11

AN ANALYSIS OF THE DETERMINANTS OF FERTILITY BEHAVIOUR IN SOUTH INDIA AT THE VILLAGE LEVEL[1]

MANISHA CHAKRABARTY AND CHRISTOPHE Z. GUILMOTO

INTRODUCTION

Regional variation patterns in the demographic transition of fertility decline and association of the processes with demographic, socioeconomic, cultural, and diffusion process of knowledge have received much attention in the demographic research of recent years.[2] These variations are related to the specific regional profiles of fertility decline that may vary in their timing and intensity across regions. Though some of the differences may be attributable to social and economic factors, there remains an important residual geographical component in these fertility variations.

India, as a country of striking demographic diversity, with enormous variations in the conditions and mechanisms of fertility transition, offers

[1] This chapter is part of the South India Fertility Project. Thanks are due to the French Institute of Pondicherry and Welcome Trust, UK, for support. The authors are grateful to Professor P.V. Srinivasan, IGIDR, Mumbai and Samarjit Das, Reserve Bank of India, Mumbai for helpful discussions and suggestions, as well as to Sattia Vingadassamy for research assistance.

[2] See Friedlander et al. (1991) and Watkins (1991) for a discussion on the determinants of fertility decline in Europe.

a rich ground for these analyses.[3] The decline in fertility in India is perceived as part of the innovation phenomena that is taking place according to diffusion mechanisms. In addition to the influence of several socio-economic factors reflecting structural changes, the decline in fertility also seems to proceed by contagion for which any form of proximity, spatial or social, is an important factor of diffusion mechanism. In fact, the strong heterogeneity of the Indian environment will generate a new form of social differentiation due to the varied manner of the diffusion process. Recent research on fertility differentials has shown that these interregional variations in fertility transition have increased over the last 40 years.[4] In south India, particularly, demographic transition has entered its last phase which is in sharp contrast with other regions in India, and this part of the country provides an ideal field for the study of determinants of demographic outcomes.[5] During the last 15 years, whereas the tempo of fertility decline remained moderate elsewhere in India, south India has registered the most spectacular decline in fertility rate as demonstrated in other chapters of this book. However, in south India itself, we are confronted with an extremely contrasted demographic landscape. A great variety is observed not only between Kerala and other states, but also between administrative blocks within a single state. This chapter is an attempt to document the details of fertility differentials and their determinants at the lowest possible scale.[6]

In our attempt to study regional patterns of fertility rate and its socio-economic determinants, we will carry out different regression analyses of fertility rates, estimated at various levels of aggregation. The data used is the census (1991) village-level data for south India. These data provide the geographically most detailed indicators of fertility ever given by Indian censuses.[7] The study of detailed village-level data is the first of its kind in India. We also broaden our analysis in several other directions. In modelling

[3] For analyses on India, see Dyson and Moore (1983), Srinivasan (1995), Malhotra et al. (1995), Murthi et al. (1995).

[4] See Guilmoto and Irudaya Rajan (2001), Guilmoto (2000).

[5] On south India, see Bhat and Irudaya Rajan (1990), Kulkarni et al. (1995), Rayappa and Prabhakara (1996), Guilmoto and Irudaya Rajan (1998).

[6] For state-level analysis see Jain (1985), Bourne and Walker (1991), Reddy and Selvaraju (1993). See the district-level analyses of fertility differentials in Rosenzweig and Schultz (1982), Gulati (1992), Kishore (1993), Malhotra et al. (1995) and Murthi et al.(1995).

[7] NFHS and SRS data are available only at the state level. Recently Bhat and Zavier (1999) carried out an analysis on NFHS data for subregions, which is a minor improvement. The two previous censuses in 1981 and 1991 also gave details for districts.

the fertility differentials, we include several supply-side infrastructure variables in order to stress on the role played by the propagation of new ideas easily achieved through this infrastructure facility. Further, we use cartographic information of each village to form spatial clusters to emphasize the role of spatial contiguity in the fertility transition, which is not expressed clearly through administrative boundaries. Finally, to examine the stage of demographic transition of these villages in south India, we use the intra-class correlation co-efficient as a statistical indicator of space-specific patterns of fertility.

This chapter is organized as follows: section 2 presents the data used in this study, section 3 explains the methodology adopted, section 4 describes the results from our various analyses, and, finally, section 5 draws the conclusions.

Data

The 1991 census data for rural India cover 70,259 inhabited villages. They include village-directory (VD) data, comprising infrastructure facility as well as primary census abstract (PCA) data, i.e., the socio-demographic characteristics of villages. This unique dataset has been prepared during the course of the SIFP.[8] It is important to stress at the outset that the number of villages is smaller than the official number for south India (71,121 inhabited villages as per the census), as many villages with inaccurate or deficient data have been eliminated from our database. Moreover, the number of villages used in our analysis may vary, as some ratios are not computable when the reference population is empty.

The VD data provides details of education, medical, drinking water, market, communication and power supply facilities, location of villages with reference to the nearest town, etc. The VD dataset consists mainly of qualitative information (dichotomous variables indicating the presence of infrastructure in each individual census village). The PCA data describes social and demographic factors pertaining to total village population, male-female population size, male-female population of age group below seven,

[8] The availability of both series on floppy discs from the Registrar General has allowed us to build a global database of south Indian villages by merging socio-demographic and infrastructure data into a unique dataset.

tribal population, literate population and population in different working professions.

South India is composed of four states (Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu) and one union territory, Pondicherry. In this chapter, the 260 villages from Pondicherry, which belong to the Tamil-speaking districts of Karaikal and Pondicherry, have merged with Tamil Nadu villages into a single state unit. These units are divided into districts and further subdivided into *taluks* (the name of this unit may vary in different states). Individual data refer to census villages, the lowest unit of our analysis. While districts across India are of comparable size, it is worth stressing that the administrative subdivision into *taluks* and villages is extremely heterogeneous in south India. Thus, while villages in Kerala generally refer to large administrative units that cover scattered settlements, with an average population above 15,000 inhabitants, villages in other south Indian subregions may pertain to units with populations below 500. Similarly, *taluks* in Andhra Pradesh refer to *Mandals* (new administrative divisions introduced during the 1980s), which are on average, 10 times smaller than *taluks* in the rest of south India. The data are described in Table 11.1, which gives an idea of number of villages, taluks and districts in each state.

Table 11.1: Preliminary Data Description for Inhabited Villages

State	Number of Villages	Number of *Taluks*	Number of Districts	Average Population per Village	Average Number of Villages per *Taluk*	Average Number of Villages per District
Andhra Pradesh	25,759	1,097	22	1,872.2	23.5	1,170.9
Karnataka	27,066	175	20	1,147.9	154.7	1,353.3
Kerala	1,384	61	14	15,475.6	22.7	98.9
Tamil Nadu and Pondicherry	16,050	178	22	2,307.9	90.2	729.5
South India	70,259	1,511	78	1,960.7	46.5	900.8

Note: Villages with deficient data have been omitted.
Source: Census of India data files.

To study fertility variation, we have computed the CWR from census PCA data, which is equal to the ratio of population under seven to women above seven. This CWR is used as an indirect indicator of fertility rate.[9] The main drawback of this indicator is the differential impact of child

[9] This definition of CWR implicitly assumes that the number of females aged 7–14 years or more than 50 years (i.e. women who don't belong to the childbearing age group)

mortality on the population below seven years: mortality variations between localities will affect the proportion of surviving children during the census. However, child mortality is low in south India and variations are moderate. Using mortality rates before three years as a proxy for below seven mortality, the computed value in south Indian districts is 61.3 per 1000, with a standard deviation of 16.1 per 1000.[10] The impact of this standard deviation of mortality on surviving population (and hence on CWR) would therefore be inferior to 2 per cent. This suggests that mortality variations may have a somewhat limited impact on fertility estimates derived from the CWR. However, more extreme levels of random variation characterize CWRs computed for villages with population below 500 and are a more a more serious cause of concern. For that reason, our statistical analysis will be weighted by the total population size of the village in order to reduce the random variation factor encountered in the smaller demographic units, and also to control for heteroscedasticity.

Each census unit is described by 135 or more variables. To reduce the dimension of the data matrix we have limited our analysis to the variables that affect the CWR most clearly. We also performed preliminary factor analyses to summarize the impact of the numerous individual infrastructure variables.[11] For instance, we used factor analysis to design five comprehensive variables pertaining respectively to education infrastructure, health infrastructure, communication networks, distance to infrastructures and postal infrastructure. These included five factor variables that are not highly correlated with each other and we have also checked for multicollinearity. Table 11.2 describes these five composite indicators of infrastructure. Factor analysis has allowed us to reduce the number of infrastructure variables from 37 indicator variables to five composite indicators.[12]

is a constant multiple of women aged 15–49. This assumption may not in general be true. However, the statistical property of the regression parameter estimate still remains the same given the assumption that errors in the definition of the dependant variable follow a white noise process and regressors are measured without error.

[10] Values are computed from the mortality estimates prepared by Irudaya Rajan and Mohanachandran (1998). A similar computation for all Indian districts yields a standard variation of mortality of 36.9 per 1000, twice larger than the corresponding value for south India.

[11] Principal components is a technique for summarizing the information contained in a large set of variables to a smaller set of indices. For a recent application to construct a household asset index, see Filmer and Prichett (1999).

[12] A few census variables on infrastructure facilities have been excluded from the factor analysis as they were found to be of poor quality.

Table 11.2: Variables Employed to Construct Infrastructure Indicators by Factor Analysis

Factor	Variable Name	Original Census Variables Used in the Factor Analysis
Education	Feduc	Primary school, middle school, high school, pre-university college, graduate college, adult literacy centre, industrial school, training school, other educational institutions.
Health	Fhealth	Hospital, maternal and child welfare centre, maternity homes, child welfare centre, primary health centre, health centre, primary health sub-centre, dispensary, family planning centre, tuberculosis clinic, nursing home, community health workers, private practitioner, subsidiary medical practitioner, other medical centres.
Range	Frange	Distances to nearest educational institution, medical institution, post facilities, communication facilities.
Post	Fpost	Post office, telegraph office, telephone facilities.
Communications	Fcomm	Bus stop, railway station, approachable by *pucca* road.

Note: Original infrastructure variables are dichotomous (or ordinal) at the village level, but in percentages in other aggregated analyses.

After preliminary modelling, we have decided to retain only 20 explanatory variables. These can be grouped into several (somewhat arbitrary) categories: socio-demographic (*Dalit*, tribals, literacy); occupational (participation rates, different working professions); agricultural development (irrigation, share of cultivators); infrastructure development (education, health, distance to nearest infrastructure facility); settlement pattern (density, village size); and urban proximity (distance from the nearest town, postal and communication infrastructure). However, a specific village-level model will allow us to decompose further the effect of some of these factors and identify the specific impact on fertility of some other individual variables (see Table 11.7).

Moreover, three state dummies have also been introduced to account for persistent regional variations. Location dummy variables, Karnataka, Kerala and Tamil Nadu, indicate the corresponding state. We use Andhra Pradesh as the reference group, since it has the largest state population.

The choice of determinants is partly guided by the preliminary analysis, but it also reflects our limitation of available information. Most variables have a significant impact on fertility in some of our models, but we also retained important village characteristics (such as the health infrastructure) that seem to surprisingly play no role on demographic behaviour. Though some relevant variables like religious practices, marriage schedule and family structure are missing from our database, it provides a good description of local facilities and some relevant socio-demographic characteristics.

In the analysis of fertility variation, this dataset also allows us to distinguish between demand and supply factors. Demand factors are related to the attitude of women and more generally of couples, towards reproduction and are often determined by the socioeconomic characteristics of the population (literacy, female role, economic activity, etc.). Supply factors in the strict sense correspond to the availability of family planning services, but here we can expand this definition to all infrastructure variables, like schooling facilities, transportation and communication.[13] All these factors are supposed to decrease fertility as they facilitate access to new services, new ideas and attitudes towards family norms in developed urban centres. The spread of new attitudes is probably a prerequisite for a sizeable fertility decline as people would be unlikely to reduce their offspring in the absence of a favourable ideological environment. It is, however, almost impossible to distinguish between the mere effect of material infrastructure and the indirect impact they have through promotion in villages of new value systems based on the benefits of education, improved access to city centres, etc. Table 11.3 gives a description of all variables included in the regression analysis.

Table 11.3: Description of Variables

Variable Name	Description
cwr	Population below 7/women above 7
littot	Literates/Population above 7
srlit	Male literates/Female literates
srtot	Male population/Female population
dalit	Dalits/total population
tribe	Tribals/total population
wtot	Workers/population above 7
srworker	Male workers/Female workers
wagri	Workers in agricultural sector/total workers
wind	Workers in industrial sector/total workers
wserv	Workers in service sector/total workers
irrig	Irrigated land/cultivated land
cultlab	Share of cultivators among workers in agricultural sector
dist	Distance from nearest town (ordinal)
feduc, frange, fhealth, fpost, fcomm	Factors described in the previous table.
dens	Population density
lpop	Individual or average village population (logarithm)

Note: Values derived from original village data from the census.

[13] As increasing village size and rural density tend to increase infrastructure availability, we have systematically included these two variables (*lpop*, *density*) in our model.

Methodology

As the CWR derives from two segments (female population > age seven and children < age seven) that rarely account for more than a third of the total population, this ratio is sensitive to the small size of the sample unit. In villages with population less than 500 inhabitants, CWR values are often highly unstable. Similarly, many other variables such as infrastructure dummies or computed percentages are unreliable when calculated from small sized units. This is the reason why the regression models used are systematically weighed using the total unit population; villages (or clusters) with smaller population are therefore less likely to disturb regression results.

However, the quality of variables may still be poor when computed on small demographic units as the lower correlation co-efficients are obtained in village-level models. It was felt that some aggregation procedure might help to increase the average unit population and reduce the disparities across regions in terms of administrative units like villages and the only administrative layer between villages and districts, viz., *taluks*. In order to get rid of the extreme heterogeneity of villages and administrative subdivisions like *taluks* in terms of population size and also to detect any geographical pattern in fertility and its socioeconomic determinants, we have performed different procedures of spatial clustering using a GIS. The basic idea is to group villages by contiguity measured in Euclidean distances between villages, instead of relying on the administrative grid based on *taluks* or districts. The unobserved cultural and other location variables affecting fertility transition can be better captured in clusters based on spatial contiguity than in heterogeneous units such as villages or *taluks*.

Moreover, clustering also serves to remove the statistical noise associated with small population statistical units.

To form the spatial clusters, we used the spatial coordinates of the villages that were also collected by the SIFP. We then use a geo-statistical technique to group together adjacent villages. The procedure is described below briefly:[14] the village map of south India has first been divided into clusters of various sizes. Each clustering is based on the spatial aggregation of contiguous villages. There are four levels of aggregation used, using a radius of 2 km, 5 km, 10 km and 20 km. These four cluster aggregations will be respectively referred to as Cl02, Cl05, Cl10 and Cl20. For example, we can aggregate the 70,259 villages in south India into 2,127 clusters (Cl10) by regrouping villages that are within 10 km of each other. From this procedure we then derive a map of clusters consisting in 2,127 Thiessen polygons that completely cover south India.[15] The data of all villages that lie within each polygon are finally aggregated. Population figures are summed and new variables such as literacy rates or SRs are computed from the total figures. Dichotomous variables such as infrastructure facilities are averaged over each cluster and the resulting mean value is used as a new variable. A new set of variables is derived for these clusters and a new factor analysis is also performed. The result is a set of seven samples, based on different clustering levels from original village data (no clustering) to district totals. We will use these samples to model fertility behaviour.

In Table 11.4, we present the results of our aggregation procedures, along with *taluk* and district aggregations. While the number of units decreases regularly as the radius of aggregation increases, the average demographic size of clusters increases. For instance, the average population size of the 23,290 2-km clusters (Cl02) is now of 5,915 inhabitants, as against 1,961 inhabitants for the original census villages. Table 11.4 also displays the co-efficients of variation of the population size after aggregation for each level of spatial clustering and for administrative levels, such as *taluks* and districts. As can be seen, this co-efficient tends to decrease regularly as the radius of aggregation increases: the larger the spatial

[14] The procedure is described in greater detail in Sébastien Oliveau's paper in this volume.

[15] Thiessen polygons (also known as Voronoi polygons) are generated from a set of aggregated villages. They offer a more spatially efficient procedure to divide a surface than other grids, like the usual rectangular grid. On the Thiessen polygon, see Okabe et al. (1994). For a detailed discussion of clustering techniques, see Murray and Estivill-Castro (1998).

aggregation, the more homogeneous the demographic units, i.e., the aggregated units. We however, notice that the administrative *taluk* clustering does not reduce very much variations in village size: the *Mandal* administrative division of Andhra Pradesh is in a large part responsible for the observed disequilibrium in *taluk* size.

Table 11.4: Size and Homogeneity for the Aggregated Cluster, *Taluk* and District Level of South India

Measures	Villages	Cl02	Cl05	Cl10	Cl20	*Taluks*	Districts
Sample size	70,259	23,290	6,841	2,127	618	1,511	78
Average population per unit	1,961	5,915	20,137	64,766	222,907	91,169	1,766,106
Co-efficient of variation	1.55	1.09	0.90	0.84	0.79	1.12	0.44
ICC	Not relevant	0.555	0.525	0.510	0.479	0.474	0.245

In order to detect the homogeneity of total population size and other characteristics determining fertility rate for adjacent units within spatial clusters and for units within administrative blocks, we also propose to compute the intra-class correlation (ICC) as a measure of homogeneity across villages in each cluster and block. This will also help to detect local, closed-distance environment effect on these population characteristics when measured at the spatial-cluster level which is constructed on the basis of Euclidean distance.

ICC refers to within class correlation. This is same as the product moment correlation, but the only difference is that both the variables measure the same characteristics. This correlation co-efficient measures how the members of a family or group are correlated among themselves with respect to some of their common characteristics. In our formulation we use this measure to detect whether any similarity/homogeneity, with regard to total population, CWR and other socioeconomic variables affecting fertility across villages within the administrative boundaries like *taluk* or districts (regarded as class) and also within spatial clusters of four levels, exists or not. The formulation for this correlation coefficient is as follows:

$$\frac{\sum_i k_i^2 \left(\bar{x}_i - \bar{x}\right)^2 - \sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2}{\sum_i \sum_j \left(k_i - 1\right)\left(x_{ij} - \bar{x}\right)^2}$$

$i=1, 2,\ldots \ldots,n$ (number of *taluks* or spatial clusters)
$j=1, 2,\ldots \ldots \ldots k_i$
k_i = number of villages in i-th *taluk*/cluster
x_{ij} denotes the measurement on the j-th village in the i-th *taluk*/cluster. We will have $k_i(k_i-1)$ pairs for the i-th class like $(x_{ij}-x_{il}), j\neq l$. So there will be $\Sigma^n_{i=1}k_i(k_i-1) = N$ pairs for all the n *taluks* or clusters. In the bivariate table x_{i1} occurs (k_i-1) times, x_{i2} occurs (k_i-1) times, …., x_{iki} occurs (k_i-1) times, i.e., from the i-th family we have $(k_i-1)\Sigma_j x_{ij}$ and for all n groups (*taluks*/clusters here) we have $\Sigma_i(k_i-1)\Sigma x_{ij}$ as the marginal frequency. The overall mean is:

$$\bar{x} = \frac{1}{N}\left[\sum_i (k_i-1)\sum_j x_{ij}\right]$$

$$N = \sum_i k_i(k_i-1)$$

and within class (i.e., within taluk or district or spatial cluster) mean:

$$\bar{x}i = \frac{\sum_j x_{ij}}{k_i}$$

This intra-class co-rrelation co-efficient is the product moment correlation but it is not symmetric around zero.

Results

As indicated in Table 11.4, the first clustering using a 2-km distance resulted in a sizeable decrease in the number of units, from 70,259 villages' original inhabited villages to 23,290, 2-km clusters. Only adjacent villages in dense areas are merged together, especially in the Mysore plateau or in the Kaveri Delta. In other regions, where human settlement is sparse and villages often distant from each other, this clustering hardly has any impact and as a result, some 2-km clusters in the Western Ghats and in other areas still have less than a 100 inhabitants. On the opposite, the last 20-km clustering performed grouped together villages that are very distant and cause sample size to decrease to 618 units. The total number of *taluks* for the whole of south India is 1,511 and the total number of districts is 78.

We know that districts of south India are of comparable size and this can be further strengthened by the lowest value of co-efficient of variation of total population size at the district level. For *taluks*, the co-efficient of variation is almost as high as that across villages, thereby representing the fact that *taluks* are not at all of comparable size. Also, from the last row of the table in which ICC is presented, we observe the quite plausible fact that only nearby villages are homogeneous with regard to population size. The ICC values, which measure the degree of homogeneity across villages within these spatial clusters or administrative grids, are higher for each cluster as compared to administrative regions like *taluks* and districts. The ICC value for a *taluk* is almost as low as the value for the villages that are 20 km apart from each other. So to get rid of the heterogeneity problem, spatial clustering does bear a meaningful smoothing mechanism as compared to administrative blocks like *taluk* or district-level aggregation because of the homogeneous pattern across units, i.e., villages within these clusters in terms of village characteristics, for example, population size. The ICC for villages is not relevant because they are the lowest available level units and we *cannot* get any further classification within the village.

Table 11.5 gives descriptive statistics of the dependent variable and of the different explanatory variables used in this chapter for south India and each separate state. From this table, we can see that average level of CWR is lowest in Kerala as expected, while it is higher in Andhra Pradesh and Karnataka than for the whole of south India. The variation in the CWR is also lowest in Kerala. From Table 11.5, we can observe that in Kerala, the SR (males per females) is lower than one and AFR level is the highest (lowest value of *srlit*) while the FWP is the lowest (as expressed through highest value of *srworker*). Only in Andhra Pradesh we can see that average level of female literacy is lower than the whole south India average and so the average FWP in this state is higher than the corresponding south Indian average. This negative correlation between AFR and FWP is quite likely given the fact that a major share of workforce are working as cultivators and agricultural labourers, which occupational categories require low levels of human capital. In Kerala also, we can see that more people are involved in the industry and service profession as compared to other states. Kerala also has a higher level of infrastructure facility than all other states.

Table 11.5: Descriptive Statistics of all Variables

	South India			
Variable	Mean	Std. Dev.	Min	Max
cwr	0.41	0.16	0.00	22.00
littot	0.42	0.19	0.00	1.04
srlit	2.51	2.51	0.00	192.00
srtotal	1.03	0.34	0.00	74.00
dalit	0.20	0.18	0.00	1.00
tribe	0.10	0.25	0.00	1.00
wtot	0.55	0.13	0.00	1.29
srworker	4.31	12.58	0.00	659.00
wagric	0.83	0.18	0.00	1.44
windust	0.05	0.09	0.00	1.00
wservice	0.08	0.08	0.00	1.00
irrig	0.31	0.31	0.00	1.00
cultalab	0.53	0.26	0.00	1.00
feduc	–0.88	0.76	–1.36	14.45
fhealth	–0.67	0.61	–1.04	20.95
frange	1.24	2.10	–0.87	14.66
dist_fr_	19.88	18.03	0.00	535.00
fpost	–0.79	1.06	–1.56	3.22
fcomm	–0.63	1.79	–3.42	4.37
dens	4.29	74.64	0.00	12,258.33
lpop	6.88	1.28	0.00	11.17

	Andhra Pradesh		Karnataka		Kerala		Tamil Nadu and Pondicherry	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
cwr	0.44	0.19	0.42	0.14	0.30	0.07	0.34	0.09
littot	0.29	0.15	0.46	0.17	0.88	0.07	0.52	0.14
srlit	3.14	3.03	2.40	2.57	1.06	0.12	1.89	0.93
srtotal	1.03	0.48	1.04	0.28	0.97	0.05	1.02	0.10
dalit	0.17	0.15	0.19	0.19	0.11	0.07	0.26	0.21
tribe	0.20	0.35	0.06	0.14	0.02	0.06	0.03	0.13
wtot	0.61	0.13	0.51	0.12	0.35	0.07	0.54	0.12
srworker	2.66	7.41	3.80	17.61	6.29	2.08	3.75	8.78
wagric	0.85	0.14	0.83	0.19	0.46	0.18	0.82	0.18
windust	0.05	0.08	0.04	0.08	0.13	0.09	0.06	0.10
wservice	0.07	0.07	0.06	0.07	0.28	0.10	0.09	0.09
irrig	0.33	0.32	0.19	0.23	0.32	0.31	0.50	0.32
cultalab	0.48	0.25	0.63	0.25	0.33	0.13	0.44	0.25
feduc	–0.94	0.58	–0.97	0.56	1.66	2.15	–0.87	0.69
fhealth	–0.68	0.45	–0.80	0.32	1.67	1.99	–0.65	0.51
frange	1.45	2.58	1.36	1.60	–0.78	0.44	0.88	1.93
dist_fr_	27.30	23.59	15.67	9.44	17.98	12.41	15.24	15.04
fpost	–0.75	0.92	–0.88	1.24	0.68	1.08	–0.83	0.78
fcomm	–0.99	1.86	–0.57	1.82	–0.24	0.26	–0.21	1.59
dens	4.72	62.34	3.43	103.53	9.77	6.54	4.57	8.01
lpop	6.92	1.26	6.47	1.22	9.47	0.68	7.28	1.04

Determinants of Fertility at Various Levels of Analysis

All the regression results presented below are based on population-weighted least square regression in order to get rid of the heteroscedasticity problem. The tables show standardized (beta) co-efficients of regression in order to allow the comparison of the results across variables and models. We will first discuss the overall results using our standard model for south India displayed in Table 11.6. Several aggregation levels are analyzed using the same regressors. There are two other sets of results that are shown further below (Tables 11.7 and 11.8) will allow us to decompose village-level results for individual states and variables.

Table 11.6 applies the standard model to different aggregation levels. The main results (column 2 of Table 11.6) are based on the entire original sample of 66,020 villages.[16] A large number of variables are shown to have a significant impact on village fertility levels, though regression co-efficients vary widely. As expected, literacy emerges as the single most important factor of low fertility. The standardized co-efficient of literacy on fertility is as low as –0.66, indicating that the increase of literacy rates translates in an almost similar decrease in fertility levels. Literacy, a standard measurement of social development, is complemented here by an index of gender bias captured by the SR of the literate population (males per females). The relative share of female literates further decreases fertility, though its impact is less visible at the village level. The total SR has a similar impact, with higher fertility levels associated with the higher masculinity of the overall village population.

Several other socio-demographic variables have also a significant impact on fertility. The share of tribals in the village population has a very distinct

[16] As indicated previously, many villages with a low population have been deleted from the sample, as it is impossible to calculate some of the ratio variables used in the analysis.

Table 11.6: Regression Results for South India (Various Levels of Aggregation)

Variable	Village	CI02	CI05	CI10	CL20	*Taluk*	District
littot	−0.660***	−0.670***	−0.552***	−0.444***	−0.212**	−0.371***	NS
srlit	0.054***	0.148***	0.293***	0.394***	0.503***	0.410***	0.690***
srtotal	0.048***	0.042***	NS	−0.032*	NS	NS	NS
dalit	0.037***	0.036***	0.035***	0.027*	NS	NS	NS
tribe	0.103***	0.104***	0.117***	0.124***	0.134***	0.145***	NS
wtot	−0.088***	−0.148***	−0.197***	−0.281***	−0.286***	−0.276***	NS
srworker	0.003NS	0.034***	0.104***	0.099***	0.115**	0.068**	NS
wagric	0.020***	0.042***	0.059***	0.108***	0.170***	0.137***	0.353*
windust	−0.011**	−0.013*	−0.023**	NS	NS	NS	NS
wservice	0.115***	0.099***	0.073***	NS	NS	0.084*	NS
irrig	−0.083***	−0.094***	−0.101***	−0.109***	−0.139***	−0.124***	−0.276**
cultalab	−0.110***	−0.153***	−0.208***	−0.249***	−0.290***	−0.249***	−0.372***
feduc	−0.089***	−0.102***	−0.172***	−0.262***	−0.385***	−0.335***	−0.614***
fhealth	0.014***	NS	NS	NS	NS	NS	NS
frange	0.048***	0.035***	0.046***	0.058**	NS	0.056*	NS
dist_fr	0.040***	0.044***	0.045***	0.054***	0.048*	0.056***	NS
fpost	−0.038***	−0.045***	−0.070***	−0.066**	NS	−0.074**	NS
fcomm	−0.015***	−0.015***	−0.023**	NS	NS	NS	−0.463*
dens	NS	NS	−0.034**	NS	NS	NS	NS
lpop	0.111***	0.162***	0.257***	0.277***	0.376***	0.313***	NS
Karnataka	0.253***	0.282***	0.305***	0.309***	0.321***	0.319***	0.307**
Kerala	0.104***	0.107***	0.055***	NS	NS	NS	NS
Tamil Nadu and Pondicherry	−0.124***	−0.118***	−0.123***	−0.135***	−0.151***	−0.149***	NS
Sample size	66,020	22,968	6,816	2,122	618	1,509	78
Adjusted R^2	0.4748	0.5569	0.6349	0.6787	0.7223	0.6853	0.7718

Notes: CWR is the dependent variable. All co-efficients are standardized. NS = non significant; * = 10% level ** = 5% level; *** = 1% level
Smaller ample after exclusion of villages with missing data or variables.

Table 11.7: Regression Models for South India (Village Level)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
			Social characteristics			
littot	−0.660***	−0.648***	−0.637***	−0.660***	−0.658***	−0.662***
srlit	0.054***	0.062***	0.055***	0.055***	0.054***	0.055***
srtotal	0.048***	0.045***	0.051***	0.049***	0.048***	0.048***
dalit	0.037***	0.009***	0.040***	0.038***	0.037***	0.038***
tribe	0.103***	0.113***	0.104***	0.104***	0.103***	0.099***
			workforce			
wtot	−0.088***	−0.179***	−0.087***	−0.088***	−0.089***	−0.090***
srworker	NS	NS	NS	NS	NS	NS
wagric	0.020***	−0.059***	–	0.018**	0.020***	0.021***
windust	−0.011**	−0.051***	–	−0.012***	−0.011**	−0.011**
wservice	0.115***	0.040***	–	0.115***	0.115***	0.117***
			agricultural development			
irrig	−0.083***	−0.155***	−0.083***	−0.083***	−0.082***	−0.083***
cultalab	−0.110***	−0.088***	−0.112***	−0.109***	−0.109***	−0.111***
			infrastructure factors			
feduc	−0.089***	−0.061***	−0.086***	−0.079***	−0.089***	−0.090***
fhealth	0.014***	0.026***	0.010**	–	0.016***	0.014***
			communication factors			
frange	0.048***	0.068***	0.047***	0.046***	0.068***	0.052***
dist_fr_	0.040***	0.023***	0.036***	0.041***	0.038***	0.041***
fpost	−0.038***	0.083***	−0.037***	−0.035***	−0.035***	−0.038***
fcomm	−0.015***	−0.015***	−0.015***	−0.015***	–	−0.014***
			settlement pattern			
dens	NS	NS	NS	NS	NS	NS
lpop	0.111***	−0.018***	0.110***	0.118***	0.113***	0.114***
			occupational details			
agricultural labour	–	–	NS	–	–	–
household industry	–	–	−0.034***	–	–	–
other industry	–	–	−0.036***	–	–	–
construction	–	–	0.027***	–	–	–
trade and commerce	–	–	0.120***	–	–	–
transportation, storage	–	–	NS	–	–	–
other services	–	–	−0.021***	–	–	–

(Table 11.7 contd)

(Table 11.7 contd)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
			health infrastructure			
primary health centre	–	–	–	−0.026***	–	–
maternity and child welfare centre	–	–	–	NS	–	–
maternity homes	–	–	–	−0.010***	–	–
child welfare centre	–	–	–	NS	–	–
family planning centre	–	–	0.008**	–	–	–
			communication			
bus facilities	–	–	–	–	0.036***	–
approachable by pucca road	–	–	–	−0.037***	–	–
drinking water facilities	–	–	–	–	–	NS
power supply	–	–	–	–	–	NS
			state dummies			
Karnataka	0.253***	–	0.238***	0.254***	0.251***	0.254***
Kerala	0.104***	–	0.086***	0.109***	0.106***	0.103***
Tamil Nadu & Pondicherry	−0.124***	–	−0.129***	−0.122***	−0.125***	−0.125***
Adjusted R^2	0.4748	0.4014	0.4796	0.4754	0.4759	0.4751

Notes: CWR is the dependent variable. All co-efficients are standardized.
NS = not significant * = 10% level ** = 5% level and
*** = 1% level significant.

Table 11.8: Regression Results for Five Different States (Village Level)

	South India	Andhra Pradesh	Karnataka	Kerala	Tamil Nadu & Pondicherry
littot	−0.660***	−0.456***	−0.502***	−0.587***	−0.283***
srlit	0.054***	0.054***	0.039***	-0.198***	0.112***
srtotal	0.048***	0.145***	0.054***	0.113***	NS
dalit	0.037***	0.020***	0.041***	NS	0.109***
tribe	0.103***	0.196***	0.047***	−0.039*	0.041***
wtot	−0.088***	−0.014*	0.123***	−0.365***	−0.247***
srworker	NS	0.034***	0.027***	0.325***	−0.020**
wagric	0.020***	−0.201***	0.073***	0.093***	−0.109***
windust	−0.011**	−0.128***	0.050***	0.041*	−0.092***
wservice	0.115***	−0.034***	0.155***	−0.175***	−0.069***
irrig	−0.083***	−0.114***	−0.059***	−0.035*	−0.058***
cultalab	−0.110***	−0.109***	−0.160***	−0.110***	0.037***
feduc	−0.089***	−0.082***	−0.019**	−0.065***	−0.064***
fhealth	0.014***	NS	−0.042***	0.089***	NS
frange	0.048***	0.043***	0.063***	0.042**	0.034***
dist	0.040***	0.052***	−0.162***	−0.334***	−0.066***
fpost	−0.038***	NS	−0.055***	NS	−0.087***
fcomm	−0.015***	0.015**	−0.037***	NS	−0.058***
dens	NS	NS	−0.038***	NS	0.093***
lpop	0.111***	0.069***	0.162***	0.130***	NS
sample size	66,020	23,359	25,657	1,383	15,621
Adjusted R^2	0.4748	0.3447	0.4295	0.6479	0.2330

Notes: CWR is the dependent variable. All coefficients are standardized.
NS = non significant * = 10% level ** = 5% level *** = 1% level
South India model includes state dummies.

impact on high fertility, even after controlling for other economic or infrastructural factors. To some extent, the tribal population constitutes the most marginalized section of the population and it is socially, economically as well as spatially secluded. The onset of fertility transition among tribals occurred later and the gap is still pronounced. This is also true for Dalits, although the relationship is less pronounced than for tribals. As a matter of fact, the Dalit population is less concentrated than the tribal population and in spite of its distinctive social and economic features, it has a closer relationship with the rest of the rural population.

Many economic variables play a somewhat moderate role in fertility variations. For instance, the masculinity of the workforce is not significant at the village level. Among broad occupational categories, it is the service sector that has a somewhat unexpected positive effect on the CWR, while the agricultural sector has a positive effect. The positive sign attached with service indicator (*wservice*) needs to be further explored.

However, village data offer precious information on agricultural development by providing the share of irrigated land. Here, we observe that irrigation causes fertility to decrease significantly. Moreover, the composition of the agricultural workforce plays a much more important role than its relative share in the entire workforce. Villages where cultivators predominate over agricultural labourers are characterized by lower fertility levels. To some extent, this variable (*cultalab*) captures the intensity of land inequality as an increased share of cultivators in the agricultural workforce is usually associated with a more equal distribution of landed property. Taken together, these two indicators (irrigation and share of cultivators) reflect the level of rural development in villages where agriculture is still the largest sector. They correspond respectively to the productivity and equality dimensions of agricultural development, which are both shown to have a depressing impact on fertility behaviour.

It is worth stressing that previous studies on the fertility-landholding relationship in rural India have often shown mixed results. Whereas the relationship used to be invariably positive in the past (with higher fertility observed among big landholders), a negative correlation has been found in some of the most recent studies such as the study by Desai and Soumya (1998) based on the NFHS data.[17] Our analysis demonstrates that fertility is actually much more responsive to irrigation and share of cultivators (vs agricultural labourers) than to landholding size and that the correlation is strongly negative. As our analysis is based on village-level data, we cannot establish with certainty that the same relationship holds also for households and that for instance, lower fertility is found among families enjoying access to irrigated land. However, as an index of individual or collective agricultural development, irrigation and share of cultivators prove to exert an undeniable negative impact on estimated fertility levels in south India.

Among the infrastructure variables, education development positively reduces the birth rate even after controlling for literacy rates. This suggests that educational amenities, apart from their direct impact on literacy levels, are more developed in low-fertility localities. It may well be also that the association between schools and low fertility works both ways, but the exact nature and direction of the causation process cannot be investigated from our data. However, the health infrastructure is shown to have a positive albeit limited impact on fertility. This result remains disturbing when taking into consideration the important family planning component of most health facilities in rural India. This puzzle will be further explored below, using some of the original variables used to construct our health infrastructure index. The standard model does however confirm the negative impact of accessibility and communication on fertility behaviour. The overall 'range factor' which captures the distance to basic amenities (from education institutions to communication facilities) and the distance to the nearest urban centre increase local fertility levels.

[17] For a very useful review article on landholding and fertility, see James (2000).

There are two other factors (*Fpost*, *Fcomm*) that display a similar negative impact on village-level fertility. This points to the possible role of the diffusion of new ideas and attitudes, related to the distance to urban areas or to the access to social infrastructures.

Table 11.6 also provides the results of the same model applied to other demographic units. As can be seen from this table, spatial or administrative aggregation causes the overall quality of the model to increase steadily. Spatial clustering is especially useful as it substitutes larger demographic units to original village data that may suffer from insufficient population size. As we increase the level of smoothness, the noise in the data is reduced and so the total variation in the CWR results in an improvement in the adjusted R^2. However, the significance level of many regressors tends to decrease as the sample size reduces.

An interesting fact is that while state dummies are all highly significant at the village level, this is no longer true as we increase the smoothness of our data by aggregating them. At the district level, no state dummy variable except for Karnataka turns out to be significant. This indicates the fact that state differentiation does not have much impact on the CWR variation at the aggregated level and that all variations in CWR after smoothing are now captured mostly by socio-demographic and agricultural development variables.

Literacy indicators exert a significant negative effect on fertility variation at all aggregation levels. However, as we move towards a more aggregated level, the impact of literacy rates reduces fast while that of the literacy SR increases. We can also observe that the effect of a few variables is more pronounced at the micro level than after several aggregations. This is the case for the *Dalit, srtotal* and *Service* variables. On the contrary, the impact of several other variables increases at larger aggregation levels. These variations suggest that the statistical correlations observed on one scale may not be the same on a different scale.[18] For example, the share of cultivators among the agricultural workforce (a proxy for land equality) appears to have to exert a stronger influence at a higher aggregation level. This means that local agrarian disparities matter less to fertility levels than the overall level of inequality measured at an aggregated

[18] This relates to the so-called 'modifiable areal unit problem' (MAUP): geographers have often observed that correlation patterns may vary for zones of different sizes, even when the area covered remains the same. In our case, the scale variation—when data are progressively aggregated into fewer and larger zones of analysis—constitutes the main source of the variation in correlation co-efficients. See for example Fotheringam et al. (2000: 237–39).

level. Interestingly enough, the same holds true for education infrastructures: the presence of schools in the village or its immediate vicinity have a lesser negative impact on fertility than a regional environment characterized by a high density of education infrastructure.

The implications of these findings are important to understand the nature of the causation processes captured by regression analyses. There appears to be important variations between local- (village) level correlations and those observed on a larger scale. Features that are more strongly associated with low fertility at the aggregated level than at the village level correspond probably to factors that exert a much more diffuse effect on reproductive norms and behaviour. The example of literacy is an interesting case in point: local fertility levels are firstly related to the proportion of literates. However, after clustering, this association reduces, while the impact of literacy SR (a proxy of gender inequality) becomes the major fertility determinant at the regional level.

The Role of Individual Variables in Determining Fertility Levels

Table 11.7 brings together regression results corresponding to the entire sample of south India with 66,020 villages. The standard model (Model 1) is compared to other models based on sets of variables. Model 1 incorporates the most significant variables that we have chosen to retain in other analyses (Tables 11.6 and 11.8). Model 2 is simply Model 1 without state dummies. The results from Model 2 indicate that the census variables fail to account for a significant share of regional variations across states. The highest regional dummy value corresponds to Karnataka, where fertility decline appears to be the most delayed. This is partly due to the high-fertility zones in the Deccan. Conversely, Tamil Nadu appears to have a lower than expected fertility level. This is the core of the famous Tamil Nadu riddle that has become even more pronounced during the 1990s with observed fertility levels much lower than what the social and economic profile of the Tamil state would imply.

The most surprising finding relates probably to the fact that the Kerala dummy is both positive and highly significant. This suggests that the state that has pioneered fertility decline in India has actually a higher CWR than our model would predict. In other terms, fertility in Kerala is rather too high in view of its social and economic characteristics. This peculiarity probably derives from the position of Kerala as the forerunner of fertility

decline in south India. While it took a very high level of social development for fertility to decrease in Kerala, other states such as Andhra Pradesh and especially Tamil Nadu have witnessed a more ‘spontaneous’ fertility decline with a far less spectacular social (or economic) development. More technically, this trait is also due to the fact that many determinants of fertility decline tends to level off once fertility approaches the replacement level. For instance, the simple linear model that we are using to relate fertility and literacy may not be appropriate for very high levels of education as obtained in Kerala. A more detailed analysis (see below) will indeed show that the relationship between literacy and fertility is actually curvilinear: the effect of increased literacy levels on fertility tends to gradually disappear as literacy rates increase.

While the reference Model 1 shows the positive effect on fertility of the share of agricultural and service sectors, Model 3 aims at isolating the effects of specific occupations. It shows that the category of trade and commerce workers has the most prominent impact on fertility levels. This relationship is not easy to interpret, as this occupational category remains very composite: it encompasses trades of various sizes and is moreover less frequent in rural areas than in towns. The regression co-efficients for other occupations are quite modest, if not insignificant as for agricultural labourers (already captured by *cultlab*).

Model 4 introduces some of the individual infrastructural details related to health facilities. We can observe that the presence of family planning centres has an almost insignificant positive impact on fertility. Other infrastructures related to reproductive health seem to have little or no bearing on local fertility levels. Even the fact that the village enjoys the presence of a primary health centre seems to have a very limited effect. Taken at a face value, these results cast a doubt on the supply factor in fertility decline.[19] The fact that localities better equipped with health facilities are not characterized by lowest fertility suggests that the presence of these facilities does not affect fertility behaviour and that villagers are not dependent on these infrastructures for information on family planning, or for contraceptive methods. As the health composite factor is still not correlated to fertility at higher aggregation levels, it is not even possible to conclude that villagers resort to infrastructure available outside the village.

[19] In a somewhat controversial paper, Pritchett (1994) considers the specific impact of family planning programmes on fertility to be exaggerated. Our analysis does not invalidate his analysis as far as the effect of infrastructure on local fertility variations is concerned.

Models 5 and 6 detail the effect of other infrastructure variables. Model 5 indicates that accessibility by *pucca* road does decrease fertility levels, while the co-efficient for a bus facility is, on the contrary, positive. This suggests that the quality of the roads may matter more than bus connections as an indicator for the volume of communication between an individual village and other urban or rural localities. Model 6 shows for the record the total absence of a relationship between fertility levels and basic amenities, such as the supply of electricity or drinking water. This is partly due to the fact that villages without these facilities are now very rare in south India.

Regional Variations in Fertility Determinants

Table 11.8 describes regression results for all states separately, using the same standard model as in previous tables. For comparison, the regression model for the whole of south India is also presented. As can be observed, Kerala data lead to a correlation co-efficient that is distinctly higher than that of other states. Due to the very large average size of villages in this state, the quality of variables used in the model is much better than in the rest of India. On the other hand, the correlation co-efficient is rather disappointing for Tamil Nadu's villages, as our regression model accounts for less than a quarter of the sample's total variance. The impact of independent variables is usually comparable across states. This is especially true about some important fertility determinants such as literacy, irrigation, distance to infrastructures (*Frange*) and education infrastructure (*Fedu*). However, several variations are noticeable across individual states and village characteristics may turn out to exert an inverse effect on fertility. We will describe the most significant differences that these comparative models demonstrate.

The table indicates that for Andhra Pradesh, only 34 per cent of the total variation in CWR is explained by the included regressors. The most distinctive feature of the Andhra Pradesh model is the strong negative role of the agricultural sector on fertility. Tribal fertility is also significantly higher in Andhra Pradesh, a state characterized by large tribal pockets in the north and a larger than average overall percentage of tribals in the total population.

In Karnataka, all broad occupational categories have, on the contrary, a higher level of fertility, with the service sector characterized by the highest positive co-efficient. Karnataka is also the only state where the health infrastructure is significantly associated with lower fertility. It is also observed that villages that are the most distant from urban centres are characterized by lower fertility. This constitutes a somewhat contradictory result in view of the observed role of urban accessibility in promoting fertility decline for the entire south Indian sample. The same unexpected negative association between CWR and distance to the nearest town is also observed for the Kerala data.

The model for Kerala displays further peculiarities when compared to the standard south Indian model. Here, higher FPW is clearly associated with lower fertility, an association that fails to appear from our entire village sample. Moreover, the share of Dalits in Kerala's population appears to have no effect on the CWR, whereas the tribal population has depressing effect, as if mass education and fertility transition had already reached out to the most underprivileged sections of the population. Similarly, the literacy SR has an unexpected negative association with fertility (lower fertility for a higher male literacy rate), although this may be due to the disturbing effect of labour migrations from the state on the search of a better job.

In the case of Tamil Nadu, the adjusted R^2 shows a value lower than 25 per cent and therefore, reflects the fact that there still remains lot of unobserved factors affecting fertility rates. Further, most variables are of plausible signs except for the distance to urban areas, SR in work and proportion of cultivators. All the infrastructure variables turn out to be significant and socio-demographic variables turn out to be more significant than those of others related to economic development, like occupation and agricultural development. The fact that economic development plays a limited role in shaping CWR variations, while the gender bias in education is the strongest in Tamil Nadu is in line the peculiar path of fertility transition in this state. On the whole however, results are disappointing, as the intensity of observed correlations is often very moderate and the use of the village database does not allow us to decompose the specific causation processes of fertility decline in Tamil Nadu.

Literacy and Fertility

We can see from the previous regression analyses that the impact of literacy on the CWR varies across states and aggregation levels. This leads us to do a separate regression of fertility rate on the female literacy rates for each state separately to capture its effect at different stages of fertility transition. Table 11.9 presents this regression result. The decrease in the degree of significance of the co-efficients and of their values indicates that female literacy loses its depressing impact on fertility in low fertility environment (as in Kerala and Tamil Nadu, where the average CWR is lower as compared to Andhra Pradesh and Karnataka). So in this stage of fertility transition, the impact of other location-specific factors may turn out to be more significant, as is seen in Table 11.8 for Tamil Nadu for the co-efficients of local supply-side variables and urban proximity variables.

Table 11.9: Regression Results for Determining the Effect of Female Literacy on CWR

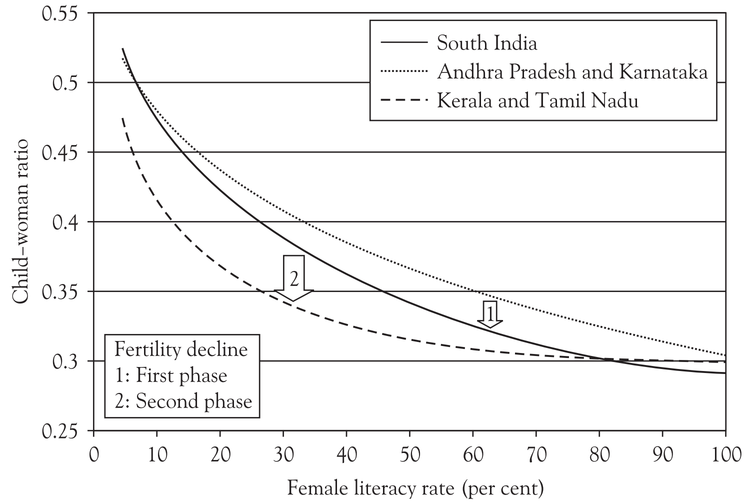
Sample	South India	Andhra Pradesh	Karnataka	Kerala	Tamil Nadu and Pondicherry
Co-efficient for female literacy rates	–0.54281	–0.43631	–0.58029	–0.39048	–0.32504
Adjusted R^2	0.2946	0.1904	0.3367	0.1525	0.1056

Notes: CWR is the dependent variable. All co-efficients are standardized.

As the literacy–fertility relationship is not exactly linear, we have attempted to provide a more refined model using polynomial factors. For that purpose, we have fitted the best fractional polynomial model to the village data. This technique uses transformations of the original literacy variable (log, square root, inverse, etc.) to identify the best curvilinear model for CWR. This modelling is applied to the entire south Indian sample, as well as to two broad northern and southern zones constituted respectively by Andhra Pradesh and Karnataka on one side, and Kerala and Tamil Nadu on the other. The results from this analysis are plotted on Figure 11.1.

In Figure 11.1, we notice the distinct shape of the northern curve vs the southern curve, with the entire south India sample lying in-between. These curves demonstrate that for a given female literacy level, Tamil Nadu and Kerala villages have significantly lower fertility than the northern states. For mid-range literacy levels, this gap amounts to more than 15 per cent of fertility level estimates in villages in Andhra Pradesh and Karnataka. This means that while literacy levels are powerful fertility determinants at the regional level, the statistical relationship between literacy rates and CWR varies from region to region. In low-fertility regions, fertility is lower for all literacy levels, though it tends to vanish when education reaches its maximum level and the curve becomes

Figure 11.1: Modelling the Literacy–fertility Relationship



horizontal indicating no change with respect to change in literacy rate. However, when combined with the general picture for south India, these curves suggest that fertility decline might be decomposed into two successive stages shown by arrows in Figure 11.1. During the first period, when fertility moves from the Andhra Pradesh–Karnataka profile to the average profile, the reduction is mainly concentrated in high-literacy villages. During the later and final period, fertility decline is, on the contrary, concentrated in low-literacy villages. It is tempting to apply the same observations to the village populations and to infer that fertility decline during the first stage affects mainly higher-educated groups, while the less educated sections of the population are influenced later.[20] This corresponds to the classical phenomenon of vertical (top-down) diffusion across social classes.

It has been routinely noticed in surveys in India and elsewhere that the higher-educated groups have been the forerunners of fertility decline.[21] Our data suggest that a parallel phenomenon be also visible at the village level: fertility decline affects firstly localities characterized by an higher educational profile, while the final stage of fertility transition reaches later to the other localities. Moreover, in a broad low-fertility environment,

[20] On the literacy-fertility relationship, see also Bhat (2000).
[21] For a broader perspective, see Jejeebhoy (1995) and Jeffery and Basu (1996).

such as the Kerala–Tamil Nadu macro region, this diffusion of fertility decline appears to have reached educationally backward localities much earlier than in Andhra Pradesh or Karnataka. While our analysis cannot examine the full implication of this observation, there are reasons to believe that regional factors are at play in accelerating fertility decline, independently of other structural conditions known to affect fertility, such as literacy or other social and economic features. These factors would probably relate to the shared social and ideological characteristics within regions.

Spatial Heterogeneity and Fertility Variations

The onset of fertility decline is as a matter of fact seldom uniform in space and time and India's experience in this regard has been very contrasted (Guilmoto and Rajan, 2001). The first phase of fertility reduction has thus resulted in an increasing degree of heterogeneity. This diversity is characteristic not only of India as a whole but also at the smallest possible geographical level. To detect homogeneity pattern in our fertility measure and its socioeconomic determinants for villages (the smallest possible units) within different administrative boundaries, we have computed the ICC of CWR and residuals in the village-level regression for each state separately and the whole of south India (from the models described in Table 11.8) by taking administrative grids such as *taluks* and districts as a class within each state and also within the whole of south India. To find out the importance of our spatial clusters which is based on closeness of villages (and therefore incorporates any space/location influence), Table 11.10 also presents the ICC co-efficient of CWR and residuals from the models described in Table 11.8 for villages within these spatial clusters. We have calculated the ICC for regression residuals (expressed as *e* in Table 11.10) in order to identify how much homogeneity patterns in CWRs across villages within spatial clusters or administrative grids is being captured by the included socioeconomic and supply-side explanatory variables.

The first row of Table 11.10 considers homogeneity pattern of villages within a *taluk*. The second row considers villages within the district. The third row considers villages within states and the following four rows describe homogeneity pattern across villages within different levels of spatial clusters. The terms in the parentheses indicate the number of classes in each case.

Table 11.10: ICC coefficient for CWR and regression residuals (e)

	South India		Andhra Pradesh		Karnataka		Kerala		Tamil Nadu and Pondicherry	
Units	CWR	e	CWR	e	CWR	e	CWR	e	CWR	e
Across villages in a family of *taluks* of same district and same state	0.359 (1,509)	0.225	0.322 (1,096)	0.191	0.307 (175)	0.144	0.772 (61)	0.431	0.323 (177)	0.234
Across villages in a family of districts of same state	0.264 (78)	0.157	0.143 (22)	0.099	0.241 (20)	0.101	0.656 (14)	0.204	0.198 (20)	0.120
Across villages in a family of states	0.083 (4)	0.005	NR	NR	NR	NR	NR	NR	NR	NR
Across villages in a family of 2 km clusters	0.427 (22,945)	0.289	0.397 (9,396)	0.264	0.362 (7,913)	0.201	0.865 (772)	0.656	0.381 (5,022)	0.285
Across villages in a family of 5 km clusters	0.390 (6,814)	0.255	0.367 (2,988)	0.236	0.315 (2,257)	0.169	0.837 (305)	0.561	0.355 (1,431)	0.263
Across villages in a family of 10 km clusters	0.370 (21,22)	0.238	0.347 (959)	0.217	0.292 (701)	0.152	0.808 (113)	0.489	0.333 (459)	0.237
Across villages in a family of 20 km clusters (619)	0.344 (618)	0.217	0.295 (296)	0.185	0.267 (212)	0.133	0.708 (38)	0.369	0.305 (142)	0.222

* NR indicates not relevant, as for individual state there is only one class, i.e., the state itself.
Terms in parentheses indicate number of classes, i.e., number of *taluks*, number of districts and number of different spatial clusters.

From this table, we observe that the erratic nature of administrative blocks like *taluks* is confirmed once more. The correlation among the villages across *taluks* in terms of fertility rate is almost close to that value for villages across 20-km clusters, i.e., villages that are not very close to each other. So, the spatial contiguity factors representing some local space-specific characteristics are not considered in the administrative boundary construction.

The cells depicting NR (not relevant) correspond to cases that cannot be considered in our analysis, as they would be only one class. From the ICC values of CWR we see that the homogeneity pattern in fertility rate increases as we move closer, i.e., consider only neighbouring villages within spatial clusters and decreases vice versa. Instead, if we consider administrative boundaries like *taluks*, districts and states, we observe that villages are not so homogeneous in the fertility pattern within these administrative blocks.

Another important observation is that in terms of fertility rate in Kerala, villages are almost similar, indicated by the high value of the ICC even among villages which are far apart. This indicates the fact that even at the village level, Kerala has reached the ultimate stage of demographic transition where homogeneity in fertility rate across villages is observed. In Kerala, *taluks* are also not so erratic in their construction.

The disturbing fact is that though the correlation in the residuals after regression reduces, it still remains high in most of the cases. So we can conclude by saying that the included socioeconomic variables, although they follow a spatial pattern similar to the regional pattern of fertility, cannot sufficiently explain the regional pattern of the fertility. There still remain some unobserved or unobservable regional or space-specific factors, which are responsible for these residuals.

Conclusions

This chapter attempts to model fertility transition at the smallest possible geographical level. In determining so we have tried to capture all possible socioeconomic determinants of fertility rate from our available list of variables. Then we have examined the issue of spatial pattern of the fertility transition, which is very much inherent in the demographic transition of

any country. This exercise shows the importance to use disaggregated data for analyzing fertility decline as high levels of local variations in observed fertility point to the role of many factors that macro-level analysis is not able to capture. At the same time, the limitations of this kind of investigations are also obvious, as the statistical analysis of micro-level data is sensitive to problems of estimation quality and small-size units.

Many observations can be derived from our attempt to analyze the socioeconomic determinants and the homogeneity in the fertility levels in south India, where demographic transition has reached its ultimate stage. First, as compared to the variables indicating economic development like workforce participation, category of profession and supply-side infrastructure variables, socio-demographic variables affect fertility rate more significantly. Especially noteworthy is the lack of impact of health (including reproductive health) amenities on fertility estimates at the village level. While there is no doubt that these infrastructures cater to the needs of the rural population for contraceptive methods, they do not seem to stimulate demand in any appreciable way and to accelerate fertility decline. The impact of other facilities (local infrastructure including education, postal and communication) is negative on fertility, although the values of the corresponding co-efficients remain moderate. A further interesting finding is that a favourable agricultural setting characterized by a smaller proportion of labourers among peasants and higher irrigated land result in distinctly lower CWR levels. This remains the most distinct impact of economic development in rural areas on fertility decline.

High values of ICC for spatial clusters indicate the importance of spatial contiguity in these population dynamics, which is missing in administrative boundary construction. The spatial smoothing (i.e., cluster formation) reduces the heterogeneous unit problem, thereby improving fertility modelling by taking into consideration also the space/location-dependent homogeneous pattern across villages. In Kerala, we have seen the highest homogeneity pattern in the fertility rate across villages, even at the administrative blocks also (high value of ICC in Table 11.10). For other states, we can say that they are also moving towards the final stage of fertility transition characterized by homogeneity across different regions, though Karnataka is lagging behind other states (as can be seen from the ICC values of village in the spatial clusters for each state in Table 11.10).

Finally, we end with a note that the technique used in this chapter is the ordinary least-square regression based on the assumption of independently and identically distributed error terms. However, in this analysis, each observational unit has spatial dimension and any type of spatial correlation between neighbouring units needs to be incorporated. Generally, the spatial autocorrelation among errors has been incorporated in fertility model in two ways. First, this has been has taken care of by using regional dummy variables (Bhat, 1996) and second, by parametrizing the error variance-covariance matrix as a function of spatial dependence parameters and either function of distance between two points (see Dubin, 1988) or a function of a spatial weights matrix indicating the contiguity between units measured by a dichotomous variable (Murthi et al., 1995). However, all these analyses are based on a sample of districts but in our framework, it is really implausible to adopt these approaches because it is not computationally tractable. Given our framework, with the number of units even at the state level greater than 20,000, a cartographic map of location of each and every village provided by GIS necessary to construct spatial-weight matrix will not solve the computational problem, as the dimension of the distance matrix or the weight-matrix will be the squared number of units, i.e., more than 20,000 × 20,000. The first approach of using regional dummy variables cannot be used in the regression analysis at the state level due to a large number of regions like *taluks*/districts/clusters, though it has been incorporated as a state dummy in the whole south India regression for each level of aggregation.

References

Bhat, P.N. Mari. 2000. 'Returning A Favour: Changing Relationship between Female Education and Family Size in India', paper presented at the workshop on Fertility Change in Developing Countries, Cambridge, UK, pp. 25–26 May.

———. **1996.** 'Contours of Fertility Decline in India: A District Level Study Based on the 1991 Census', in K. Srinivasan (ed.), *Population Policy and Reproductive Health*, Hindustan Publishing Corporation, New Delhi, pp. 96–180.

Bhat, P.N. Mari and **Irudaya Rajan. 1990.** 'Demographic Transition in Kerala Revisited', *Economic and Political Weekly*, 25: 35–36, pp. 1957–80.

Bhat, P.N. Mari and **Francis Zavier. 1999.** 'Findings of National Family Health Survey: Regional Analysis', *Economic and Political Weekly*, 34: 42–43, pp. 3008–32.

Bourne, K. and **G.M. Walker. 1991.** 'The Differential Effect of mother's Education on Mortality of Boys and Girls in India', *Population Studies*, 45: 2, 203–19.

Desai, S. and **A. Soumya. 1998.** 'Land Redistribution: A Population Stabilisation Strategy?', *Economic and Political Weekly*, 33: 10, pp. 533–36.

Dubin, R.A. 1988. ' Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms', *The Review of Economics and Statistics*, 70: 3, pp. 466–74.

Dyson, T. and **M. Moore. 1983.** 'Kinship Structure, Female Autonomy and Demographic Behaviour in India', *Population and Development Review*, 9: 1, pp. 35–60.

Filmer, D. and **L. Pritchett. 1999.** 'The Effect of Household wealth on Education Attainment: Evidence from 35 countries', *Population and Development Review*, 25: 1, pp. 85–120.

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

Friedlander, D., J. Schellekens and **E. Ben-Moshe. 1991.** 'The Transition from High to Low Marital Fertility: Cultural or Socioeconomic Determinants?', *Economic Development and Cultural Change*, 39: 2, pp. 331–51.

Guilmoto, C.Z. 2000. 'The Geography of Fertility in India (1981–1991)', in C.Z. Guilmoto and A. Vaguet (eds), *Essays on Population and Space in India*, French Institute, Pondicherry, pp. 37–56.

Guilmoto, C.Z. and **I.S. Rajan. 1998.** 'Regional Heterogeneity and Fertility Behaviour in India', Working paper, No. 290, CDS, Kerala.

Guilmoto, C.Z. and **S.I. Rajan. 2001.** 'Geographic Patterns in Fertility Change', in K. Srinivasan and M. Vlassoff (eds), *Population-Development Nexus in India. Challenges for the New Millennium*, Tata McGraw-Hill, New Delhi, pp. 88–109.

Gulati, S.C. 1992. 'Development Determinants of Demographic Variables in India: A District Level Analysis', *Journal of Quantitative Economics*, 8: 1, pp. 157–72.

Irudaya Rajan, S. and **P. Mohanachandran. 1998.** 'Infant and Child Mortality Estimates, Part I', *Economic and Political Weekly*, 33: 19, pp. 1120–40.

Jain, A.K. 1985. 'Determinants of Regional Variations in Infant Mortality in Rural India', *Population Studies*, 39: ????.

James, K.S. 2000. 'Landholding and Fertility in India. A Review of Theoretical and Empirical Evidence', *Genus*, July–December, pp. 61–80.

Jeffery, R. and **A.M. Basu (eds). 1996.** *Girl's Schooling, Women's Autonomy and Fertility Change in South Asia*, Sage Publications, New Delhi.

Jejeebhoy, S.J. 1995. *Women's Education, Autonomy, and Reproductive Behaviour. Experience form Developing Countries*, Clarendon Press, Oxford.

Kishore, S. 1993. '"May God Give Sons to All" Gender and Child Mortality in India', *American Sociological Review*, 58, pp. 247–65.

Kulkarni, P.M., S. Krishnamoorthy and **Audinarayana. 1995.** 'Review of Research on Fertility in Tamil nadu', Report for the UNDP project on *Fertility Transition in Selected Districts of Tamil Nadu*, CDS, Kerala.

Malhotra. A., R. Vanneman and **S. Kishore. 1995.** 'Fertility, Patriarchy and Development in India', *Population and Development Review*, 21: 2, pp. 281–305.

Murray, A.T. and **V. Estivill-Castro. 1998.** 'Cluster Discovery Techniques for Exploratory Spatial Data Analysis', *International Journal of Geographic Information Systems*, 12: 5, pp. 431–43.

Murthi, M., A.C. Guio and **J. Dreze. 1995.** 'Mortality, Fertility and Gender Bias in India', *Population and Development Review*, 21: 4, pp. 745–85.

Okabe, A., B. Boots. and **K. Sugichara. 1994.** 'Nearest Neighborhood operations with Generalized Voronoi Diagrams: A Review', *International Journal of Geographic Information Systems*, 8: 1, pp. 43–71.

Pritchett, L.H. 1994. 'Desired Fertility and the Impact of Population Policies', *Population and Development Review*, 20: 1, pp. 1–56.

Rayappa, P.H. and **Prabhakara. 1996.** 'Demographic Transition in the South: A Regional Perspective', ISEC, Bangalore.

Reddy, K.N. and **V. Selvaraju. 1993.** 'Determinants of Health Status in India: An Empirical Verification', mimeo, National Institute of Public Finance and Policy, New Delhi.

Rosenzweig, M. and **T.P. Schultz. 1982.** 'Marker opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in Rural India', *American Economic Review*, 72: 4, pp. 803–15.

Srinivasan, K. 1995. *Regulating Reproduction in India's Population Efforts, Results and Recommendations.* Sage Publications, New Delhi.

Watkins, S.C. 1991. *From Provinces into Nations. Demographic Integration in Western Europe 1870–1960*; Princeton University Press, Princeton.