Spatial Variation of Juvenile Sex Ratios In the 2000 Census of China

William Lavely
University of Washington
lavely@u.washington.edu

CAI Yong
University of Washington
caiyong@u.washington.edu

For presentation at the annual meeting of the Population of Association of America, Boston, April 1-3, 2004. An earlier version was presented at the Workshop on Population Changes in China at the Beginning of the 21st Century, The Australian National University, Canberra, Australia 10-12 December 2003. Address correspondence to: William Lavely, Department of Sociology, University of Washington, Seattle, WA 98195. E-mail: lavely@u.washington.edu.

1 Introduction

Juvenile sex ratios in China underwent an extraordinary rise in the last two decades of the twentieth century, but just as dramatic was the emergence of immense regional differences. The ratio of boys to 100 girls age 0-4 rose from 107.1 in 1982 to 120.2 in 2000, but the inter-provincial variance grew from 3.5 to 83.5 in the same period. By 2000, provincial ratios ranged from a low of 101.4 in Tibet Autonomous Region to a high of 136.0 in Hainan Province (Table 1). Among the several studies of rising sex ratios and "missing girls" in China (e.g., Hull 1990; Johannson and Nygren 1991; Zeng et al. 1993; Coale and Banister 1994; Cai and Lavely 2003), none address the problem of regional variation. Yet, spatially-disaggregated data are essential for understanding the causes of high sex ratios and for projecting their consequences. This paper describes regional variation in juvenile sex ratios, which we define as the ratio of males to 100 females at ages 0-4, and offers a preliminary interpretation.

For 2000 census data we rely on a county-level data set and GIS base map provided by the University of Michigan China Data Center (ACMR 2003a and 2003b).[1] The data set contains 2,367 county-level units (counties, cities, and merged urban districts) with non-missing data. The frequency histogram of counties by sex ratio of population age 0-4, shows a wide variation and a distribution skewed to high ratios (Figure 1). The juvenile sex ratio ranges from a low of 89.7 in Tibet's Zhada Xian [county] to a high of 197.3 in Hubei's Wuxue Shi [city]. The mean and median of counties are both below the national mean, indicating that high sex ratios tend to be concentrated in counties with larger populations.

Random variation could account for only a minor fraction of the observed inter-county variance. The average county contains approximately 30,000 persons age 0-4, implying a fairly restricted range of random variability. Assuming 30,000 children age 0-4 and a hypothetical "true" sex ratio of 105 males per 100 females, the 95 percent confidence interval would range between 102.65 and 107.40. The mainly Han counties of China Proper tend to have larger populations, so we would expect the greatest variability among the sparsely populated, mainly non-Han counties of the west.

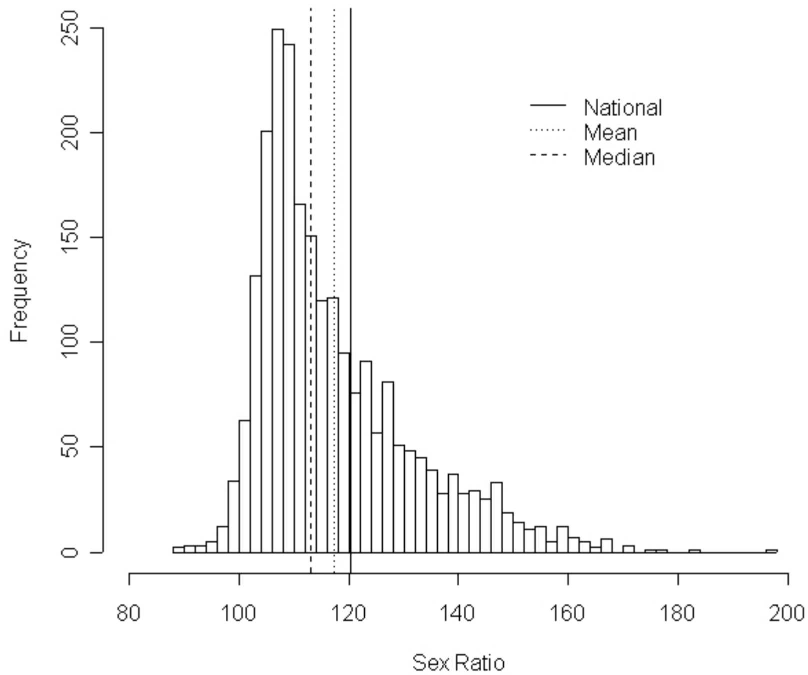
Map 1 displays the sex ratio of children age 0-4 in 2000 in choropleth form, with darker reds indicating higher sex ratios. County and provincial boundaries are shown for reference. The map makes manifest some basic facts. First, as observed in the province data, juvenile sex ratios are not homogeneous in space. There appear to be distinct regional clusters of high and low sex ratios. Second, the clustering does not correspond to provincial boundaries. Clusters cross province boundaries and high and low clusters sometimes occur within a single

[1] Yuanjiang Shi (GB 430981) was split into two polygons in the original map because Dongting Lake divides the city in two. For technical reasons related to the calculation of the Local *Moran I* indicator, these polygons are joined and treated as a single unit in our analysis. No data were provided for Mangya Dachaidan (GB 632824), a recently created unit in Qinghai Province, although it is shown on the map. No data were provided for Taiwan or Hong Kong.

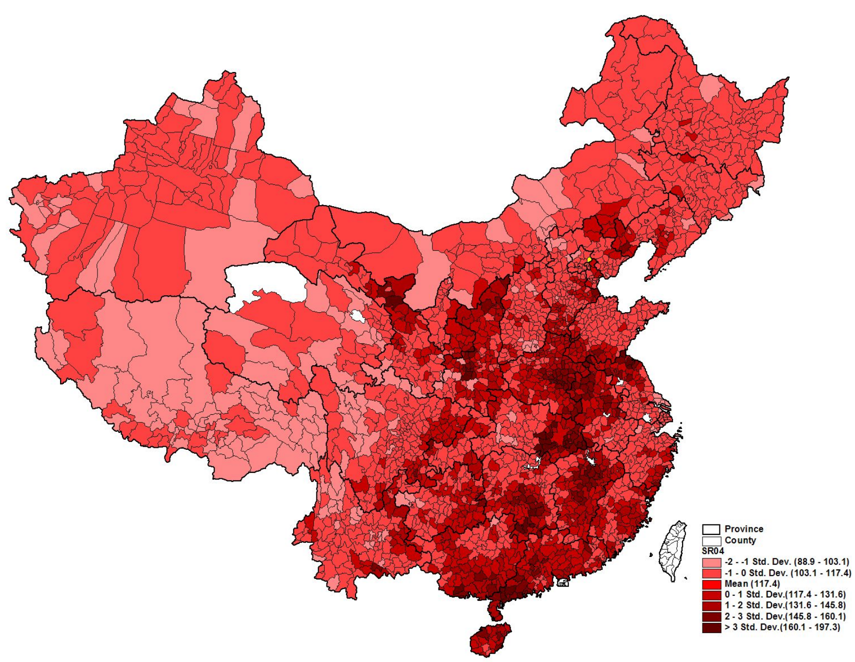
Table 1: Sex Ratio at Age 0-4 by Province: China 1982 and 2000

	Census 1982		Census 2000	
Province	Population 0-4	Sex Ratio 0-4	Population 0-4	Sex Ratio 0-4
Beijing	667,469	107.3	442,578	110.8
Tianjin	607,798	106.5	376,002	112.6
Hebei	5,271,448	107.1	3,611,922	115.8
Shanxi	2,353,896	108.7	2,157,761	110.9
Neimeng	1,991,353	105.8	1,182,190	109.6
Liaoning	3,183,992	106.3	1,757,164	113.1
Jilin	2,153,561	105.8	1,061,361	110.8
Heilongjiang	3,272,890	104.9	1,530,908	108.8
Shanghai	780,627	105.9	489,978	110.3
Jiangsu	4,844,751	107.3	2,986,808	122.7
Zhejiang	3,204,255	108.4	2,259,449	113.8
Anhui	4,749,163	109.9	3,278,183	129.8
Fujian	2,756,149	106.2	1,623,227	123.8
Jiangxi	3,671,568	106.7	2,747,788	132.6
Shandong	6,383,914	108	4,579,809	114.4
Henan	7,357,828	108.3	5,276,801	132.4
Hubei	4,451,003	106.1	2,444,737	129
Hunan	4,905,074	106.4	3,199,237	124.5
Guangdong	6,445,286	109.2	5,532,309	129.6
Guangxi	4,404,858	108.8	2,797,864	127.9
Hainan			527,309	136
Chongqing			1,774,488	116.7
Sichuan	6,975,540	106.8	4,871,752	115.3
Guizhou	3,331,338	105.9	3,227,009	114
Yunnan	3,804,230	104.2	3,387,080	112.9
Tibet	232,753	101.9	248,463	101.4
Shaanxi	2,599,551	108.4	1,821,527	126
Gansu	1,794,205	105.5	1,617,252	119.3
Qinghai	463,356	103.7	371,531	108.6
Ningxia	510,286	104.2	452,495	109
Xinjiang	1,536,219	103.7	1,343,392	105.6
National Total	94,704,361	107.1	68,978,374	120.2
Mean of provinces		106.5		117.7
Inter-provincial variance		3.5		83.5

Figure 1: Frequency Histogram of Counties by Sex Ratio at Age 0-4, China 2000 (N=2,367)



Map 1: Sex Ratio at Age 0-4, China 2000



province. Finally, sex ratios tend to be higher in the mainly Han areas of China Proper and low in the mainly non-Han areas of the north and west. Beyond these simple points, the picture is complex.

2 Local Indicators of Spatial Association (LISA)

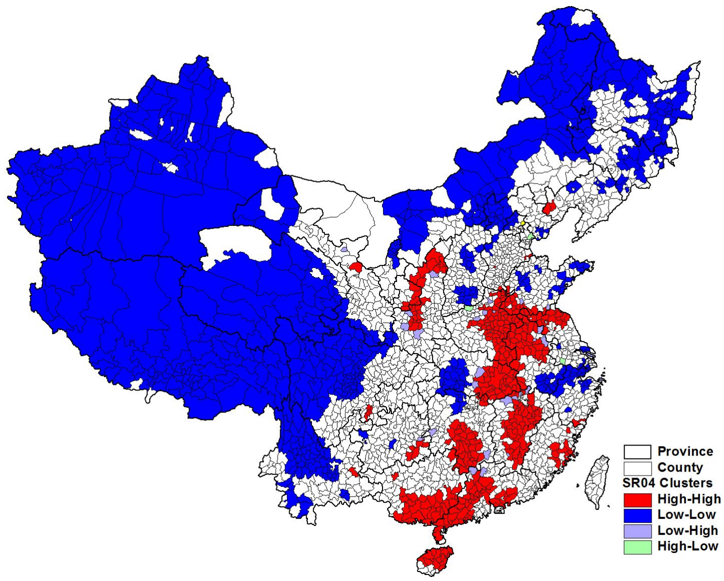
Because spatial clusters could occur by chance, spatial statisticians have developed tests to assess the likelihood that characteristics are distributed randomly in space or, alternatively, that observed clustering indicates a true underlying spatial structure. We utilize a measure developed by Luc Anselin (1995) known as Local *Moran* I_i, one of a set of exploratory spatial analysis statistics together referred to as Local Indicators of Spatial Association or LISA. We use Anselin's implementation of Local *Moran* I_i contained in his suite of spatial analysis programs known as GeoDa (Anselin 2004). As the name implies, Local *Moran* I_i is a measure of local spatial association, that is, it is a measure of correlation between a spatial unit and its neighboring units.[2] When two or more neighboring spatial units have significant Moran values and share a similar tendency, a cluster is defined. In Map 2, the raw data portrayed in Map 1 have been rendered in *Moran* I_i defined clusters.

Map 2 portrays counties in five categories. County units for which *Moran* I_i is not significant, indicating no association with neighboring counties, are shown in white. The other four color shadings indicate that *Moran* I_i is statistically significant at the .05 level,[3] but in a four-fold classification of counties by sex ratio tendency and the nature of association with neighbors. The high-high category refers to counties with high sex ratios that neighbor counties with high sex ratios. The low-low category refers to counties with low sex ratios that neighbor counties with low sex ratios. The red high-high areas thus represent clusters of high sex ratios and the blue low-low areas represent clusters of low sex ratios. It is important to note that the "high and low" are relative terms. The low values are low relative to the mean county in China, which has a sex ratio age 0-4 of over 117 (see Figure 1); thus even counties with above-normal sex ratios may qualify for membership in "low-low" clusters. The remaining categories represent anomalous counties that do not conform

[2]Local *Moran* I_i is a local version of the global spatial autocorrelation measure Moran's *I*. Given a set of observations, each with a predefined neighborhood structure, Local *Moran* I_i is the (weighted) correlation between each observation and its neighbors. For example, Wuhan has four neighbors. If we weight each neighbor equally, the Local *Moran* I_i is the correlation between Wuhan's deviation from the global mean and the deviation of its four neighbors from the global mean. If Wuhan is similar to its four neighbors, *Moran* I_i is positive at Wuhan. If Wuhan is dissimilar from its four neighbors, *Moran* I_i is negative at Wuhan. Combining this information and the observed values, it is possible to identify so-called hot spots. Counties in the high-high category are those with high observed values and a significant positive I_i, indicating affinity with its neighbors. High-low clusters are those with high value observed values and a significant negative I_i, indicating contrast with its neighbors.

[3]We use randomization test with 9999 permutations. Same applies to all other *Moran* I_i significant tests in this article.

Map 2: Local *Moran* I_i Defined Clusters: Sex Ratio at Age 0-4, China 2000

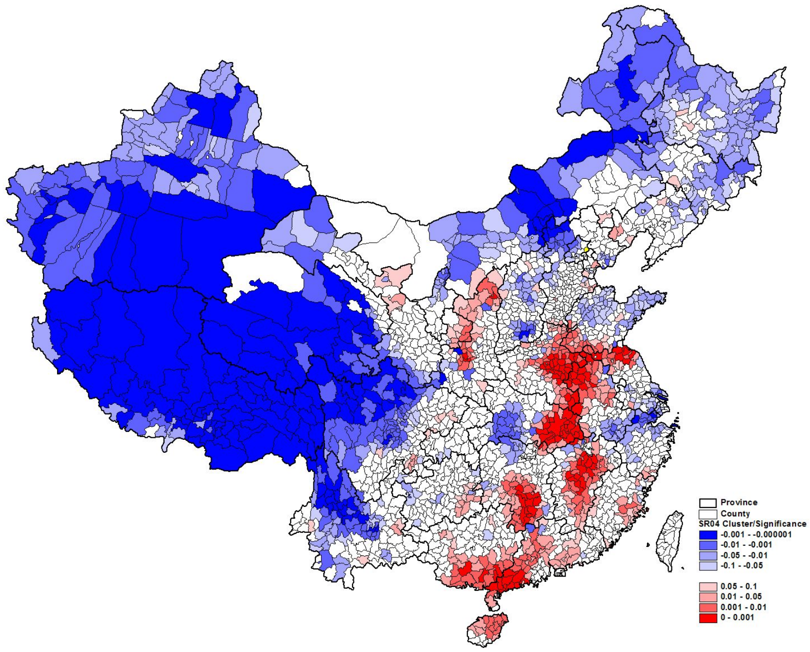


to the tendency of the neighboring counties. High-low counties (in pink) are those with high sex ratios neighboring counties with low sex ratios, and low-high counties (in light blue) are those with low sex ratios neighboring counties with high sex ratios.

The advantage of using LISA-defined clusters is the knowledge that the association among units is unlikely to be due to chance. These statistical *bona fides* notwithstanding, it is important to understand that inclusion in a cluster is a function of the level of significance selected and the definition of a neighboring unit adopted. In the case of Map 2, Moran values are significant at the .05 level and neighboring units are defined as adjacent units, including (in chessboard terminology) both "rook neighbors"(sharing common boundaries) and "bishop neighbors"(sharing a common point), the combination of the two known as "queen neighbors". It is possible to define neighboring units less restrictively, for example, by second-order adjacency (adjacent to adjacent) or third-order adjacency (adjacent to adjacent to adjacent). Adjacency could also be defined by distance rather than contiguity.

The same point can be made with regard to the level of significance. To accept a higher probability that neighboring units are associated by chance increases the extent of the cluster. Map 3 classifies the sex ratio of county units simultaneously by the level of significance of the *Moran* I_i statistic and by the sex ratio tendency. There are four levels of significance for counties in low sex ratio clusters, shown in shades of blue, and four levels of significance for high sex ratio clusters, shown in shades of red. Counties shaded dark blue represent clusters of low sex ratios with Local *Moran* I_i statistics significant at the .001-level, and counties shaded dark red represent correspondingly significant clusters of high sex ratios. The fact that counties at lower levels of significance tend to be continguous to those with higher significance once again seems to confirm that we are taking the measure of a true underlying spatial structure.

Map 3: Significance of Local *Moran* I_i: Sex Ratio at Age 0-4, China 2000



3 Major Regional Features

This section briefly describes the major features observed in Map 3. For the purpose of describing regional space we adopt the naming conventions of G. William Skinner's decomposition of China into macroregional systems (Skinner 1977; Skinner et al. 2000, see Appendix). Beginning with the low sex ratio clusters shaded in blue, there is an obvious division between China Proper and the Inner Asian frontiers of the West, Northwest, and Manchuria. A band of low ratios describes much of China's non-Han frontier, extending from central Yunnan onto the Tibetan Plateau and through the west, but excluding all but the westernmost counties of Gansu. The band resumes in central Inner Mongolia and stretches into Manchuria, encircling a corridor that extends southward from Harbin. Low sex ratio clusters also occur in the mainly Han regions of China Proper, and we note four in particular: in the Chengdu Plain in the western Sichuan Basin of the Upper Yangzi; in central Hubei and extending into northern Hunan in the Middle Yangzi; in the northern Zhejiang-southern Jiangsu-southern Anhui border region in the Lower Yangzi; and on the southern end of the North China Plain and extending across the Shandong Penninsula.

There are six major red-shaded high sex ratio clusters, although it should be noted, four of these are weakly linked. We begin in the northwest and move clockwise.

1. A high ratio cluster follows a north-south corridor running through the north Shaanxi Plateau and crossing through the Wei River Valley in Northwest China.

2. The largest single high ratio feature extends from the North China Plain into the Middle Yangzi, including parts of southern Shandong, northern Jiangsu, northern Anhui, eastern Henan and extending into eastern Hubei centering on Wuhan.

3. Separated from the previous cluster by only a single tier of counties in the Boyang Lake area, another cluster extends southward into central Jiangxi, in what Skinner refers to as the Gan River subsystem of the Middle Yangzi.

4. An enclave of high ratios is contained in the central coast of Fujian, centered on Quanzhou in the Southeast Coast region.

5. A major southern cluster includes most of Hainan, except for its southern tier of counties, and extends into western Guangdong and southeastern Guangxi in the Lingnan Macroregion. It includes central Guangdong except for a coastal enclave south of Guangzhou.

6. Weakly connected to the Lingnan cluster is a northward extension into south-central Hunan in the Middle Yangzi, including the upper reaches of the Xiang River but not including the Xiang Valley north of Hengyang City.

4 Interpreting Sex Ratio Variation

We assume that child sex ratios reflect sex-biased behaviors guided by preferences and constraints. The reported sex ratio of children is determined by several factors, including random variation in the sex ratio at birth, the effect of sex-selective abortion on the sex ratio at birth, sex differences in infant and child death rates, and sex-differential census undercount, presumably the result of parental subterfuge. We assume that sex-specific migrations are unlikely to be a factor in this age range. With the exception of the random variation in the natural sex ratio of births, which varies within a narrow range, all of these factors reflect the sex-biased behaviors of parents.

Sex-biased preferences for children are influenced by cultural and socioeconomic contexts, the most encompassing of which is the family system, but which can include aspects of the local economy, social position, and the micro-demography of the family (Skinner 1997; Lavely et al. 2001). Sex preference reflects the value of children to their parents, which varies by context. Preferences are constrained by the costs of achieving a desired family configuration. These include the costs of childbearing (including costs imposed by fertility control policies), and the costs of sex selection, which, for example, could be increased by social or legal norms against infanticide, or reduced by the availability of fetal sex-determination technologies. Both preferences and constraints can be of long standing, and both can change, but in the contemporary Chinese context, it is evident that constraints, in the form of fertility limitation and sex-selective technologies, have been changing more rapidly than preferences.

The contingencies are complex, but it is possible to set out some broad propositions. For example, we would expect the patrilineal joint family system of the Han Chinese to be more conducive to son preference than the family systems of western and southern non-Han minorities (Skinner 1997). And we would expect tighter birth planning regulations, because they raise the cost of an additional child, to be conducive to sex-selective behavior. In the case of non-Han minorities, then, there are two powerful contingencies conducive to low sex ratios – weak son preference and more relaxed constraints on fertility.

In Han counties of China Proper we would expect son preference to be weaker where economies are more developed, such as more urbanized areas. Within rural areas we would expect stricter birth planning policies to be associated with sex preferential behaviors. And we anticipate that local variations in family system norms, which define women's status and the value of children by sex, to influence sex selective behaviors. Skinner and colleagues (1994, 2000) have demonstrated that socioeconomic levels and fertility change are spatially patterned by the urban hierarchy within macro-regional systems, an underlying spatial structure they refer to as Hierarchical Regional Space or HRS. They argue plausibly that parental sex preferences, birth planning administration, and the availability of ultrasound also co-vary with HRS, although a lack of systematic local data on sex preference, birth planning administration, or the availability of ultrasound make these propositions difficult to test directly. Nonetheless, a comprehensive spatial model of sex ratio variation would take account of this underlying spatial structure and would reflect some of the complex interactions between the urban hierarchy, ethnic status, and birth planning regimes.

5 Modeling Spatial Variation

Such a comprehensive modelling project being beyond our present means, we have adopted a simple approach, using additive OLS regression models to test the influence of available measures that are plausibly related to sex ratio variation. We narrowed the field to nine variables that are related to the sex ratio in theoretically-grounded and interpretable ways. These variables and listed and described in Table 2.

We conceive of these variables as falling into three sets, representing respectively the culture of sex preference, the birth planning context, and the only measurable instrumentality of sex selection, sex-differential mortality. In practice, these three categories are far more interrelated than this clearcut conceptualization implies. The first measure, designated **difschool**, is the absolute difference between the average years of male and female education of the population age 6 and above in the county. This variable, is assumed to reflect gender-biased investment in children, and because it includes cohorts from the youngest to

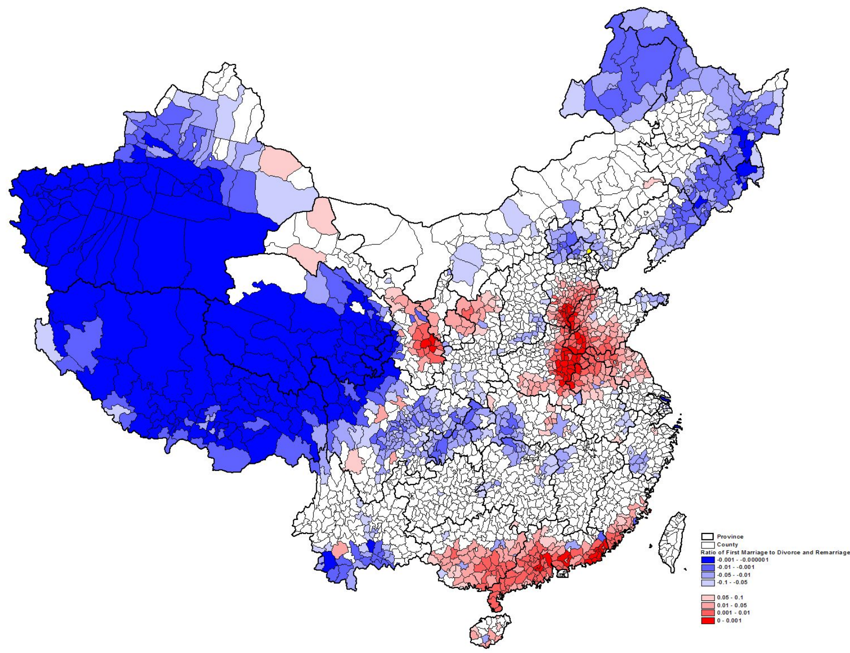
Table 2: Variable Definitions and Descriptive Statistics, County-level units: China 2000 (N=2,367)

Variable	Definition	Mean	S.D.	Min.	Max.
sr04	Males per 100 females at age 0-4	117.4	14.2	89.7	197.3
difschool	Male average school years minus female	1.1	0.5	0.0	3.1
drratiof	First marriage females per 100 divorced and re-married	27.6	14.2	1.0	172.0
tfr	Total Fertility Rate	1.4	0.5	0.5	5.5
pminority	Percent non-Han	18.6	31.1	0.0	99.8
pnonagri	Percent non-agricultural	20.6	17.5	2.1	97.0
${}_1q_0^m$	Male ${}_1q_0$*1000	25.5	22.5	0.0	238.9
${}_1q_0^f$	Female ${}_1q_0$*1000	32.3	31.6	0.0	330.8
${}_4q_1^m$	Male ${}_4q_1$*1000	6.9	7.2	0.0	76.6
${}_4q_1^f$	Female ${}_4q_1$*1000	6.9	7.8	0.0	86.5

the oldest, it can be assumed to reflect gender bias of long standing. The next measure, designated **drratiof**, is the ratio of women in their first marriage to divorced and re-married women in the county. This measure is designed to indicate family system norms. Divorce and re-marriage is far more common in non-Han areas than in Han dominated areas where patrilineal joint family system norms prevail. In Han areas, we would expect low rates of divorce and remarriage to reflect conservative, patriarchal norms deleterious to the status of women.

The first marriage to divorced and remarried ratio has a story to tell about the regional structure of family system norms. Map 4 shows the distribution of high and low values of the ratio by level of *Moran* I_i significance. Blue shaded areas indicate areas where low ratios cluster, indicative of more relaxed marriage norms; red shaded areas indicate clusters of high ratios, where marriage norms are presumably more rigidly patriarchal. There are some broad correspondences with the non-Han areas of the west, Tibet and Xinjiang, in particular, where low ratios predominate. There are also clusters of low ratios stretching across the Sichuan Basin and into the Middle Yangzi region, and in the Manchurian border regions. High ratios cluster in the southern tier of the North China Macroregion, in particular in Henan and the periphery dividing the North China, Lower Yangzi, and Middle Yangzi. The other major cluster of high ratios covers much of Lingnan and stretches up the Southeast Coast. An explanation of these regional clusters would be an ambitious project in itself. Suffice it to say that there is some considerable overlap between regional marriage patterns and juvenile sex ratios.

Map 4: Significance of Local *Moran* I_i: Ratio of First Marriage to Divorce and Remarriage, China 2000



The next set of variables is intended to represent variation in birth planning regimes. The Total Fertility Rate is taken as a measure of fertility constraint. The TFR is measured in the year prior to the census, while our sex ratio measure represents the five cohorts born prior to the census. While it is not a perfect temporal match, we assume the persistence of fertility levels over the five-year span. We observe a curvilinear relationship between TFR and the juvenile sex ratio. The TFR is positively associated with the sex ratio up to TFR=2.0, above which there is a negative association. We explain this by the gradient of fertility through three socioeconomic strata, urban China, the rural core, and the rural periphery, describing a progressively more relaxed fertility regime – approximately one child per woman in urban China, between one and two in the rural core, and more than two in the periphery. Sex bias is lowest in urban China; it is highest in the rural core; and it is high in the periphery but less subject to sex-selective behavior due to the relatively weak constraints on fertility.

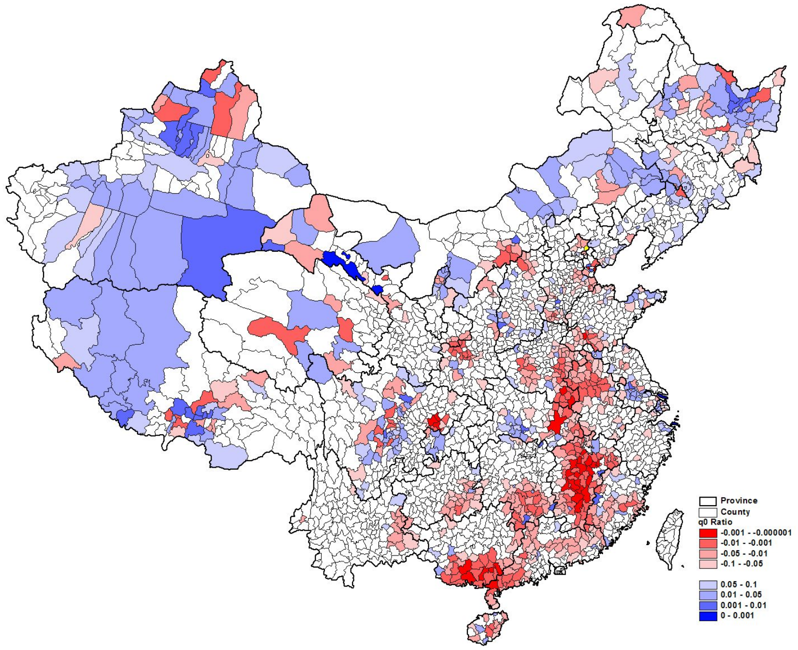
We use the percent non-Han, designated **pminority**, and the percent non-agricultural, designated **pnonagri**, as additional proxies for birth planning regimes. Although there are important exceptions, birth planning regulations tend to be more relaxed for minority nationalities, particularly those in the west and southwest. The rural-urban divide is the other major dimension of birth planning policy and and implementation. Although it is a blunt instrument compared to a decomposition into Hierarchical Regional Space, the percent of the population holding a non-agricultural registration status is a parsimonious indicator of the rural-urban continuum. While grouping these measures with TFR to represent birth planning practice, we readily concede that the percent non-Han and the proportion non-agricultural could also reflect cultural and socioeconomic dimensions related to gender bias.

The final set of variables are sex-specific measures of infant and child mortality, representing the instrumentality of sex selection. Sex-selection in China is implemented either pre-natally, through sex-selective abortion, or post-natally, through infanticide or more subtle sex-biased behaviors. An influential stream of scholarship maintains that sex-selective abortion is the major instrument of sex selection in contemporary China, and maintains that infanticide is the rare exception (e.g., Zeng et al. 1993; Gu and Roy 1995; Chu 2001). It must be said that the evidence for either hypothesis is thin. Although scattered data exist on the prevalence of ultrasound-B technology and the incidence of sex-selective abortion, no systematic data on either are available. As for sex differential mortality, no spatially disaggregated sex-specific mortality data from the 1982 or 1990 censuses were ever officially released, making it difficult to test the proposition that high sex ratios are associated with sex differential mortality. But with the release of county-level data from the 2000 Chinese census, the proposition can be tested directly.

Life table parameters by sex were calculated for all county-level units (Cai 2004). To capture infant and child mortality, we utilize ${}_1q_0$ and ${}_4q_1$, separately by sex, which we designate as ${}_1q_0^m$, ${}_1q_0^f$, ${}_4q_1^m$ and ${}_4q_1^f$. It should be noted that these measures only represent the mortality experience in the year prior to the census and thus are not synchronized with the five cohorts represented in the juvenile sex ratio. We perforce assume that sex differentials in mortality are relatively persistent over the five year period. Inclusion of the mortality rates of both sexes in the model permits us to observe the effect of the mortality of each sex while controlling for the mortality of the other. For the purposes of this exposition, we map the ratio of male, again rendered in categories of *Moran* I_i significance. The result is shown in Map 5.

Blue shaded areas in Map 5 represent regional clusters of high ratios of male to female infant mortality while red shaded clusters represented clusters of low ratios of male to female mortality. The relative survival advantage of infant females in the far west and parts of Manchuria are obvious features. Relatively low survival of females appears in clusters that begin in the southern periphery of North China, extend through the Middle Yangzi core and through the Gan Yangzi, extending into Lingnan. Map 5, it must be noted, is a remarkable document, possibly the first ever systematic mapping of sex differential mortality for Chinese society, and meriting an epidemiological investigation in its own right. For the present purpose, we merely note that it portrays spatial patterns of female survival disadvantage that bear a striking resemblance to the juvenile sex ratio patterns observed in Map 3.

Map 5: Significance of Local *Moran* I_i: Ratio of Male q_0 to Female q_0, China 2000



6 Results

Three regression models of the juvenile sex ratio are contained in Table 3. The first model, representing sex preference and family system effects, includes the difference in male and female years of school and the ratio of first marriage to divorce and re-marriage. Both variables are positively and significantly related to the sex ratio, the former demonstrating an anticipated linkage between differential investments in children's education and the sex ratio, and the latter demonstrating an association between conservative family system norms and the sex ratio. These two variables alone account for a quarter of the inter-county variation in the juvenile sex ratio as measured by R^2, with the marriage-divorce ratio by far the more powerful of the two.

The next model adds the three birth planning-related variables, the transformed TFR, the percent minority, and the percent non-agricultural. To capture this curvilinear relationship we utilize an arithmetic transformation of the TFR, **tfr2**, which reverses the progression of TFR values above 2.0.[4] The fertility rate is related to the sex ratio as hypothesized, as is the percent minority. The coefficient of the percent non-agricultural is not statistically significant, but is in the hypothesized direction. The addition of these variables raises the explained variance to 30 percent.

[4]We fit this as one linear term in the model, assuming that the positive and negative relationship of TFR and sex selective behavior is symmetric at TFR=2. A squared term of TFR also picks up the reversal of the relationship and is an improvement over a simple (untransformed) linear term, but fits less well than our transformation.

Table 3: OLS Regression Statistics, Sex Ratio Age 0-4, China 2000 (N=2,367)

Variable	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
difschool	3.272***	0.547	3.472***	0.568	1.955***	0.551
drratiof	.471***	0.018	.369***	0.019	.324***	0.018
tfr2†			7.348***	0.871	3.984***	0.816
pminority			-.109***	0.009	-.074***	0.009
pnonagri			-0.018	0.017	-.033*	0.015
${}_1q_0^m$					-.331***	0.018
${}_1q_0^f$					.271***	0.012
${}_4q_1^m$					-.124*	0.061
${}_4q_1^f$					.146**	0.055
constant	100.630***	0.779	95.975***	1.524	102.520***	1.424
R^2	0.246		0.306		0.430	

*** denotes p <.001; ** denotes p <.01; * denotes p <.05.

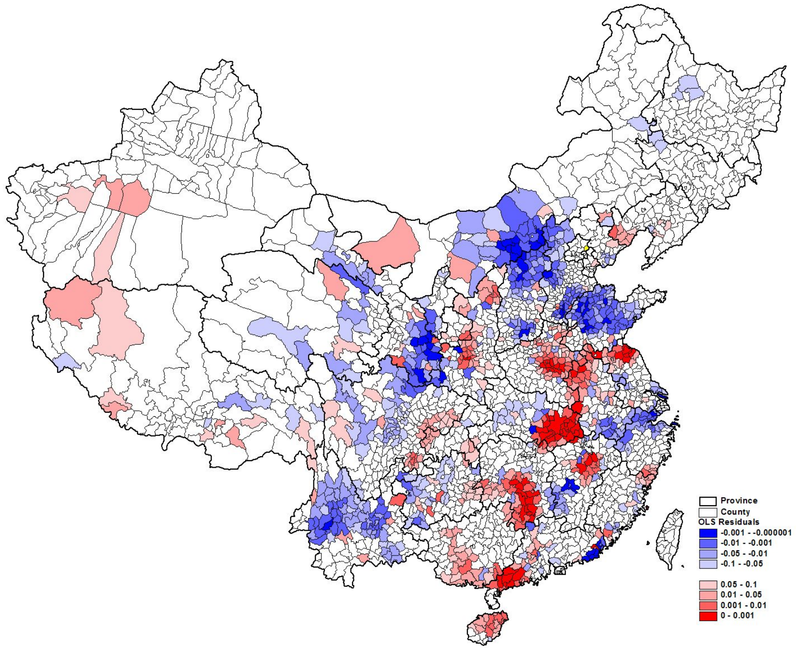
†tfr2 is a transformation of tfr. See text.

The third model includes all of the above, plus the four sex-specific infant and child mortality rates. As expected, higher male mortality is associated with lower sex ratios, and higher female mortality with higher sex ratios. The infant mortality measures ${}_1q_0^m$ and ${}_1q_0^f$ are far more influential than the child mortality measures, an expectable outcome given that child mortality is a relatively minor part of mortality under age 5. The mortality measures raise the explained variance to 43 percent. The two infant mortality measures standing alone explain nearly 26 percent of the inter-county variance.

The model residuals are portrayed in Map 6, again classified by *Moran* I_i significance. Comparison of the residual clusters with the corresponding sex ratio clusters observed in Map 3 indicates where the model succeeds and where it fails. Success in this case means that the model residuals do not cluster in space, even if much of the variance is left unexplained. It is evident that the model does quite well explaining the cluster of low sex ratios in the west and northeast. These clusters have essentially disappeared. The model does far less well in China Proper. Although the high sex ratio clusters are greatly reduced in extent, major remnants persist in the periphery between north China and the Middle Yangzi, in the Middle Yangzi core centered on Wuhan, the Gan Yangzi, the upper Xiang River valley, and western Lingnan. At the same time, low residuals cluster in the the North China region, particularly in Hebei and Shandong. Low residuals also cluster in the Lower Yangzi core, in Northwest China, particularly in southern Gansu, and in the Yungui Macroregion, in much of northern and western Yunnan.

When residuals are correlated, OLS produces inefficient estimators. For example, in the

Map 6: Significance of Local *Moran* I_i: OLS Regression Residuals, China 2000



presence of positive auto-correlation, the standard errors of the coefficients reported by OLS are underestimated, thus should be interpreted with caution. To address this problem, we estimated two spatial regression models: a spatial lag model, and a spatial error model [not shown]. The spatial error model successfully controls the correlation in the residual and yields the same pattern as the OLS model.

7 Discussion

Locally-disaggregated data offer insight and analytic possibilities not available from provincial averages. We now know, for example, that China has sizable regions where there are more than 145 boys for every 100 girls (see Map 1). If high sex ratios are going to have social consequences, these are places to watch. Although the Local Moran statistic is blind to known spatial structures, terrain, transport, language, or local history, it provides a useful distillation of a complex landscape. The fact that Moran-defined low sex ratio clusters are only low in a relative sense is not a drawback because in the context of contemporary rural China, even normal sex ratio clusters present an intellectual problem as confounding as the ultra high. We might understand high clusters better if we could understand the low.

The residual clusters confront us with a problem on which there is much lore but few systematic measures: the persistence of sub-ethnic cultural variation within Chinese society. The recency and rapidity of the rise of sex ratios makes it tempting to conclude that spatial clusters merely reflect recent and idiosyncratic differences in birth planning administration or the availability of ultrasound. But if the rise of sex ratios is a recent phenomenon, the spatial patterns that underlie them appear to be of long standing. Clusters spanning major regions and province boundaries cast doubt on purely administrative explanations and suggest the role of primordial sub-cultural variation. Within Han society, a paucity of divorced and remarried women relative to the first married may reflect the "salience of the patriline"(Skinner et al. 2000: 647), that is, conservative family system norms that reduce the status of women and underpin son preference.

The meaning of the strong relationship between infant mortality and sex ratios is open to interpretation. It might be argued that sex differentials in mortality are themselves proxies for sex preferences and other related behaviors. It is possible, for example, that poor survivorship of females goes hand in hand with sex-selective abortion, or with the practice of hiding females from the view of census takers. Since we lack any measure of sex-selective abortion or of the incidence of "hidden" girls, we can never be certain that the sex ratio is not shaped by unmeasured behaviors that are spuriously related to mortality. But an argument that minimizes the effect of mortality itself would be founded more on faith than on evidence. Sex differential mortality is a direct and efficient explanation for sex ratio variation, and the strength of the mortality effect is all the more surprising given the attendant problems of measurement, the serious underreporting of infant mortality in particular. On the questions of whether sex-selective abortion is a major modality of sex selection, or whether hidden girls account for much of the variation in sex ratios, we can provide no evidence. But the evidence at hand strongly suggests a linkage between infant mortality and juvenile sex ratios.

In summary, regional clusters of juvenile sex ratios in China do not occur by chance. They reflect real underlying spatial structures of ideology, administrative practice, and behavior. Much of the variation in sex ratios is explained by underlying cultural patterns represented by major ethnic divisions and by family system norms. Birth planning regimes appear to play a role. Parental interventions affecting infant survival appear to be instrumental in producing the observed pattern of sex ratios. While our model explains nearly half of the inter-county variance, it does least well in explaining the clustering of behavior in predominantly Han China Proper. The failure to account for these clusters means that there is still much left to do.

References

ACMR [All China Marketing Research Co., Ltd.]. 2003a. 2000 China County Population and Socioeconomic Indicators with County Maps. CDC-S-2003-136.

ACMR. 2003b. The Complete Collection of China 2000 Provincial Population Census Data Assembly. CDC-S-2003-201.

Anselin, Luc. 1995. Local indicators of spatial autocorrelation – LISA. *Geographical Analysis* 27,2: 93-115.

Anselin, Luc. 2004. *GeoDa 0.9.5-i.*. http://sal.agecon.uiuc.edu/csiss/geoda.html

Cai Yong. 2004. County and district level life tables based on the 2000 Census of China. Unpublished ms., March 10, 2004.

Cai Yong and William Lavely. 2003. China's missing girls: numerical estimates and effects on population growth. *The China Review* 3,2: 1-17.

Chu Junhong. 2001. Prenatal sex determination and sex-selective abortion in rural central China. *Population and Development Review* 27,2: 259-281.

CITAS [China in Time and Space]. 1997. Census and Regional Systems Data, 1990 [ChinaA]. Compiled by G. William Skinner. http://citas.csde.washington.edu

Coale, Ansley and Judith Banister. 1994. Five decades of missing females in China. *Demography* 31,3: 459-79.

Gu, Baochang; Roy, Krishna. 1995. Sex ratio at birth in China, with reference to other areas in East Asia: what we know. *Asia-Pacific Population Journal* 10, 3: 17-42.

Hull, Terence H. 1990. Recent trends in sex ratios at birth in China. *Population and Development Review* 16: 63-83.

Johansson, Sten and Ola Nygren. 1991. The missing girls of China: A new demographic account. *Population and Development Review* 17: 35-51.

Lavely, William, Li Jianke, Li Jianghong. 2001. Sex preference for children in a Meifu Li community in Hainan, China. *Population Studies* 55:319-329.

Skinner, G. William. 1977. Regional urbanization in nineteenth-century China. In G. William Skinner, (ed.), *The City in Late Imperial China*. Stanford: Stanford University Press.

Skinner, G. William. 1994. Differential development in Lingnan. Pp. 17-54 in Thomas P. Lyons and Victor Nee (eds.), *The Economic Transformation of South China: Reform and Development in the Post-Mao Era*. Ithaca: Cornell University East Asia Program.

Skinner, G. William. 1997. Family systems and demographic processes. Pp. 53-95 in David I. Kertzer and Thom Fricke (eds.), *Anthropological Demography: Toward a New Synthesis*. Chicago: University of Chicago Press.

Skinner, G. William, Mark Henderson, and Yuan Jianhua. 2000. Chinas fertility transition through regional space. *Social Science History* 24,3: 613-652.

Zeng Yi, Tu Ping, Gu Baochang, Xu Yi, Li Bohua, Li Yongping. 1993. Causes and implications of the recent increase in the reported sex ratio at birth in China. *Population and Development Review* 19, 2: 283-302

Appendix

China's Macroregional Systems in Relation to Provinces[5]

[5]Skinner et al. 2000, http://muse.jhu.edu/journals/social_science_history/v024/24.3skinner_map01.html