CEPED-CICRED-INED Seminar on Female Deficit in Asia: Trends and Perspectives, Singapore, 5-7 December 2005. **Draft paper before publication; please do not use it as reference.**

Spatial regression and determinants of juvenile sex ratio in India

CZ Guilmoto
LPED-IRD

Introduction

The rapid masculinization of India's child population during the last two decades runs parallel to the trends observed elsewhere in Asia.[1] The circumstances of India's recent rise in juvenile sex ratio are however complex, because of the combination of a rapid improvement in female life expectancy -which caught up recently with male life expectancy- and of a significant degradation of the child sex ratio during the last decades. While the sex ratio[2] of the whole population has remained almost stable since 1971, oscillating between 107 and 108 men per 100 women, child sex ratio has risen from less than 98 boys for 100 girls below 5 years before 1941 to more than 107 today. These conflicting trends have been the source of confusion among observers unfamiliar with India's demographic history and have contributed to a serious delay in the analysis of rising juvenile masculinity.[3]

This paper will focus on the child sex ratio in 2001 across Indian districts and its determinants. The gradual rise in the proportion of male child population is one of the most visible changes in the sex discrimination regime witnessed in modern India. The contemporary trend is no doubt reminiscent of the old patterns of patriarchal discrimination towards women prevailing during the colonial period. However, the last two decades have clearly heralded a new era of sex discrimination that differs from the *ancien régime* by several of its characteristics such as its unusual efficiency and resulting demographic intensity, its wider social and geographical spread and the new methods employed. To a large extent, the interpretation given to this process of masculinization remains incomplete. Sex discrimination was initially associated to fertility decline as is commonly done for China, but this hypothesis had to be abandoned most notably because of the absence of close correspondence between masculine areas and low fertility areas.[4] Similarly, the trend has attributed to other factors already linked to fertility decline, such as patriarchal systems, economic growth and market-driven modernization, urbanization or progress of education. The recent dowry inflation has also been advanced as an explanatory factor, but it must be noticed that the very propagation of dowry across regions and social groups in independent India has in itself not been satisfactorily explained. One of the major factors behind the trends in the juvenile sex ratio is probably the recent spread of new prenatal sex-detection techniques. This is however only a partial account, as it does not explain why some regions and social groups were more receptive to this technological revolution than others.

For simplicity's sake, the analysis presented here remains strictly synchronic and is based only on the recent 2001 data. It does not therefore take into account the continuing change observed over the last decades. But our analysis will try to focus a large array of factors that can be associated with contemporary sex discrimination and aims at clarifying their relative explanatory power. In particular, this paper intends to examine systematically the specific role played by geographical determinants in sex ratio variations. The regional aspects of child sex ratio in India has often been treated by scholars as a characteristic of anecdotal significance even tough State-level differentials are invariably used to illustrate the

---

[1] See Croll (2000) and Attané and Véron (2005) for a broader perspective.
[2] Sex ratio will be computed here as the number of men per 100 women. Child sex ratio is based on the 0-6 population.
[3] See Mayer (1999) or Griffiths *et al.* (2000). For a disaggregated analysis of sex ratio trends during the 20th century see Bhat (2002).
[4] See however the recent discussion in Drèze and Murthi (2001) and Bhat and Zavier (2003).

recent degradation of child sex ratios in India. We will attempt here to incorporate the spatial factor into our analysis and confront it with most of the social or cultural correlates of sex ratio that are usually put forward. To do this, we will use tools that are still not often used in social sciences, i.e. measurements of spatial dependence and spatial regression. These techniques, which are more commonly implemented in the field of spatial econometrics, are employed here to incorporate in the modelling process both the socio-economic independent variables -as in the traditional statistical analysis- and the spatial autocorrelation of sex ratio that may be observed. From a statistical point of view, such methods aim at reducing the impact of any possible heteroscedasticity that may be due to the geographical patterning that may be observed. Spatial regression takes into account the effect of spatial autocorrelation of the error term in order to improve the quality of the model and to correct erroneous specifications, in particular the computation of regression coefficients. Owing to the often highly spatialized pattern of demographic phenomena (migration, mortality, etc), these new methods represent a promising addition to demographic analysis (Voss *et al.* forthcoming) and our analysis illustrates some of their potentialities.

In a spatial model, geographical proximity is not envisaged simply as a dimension inherited from other characteristics that would themselves follow specific spatial patterns. The geographical patterning when it exists is tested against the other, “aspatial” determinants and the modelling allows confirming the existence of specific spatial residuals left unexplained by other independent variables. It points to the possible role played by spatial configurations such as the impact of geographical contiguity. Proximity and its effects on demographic behaviour point to exchange mechanisms that connect individuals, families, social groups or localities. Each field of social sciences such as demography, geography, sociology or economics tends to express these interaction mechanisms in a particular way (e.g. diffusion, convergence, social interaction, homogenization, etc.) based on various concepts such as epidemic propagation, spill-over, spatial or social diffusion, imitation effect or social capillarity. These ideas will help us to explore at the end of the paper the implications of our results for explaining some of the salient features of sex ratio differentials in India.

A few words about the contents of this paper. It starts with an (necessarily) brief review of the data available for studying gender-based discrimination among the children in India and it discusses the advantages and limitations of the census data. The following part presents findings from least-squares models of sex ratio variations across districts and leads to a re-evaluation of the role played by social and economic factors in shaping sex ratio imbalances. The next section aims at addressing the issues of spatially autocorrelated residuals derived from traditional regression analysis by using a spatial regression model. It brings to the fore new hypotheses on the role of spatial processes. The discussion at the end of the paper offers a summary of the main findings and broadens the debate on the respective function of socio-economic, anthropological and geographical factors in explaining sex ratio differentials in India.

Data to study sex discrimination

The data used here are based on the age and sex distribution of the population during the last 2001 census in India. While age data are far from perfect in India, it can be easily observed from the last single age data released by the Census of India that 1) age distortions are far less acute among children below ten and that 2) sex ratio by age is much more regular among children than adults. In a nutshell, the reason for this lies in the fact that age heaping is the primary cause for distortions in the age distribution and is in turn less acute among children than among other age group.

At the same time, estimating sex ratio at birth in India is made almost impossible by the lack of adequate civil registration statistics. While estimates derived from sample surveys (such as the *Sample Registration System* and the *National Family and Health Surveys*) exist, sex ratio at birth is computed on limited samples and are therefore unreliable whereas sex ratio levels for various age groups are based on exhaustive census counts. Similarly, estimates of excess girl mortality, which is the second component of girl deficit, suffer severe limitations and are not available at a fine geographical scale. A further advantage in using sex ratio computed on the child population rather than at birth or after birth is precisely that this measure sums up the joint impact of both antenatal and postnatal discriminatory practices such as feticide, infanticide or excess mortality among girls.

The level of sex ratio at birth in India is comparable to what can be observed elsewhere with minor differences that are however unlikely to generate any significant increase in the sex ratio of the general population. It has been recently argued that Hepatitis B infection might be responsible for some of the deficit in women observed in the past in Asia. However, whatever the accuracy of these hypotheses that have been already been disputed[5], they point to a limited impact of hepatitis B on the sex ratio at birth in India compared to what has been alleged for China.

As a result, the lack of established biological imbalance in the sex ratio at birth among Indian populations and the presumably negligible impact of underestimation differentials signal that most of the deficit of girls measured from the census age statistics derive from social manipulations of the sex composition of children. Three main discriminatory mechanisms can be listed here:

- Sex-selective abortion: termination of pregnancy following detection (by amniocentesis or ultrasound scan) of a female foetus.

- Female infanticide: killing of girls (often immediately) after birth

- Excess girl mortality: excess mortality related to neglect (lack of proper care or food to girls).

The effect of other mechanisms related to intra-uterine mortality or stillbirth on sex ratio at birth is undoubtedly modest. This is probably also the case of the sexual selection before conception (e.g. through sperm separation). While this method is not absent from India and is offered by some professionals in states like Punjab, it is still too expensive to become common. However, the estimation of the respective share of each method to the overall girl deficit is beyond the scope of this paper, in spite of its importance for understanding the trends in discriminatory behaviour.[6]

[5] See Olster (forthcoming) and a detailed critique in Klasen (2005).
[6] See however Arnold *et al.* (2002).

Modelling sex ratio variations

The social and geographical heterogeneity of sexual discrimination in India

Reports on the juvenile masculinity tend to give for each country a somewhat homogeneous image as if the rise in sex ratio levels had affected all regions and social categories in a uniform manner. From such a perspective, it is but natural to look for global or structural phenomena at the core of sex ratio degradation. However, such an approach not only misses the social or geographical details of the phenomenon, but also deprives us of crucial information to attempt a systematic interpretation of demographic change and its causality. In India as well as in China, a considerable level of heterogeneity between social groups or between areas are visible in terms of sex ratio levels and these differentials do offer invaluable keys for the comprehension of the discriminatory mechanisms and its main actors (Guilmoto 2005).

In India, the debate fuelled by the 2001 census results that showed the concentration of sexual discrimination in the North-West of the country has gradually moved to a non-geographical interpretation focusing on the Sikh population. Sikhs in India are indeed the largest community in the North-Western region that is otherwise the most affected by girl deficit. The regional dimension has given way to a more religious-oriented interpretation blaming Sikh communities for the degradation in child sex ratio. The census of 2001 even published for the first time age and sex data cross-classified by religion. Religious data in this format had never been published before, except during the colonial period when the British were trying to break down India's society into caste and religious subsets. These figures (Table 1) do confirm the exceptionally high sex ratio among Sikhs and, to a lesser extent, that observed in the Jain community. Within the other minority religious groups like Muslims or Christians, discrimination towards young girls appear to be less pronounced than in the rest of the population. Moreover, the data also highlight the slightly urban aspect of the phenomenon. Unfortunately, the sex ratio of the children cannot be computed for other social variables such education level or economic status dimensions, mostly because these variables are not collected for children.[7]

Insert Table 1

Census statistics offer however extremely detailed information on child sex ratio by administrative unit as figures are available at the lowest administrative units (such as the rural "census village" or the urban "wards"). This shows that not only are the variations between Indian States significant, but the mean State values themselves tend to conceal sizeable intra-regional disparities. Thus, while the average sex ratio in Tamil Nadu is lower than in the rest of India (101 boys per 100 girls less than 7 years against 108 in India), several of its districts record abnormal values above 112. Figure 1 shows the distribution of the values by district.[8] It may be noticed that whereas 148 districts exhibits sex ratio levels of masculinity that are higher than 110, with 39 among them displaying extreme values exceeding 120, several Indians districts are on the contrary characterized by a female preponderance before 7 years, which in itself also seems a rather unusual phenomenon.

[7] Household data derived from the NFHS or the RCH surveys can be used to relate in more details sex ratio to household characteristics.

[8] India is divided into 35 States and 593 districts.

Insert Figure 1

This map (Figure 1) shows obvious regional contrasts, but it also indicates a strong level of geographical regularity in the distribution of the values.[9] There are several compact pockets of high sex ratio values, the first among them consisting Punjab and Haryana and stretching towards the adjacent States of Rajasthan to the South and Madhya Pradesh and Uttar Pradesh to the East. Another core area is visible in central Gujarat (not far from Ahmadabad city) and extends to the South towards Maharashtra. However, a subregion with lower levels of child sex ratio is located at the border of the states of Gujarat, of Rajasthan and Madhya Pradesh; this area corresponds to rural tracts where the population is mostly tribal (with the *Bhils* as the major tribal group), characterized like often elsewhere in India tribal by a sex ratio close to the normal. It is important to stress that the remainder of the country is hardly affected by sex ratio unbalances. Except for some isolated pockets in coastal Orissa and in Tamil Nadu, child sex ratio seldom exceeds 105 outside the core areas previously described

The ultimate determinants of sex ratio variations remain therefore open to debate as regional and social or religious dimensions tend to form superimposed layers that are at times difficult to unravel. The respective role of factors related to the composition of population and interpretations is often ambiguous. Is indeed Punjab male-dominated because it is mostly Sikh or because it is one of the most prosperous and economically dynamic State? Is discrimination pronounced among the elite groups because of their caste or religious membership, because of their education levels or because of their very standard of living? Conversely, should we attribute the low sex ratio observed across tribal India to the underdevelopment of the tribal-inhabited areas, to their relative geographic isolation from the rest of India or to the specific anthropological characteristics of tribal communities? These issues are often very sensitive in India as the recent debate on the causes for Muslim high fertility in India has indicated: scholars find it difficult to tell apart the social and economic determinants of high fertility from a hypothetical specifically Muslim factor. To investigate these issues here, we will use a multilinear approach in order to model inter-district variations in child sex ratio using a large array of independent variables.

Regional variations

Modelling regional variations

District figures by district will be used here for various models, starting with a straightforward linear regression model.[10] The variables used to model district differentials in child sex ratio falls into three main categories:

- demographic indicators: density, fertility, child mortality
- socio-economic indicators: urbanization, literacy, participation rates, agriculture, infrastructures
- religious and social composition: religion, Dalits (Scheduled castes, ex-untouchables), tribals

[9] For the proper definition of “spatial heterogeneity” and “spatial dependence”, see Voss *et al.* (forthcoming).
[10] Data used here pertain to the district as very few data are available at the lower administrative level (5564 tahsils or subdistrict units)

The distance to Amritsar (the religious capital of Sikhism located in Punjab) will also serve a preliminary geographical indicator. These variables are presented in Table 2, with a brief statistical description. It may be noted that some variables do not come from the Census such as distance measurements (computed between Amritsar and district headquarters with a GIS), fertility and mortality estimates (Guilmoto and Rajan, 2001) and the district infrastructure index for 1996-97.[11] Some variables had to be discarded for lack of statistical significance (participation rates, employment sectors, specific religious groups) or due to risk of strong colinearity (proportion of Hindus, other development indicators from the CMIE). Moreover, some data are unfortunately missing and many of them (such the average household income, the incidence of poverty, the family structure or the prevalence of dowry) would definitely have enhanced our analysis had they been available at the district level.

Insert Table 2

Independent variables were gradually introduced into the statistical model, starting with demography (model A), followed then by socio-economic characteristics (model B) and social composition (model C). A synthesis model (model D) brings all of them together them. The results from these four models are shown in Table 3.

Model A shows the absence of any straightforward link between the regional demographic characteristics and high sex ratio, except for the positive correlation observed with population density. It is worth stressing once again that high sex ratio is not mechanically associated with fertility decline, especially because lower fertility areas include both low sex ratio districts (in South India) and the highest sex ratio districts (In Punjab). Model B similarly indicates a rather weak correlation between juvenile masculinity and the various socio-economic variables used here such as urbanization or literacy levels. The only really substantial effect of these variables is visible for the infrastructure index: better developed districts are characterized here by a sex ratio higher than other districts. While the first two models examined explain between 6% and 12% of the overall variations across districts, the following model C based social composition variables proves much more significant and its variables account for more than 46% of the variations. As expected, the proportion of Sikhs and Jains in each district's population is a major determinant of sex ratio differentials. Conversely, the presence of Muslims and of tribal population causes a substantial reduction in the child sex ratio.

Insert Table 3

The results of these separate models are now pooled into a single model, which explains now more half of the district differentials in child sex ratio with no less than nine significant variables identified (at 5% confidence level). Model D shown on Table 3 has the great advantage of eliminating several instances of spurious correlations while confirming ambiguous links. For instance, after adding to the model other variables such as proportion of Sikhs in the population, child sex ratio appear to be strongly related to high fertility. [12] Similarly, the previously unnoticed correlation with literacy is now evident and reinforces the link between masculine sex ratios and development indicators that was already detected

[11] The CMIE infrastructure index is based on various dimensions such as electrification, road, railways, irrigation, bank facilities, post offices, schools and health centres. See CMIE (2000). This index is however not available for many district of North-East India.

[12] The previous negative link between literacy and discrimination towards girls reported by Murthi et al. (1995) is due most probably to the fact that these authors analysed excess child mortality, a somewhat archaic method of sex discrimination. Modern discrimination techniques mostly affect the sex ratio at birth.

through the infrastructure index. Three positive correlations coefficient point to one of the most puzzling aspects of increasing sex ratios in India: their relationship with on the one hand high fertility and on the other hand with development indicators that are otherwise negatively associated with fertility.[13] It is worth noting at this juncture that contrary to other univariate analyses (Agnihotri, 2000), female employment does not reduce sex ratio, in contradiction with the original hypothesis linking female participation rates to lower levels of discrimination (Bardhan 1974).

However, the overall impact of these social or economic factors is of lesser relevance than that of the anthropological variables introduced in model C. The impact of social composition is significant for all these variables except for Buddhists.[14] According to model D, a district's sex ratio would drop by 7 boys for 100 girls if it were entirely tribal (or Muslim, Dalit), but would rise by 28 if it were populated only by Sikhs. Thus, the religious or caste composition of the population appears to play a larger role in explaining sex ratio differentials than purely economic factors. This central feature of the geography of discrimination in India seriously undermines explanations that used structural factors such as capitalist development, social modernization or the impact of increased monetization. While these global social and economic trends affecting a rapidly modernizing society like India's probably matter in the recent rise of juvenile masculinity, they appear to be filtered through the social fabric of society built on traditional caste and religious divisions. Another way to put it is to suggest that castes and communities are still in India the primary matrix determining behaviour related to sex discrimination whereas the effect of economic change on is still to be felt.

Space matters

Having said that, do we mean to imply that once again, the geographical clustering that is so obvious on maps at various scales is nothing but an offshoot of other factors such as the social and religious make-up of each district's population ? It would definitely be tempting to stop at this point and consider that our model does explain a large part of the observed disparities across India. However, it would be wiser to stress the fact that there remains a large grey zone of unaccounted variance in our modelling of sex ratio variations (more than 45%) despite the large number and diversity of variables that have been probed.[15] There is little probability that we would achieve a much better modelling by using missing variables such as income levels (closely related to literacy or infrastructures) or dowry prevalence (the latter variable being probably endogenous).

A first examination that is rather unsophisticated in its geographical formulation consists in introducing geographical clustering around the North-Western region by a simple distance measurement. We computed this variable by measuring the distance from each district to Amritsar city, home to the holiest shrine of Sikhism (the *Harmandir Sahib* or Golden Temple). This variable turns out to be very strongly correlated to sex ratio as Model E shows in Table 4 and accounts for more than a third of the overall sample variance.[16] In the following model F, derived from the synthetic model D previously examined, we have added

---

[13] See also Agnihotri (2000, 2003) for the analysis of the discrimination-prosperity nexus.

[14] Buddhists represent a rather disparate category in India comprising former Dalits as well as Himalayan populations.

[15] Other models of discrimination variations across district do not yield better results when they don't make use of additional spatial information. See Kishor (1993), Murthi et al. (1995) or Drèze and Murthi (2001).

[16] Note that other locations such as Chandigarh city would yield similar results.

this distance measurement to the other social and economic variables; it becomes now the most powerful variable of the model as if proximity to Amritsar were the first determinant of masculinity in India. The overall quality of our model as reflected by the correlation coefficient has also shot up to more than two thirds of the overall variance of Indian district sex ratios. We may also observe that adding this variable cancels out almost entirely some of the statistical links previously shown. This is true of most social or economic variables which are no more significant. Worth noting is also the significant reduction in the explanatory power of the proportion of Sikhs in the population. It may be time to remember that all Punjabis are not Sikhs and also, for that matter, that a majority of Sikhs lives outside Punjab.

The distance to Amritsar remains a rather crude indicator of the spatial patterning of child sex ratio in India. It does not, for instance, take into account other core areas of masculinity such as Gujarat. Punjab is not the only eye of the masculinity in India. Furthermore, computed geometric distances between localities ignore intervening obstacles that geographers label as barriers, be they physical barriers or cultural discontinuities. Instead of trying to improve on our distance variables, let us come back in more detail to the unexplained part of our previous strictly aspatial models.

Insert Figure 2

To do that, we turn to the inspection of residuals derived from our previous synthesis model. These values are computed from Model G [17] for all Indian districts and then plotted on a map (Figure 2). Only districts with absolute values above 2.5 boys per 100 girls are shown. Positive residuals (when actual sex ratio values are higher than predicted values) are prominent and point to several core zones that have been already mentioned such as Gujarat or more isolated areas like Salem in Tamil Nadu. Similarly, the North-West region centred on Delhi and stretching from Jammu-Kashmir to Gwalior in Madhya Pradesh remains visible with the highest positive residuals, with the exception of Punjab where the presence of Sikhs has cancelled out high sex ratio residuals in our model. While negative residuals are of lesser value and slightly less concentrated, they nevertheless display several geographical features as the examples of Kerala, coastal Tamil Nadu or the Bihar-Jharkhand region demonstrate.

Spatial homoscedasticity and its solutions

This examination means that our model has hardly eliminated the prominence of the spatial patterning of sex ratio. Moreover, this feature is a source of a major statistical problem, because it contradicts the hypothesis of homoscedasticity that is central to our regression models. The assumption of homoscedasticity is that the variance of residuals (or error terms) is the same for all predicted values. In fact, this residuals map suggests that data that do not meet the assumption of homoscedasticity because of spatial autocorrelation of the residuals. Spatial autocorrelation reflects the fact that residuals of similar values are not randomly distributed as they are seen to cluster on the map. While slight heteroscedasticity has little effect on significance tests, marked heteroscedasticity can lead to serious distortion of estimates and therefore weakens the analysis. Fitting a model by ordinary least-squares when the residuals are autocorrelated causes the sample variance to be underestimated and the correlation coefficient to be overestimated. Beyond the intuition derived from the inspection of their skewed geographical distribution on the map, residuals can be tested for homoscedasticity through a test like that of Breusch and Pagan. When applied to our synthetic

[17] Model G is similar to synthesis model D except for the removal of the CMIE infrastructure variable (missing for 71 districts).

model (Model G), the test confirms the heteroscedasticity of the residuals with the statistic taking a value of 76,76, significant at 0.1% level .[18]

Spatial econometrics and quantitative geography have examined issues of spatial heteroscedasticity ever since the measurement of spatial dependence has become common. Various solutions have been proposed, among which the simplest consists in introducing dummy variables to correct the regional effects. The addition of binary indicators related to regions («southern», «Bimaru», etc.) or States («Kerala», «Punjab», etc.) is a widespread expedient procedure to improve regression results. However, such an ad hoc solution tends to transform complex regional effects into simple dummy variables that reduce spatial proximity to a binary indicator. The meaning of such dummy variables remains completely unspecified and contributes to freeze the investigation by giving an apparent clear-cut regional explanation to a problem which is geographically much more intricate. Thus, "Punjab" can become an explanatory variable whereas the regional effect visible in North-Western is far from reduced to specific States and tends to spread gradually from core areas to their surroundings (like western Uttar Pradesh or North Rajasthan). Moreover, such a method is unlikely to exhaust the effect of spatial dependence which is also present in areas with lower sex ratio (such as the so-called tribal belt ranging from Gujarat to Jharkhand).

More robust methods tackle the problem of heteroscedasticity by resorting to geostatistical modelling of the spatial autocorrelation. Here, we propose to correct the effect of spatial autocorrelation of child sex ratio by incorporating this spatial dependence in the model itself. Two major types of methods of spatial regression exist to do this.[19] The first technique is the model of *spatial lag*. It basically consists in assuming that a value in a given district is influenced by the values observed in adjacent districts. It is based on a model combining independent variables and an autoregressive lagged component (with a contiguity matrix to identify "adjacent districts"). This specification is useful when the phenomenon under study is directly affected by values observed in the surroundings. This model was applied to our Indian data, but proved rather inconclusive.[20] Results will therefore not been presented here for the sake of brevity.

The second type of model for spatial regression is the *spatial error* model (or *spatial autoregressive error* model). It presupposes that only the term of error (the residual) is spatially autoregressive: this residual term is supposed to capture unobserved factors, which on the one hand influence the dependant variable and on the other hand are spatially autocorrelated and therefore responsible for the autocorrelation of the residuals. The following section portrays in more details the model's specification.

We start with an ordinary linear model of child sex ratio (*CSR*), determined partly by exogenous variables (**X**). This model is given by (1).

(1) $CSR = \mathbf{X}\beta + \varepsilon$

with *CSR:* child sex ratio

X vector of independent variables

[18] Similar results can be derived from the White test.
[19] Methods described by Bailey and Gatrell (1995), Haining (2003) or Fotheringham *et al.* (2000). The computing procedure is given in Anselin (2003).
[20] Results from this spatial lag model have been tested using the overall correlation coefficient and the value of the maximum likelihood test.

β parameters

ε, error term with mean 0 and standard deviation σ

However, in the model with error term, the residual (error) term U is spatially autocorrelated, i.e. it depends in part on values of the error terms in neighbouring observations as shown in (2). This specification is shown in (3).

(2) $CSR = \mathbf{X}\beta + U$

(3) $U = \lambda\, \mathbf{W}\, U + \varepsilon$

with U vector of spatially autocorrelated error

λ parameter of the error term

W contiguity matrix (valued 1 for pairs of contiguous districts and 0 elsewhere)

The matrix **W** defines the pairs of adjacent observations and is different from 0 only when geographical units are regarded as «neighbours». The error term U is then modelled in an autoregressive way. The model leads to a system of simultaneous equations, but it will be observed that when λ= 0, the equation is that of an ordinary (aspatial) regression. These models with spatially autocorrelated error terms are in particular applicable when unobserved characteristics, present in the residual error term, are spatially autocorrelated (Haining 2003). Such an assumption in our case would for example suggest that fine indicators of community membership, which are strongly spatially correlated, account for residual variations not explained by our independent variables. We will come back later to this point.

We can apply the model with spatial error described above after computing a contiguity matrix. To do so, we used common district boundaries to determine "contiguity districts", a method formally know as "first order queen contiguity". However, higher orders of contiguity have been also used and give similar results. Results of this regression are shown as Model H and are to be compared with the corresponding synthetic model G also reproduced in Table 4.

Insert Table 4

Results and interpretation of the spatial regression

The results of our spatial regression are remarkable in several respects and call for various observations. At the outset, it should be noted that the goodness of fit of model H has considerably increased from 0.52 to 0.80. This large improvement is entirely attributable to the autoregressive error factor introduced in the model: the parameter λ takes the value of 0.84, a value close to unity that indicates a maximum level of correlation (t = 35.4). As the previous formula (3) shows, λ is the coefficient of the autoregressive effect of contiguity on the distribution of the residuals. We may also observe the impact of spatial specification on other estimated coefficients for the model's independent variables. Their values tend to drop almost systematically. Accordingly, several indicators that were apparently closely associated to regional variations in sex ratio in the previous model (model G), cease being significant (e.g. fertility) or record a clear reduction in their correlation (literacy, tribal population, Jains). The variables that seem to better resist to the introduction of the spatial specification are the proportion of Sikhs and Muslims in the population.

The contribution of spatial dependence represents thus a major improvement for the modelling of child sex ratio variations and it widens the interpretative prospect beyond social or demographic dimensions usually resorted to. Spatial modelling has three principal consequences: it corrects misspecified regression coefficients pertaining to several socio-economic correlates, it confirms the centrality of the social composition variables to the understanding of the geographical structure of sex ratio and it underlines the pre-eminence of geographical structuring in the causal diagram. Let us re-examine here this last point that is seldom thoroughly scrutinized. The spatial factor proved to be more significant than any other single social, anthropological or economic factor examined so far. While the model used refers to a “spatial error”, this is something of a misnomer as the mechanisms at work are unlikely to be random or fortuitous errors. The error term must consequently be interpreted differently, as a residual corresponding to unobserved regionalized factors. If these factors are independent of other variables, it may be surmised that they are related to spatial mechanisms of propagation of discriminatory behaviour, operating in particular from specific core areas. Proximity of districts to these “hot spots” of sex discrimination determines changes in behaviour towards girls that are in part independent from the conditions that are usually associated to high sex ratio.

What our findings suggest is that there remain probably various objective conditions that enhance gender-based discriminations. For instance, it is likely that mothers’ literacy promote the spread of information on new technologies, that economic status facilitates access to these technologies or that high child mortality exacerbates boy preference. But as our statistical analysis unambiguously shows, these factors are far from explaining the severely skewed spatial distribution of high sex ratio in India. Instead, the strong spatial dependence that can observed and measured hint at the presence of factors directly related to the geographical distribution of communities and to additional spatial processes. The primary level at which discriminatory behaviour is shaped is probably that of communities, i.e. mostly caste (*jati*) groupings. One of the few generalizations that can be made about caste distribution in India is that they display a high level of geographical concentration within historical regions. The vast majority of caste groups inhabit a narrow territory and modern migratory redistribution has hardly modified the picture except perhaps in the largest urban agglomerations. Discriminatory behaviour towards female offspring derives from patriarchal value systems tend to be enforced within caste groups and are therefore relatively homogenous within specific communities in spite of other social and economic differentials. Geographical concentrations of specific communities and the importance of these institutions in shaping discriminatory behaviour among their members explain then to a large part why child sex ratio as an index of discrimination towards girls displays strong spatial patterning. However, this explanation is based on a somewhat socially immobile society where behaviour is assumed to be unchanging over the years. The issue of increasing sex ratio has however emerged because of a rapid change in the proportion of girls among the children in certain regions, demonstrating the suddenness and the range of social transformations and changing attitudes. We will re-examine these issues and their implications in the final concluding section.

Culture vs. Space

This paper has tried to explore to describe the singular excess masculinity among children in India from various angles. For reasons of space, our attempt had to be restricted to a synchronic approach focusing on the 2001 figures in order to fully explore the geographical and statistical differentials in sex ratio. Our approach aims at offering a renewed interpretation of sex ratio degradation, which is often attributed to the persistence of the patriarchal regimes in modern India, regimes which are sometimes indirectly gauged by related indicators such as high fertility, joint family, or various dimensions of female marginalization (violence against the women, lack of economic or social autonomy, dowry, etc).

Our analysis does indicate that some of the major factors associated with skewed sex ratio pertain primarily to the social composition of the population. Castes or religious groups offer the primary interpretative framework, suggesting that socio-economic status matter less than community membership when it comes to discriminatory practices. While child sex ratio may be slightly higher among the affluent, the key differentials across the population point more to religious, caste and regional effects than socio-economic effects. But it would be probably fair at this point to mention some of the limits of our anthropological analysis, based at it is on somewhat imperfect categories. This might the case for the category of "Muslims", a population of no less than 138 million inhabitants that encompasses scores of communities that are distinct from each other by language, history or sectarian orientation. Similarly, the category of "tribal population" brings together many groups that may share little between themselves except for a high level of economic deprivation and their relative isolation from the rest of the society. Even a group like that of the Sikhs that may appear socially and spatially well-defined happens to be internally segmented: it comprises various caste groups such as the *Jats* (also found among Hindus) as well as the subaltern groups of Dalit origin like Sikh *Mazhabis.* Without indulging in the now ritual critic of the classifying project of the Census of India, it would be nevertheless easy to show how each individual group captured by census definitions and used in this analysis may actually be internally heterogeneous.

It is therefore all the more amazing to observe that these sociologically questionable categories turn out to be statistically significant in our models at the pan-Indian level. This suggests that had we been able to use finer categories (like Brahmins, Urdu-speaking Muslims, most backward classes, etc.) or individual subcaste groups (i.e. *jatis*), results would have been even more significant. An illustration of the potential derived from the use of better defined social categories is given by one of the numerically smallest groupings employed here, i.e. that of the Jains accounting for a mere 0.4% of the total Indian population and whose sex ratio is however consistently higher in our models. Work conducted at the regional level such as Vella's research on the Salem area (2005) demonstrate the role plaid by specific local groups such as the dominant castes of *Kongu Vellalar Kavuntar* in introducing a systematic bias towards female offspring formerly based on common female infanticide.[21]

They are institutional mechanisms within caste groups that tend to reinforce a value system biased against girls and women. The marriage system is no doubt closely related to gender-based discrimination and dowry arrangements are a clear illustration of that feature. But can we consider that castes and communities constitute the basic units to understand the unequal distribution of child sex ratio in India? Should cultural indicators such as caste or community entirely replace the social or economic household characteristics as predictors of discriminatory behaviour? Probably not, as I believe it would be misleading to envision India simply as the collection of thousands of caste atoms, each of them constituting a discrete segment with specific demographic behaviour. All things considered, if the community framework were the ultimate key to understand sex ratio diversity, the resulting map would look like a fragmented mosaic following local caste contours rather than the well-ordered distribution observed on the map. Instead of the relative patchwork observed in Eastern China

[21] For another local analysis of the caste configurations, see for example the study by George (1998) on Haryana.

(see for example the maps by Lavely and Cai, 2004), India offers a very coherent geographical pattern and the presence of individual communities (that rarely number more than a few millions people) are subsumed under larger regional trends. This spatial dependence does not link only members of specific communities to each other, but also contiguous caste groups and communities each other. It seems obvious that the spatial logic takes here precedence over other social determinants. To give a simple illustration of this, let's examine the child sex ratio of Muslims in Punjab (115 boys per 100 girls) which turns out to be much higher than that of Hindus in Kerala (104), even when we have shown that on a *ceteris paribus* basis, gender-based discrimination is systematically weaker among Muslims. But no more so in Punjab where Muslim sex ratio is rather high, even if lower than that of other religious groups. A parallel illustration could be derived from the significantly lower child sex ratio observed among Sikhs living outside Punjab or Haryana.

The anthropological model of discriminatory behaviour should then be qualified in two ways. First, we have to ensure that units for the analysis are localized communities rather than larger heterogeneous groupings spread over a large territory. These local groups may function as cultural units with their own culturally formed systems of values and institutions and therefore determine homogenous gender attitudes. But at the same time, these groups don't live in complete closure and their permeable system of norms and representations is affected by that of their neighbours through a compel pattern of interaction. The above-mentioned research by Vella on Salem area shows for instance how dominant groups can locally impose their value system that used not long ago to legitimate female infanticide and contribute to its diffusion to other, lower-status groups. The resulting map of sex ratio clearly displays features familiar to geographers as they point to diffusion mechanisms whereby behaviour tends to spread from core, pioneer areas towards peripheral zones. Two main mechanisms are however at work. The first is *vertical (social, internal, hierarchical) diffusion* within the locality towards other social groups and follows often a top-down model. This gets primarily reflected by the chronological intensification of sex ratio distortion as more and more groups practice active sex selection (through abortions) in given areas. The second mechanisms is that of *horizontal (spatial, external) diffusion* towards "neighbours" (adjacent localities etc.) and its pattern appears clearly on the successive maps of child sex ratio that have not been examined here for lack of space. High sex ratios affect contiguous areas after a specific time lag through a visible process of spatial diffusion.

Our analysis does not allow us to envision what the exact ingredients of these diffusion processes are. It would be indeed relevant to clarify whether these innovations consist of ideas spreading across social groups or of new techniques and know-how. Recent measurements of sex bias from the NFHS surveys suggest that preference for boys may have rather slightly declined during the last decade. The ideational contents of diffusion processes may therefore be less important than that of other factors. In particular, the rapid diffusion of new sex-detection techniques based on ultrasound scan starting during the 1980s suggest that this technological breakthrough has suddenly enabled families in traditionally patriarchal areas to carry out a more systematic sexual screening of their offspring. This line of interpretation offered by Bhat and Zavier (2003) who compare it to an epidemic appears to fit our observations. We would simple need to add to it a time perspective to explain how new discriminatory techniques have gradually spread to neighbouring groups where preference for boys had previously been less noticeable. Sex-determination techniques have become cheaper over the years, allowing lower sections of the population to avail themselves of the new technology in rural areas. At the same time, new, modern norms of boy preference based on foeticide (rather than older strategies such as infanticide) have probably affected new communities in the same way as dowry has spread across communities that formerly followed the bride-price system. Consequently, the propagation of sexual selection as a rational family strategy to optimize the "quality of children" (to use Gary Becker's terminology) has still a long way to go before exhausting its potential in India. The large number of areas and communities still unaffected by it point to its growth and diffusion potential across new territories.

Bibliography

Agnihotri, S B, 2000. *Sex Ratio Patterns in the Indian Population – A Fresh Exploration*, Sage, New Delhi.

Agnihotri, S B, 2003. "Survival of the Girl Child. Tunnelling Out of the Chakravyuha", *Economic and Political Weekly*, October 11, 4351-4360.

Anselin, L., 2003, *GeoDa 0.9 User's Guide*. Spatial Analysis Laboratory (SAL), University of Illinois at Urbana-Champaign.

Arnold, F. *et al.*, 2002, 'Sex-selective abortions in India', *Population and Development Review*, Vol.28, No. 4, 759-785.

Attané, I., Véron, J., 2005, *Gender issues at the early stage of life in South and East Asia,* Collection Sciences Sociales n° 9, IFP-CEPED, Pondicherry.

Bailey, T.C. and Gatrell, A.C., 1995, *Interactive Spatial Data Analysis*, Longman, Harlow.

Bardhan, P., 1974. "On Life and Death Questions", *Economic and Political Weekly*, August, 1293-1304.

Bhat, PN Mari and AJF, Zavier, 2003, 'Fertility Decline and Gender Bias in Northern India', *Demography*, Vol. 40, No. 4, 637-657.

Bhat, PN. Mari, 2002, 'On the Trail of 'Missing' Indian Females', *Economic and Political Weekly*, December, 105-5118 and 5244-5263.

CMIE, 2000, *Profile of Districts*, Centre for Monitoring Indian Economy, Mumbai.

Croll, E. 2000, *Endangered Daughters: Discrimination and Development in Asia*, London and New York: Routledge.

Das Gupta, Monica, and Bhat, P.N. Mari 1997. "Fertility decline and increased manifestation of sex bias in India". *Population Studies,* 51, 3. 307-315.

Drèze, Jean and Murthi, Mamta. 2001. Fertility, education and development: Evidence from India. *Population and Development Review,* 27, 33-63.

Fotheringham, S., Brundson, C., and Charlton, M., 2000, *Quantitative Geography. Perspectives on Spatial Data Analysis,* Sage, London.

George, Sabu M, and Dahiya, Ranbir S., 1998, "Female Foeticide in Rural Haryana", *Economic and Political Weekly*, August, 33, 32, 2191-2197.

Griffitths, Paula, Zoe Matthews and Andrew Hinde, 2000. 'Understanding the Sex Ratio in India: A Simulation Approach', *Demography,* 37, 4. 477-88.

Guilmoto, C. Z. and S. Irudaya Rajan, 2001, "Spatial Patterns of Fertility Change in Indian Districts", *Population and Development Review*, 27, 4, 713-738.

Guilmoto, C. Z. and S. Irudaya Rajan, 2002, "District Level Estimates of Fertility from India's 2001 Census", *Economic and Political Weekly*, 37, 7, 665-672.

Guilmoto, CZ, 2005. "A spatial and statistical examination of child sex ratio in China and India", in Attané, I., and Véron, J., eds., *Gender discriminations among young children in Asia*, French Institute, Pondicherry, IFP-CEPED, Pondicherry, 137-167.

Haining, Robert, 2003, *Spatial Data Analysis: Theory and Practice*, Cambridge University Press, Cambridge.

Kishor, S., 1993. "«May God Give Sons to All»: Gender and Child Mortality in India", *American Sociological Review*, 58, 2, 247-265.

Klasen, S., 2005, "Comments on Paper by Emily Oster: Hepatitis B and the Case of the Missing Women", accessed from http://www.iai.wiwi.uni-goettingen.de/

Lavely, William and Cai Yong, 2004. "Spatial Variation of Juvenile Sex Ratios In the 2000 Census of China", presentation at the annual meeting of the Population of Association of America, Boston, April 1-3

Mayer, P. 1999, 'India's Falling Sex Ratios' *Population and Development Review*, Vol. 25, No. 2, pp. 323-343.

Murthi, M. *et al.*, 1995, 'Mortality, Fertility, and Gender Bias in India: A District-Level Analysis', *Population and Development Review*, 21, 4, 745-782.

Oster, E., Forthcoming, «Hepatitis B and the Case of the Missing Women», *Journal of Political Economy*

Vella, S., 2004, "Low Fertility and Female Discrimination in South India: The Puzzle of Salem District, Tamil Nadu", in C. Z. Guilmoto and S. Irudaya Rajan, eds., *Fertility Transition in South India*., Sage, New Delhi.

Voss, Paul R., Katherine J. Curtis White, and Roger B. Hammer, forthcoming, "Explorations in Spatial Demography." in William Kandel and David L. Brown (eds.), *The Population of Rural America: Demographic Research for a New Century.* Dordrecht, The Netherlands: Springer.

Table 1: Child sex ratio by religion, rural and urban, 2001

Religions	Total	Rural	Urban	*Share in the total population*
Hindus	108.1	107.3	111.4	*80.5%*
Christians	103.7	103.6	103.7	*13.4%*
Muslims	105.3	104.7	106.6	*2.3%*
Other	102.5	102.5	103.4	*0.6%*
Sikhs	127.2	126.9	128.5	*1.9%*
Buddhists	106.2	105.8	107.0	*0.8%*
Jains	114.9	115.1	114.9	*0.4%*
Inde	**107.9**	**107.1**	**110.4**	**100.0%**

Source : Census of India 2001

Table 2: Variables used in the models

Variable	Obs	Unit	mean	SD	Min	Max	Source
Density (log)	591	log(density)	5.79	1.19	0.69	10.29	**Census**
Urbanisation	591	proportion	0.24	0.20	0.00	1.00	**Census**
Demographic growth 1991-2001	591	%	22.54	11.6	-3.5	95.01	**Census**
Fertility	591	TFR	3.30	1.01	1.33	5.79	**CZG SIR**
Child survival	591	L_{0-6}	0.92	0.03	0.83	1.00	**CZG SIR**
Literacy	591	Proportion	0.64	0.13	0.30	0.97	**Census**
Literacy (m)	591	Proportion	75.13	11.2	39.59	97.59	**Census**
Literacy (f)	591	Proportion	53.10	15.5	18.49	96.06	**Census**
Literacy (m/f)	591	Sex ratio	1.49	0.29	0.91	2.53	**Census**
Participation (m/f)	591	Sex ratio	2.32	1.47	0.94	12.30	**Census**
Participation (m)	591	proportion	.62	.05	.4304	.99	**Census**
Participation (f)	591	proportion	.34	.14	.049	.73	**Census**
Agriculture	591	proportion	0.60	0.21	0.00	0.90	**Census**
infrastructures	520	Index CMIE (India=100)	103.5	38.8	57.4	625.7	**CMIE**
Buddhists	591	proportion	0.02	0.07	0.00	0.77	**Census**
Christians	591	proportion	0.07	0.19	0.00	0.98	**Census**
Jains	591	proportion	0.00	0.01	0.00	0.05	**Census**
Muslims	591	proportion	0.12	0.15	0.00	0.98	**Census**
Sikhs	591	proportion	0.02	0.11	0.00	0.86	**Census**
Dalits	591	proportion	0.15	0.09	0.00	0.50	**Census**
tribes	591	proportion	0.16	0.26	0.00	1.00	**Census**
distance to Amritsar	591	gradients	11.92	6.06	0.00	28.68	**SIG**

Notes: some variables are not retained in the final models presented here

Sources: Census = census of India; CZG-SIR= Guilmoto and Rajan (2001, 2002): CMIE= CMIE (2000) ; GIS= geographic information system.

Table 3: Regressions of child sex ratio variations

	model A demography			model B socio-economic data			model C social composition			model D synthesis		
Variable	**Coef**	***t***	**Prob**	**Coef**	***t***	**Prob**	**Coef**	***t***	**Prob**	**Coef**	***t***	**Prob**
Density (log)	0.012	5.950	0.000							-0.008	-2.490	0.013
Fertility	-0.001	-0.400	0.691							0.021	5.880	0.000
Child survival	0.037	0.390	0.698							0.069	0.740	0.459
Urbanisation				0.038	1.750	0.081				0.013	0.730	0.464
Literacy				0.071	1.730	0.084				0.128	3.680	0.000
Literacy (m/f)				0.005	0.280	0.781				0.002	0.150	0.882
Participation (m/f)				0.004	2.240	0.026				0.003	1.670	0.096
Agriculture				0.028	1.190	0.235				-0.038	-1.920	0.055
infrastructures				0.000	5.120	0.000				0.017	2.730	0.007
Buddhists							-0.006	-0.200	0.841	-0.038	-0.860	0.391
Christians							0.008	0.600	0.548	-0.215	-5.640	0.000
Jains							1.746	5.470	0.000	0.614	1.710	0.089
Muslims							-0.055	-4.170	0.000	-0.077	-5.040	0.000
Sikhs							0.297	16.17	0.000	0.283	13.96	0.000
Dalits							-0.002	-0.070	0.943	-0.070	-2.120	0.035
tribes							-0.078	-6.720	0.000	-0.075	-5.170	0.000
constant	0.976	10.24	0.000	0.953	17.12	0.000	1.085	157.2	0.000	0.931	9.340	0.000
observations	*591*			*520*			*591*			*520*		
R^2	***0.062***			***0.125***			***0.465***			***0.546***		

Notes: Coef = coefficient; Obs= observations; Prob= Probability(*t*=0)

All models are least-square models

See Table 2 for definition of the variables

Table 4: Spatial regressions of child sex ratio variations

Variable	model E distance to Amritsar			model F synthesis + distance			model G synthesis 2			model H with spatial error		
	Coef	***t***	**Prob**	**Coef**	***t***	**Prob**	**Coef**	***t***	**Prob**	**Coef**	***t***	**Prob**
Density (log)				0.002	0.800	0.425	-0.003	-1.229	0.220	0.000	-0.052	0.958
Fertility				0.001	0.350	0.729	0.017	5.510	0.000	0.001	0.203	0.839
Child survival				0.103	1.310	0.191	0.036	0.446	0.656	0.030	0.384	0.701
Urbanisation				-0.008	-0.54	0.589	0.018	1.130	0.259	0.001	0.110	0.913
Literacy				0.002	0.070	0.943	0.114	3.852	0.000	0.065	2.353	0.019
Literacy (m/f)				-0.031	-2.33	0.020	0.008	0.615	0.539	0.015	1.342	0.180
Participation (m/f)				0.002	1.490	0.137	0.004	2.104	0.036	0.004	2.464	0.014
Agriculture				-0.033	-1.93	0.054	-0.020	-1.192	0.234	-0.011	-0.815	0.415
infrastructures				0.000	0.720	0.470						
Buddhists				-0.072	-1.90	0.058	-0.021	-0.706	0.481	-0.004	-0.165	0.869
Christians				-0.055	-1.60	0.111	-0.020	-1.343	0.180	-0.023	-1.532	0.126
Jains				0.349	1.150	0.252	1.112	3.290	0.001	0.761	2.454	0.014
Muslims				-0.102	-7.80	0.000	-0.065	-4.526	0.000	-0.062	-3.835	0.000
Sikhs				0.156	8.050	0.000	0.304	15.881	0.000	0.162	6.408	0.000
Dalits				-0.089	-3.20	0.001	-0.009	-0.300	0.765	-0.030	-1.167	0.243
tribes				-0.076	-6.24	0.000	-0.063	-4.698	0.000	-0.024	-2.018	0.044
distance to Amritsar	-0.006	-19.11	0.00	-0.006	-14.2	0.000						
spatial autocorrélation (λ)										0.839	35.474	0.000
constant	1.153	262.34	0.000	1.127	13.190	0.000	0.932	10.086	0.000	0.986	11.747	0.000
observations	*591*			*520*			*591*			*591*		
R^2	***0.383***			***0.676***			***0.520***			***0.809***		

Notes: Coef = coefficient; Obs= observations; Prob= Probability(*t*=0)

All models are least-square models except Model H (maximum likelihood).

See Table 2 for definition of the variables

Figures

Figure 1: Juvenile sex ratio, Indian districts, 2001

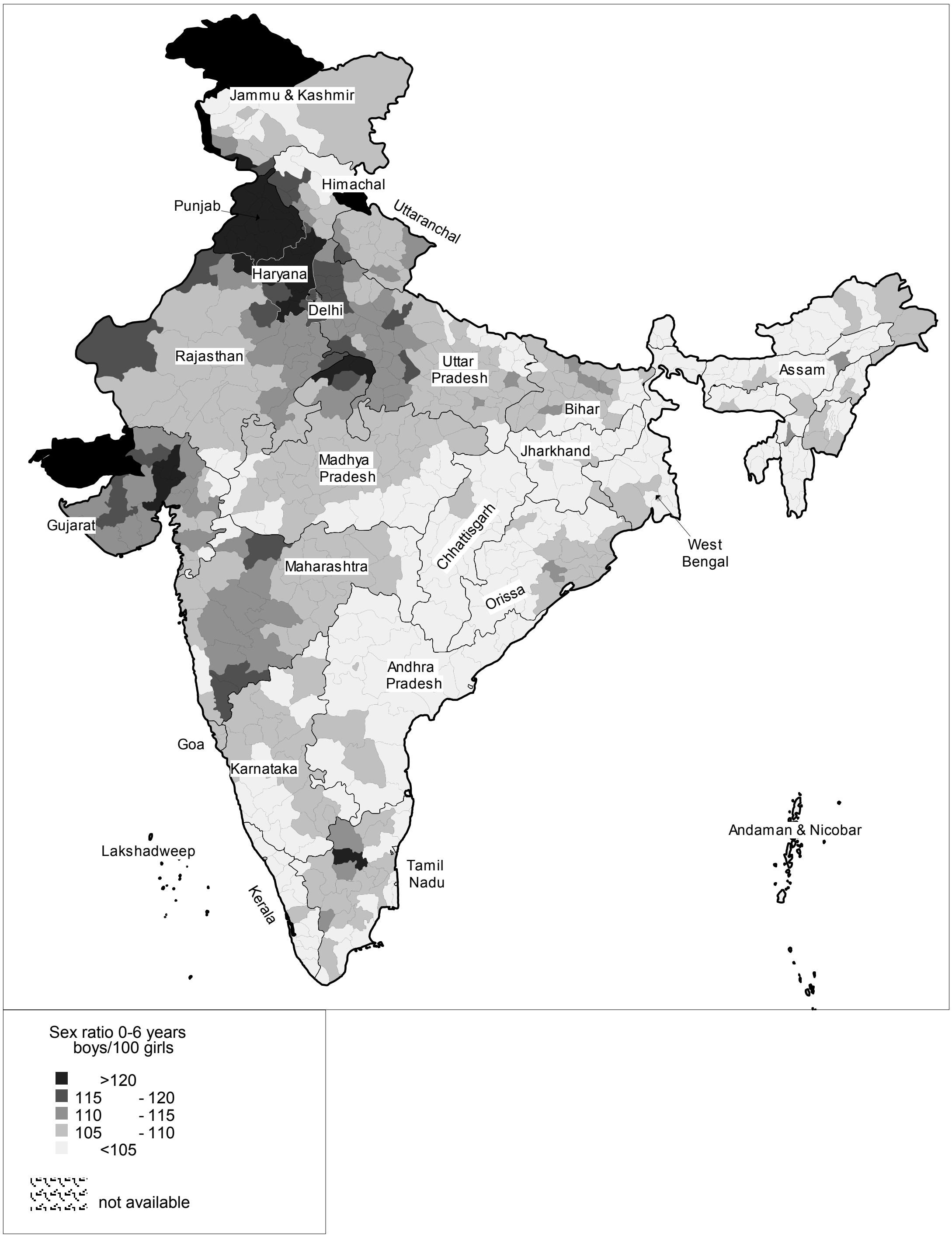

Figure 2: Residuals of regression model G