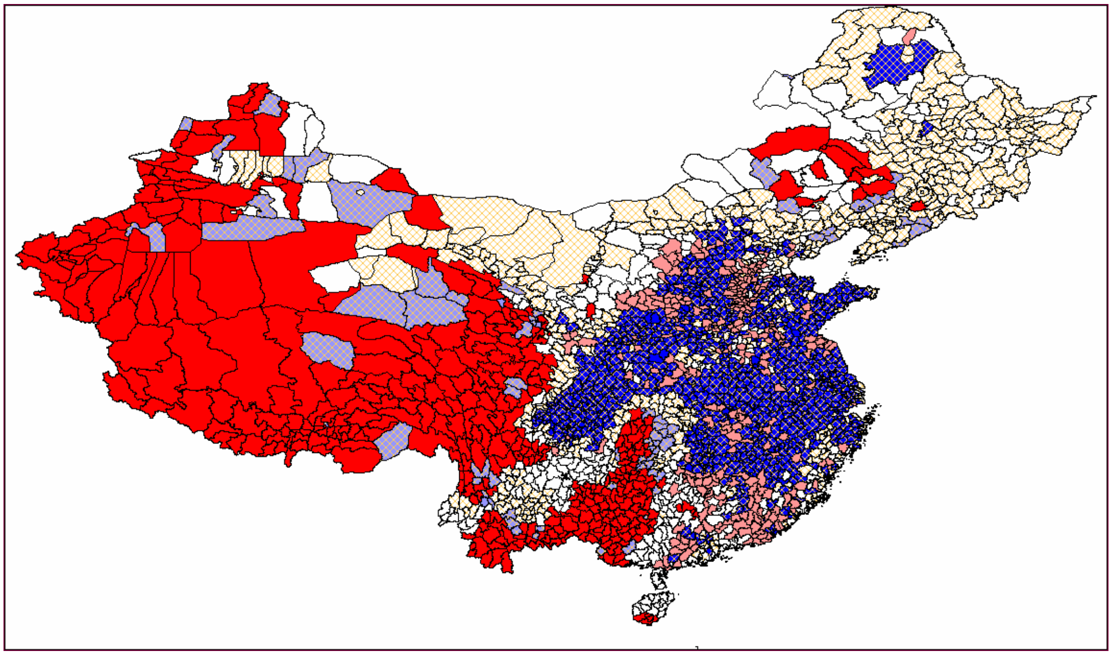


Figure 1: LISA plot of percent minority and total fertility rate



	% Minority	Birth Rate
Red	High	High
Pink	High	Low
Blue	Low	High
Navy	Low	Low

Figure 2: Moran scatterplot of log GFR and spatial lag of log GFR

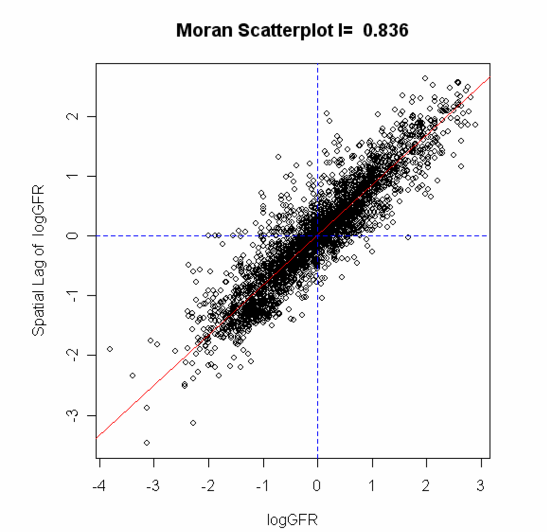


Table 1:

	Variables	OLS		OLS Log Birth Rate		OLS + Spatial Lag		OLS + Spatial Error	
1982	Pop. Density	-0.0002	*	-0.0003	***	0.0000		0.0000	
	% Age 15 to 49	-0.6613	***	-0.1127	***	-0.1637	***	-0.1462	***
	Birth Rate	0.1132	***	0.0034		0.0416	***	0.0454	***
	Death Rate	0.0658	***	-0.1090	***	0.0348	*	0.0084	
	Infant Mortality	0.0020		0.0117	***	-0.0007		0.0003	
	College/10,000	0.0104	***	0.0086	***	-0.0026	*	0.0036	**
	Middle School/10,000	0.0001		-0.0002		0.0000		0.0000	
	% Illiterate	0.0379	***	0.0181	**	0.0098		0.0443	***
	% Agricultural	-0.0097		0.0066		-0.0218	**	-0.0060	
	% Employed	0.0580	***	-0.0142		0.0008		0.0128	
	Industrial Employment	0.0326	*	0.0267	***	-0.0014		0.0160	*
	GDP per capita	-0.0001		-0.0001	*	0.0001		0.0000	
1990	Birth Rate	0.1008	***	-0.0162	***	0.0274	***	0.0236	***
2000	% Minority	0.0257	***	0.0175		0.0054	***	0.0257	***
	% Non-agricultural	-0.0348	***	0.0004		-0.0264	***	-0.0313	***
	Ave. Years Schooling	0.0369	***	0.0005	*	0.0111	*	0.0364	***
Spatial parameter						0.7656	***	0.8852	***
Constant		18.0691	***	-1.3816		6.6228	***	11.5661	***
R^2		0.6083		0.2352		0.8638		0.8818	
df		3346				3345		3346	
Robust LM (lag)		223.8491	***						
Robust LM (error)		795.9973	***						
Heteroskedasticity (B-P)		598.9019	***	189.4929	***				
Likelihood		-7966.7400		-7072.0600		-6416.4400		-6304.0007	
AIC		15967.5000		14178.1000		12868.9000		12642.0000	
Moran's I (residuals)		0.6476		0.6951		0.0270		-0.0686	

REFERENCES

Anselin, L. 1995. "Local indicators of spatial autocorrelation." *Geographical Analysis* 27: 93-115.

Anselin, Luc. 1996. "The Moran scatterplot as an ESDA tool to assess local instability in spatial association." In M. Fischer, H. Scholten, and D. Unwin (eds.), *Spatial Analytical Perspectives on GIS*. London: Taylor and Francis, pp. 111–125.

Anselin, L. 2005. *Spatial Regression Analysis in R: A Workbook..* http://www.sal.uiuc.edu/stuff/stuff-sum/pdf/rex1.pdf

Anselin, Luc, J. Le Gallo, and H. Jayet. Forthcoming. "Spatial Panel Econometrics." In L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice* (3rd edition). Dordrecht, Kluwer (forthcoming).

Loftin, Colin and Sally K. Ward. 1983. "A Spatial Autocorrelation Model of the Effects of Population. Density on Fertility." *American Sociological Review* 48: 121-128

Moran, P. 1950. "Notes on continuous stochastic phenomena." *Biometrika* 37:17-23.

Peng, Xizhe. 1989. "Major determinants of China's fertility decline." *China Quarterly* 117: 1-38.

Poston, Dudley L. Jr. 2000. "Social and economic development and the fertility transitions in mainland China and Taiwan." *Population and Development Review* 26: 40-60.

Retherford, R. D., Minja Kim Choe, Jiajian Chen, Li Xiru, and Cui Hongyan. 2005. "Fertility in China: How Much Has It Really Declined?" *Population and Development Review* 31:57-84.

Riley, Nancy E. 2004. "China's population: new trends and challenges." *Population Bulletin* 59: 3-36.

Short, Susan E. and Zhai Fengying. 1998. "Looking locally at China's one-child policy." *Studies in Family Planning* 29: 373-87.

Skinner, G. William, Mark Henderson, and Yuan Jianhua. 2000. "China's Fertility Transition through Regional Space: Using GIS and Census Data for a Spatial Analysis of Historical Demography. *Social Science History* 24,3: 613-652.

Xie, Yu and Emily Hannum. 1996. "Regional Variation in Earnings Inequality in Reform-Era Urban China." *American Journal of Sociology* 101:950-992.

Zhang, Guangyu and Zhongwei Zhao. 2006. "Reexamining China's fertility puzzle: data collection and quality over the last two decades." *Population and Development Review* 32: 293-321.