CPRC-IIPA WORKING PAPER

# Spatial Inequality in Rural India: Do Initial Conditions Matter?

*Puja Vasudeva Dutta*
*Hari K. Nagarajan*

*Working Paper 29*

Chronic Poverty Research Centre

Indian Institute of Public Administration

An earlier version of this paper was presented at the Seminar entitled “Chronic Poverty: Emerging Policy Options and Issues” that was held at IIPA on 29th and 30th September, 2005. The workshop was organised by IIPA and CPRC.

The CPRC-IIPA Working Paper Series disseminates the findings of work in progress to encourage exchange of ideas about poverty related issues. The findings are being placed in the public domain at the earliest with a view to generate wider debate.

*For information about the working paper series, contact:*

**Professor Aasha Kapur Mehta**

*Professor of Economics*

or

**The Registrar/Deputy Registrar (Academic Support)**
Indian Institute of Public Administration
Indraprastha Estate, Ring Road
New Delhi – 110002

CPRC-IIPA Working Paper No. 29

# Spatial Inequality among Indian Villages Do Initial Conditions Matter?

*Puja Vasudeva Dutta and Hari K. Nagarajan*

**Indian Institute of Public Administration, New Delhi**

**Chronic Poverty Research Centre**
**2005**

**Puja V. Dutta** has an MA and D.Phil from University of Sussex. She is an economist currently working as a consultant at the National Council of Applied Economic Research (NCAER) in India and is a research affiliate at the Poverty Research Unit, University of Sussex. Her area of specialisation is micro-econometrics, with a focus on labour and development economics. Her doctoral research was concerned with an empirical examination of the relationship between trade liberalisation and the distribution of wages in India. She has also worked on the impact of Trade liberalisation on household poverty dynamics in developing countries. Her current research interests include the evolution of spatial and local inequality among Indian villages; and the links between ethnicity, labour market discrimination and social exclusion.

**Hari K. Nagarajan** is a senior fellow with the National Council of Applied Economic Research (New Delhi, India). He holds a Ph.D. from the University of Oklahoma (Norman). His current research interests include, household behavior (with emphasis on mutual insurance and, risk aversion), income dynamics, and, spatial behavior of agricultural prices. He has published in journals like European Economic Review, Journal of Development Economics and, Agricultural Economics.

# Spatial Inequality in Rural India: Do Initial Conditions Matter?[1]

***Puja Vasudeva Dutta and Hari K. Nagarajan***

## 1. Introduction

Disparities in income and living standards across countries and between regions within countries (spatial inequality) have been the subject of much debate and research in recent years. Spatial inequality is a construct arising out of variations in economic endowments, geography and, socio-political structure across the relevant economic space. It is typically measured as an outcome of differences in mean income or consumption levels across the economic space. The extant literature has examined some of its causes. These include globalisation, variations in availability and quality of infrastructure, and, persistent conflicts. If a significant proportion of overall inequality is spatial in nature then this can produce the preconditions for chronic poverty. Persistent spatial inequality reduces household level mobility in terms of income, occupation etc. Policies aimed at reducing chronic poverty will then have to focus on structural rather than household specific factors.

The literature has typically measured spatial inequality (as proxied by variations in mean income or consumption across the relevant economic space) in terms of aggregate state domestic product using

[1] This paper is part of a larger project funded by the Chronic Poverty Research Centre. We are grateful to our colleagues at NCAER, Dr. Shashanka Bhide, and at the CPRC, India, Dr. Aasha Kapur Mehta, as well as participants at the CPRC-IIPA Seminar on "Chronic Poverty: Emerging Policy Options and Issues" on 29-30 September, 2005 for their comments particularly Professor David Hulme, Dr. B.K. Pradhan and Dr. Lant Pritchett. We would also like to thank Anuj Sharma, Reema Kapoor and Swati Bajaj for providing research assistance.

national accounts data or, average income for all households in a given region using survey data. In this paper spatial inequality is construed as returns to location that derive from local village characteristics including institutions, initial conditions (such as economic endowments and level of activity specialisation) and, the extent of integration of the local economy, with the rest of the economic space. Measures based on variations in aggregate state domestic product or the average aggregate income of the households using survey data do not constitute 'pure' returns to location since a significant part of these variations can be explained by differences in household characteristics. In this paper returns to location are proxied by village income premia (i.e., relative village incomes purged of household characteristics) and local economic activity specialisation patterns (i.e., the true location quotient under condition of perfect mobility of resources). The instruments of village income premia and activity specialisation then jointly determine spatial inequality. This paper explains spatial variations in income as a result of pure returns to location plus possible returns to unobserved characteristics caused by the interaction of household level variables with spatial variables. An implication of this line of investigation will be that where such conditions produce externalities related to location, even a poor household can take advantage of possible positive externalities and have better long-term prospects. Alternatively, poor households located in villages where negative externalities generate negative returns to location, will have worse long-term prospects.

This paper contributes to the literature in number of ways. First, it helps focus attention on the fact that meaningful comparison of incomes across space can be done in the context of representing these incomes as pure returns to location. Differences in income can then be described as resulting from different neighbourhood structures. Second, it is able to control for a wide range of village-level structural variables, thereby overcoming the common problem of unobserved heterogeneity. In the cross-country and cross-region inequality among broad geographic entities, such as states or districts Indian villages mimic small countries that operate in relative autarky (Foster and Rosenweig 2003). Finally, this paper shows that local endowments matter in so far as the local economies interact with a wider geographic space. Such interactions could produce positive as well as negative externalities, some of which are responsive to policy.

This paper exploits a unique dataset covering about 208 Indian villages spread across 14 states at two points of time – 1981-82 and

1998-99 – during which the Indian economy underwent considerable change. These data represent an under-researched resource and are unique in that there is information on households in each village, thereby making it possible to estimate the income distribution for each village. The paper is organised as follows. Section 2 discusses the extant literature and, the background underlying the empiric of the paper. Sections 3 and 4 describe the data and methodology respectively. The results are presented in Section 5. Section 6 offers some conclusions.

## 2. Literature and Background

The motivation for examining the determinants of spatial inequality in India lies in the extant literature and the data where such inequality is not only persistent and significant but also increasing over time. Table 1 reports the decomposition of overall inequality in real annual household income into what is referred to in the literature as between and within group at different levels of geographic disaggregation (by rainfall region and by village).

**Table 1: Inequality decomposition by geographic sub-group**

| | **1982** | | **1999** | |
|---|---|---|---|---|
| **Geographic sub-group** | **MLD** | **Theil** | **MLD** | **Theil** |
| Overall inequality | 0.3533 | 0.3840 | 0.4228 | 0.4639 |
| ***by region:*** | | | | |
| Within-group inequality | 0.3142 | 0.3450 | 0.3295 | 0.3673 |
| Contribution (%) | (88.93) | (89.83) | (77.94) | (79.18) |
| Between-group inequality | 0.0391 | 0.0390 | 0.0933 | 0.0966 |
| Contribution (%) | (11.07) | (10.17) | (22.06) | (20.82) |
| ***by village:*** | | | | |
| Within-group inequality | 0.2385 | 0.2774 | 0.2582 | 0.2957 |
| Contribution (%) | (67.51) | (72.23) | (61.06) | (63.75) |
| Between-group inequality | 0.1148 | 0.1066 | 0.1646 | 0.1682 |
| Contribution (%) | (32.49) | (27.77) | (38.94) | (36.25) |

***Source:*** *Authors' calculations from ARIS-REDS listing sheets (see data section).*

*Notes: A. Rainfall regions are climatic zones as defined by the Indian Meteorological Department.*

*B. Figures in parentheses are the contribution (%) of each component to overall inequality.*

We find that while within-group inequality provides the major contribution to overall inequality (indicating considerable heterogeneity among households at the village-level), the contribution of between-group inequality rises with the level of disaggregation (see Elbers *et. al.* 2005). The latter is essentially the variation in mean income across geographic regions or villages. In 1982, inequality between villages in the sample explains between 27 to 33 per cent of overall inequality depending on the inequality index used.[1] The noteworthy fact however, is the rise in overall inequality between 1982 and 1999 where, almost the entire increase can be accounted for by the rise in between-group inequality (whether at the regional or village level).

Several studies have documented the persistence and growth of spatial inequality between countries and across regions within countries. While there is consensus about both its existence and importance, the causes of spatial inequality have been variously explained. One strand of this literature attributes persistent spatial inequality to structural factors. These factors include ethnicity (Anderson and Pomfret 2003), diversity of institutions (Fields 1980; Rainwater 1992), and absence of well functioning credit markets (Galor and Zeira 1993; Perotti 1993). Variations in household borrowing and repayment capacity (see Fleisher and Chen (1996), and Hare and West (1999) for China) and the presence of social networks for mutual insurance (Munshi and Rosenzweig 2005) have also been cited as causes of spatial inequality.

Another strand of the literature has documented the importance of initial conditions (Alesina and Rodrik 1994; Deininger and Olinto 2000; Deininger and Squire 1998). The initial regional endowment and distribution of income, land, and other asset holdings can have an impact on the long-term growth prospects.

A third strand of the literature that has its origins in economic geography explains the persistence of spatial inequality as arising out of market thickness so that variations in mean consumption levels can be explained by location (see Fujita, Krugman and Venables 1999). This then suggests that economic space will matter in determining the magnitude of spatial inequality (Davis and Weinstein 1999; Glaeser, Kallal, Scheinkman and Shleifer 1992; Henderson, Shalizi and Venables 2001). If space is to matter for reasons other than purely jurisdictional, then spatial inequality is both, a function of purely local characteristics and the characteristics of the wider region. Spatial interactions between agents in differently endowed regions can contribute to persistent and even increasing spatial income inequality (Nazara 2003; Nazara and Hewings 2003). Eeckhout and Jovanovic (2001), Mookherjee and Ray (2002), Escobal and Torero (1999) and Ravallion (2002) have shown that geography can influence the evolution of household welfare over time. That is, positive externalities associated with geography such as local public goods and endowments of private goods imply that a household (even if poor) associated with such externalities could be better off in the long run. Similarly, Balisacan and Fuwa (2003) have estimated the magnitude of inequality arising out of household-specific and location-specific factors. They have shown for example, that households located in areas with connectivity to urban areas are able to diversify their activities and consequently are able to insure themselves against shocks that might impinge on certain sectors. Jalan and Ravallion (1998) and Engerman and Sokoloff (1998) have used this argument to show how policy should endogenise heterogeneity arising out of geography to reduce spatial variations in outcomes.

Other explanations for both persistence and increase in spatial inequality include government policies, fiscal federalism, migration and, trade. The magnitude of spatial inequality can also be influenced by the specific policies that have been formulated to influence the growth process (Fields 1980), in particular the Chinese reform process (Raiser 1998). Fiscal federalism and the relationship between the central government and the regional governments have been found to influence spatial inequality in Indonesia (Tadjoeddin, Suharyo and Mishra 2003). Yet another strand of the literature highlights the impact of activity specialization and the consequent composition of the labor force in various regions. It focuses on factors, such as migration (see Özmucur and Silber 2002 for Turkey), and trade (see Andalón and López-Calva 2002 for Mexico, and Zhang and Zhang 2003 for China), that influence these patterns of specialization and the relative importance of these regions.

Spatial inequality has been the focus of research in India as well and a number of explanations, drawing on several of the strands mentioned above, have been offered to account for this disparity. Noorbakhsh (2003) finds that inequalities in production and consumption, growth in such inequalities, and the relatively fewer number of convergence clubs, have together contributed to spatial polarization. Ghosh and De (2000) bring out the link between variable infrastructure development and growth in income - areas with better infrastructure have experienced faster income growth. Datt and Ravallion (2002) have pointed out the extremely variable initial conditions in rural development and human capital development that are obtained in rural India as causes for persistent and often increases in spatial inequality. Ahluwalia (2002) has found that variation in flows of private investment leads to spatial variation in income. Shand, Kalirajan and Rao (1999) show that initial conditions of capital formation and human capital endowments help explain the rate of convergence of incomes across regions. Cashin and Sahay (1996) have explained spatial inequality and convergence as an artefact of internal migration. Finally Sachs, Bajpai and Ramiah (2002a) show that social and geographic externalities matter in mediating the rate of change of spatial inequality.

In this paper we draw on several strands of the literature in development, economic geography, income inequality and labor economics in order to identify the structural determinants of spatial inequality among Indian villages. Existing Indian studies measure spatial inequality as the variation of mean income or consumption across geographic units (usually states, districts or groups of contiguous districts). The approach in this paper however, differs in three ways. First, spatial inequality has been measured as resulting from pure returns to location. Such returns to locations have been proxied using relative village income premia, (which are purged of household characteristics) and local economic specialization patterns (the true location quotient under conditions of perfect mobility). Thus, spatial inequality has been conceptualized as resulting from pure returns to location plus returns to unobserved characteristics caused by the interaction between household and spatial variables. In this sense this paper belongs to the literature identified by Heltberg (2003), Kanbur and Zhang (2005) and Sachs (2003). Second, one of the common criticisms of the cross-section (usually cross-country or cross-state) inequality literature is the presence of unobservable heterogeneity. In this paper, these have been controlled for using a range of structural village-level characteristics that capture local conditions. Finally, the village has been used as the unit of analysis. Villages in India typically operate as autarkic small countries (Foster and Rosenzweig (2003).

## 3. Data

The paper uses data from the various rounds of the ARIS/REDS surveys conducted by the National Council of Applied Economic Research (NCAER). These surveys were conducted at five points in time, viz., in 1968-69, 1969-70, 1970-71, 1981-82 and 1998-99. The objective of the original rounds in 1968-71 was to determine the performance of cultivators of high-yielding varieties relative to cultivators of traditional varieties of crops and the consequences for income inequality. Approximately two-thirds of the entire sample were selected from villages covered by the Intensive Agricultural Development Program (IADP) or the Intensive Agricultural Area Program (IAAP). In order to maintain the panel characteristic, the same villages were tracked in the subsequent rounds in 1981-82 and 1998-99.

Each round of the survey contains three parts. The first part is the "listing sheet". This contains information on household income and demographic variables. The second part is the "village questionnaire". This is the source of information on village-level characteristics such as agricultural production and land use, irrigation facilities, agricultural prices and wage rates, access to markets, local political structure, land tenure systems and, availability of public goods such as schools and medical centers. The third part is the "household questionnaire" where, data on a range of variables relating to a select sample of households has been collected.

The listing sheets are used to select the households to be surveyed. These contain information on several household-level characteristics such as the age, gender and occupation of the head of the household, household income, family size, and number of earners. The income data is based on a single question on total household income from all sources. This data represents a valuable resource in estimating the distribution of household incomes at the village level.[2] In the initial round, the true income distribution can be identified for almost 50 per cent of the villages in which all or at least 80 per cent of resident households (as reported in the Census) have been listed. For some of the larger villages, only a random sample was listed. By the 1998-99 round, the proportion of villages with over 80 per cent of resident households listed fell to about 40 per cent. However, in all the rounds, at least half the resident households are listed in about three-quarters of the villages.

In this paper, we have used the data from 1982 and 1999 rounds. The size of the sample consists of 39,541 and 53,562 households in 1982 and 1999 respectively spread across 138 *taluks*[3] in 86 districts in 14 states. The average number of households per village is 19 ranging from a minimum of 34[4] to a maximum of 600 in 1982. The corresponding figures for 1999 are 258, 51 and 1167 respectively. The 208 villages in our sample display considerable heterogeneity with respect to average household income and land as well as the distribution of and trends in these variables[2].

The nominal annual household income is trimmed by dropping households that reported incomes less than the average monthly village agricultural wage (as reported in the village surveys).[5] These are then converted to real income by deflating to 1999 prices using the state-level consumer price index for agricultural laborers (CPIAL). The variables used to explain household income are available in the listing sheets and include household demographic information such as the age and gender of the household head, household size, the number of earners, household land,[6] 11 occupation categories[7] and 208 village dummy variables.

The listing sheet and the village level data were combined with data from other secondary sources such as the National Census and the NCAER rainfall database. The variables from the village survey include factor endowments, local governance and integration with the larger economy (the sources and definitions of these variables are outlined in Table A1 in the Appendix). The variables capturing activity specialization patterns – the indices of specialization for the village with respect to the taluk and for the taluk with respect to the rainfall region, dependency ratios and population-to-basic employment ratios - have been computed using the data from the Village and Town Primary Census Abstract (Part XII-B) of the District Census Handbook and Economic Tables (Part II-B(i)) of the Census of India for 1971, 1981, 1991 and 2001.

## 4. Methodology

The paper draws on a number of different strands in the development, economic geography, inequality, and labor economics literature. Using the economic geography literature (see Miller, Gibson and Wright 1991), indices of local specialization for the village are computed with respect to the *taluk* and for the *taluk* with respect to the rainfall region. Based on the trade and labor literature, a two-stage

---

[2] The 1971 survey is not used, as comparable data on household characteristics (e.g., land owned, household size, dependency ratio, age and occupation of the household head) are not reported in the listing sheets for this year. In addition, four states – Assam (surveyed only in 1999), Jammu and Kashmir (surveyed only in 1982), Madhya Pradesh and Maharashtra – were not surveyed in 1971. There are fifteen states common to 1982 and 1999. However, the 1982 listing sheets for Madhya Pradesh were not available. Therefore, the final data comprises information from 14 states.

methodology is employed (first applied by Gaston and Trefler (1994) to investigate the link between trade and industry wage premia for the United States). In the first stage the reduced-form household income regression is estimated with controls for village residence. The estimated village coefficients are transformed into village income premia. In the second stage, the simultaneous evolution of these village income premia and the activity specialization patterns is investigated.

### *Activity specialization patterns*

Patterns of activity specialization are examined across economic space at four points in time – 1971, 1981, 1991 and 2001 – spanning four decades. Spatial variations in income can be explained in the context of differences in the patterns of activity specialization, concentration and clustering of activities across the economic space (Fan and Casetti 1994). One of the indicators of specialization is the index of local specialization (hereafter ILS). This has been used in the literature to both measure the productivity of a particular activity in the context of clustering as well as an indicator of the location premium associated with a specific activity. When all things are equal, i.e., under conditions of perfect mobility of resources, ILS is the true location quotient. A given household under such conditions will be free to locate itself anywhere in the economic space. Hence ILS becomes a perfect proxy for location premium associated with continuing with that activity.

The ILS for various activities is constructed as follows for the village with the *taluk* as the benchmark area and for the *taluk* with the rainfall region as the benchmark:

$$ILS^{i}_{j,k} = \frac{\left(E^{i}_{j} / E_{j}\right)}{\left(E^{i}_{k} / E_{k}\right)} \qquad (1)$$

Where ILS is the index of local specialization with the subscripts *j* representing the local area, and k the benchmark area (i.e., if *j* = village, then *k* = *taluk* and if *j* = *taluk*, then *k* = rainfall region). The superscripts *i* representing different activities (1 = cultivators, 2 = agricultural labor and 3 = household industry) (time subscripts suppressed).[8] $E^{i}_{j}$ is the number of main workers employed in activity *i* in the area *j* and $E_j$ is the total number of main workers employed in the area *j*.[9] An ILS value for a particular activity that is greater than one indicates specialization in that activity.

### *Household income regression models*

Household income can be considered to be deriving from the level endowments of labor, land and capital (human and physical) as well as the returns to these factors. Household income regression is estimated in the semi-logarithmic form with household endowments of labor and land as explanatory variables along with geographical location controls that capture the effects of the local village economy (including the returns to labor and land) on household income. The semi-logarithmic specification is chosen in keeping with the standard Mincer earnings equation and also as the distribution of household incomes in our sample follows approximately a log-normal distribution.[10] The household income regression models can be expressed as follows:

$$y_i = \alpha + x_i'\beta + v_i'\delta + \mu_i \qquad (2)$$

Where $y_i$ is the natural log of the real annual household income, $\alpha$ is the intercept term, $x_i$ comprise exogenous explanatory variables, $v_i$ the vector of 207 village residence dummy variables (one village is omitted being the reference village), the $i$ subscripts for individuals (suppressing the time subscript for 1982, 1999), and $\mu_i$ is a random error term.

The vector of explanatory variables, $x_i$, include the age and gender of the household head (expressed as a quadratic function), household size, number of earning members, household land, and 10 occupation categories. While it would be desirable to include a measure of the educational attainment of the household (see for example Wan and Zhou (2005)) these data are not available for both years. The village dummies, $v_i$, capture the effects of location, including market access, infrastructure, governance, geographic conditions and resources, and local culture. Equation (2) is estimated using Ordinary Least Squares with robust standard errors corrected for heteroscedasticity using the Huber-White correction.

### *Village income premia*

The methodology introduced by Krueger and Summers (1988) is adapted in the context of industry wage premia, and applied to transform the estimated village coefficients in the household earnings function into deviations from the size-weighted mean income differential as follows:

$$\hat{\delta}^* = (I - e \times s')\hat{\delta} \qquad (3)$$

Where, $\hat{\delta}^*$ is a (K+1x1) column vector of village income premia, *I* is (K+1xK+1) identity matrix, *e* is a (K+1x1) vector of ones, $\hat{\delta}$ is the (K+1x1) vector constructed by stacking the (Kx1) vector of village coefficients estimated from the income regression models above a (1x1) matrix with zero as the single element,[11] and *s* is a (K+1x1) vector of village size weights with each element $s_k = n_k \Big/ \sum_{k=1}^{K} n_k$ where $n_k$ is the size of village *k* for *k*=1,..,K+1 villages. The weights can relate to village size either in terms of area or population. The adjusted variance-covariance matrix $V(\hat{\delta}^*)$ is computed as suggested by Haisken-DeNew and Schmidt (1997) and can be expressed as follows:

$$V(\hat{\delta}^*) = (I - e \times s')V(\hat{\delta})(I - e \times s)' \qquad (4)$$

The resulting income premia represents the difference in the income received by a household in village *j* to the average household across all villages in our sample.

### *Determinants of village income premia and activity specialization patterns*

Simultaneous determination of the village premia along with activity specialization is a hypothesis to be tested. Towards this end the following system of equations has been estimated.

$$y_m = X_m' \beta_m + \varepsilon_m \qquad m = 1,2,3,4 \qquad (5)$$

Hence, the covariance matrix of the error terms is given by $V = E\left[\varepsilon_i \varepsilon_j^{'}\right] = \sigma_{ij} I$ (where *i, j* = 1,2,3,4). Clearly, $V = E \otimes I$ with $E = \left[\sigma_{ij} I\right]$ where $\otimes$ is the Kronecker product.

As a result, the generalized least squares estimator can be written as follows:

$$\overline{\beta} = \left[X'V^{-1}X\right]^{-1} X'V^{-1}y == \left[X'\left(E^{-1} \otimes I\right)X\right]^{-1} X'\left(E^{-1} \otimes I\right)y \qquad (6)$$

This estimator is consistent, unbiased and efficient in contrast to the OLS estimator, which is only consistent. Whether E is diagonal can be tested using the Breusch-Pagan test (Greene 1990).

**5. Empirical results**

## *Activity specialization patterns*

There is some evidence of persistence in activity specialization and clustering over time as initial ILS values seem to determine long-term patterns Table 2. Such persistence could be attributed to factors such as culture, technology, preferences, etc. For instance, areas specializing in labor-intensive agriculture typically have tended to retain this character.

At the same time, there is evidence of some changes in the level of ILS values over time indicating some changes in activity specialization during this period. The level of specialization as indicated by the value of the ILS is a function of both household level and local economic variables associated with the village. Changes in the value of the ILS indicate the propensity of the households to be mobile and/or changes in the returns to the specific activities. However, the value of the ILS has to be read along with the population to basic employment ratio.[12] The latter indicates the carrying capacity of a member employed in the sector in question. Other things remaining constant, a relationship between ILS and the population to basic employment ratio is expected, such that higher values for this ratio would be accompanied by lower values of the ILS. The logic is that a higher population to basic ratio should result in activity diversification failing which location premium associated with the given activity will collapse. This expected relationship between the population to basic employment ratio and ILS values is identified for the three different activities.

## *Village income premia*

The first-stage household income regression models are reported in Table 3. The explanatory power of the variables in both years is quite high – explaining about two-thirds of the variation in log household income in each year. The effects are generally plausible and significant at the 1 per cent level or better.

The age-earnings profile follows the conventional inverted U-shape. Male-headed households have incomes about 11 per cent higher than female-headed households in 1982, rising to 15 per cent by 1999 (t-statistic = 2.58). The variables that capture the endowments of productive factors - namely, household size, number of earners, and land - have a positive and significant effect on household income in both years. Borooah (2005) and Wan and Zhou (2005) also find that factors such as age, education, labor and land endowments as well as location are the key determinants of household income in their samples of rural households in India and China respectively.

The income elasticity of land owned is unchanged in the two years at about 6-7 per cent while that of the household size has fallen marginally from about 29 per cent in 1982 to 26 per cent in 1999 (t-statistic = -1.81) and that of the number of earners has risen from about 11 per cent to 20 per cent between the two years (t-statistic = 10.96). This could indicate a decline in the importance of agriculture as a primary source of household income.

Table 2

## Table 3: Household income regression models

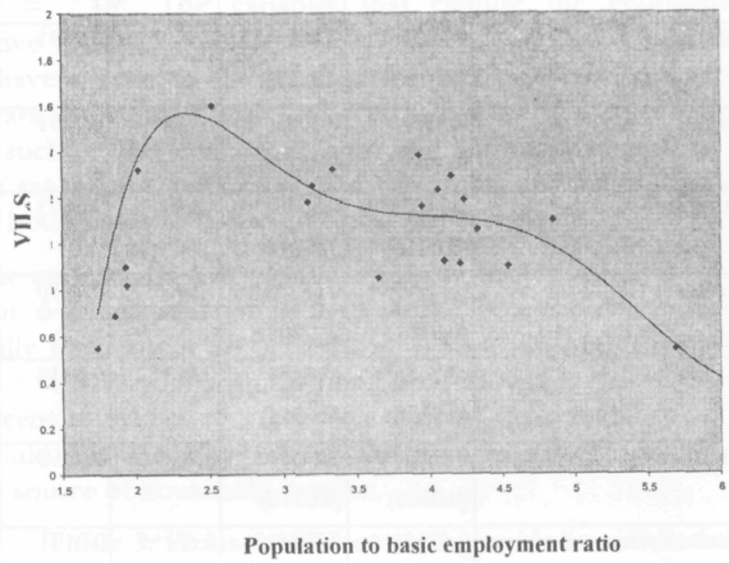
| | **Income regression models** | | **t-test** | **Summary statistics** | |
|---|---|---|---|---|---|
| | **1982** | **1999** | **(1999-82)** | **1982** | **1999** |
| ***Household demographics:*** | | | | | |
| Household head age | 0.0080*** (0.0012) | 0.0116*** (0.0011) | 2.21** | 46.87 (13.41) | 49.47 (13.88) |
| Household head age squared | -0.0001*** (0.00001) | -0.0001*** (0.00001) | 0.00 | 2376.95 (1324.73) | 2640.02 (1438.45) |
| Male household head | 0.1102*** (0.0116) | 0.1511*** (0.0108) | 2.58*** | 0.9294 | 0.9306 |
| Household size | 0.0489*** (0.0013) | 0.0457*** (0.0012) | -1.81* | 5.9815 (3.2810) | 5.6650 (3.0866) |
| Number of earning members | 0.0609*** (0.0037) | 0.1231*** (0.0043) | 10.96*** | 1.8042 (1.1337) | 1.6068 (0.9552) |
| ***Household assets:*** | | | | | |
| Land (acres) | 0.0219*** (0.0017) | 0.0211*** (0.0011) | -0.40 | 3.2244 (6.5340) | 2.7759 (6.1050) |
| ***Household head occupation:*** | | | | | |
| Marginal farmer | 0.0687*** (0.0100) | 0.1614*** (0.0082) | 7.17*** | 0.0791 | 0.1107 |
| Small farmer | 0.3049*** (0.0086) | 0.3035*** (0.0079) | -0.12 | 0.1191 | 0.1231 |
| Medium farmer | 0.5422*** (0.0116) | 0.5079*** (0.0094) | -2.30** | 0.1593 | 0.1298 |
| Large farmer | 0.7729*** (0.0287) | 0.6873*** (0.0204) | -2.43** | 0.0910 | 0.0631 |
| Fishing | -0.2039*** (0.0578) | -0.0484 (0.0361) | 2.28** | 0.0036 | 0.0027 |

| | Income regression models | | t-test | Summary statistics | |
|---|---|---|---|---|---|
| | **1982** | **1999** | **(1999-82)** | **1982** | **1999** |
| Animal husbandry | 0.3401***<br>(0.0305) | 0.4439***<br>(0.0516) | 1.73* | 0.0058 | 0.0018 |
| Non-agricultural white-collar labor | 0.6759***<br>(0.0256) | 0.8088***<br>(0.0259) | 3.65*** | 0.0191 | 0.0133 |
| Non-agricultural blue-collar labor | 0.4671***<br>(0.0102) | 0.7111***<br>(0.0107) | 16.51*** | 0.1177 | 0.1113 |
| Non-agricultural business | 0.5028***<br>(0.0150) | 0.5860***<br>(0.0134) | 4.14*** | 0.0423 | 0.0503 |
| Transfer income | 0.0469<br>(0.0328) | 0.3430***<br>(0.0184) | 7.87*** | 0.0157 | 0.0403 |
| Constant | 8.8567***<br>(0.0523) | 8.7501***<br>(0.0485) | | | |
| Number of observations (N) | 39,541 | 53,562 | | 39,541 | 53,562 |
| R-squared | 0.6419 | 0.6410 | | | |
| $F_{207,N}$ for village dummy variables | 176.89 | 246.15 | | | |

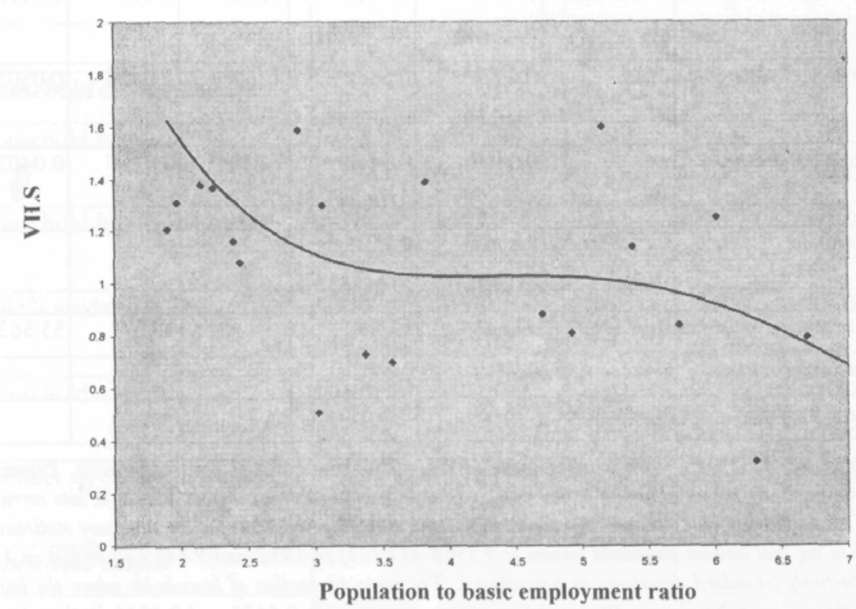
*Notes: A. *, **, *** denote significance at the 10%, 5% and 1% level respectively. Figures in parentheses are robust standard errors corrected for heteroscedasticity using the Huber-White correction for the coefficients of the income regression models and standard deviations for the summary statistics. B. Mean log real annual household income is 9.5968 (0.8773) in 1982 and 9.8419 (0.8180) in 1999 respectively (standard deviations in parentheses). The mean proportion of households where the head is engaged in agricultural labor (the occupation reference category) is 0.3473 and 0.3535 in the two years respectively. C. 207 village dummies (Keshopur village in Uttar Pradesh is omitted) are included in the regression models; their coefficients are not reported here; some selected transformed village income premia are reported in Table 4 below. D. Dependent variable: Natural log of real household annual income*

**Fig. 1 Relationship between VILS and Population of Basic Employment Ratio**

A. Cultivators (1982)

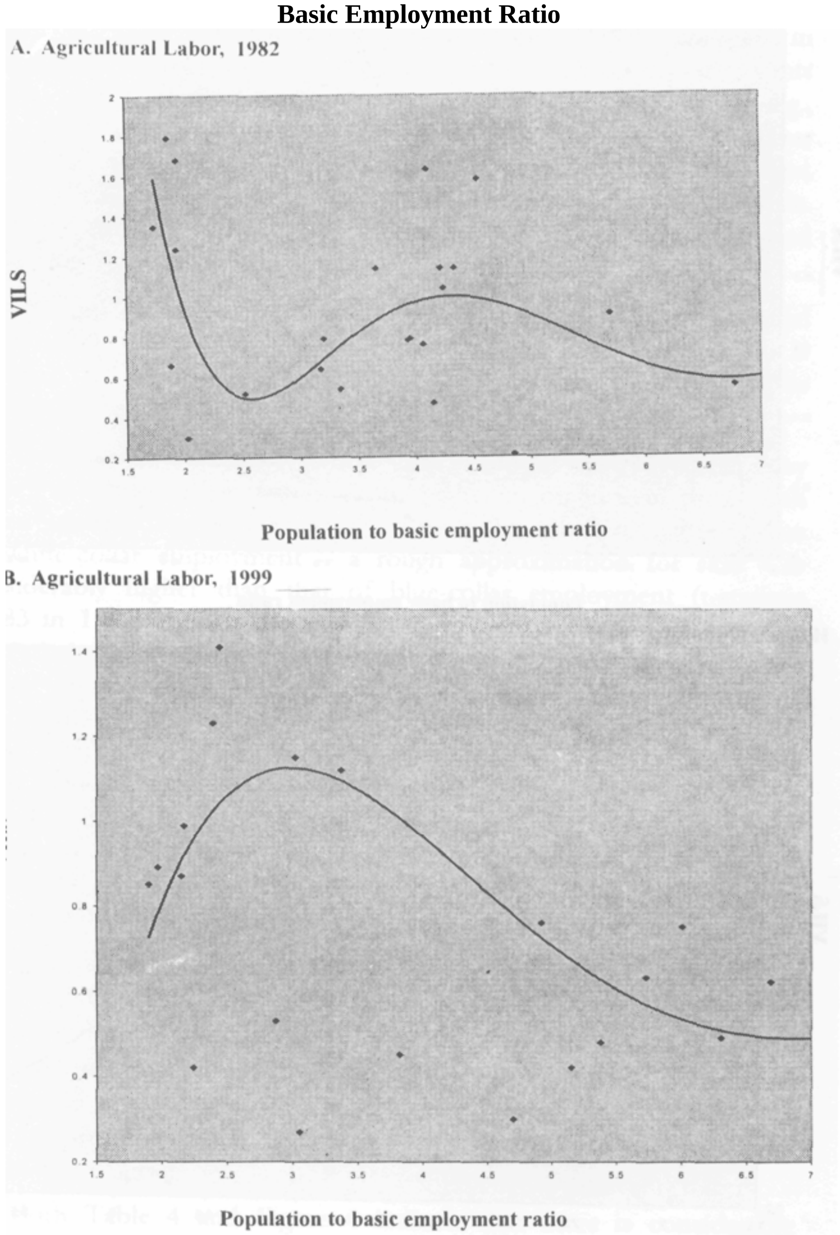



B. Cultivators (1999)



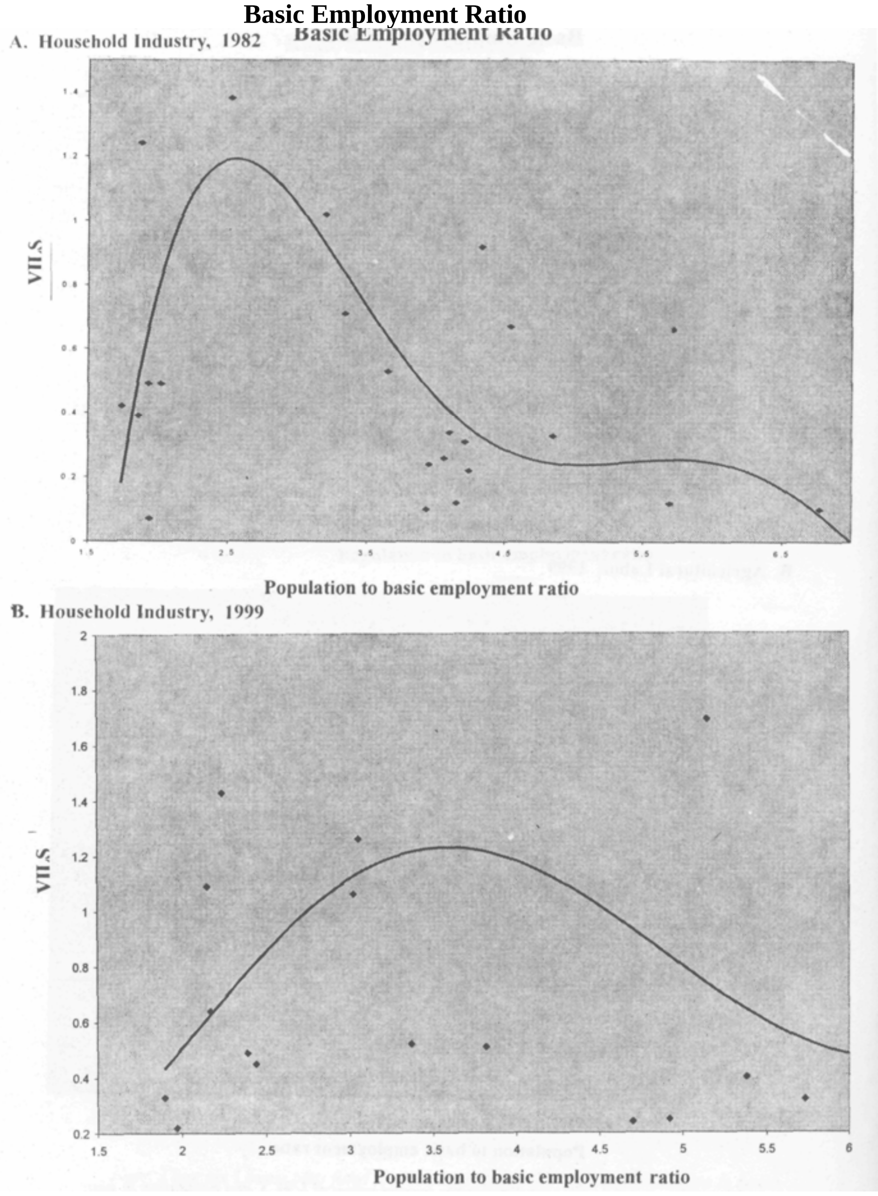

*Note: These are Lowess plots fitted to the average values of VILS and population to basic employment ratios for the rainfall regions*

**Fig. 2 Relationship between VILS and Population to Basic Employment Ratio**




*Note: These are Lowess plots fitted to the average values of VILS and population to basic employment ratios for the rainfall regions*

**Fig. 3 Relationship between ILS and Population to Basic Employment Ratio**




*Note: These are Lowess plots fitted to the average values of VILS and population to basic employment ratios for the rainfall regions*

The occupation of the household head has a significant impact on the household income. Cultivating households tend to have significantly higher incomes than households where the head is engaged in agricultural labor. In addition, the size of land holdings matters as this positive effect is monotonically rising with average farm size – e.g., marginal farmers earn about 7 per cent more than agriculture labour households in 1982 whereas large farmers earn as much as 77 per cent more. This indicates the presence of significant returns to scale. However, these scale economies do not seem to have been accompanied by any significant technological change.

Technological changes appear to have been adopted by marginal farmers (who tend to cultivate their land more intensively) and this is reflected in the rise in their income advantage with respect to agricultural labor.

Similarly, households where the head is employed in any occupation (other than fishing) earn higher incomes than those with agricultural laborers as the household head. The positive income effect of white-collar employment – a rough approximation for skill – is considerably higher than that of blue-collar employment (t-statistic =7.83 in 1982) though this gap has narrowed over time. It was also noticed that a large and significant increase in the income advantage arising out of transfer income. During this period, a significant change was found in the age and occupational structure of the heads of the households that could be tracked in both years. Also, there has been a sharp increase in the dependence on remittances in states like Kerala (see also Prakash 1998). In summary, the observed changes in the returns to occupation are a reflection of the shift in the economic conditions in the villages and could, in turn, be the drivers of an increasing diversification of activities.

The village coefficients estimated in the income regression models are all significant at the 1 per cent level or better. These are transformed into deviations from the area-weighted mean differential.[13] Selected village premia are reported in Table 4 below for villages that remained in the top 40 lowest or highest income premia bracket in both years.[14] Figure 4 presents the correlation of village premia over time through a scatter plot diagram.

Both Table 4 and Figure 4 indicate that there is considerable variation in the village income premia and that there is some persistence in the income ranking of villages. For instance, village Kechaki in Jharkhand has a mean income that is about 81 per cent less than that for the average village in the sample while village Kuttoor in Kerala has a mean income about 31 per cent higher than the sample average. In general, villages in Karnataka, Orissa and Maharashtra had the lowest relative incomes, while villages in Kerala and Punjab had higher relative incomes. In both the years, villages in the highest quintile with respect to income premia were found to be associated with the highest proportion of non-cultivating households, irrigated area (relative to gross cropped area) and of the use of high-yielding varieties seeds. These villages were also the best connected with the wider economy in terms of distance to the nearest town, administrative center, facilities such as schools, hospitals and banks, agricultural markets. This greater integration is further reflected in the high literacy rates in these villages.

**Fig. 4: Distribution of village income premia, 1982 to 1999**

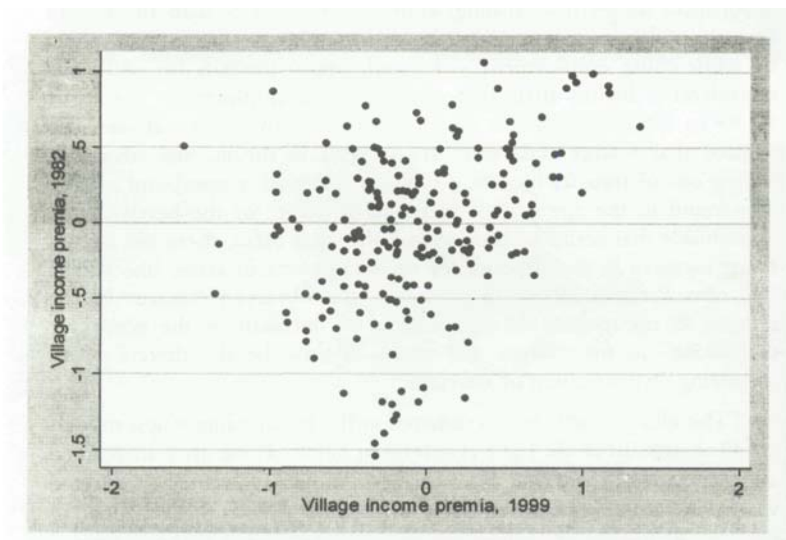



**Table 4: Income premia for selected villages**

| Region | Village | Income equation coefficients | | Village income premia | |
|---|---|---|---|---|---|
| | | 1982 | 1999 | 1982 | 1999 |
| *Low income premia:* | | | | | |
| Jharkhand | Kechaki | -0.8056*** (0.0509) | -1.4622*** (0.0558) | -0.4800*** (0.0264) | -1.3427*** (0.0389) |

| | | | | | |
|---|---|---|---|---|---|
| Gujarat region | Vanki | -0.9944*** (0.0662) | -1.0031*** (0.0435) | -0.6688*** (0.0500) | -0.8837*** (0.0191) |
| South Interior Karnataka | Guntappalli | -1.1210*** (0.1080) | -0.8872*** (0.0613) | -0.7954*** (0.0985) | -0.7678*** (0.0471) |
| Madhya Maharashtra | Baburdi | -0.9317*** (0.0567) | -1.0129*** (0.0566) | -0.6061*** (0.0361) | -0.8935*** (0.0403) |
| Orissa | Padhan | -0.8248*** (0.0634) | -0.8187*** (0.0571) | -0.4991*** (0.0465) | -0.6993*** (0.0417) |
| *High income premia:* | | | | | |
| Haryana | Nayabans | 0.1257** (0.0558) | 0.7088*** (0.0498) | 0.4513*** (0.0355) | 0.8283*** (0.0310) |
| Kerala | Kuttoor | 0.3093*** (0.0598) | 1.2497*** (0.0509) | 0.6349*** (0.0410) | 1.3691*** (0.0327) |
| Punjab | Satowali | 0.3706*** (0.0678) | 0.6253*** (0.0559) | 0.6962*** (0.0523) | 0.7448*** (0.0401) |
| East Rajasthan | Kheda | 0.1329** (0.0591) | 0.7507*** (0.0557) | 0.4585*** (0.0402) | 0.8701*** (0.0397) |
| West Uttar Pradesh | Lohlora | 0.3400*** (0.0579) | 0.4216*** (0.0476) | 0.6656*** (0.0389) | 0.5410*** (0.0276) |

*Notes: A. *, **, *** denote significance at the 10%, 5% and 1% level respectively. Figures in parentheses are standard errors. B. The income premia and their standard errors are computed as described in equations (4) and (5)respectively. C. The weights are village area. D. This is a summary measure of the overall variability in wages across industries as described in endtnote (xvii).*

The magnitude of these premia has risen over time[15] – both the income disadvantage for low-income villages and the income advantage for high-income villages has increased between 1982 and 1999. It is also apparent that the village premia are positively correlated over time and that there is a reasonable degree of rank correlation - Spearman's rank correlation coefficient between 1982 and 1999 is 0.3299*** (rejecting the null hypothesis of independence at the 1 per cent level of significance or better). The Kruskall-Wallis rank test of equality of populations ($\chi_1^2$ =1.0370) cannot reject the null hypothesis that the rankings in the two years are obtained from the same population. At the same time, bivariate regression models of the income premia between the two years reveal persistence over time – raising the premia in 1982 by about one percentage point raises the premia in 1999 by about 0.35 of a percentage point. This supports the notion of some persistence in village premia or, in other words, of spatial inequality.

The results also suggest that the dispersion of income premia across villages rose during this period. The overall variability in village income premia is summarized using the size-weighted standard deviation of the inter-village income premia adjusted for sampling variance.[16] The variability in income premia rose from about 48 per cent in 1982 to about 57 per cent in 1999. This increasing dispersion indicates a rise in the importance of returns to location. This increase is likely to have driven the increase in the contribution of between-village inequality to overall inequality (see Table 1).

### *Simultaneous evolution of village income premia and activity specialization patterns*

It is posited that spatial inequality is simultaneously explained by the income premia associated with location and activity specialization patterns. The village income premia represent the 'pure' returns to location, purged of household characteristics. The VILS values for each of the three categories of activity specialization also proxy location premia arising out of local culture, preferences, and customs. The village income premia and VILS are expected to be jointly determined. However, this is a hypothesis to be tested. Hence, Zellner's seemingly unrelated regression model has been estimated for the system of four equations with the village income premia, village-to-*taluk* ILS for cultivators, agricultural labor and household industry as the dependent variables. The results of this estimation are reported in Table 5.

**Table 5: Seemingly Unrelated Regression (SUR) Model**

| | **(1)** | **(2)** | **(3)** | **(4)** | |
|---|---|---|---|---|---|
| | **Income premia** | **VILS cultivators** | **VILS agr. labor** | **VILS hhd. industry** | **Mean** |
| ***Initial conditions (1982):*** | | | | | |
| Lagged dependent variable [a] | 0.1294**<br>(0.0630) | 0.4364***<br>(0.0649) | 0.2351***<br>(0.0544) | 0.3530***<br>(0.0469) | |
| Literacy rate | 0.4305<br>(0.2802) | -0.3492<br>(0.2519) | 0.4169<br>(0.4532) | 2.0104***<br>(0.6594) | 0.3872<br>(0.1716) |
| Mean household size | -0.0407<br>(0.0372) | -0.0690**<br>(0.0321) | 0.1573**<br>(0.0760) | -0.1132<br>(0.0901) | 5.9626<br>(1.0594) |
| Mean age of household head | -0.0069 | -0.0016 | -0.0029 | -0.0029 | 46.7391 |

| | (1) | (2) | (3) | (4) | |
|---|---|---|---|---|---|
| | **Income premia** | **VILS cultivators** | **VILS agr. labor** | **VILS hhd. industry** | **Mean** |
| | (0.0118) | (0.0104) | (0.0191) | (0.0278) | (2.6362) |
| Population of nearest town (log) | -0.0907*<br>(0.0501) | 0.0002<br>(0.0436) | -0.0943<br>(0.0811) | 0.3456***<br>(0.1212) | 10.3539<br>(1.1739) |
| School within 2 km | -0.0540<br>(0.0937) | | | | 0.8894<br>(0.3144) |
| Health facility within 2 km | 0.1438**<br>(0.0653) | | | | 0.6442<br>(0.4799) |
| Mean rainfall (mm) last 10 years | | 0.0131**<br>(0.0067) | | | 11.0287<br>(5.6862) |
| Std. dev. of rainfall last 10 years | | -0.0005<br>(0.0005) | | | 169.7119<br>(69.6323) |
| Prop. of irrigated area | | 0.0165*<br>(0.0087) | 0.0180<br>(0.0149) | | 0.7261<br>(3.0016) |
| Mean size of land holdings | | -0.0174<br>(0.0110) | 0.0148<br>(0.0135) | | 5.7213<br>(4.0783) |
| Tenant cultivation | | 0.0324<br>(0.0535) | 0.3007<br>(0.5172) | | 0.3798<br>(0.4865) |
| Share cropping | | | 0.7059<br>(0.5209) | | 0.5000<br>(0.5012) |
| Share cropping x Mean hhd. size | | | -0.1269<br>(0.0865) | | 2.9934<br>(3.0836) |
| Tenant cultivation x Mean hhd. size | | | -0.0261<br>(0.0860) | | 2.1942<br>(2.8976) |
| Prop. of cultivating households | -0.3382**<br>(0.1435) | | | | 0.4833<br>(0.2574) |
| Prop. of large farmers | | 0.9180***<br>(0.3262) | | | 0.1002<br>(0.1309) |
| Prop. of marginal farmers | | | 1.3225***<br>(0.4935) | | 0.0851<br>(0.1132) |
| Prop. of landless households | | | 0.6231***<br>(0.2127) | -0.4808*<br>(0.2857) | 0.4260<br>(0.2698) |
| Prop. of non-agricultural | 0.5811** | | | | 0.1585 |

| | (1) | (2) | (3) | (4) | |
|---|---|---|---|---|---|
| | **Income premia** | **VILS cultivators** | **VILS agr. labor** | **VILS hhd. industry** | **Mean** |
| workers | (0.2563) | | | | (0.1463) |
| TILS for cultivators [b] | 0.4002*** (0.1176) | | | | 1.0009 (0.2953) |
| TILS for agricultural labor [b] | 0.0676 (0.0518) | | | | 1.0953 (0.6512) |
| TILS for household industry [b] | 0.0138 (0.0337) | | | | 1.0813 (0.9371) |
| Elected panchayat | -0.0867 (0.0688) | | | | 0.4904 (0.5011) |
| ***Current conditions (1999):*** | | | | | |
| Literacy rate | 0.0402 (0.3034) | 0.3187 (0.2497) | -01.2967*** (0.4745) | -2.1850*** (0.6917) | 0.6245 (0.1539) |
| Mean household size | 0.0941*** (0.0359) | 0.0728** (0.0301) | -0.3085*** (0.0963) | -0.0031 (0.0841) | 5.6504 (1.1324) |
| Population of nearest town (log) | 0.0966* (0.0500) | -0.0348 (0.0443) | 0.0383 (0.0813) | -0.3025** (0.1218) | 10.7529 (1.1613) |
| Elected panchayat | 0.2123** (0.0870) | | | | 0.8269 (0.3792) |
| Distance to district HQ | -0.0013 (0.0009) | | | | 50.4952 (38.1052) |
| Distance to wholesale market | 0.0082*** (0.0022) | 0.0021 (0.0020) | | | 10.1490 (13.3540) |
| Distance to retail market | | | | -0.0032 (0.0073) | 8.4808 (9.4455) |
| Distance to bus stop | -0.0114** (0.0050) | | | -0.0140 (0.0118) | 3.3606 (5.7576) |
| Distance to railway station | -0.0041*** (0.0013) | | | | 22.5192 (24.0816) |
| Tenant cultivation | | 0.1095** (0.0525) | -1.0415** (0.4893) | | 0.4808 (0.5008) |
| Share cropping | | | -1.3338*** | | 0.5577 |

| | (1) | (2) | (3) | (4) | |
|---|---|---|---|---|---|
| | **Income premia** | **VILS cultivators** | **VILS agr. labor** | **VILS hhd. industry** | **Mean** |
| | | | (0.5066) | | (0.4979) |
| Share cropping x Mean hhd. size | | | 0.2229** (0.0886) | | 3.2907 (3.0536) |
| Tenant cultivation x Mean hhd. size | | | 0.1879** (0.0835) | | 2.6358 (2.8625) |
| Mean yield rate | | 0.0002 (0.0020) | | | 16.9320 (12.1171) |
| Mean yield x distance to market | | -0.0083* (0.0043) | | | 1.4468 (5.5545) |
| Number of factories | | | | -0.0098*** (0.0035) | 10.3173 (23.1983) |
| Number of shops/service units | | | | 0.0083*** (0.0011) | 54.8317 (74.8044) |
| Constant | -0.8004 (0.6840) | 0.8218 (0.5572) | 2.4140** (1.1553) | 1.6257 (1.4953) | |
| Observations | 208 | | | | |
| Log-likelihood value | -638.25 | | | | |
| Breusch-Pagan test | 32.53 | | | | |

*Notes: A. Figures in parentheses in columns 1-4 are standard errors and standard deviations for the mean values of the explanatory variables in col. 5. * , **, and *** denote significance at the 10%, 5%, and 1% level respectively. B. The lagged dependent variable is the 1982 value of village income premia, VILS for cultivators, VILS for agricultural labor and VILS for household industry in columns 1-4 respectively. VILS refers to the village-to-taluk index of labor specialization. The means (standard deviation) of the lagged (1982) values are -0.0452 (0.5132), 1.0147 (0.3613), 1.0665 (0.7890) and 1.2678 (1.4739) for these variables respectively. The mean (standard deviation) of the current (1999) values are –0.0657 (0.5044), 1.0240 (0.4066), 1.0014 (0.7083) and 1.0601 (1.2885) for the four dependent variables respectively. C. TILS refers to the taluk-to-rainfall region index of labor specialization.*

The Breusch-Pagan test rejects the null of the independence of the equations (i.e., whether the covariance matrix of the errors is diagonal) ($\chi^2$ = 32.53) thus supporting simultaneous estimation using SUR rather than separate OLS equations.[17]

Initial conditions explain inequality in general and spatial inequality in particular (see for example Deininger and Squire (1998) and Datt and Ravallion (2002)). Initial conditions with respect to a range of structural characteristics have been included: the initial (1982) stock of endowments of labor (as proxied by average age, household size, literacy rate), land (as proxied by average holdings, land tenure systems and structure of land holdings with respect to farm size), infrastructure (as captured by education, health and irrigation facilities), institutions (i.e., local governments) and rainfall patterns. The geographic endowment of the village in economic space as measured by the relative diversification of the *taluk* with respect to rainfall region and the population of the nearest town are also incorporated. Only a subset of these variables is expected to affect all four dependent variables. These include literacy (both years), mean household size, mean age of the household head, as well as the population of the nearest town (both years). The lagged value of the village income premia and the village-to-*taluk* ILS values for cultivators, agricultural labor and household industry are also used as explanatory variables in their respective equations in order to capture historical patterns. Variables capturing the level of current integration in the wider economy and local village economic conditions have also been taken into account. The dependent variables are the village income premia for 1999 and the VILS values for the three categories for 2001.

The results show that the initial values of the dependent variable are significant determinants of the current values indicating considerable persistence in location premia and VILS values over time (see also Table 2 and Figure 4). Initial conditions with respect to other structural variables are also found to play an important role.

Government policy in India has focused on reducing spatial inequality. These policies include land reform, providing access to rural non-farm employment through training, credit extension services for setting up non-agricultural businesses and encouraging industrial location in rural areas, among others. The role of such policies in poverty alleviation has been widely stressed in the literature and needs no repetition here (see for example Datt and Ravallion 1998). In keeping with this literature, the results show that a greater proportion of non-agricultural households in a village increases the relative income of the village. Conversely, a higher proportion of cultivating households reduces the village income premia. The effect of literacy rates and the size of the nearest town can also be viewed in the same light. The former can be viewed as a retardant to entry into agricultural labor force and could also, with a time lag, affect cultivation especially if agriculture is labor-intensive. As the barriers to entry into and exit from agricultural labor are low, the current literacy rate has a significant negative impact on the VILS for agricultural labor as opposed to initial conditions. The reverse is true for the VILS for household industry and cultivators.[18]

The size of the nearest town can be viewed both as a market for village household industry and as a source of alternative employment opportunities. The current and initial population of the nearest town have two different roles to play, possibly due to the rise in the other non-agricultural activities (including manufacturing, services, etc.) that are not limited to household industry, which is apparent in several of these villages. As a result, a large town may have been perceived as a market and encouraged specialization in household industry in 1982. By 1999, however, with the rise in alternative employment opportunities, proximity to a large town would have tended to encourage specialization in other non-agricultural and non-household industry activities. The same explanation applies to the effects on the income premia.

Infrastructure in the form of health and educational facilities is expected to affect the village income premia directly, while that in the form of irrigated facilities (i.e., proportion of cropped area irrigated) is expected to affect the income premia indirectly through its effect on the VILS for cultivators and agricultural labor. Infrastructure as captured by health facilities has a significant positive effect on the income premia though the proximity of schools have an insignificant effect.[19] The proportion of irrigated area – a measure of the availability of irrigation facilities – has the expected positive effect on the VILS for cultivators. Regional rainfall patterns are also significant determinants for the VILS for cultivators.

Current land tenure systems affect both cultivating households and agricultural labor. Conferring tenancy rights is likely to increase the specialization in cultivation and reduce that in agricultural labor as typically tenants are marginal and small farmers. The same argument applies to the prevalence of share cropping in a village. With increasing household size households engaged in share cropping and tenancy arrangements are likely to diversify away from cultivation and participate instead in the agricultural labor market. Initial values of land tenure systems are found to have no significant effect, possibly reflecting the usual long time lag between legislation and implementation.

The structure of land holdings with respect to farm size also has a significant effect on activity specialization patterns in agriculture. A greater proportion of large farmers increases the VILS for cultivators as these are often historically determined patterns of land ownership so that the presence of several large farmers creates a culture of specialization in cultivation in the village. On the other hand, a greater proportion of marginal farmers increases the VILS for agricultural labor as these farmers are particularly vulnerable to shocks and typically move into agricultural labor in times of low income and stress. Similarly, the proportion of landless households increases the VILS for both agricultural labor and household industry as these households turn to activities other than cultivation given their lack of access to land, though this effect is mitigated somewhat by the presence of share cropping.

Local geographic endowments are also important as these can give rise to externalities with respect to networks, supply chains, markets, etc. A village located in a high-performing cultivating area tends to have a higher relative income. The current conditions of better connectivity, with respect to proximity to bus and railway stations and administrative head quarters, are instrumental in raising the village income premia. It seems as though the supply elasticity of the output sold by these villages decreases with distance and that buyers are generally distant buyers (e.g., the government). This is reflected in the positive impact of the distance to wholesale markets in the income premia equation as farmers are able to discriminate against distant buyers.[20] It is possible to discern that current yield rates are likely to be sub-optimal as cultivators would be negatively affected if the scale at which they can sell in distant markets is small. Thus, local supply conditions matter for cultivating households. Additionally, the local village labor demand conditions as captured by the number of factories and shops and/or service units are significant determinants of the VILS for household industry. The former provide a source for alternative employment and thus reduce the VILS for household industry while the latter are an indication of the demand for the products of household industry and are found to increase the VILS for household industry.

Finally, do institutions for local governance matter; specifically, do elected institutions matter? The 73rd Amendment which gave logic, form and constitutionality to local governments came into force only in 1996. Correspondingly, it was observed that the presence of elected *panchayats*[21] does significantly raise the income premia as this form of local government is more likely to provide an appropriate mix of public goods, depending on the representation of the local populace in the government.

To summarize, there is some evidence that spatial inequality among our sample villages is jointly determined by the instruments of village income premia and the village index of labor specialization. The results suggest a high degree of persistence – low-income villages relative to the average village in 1982 will tend to remain low income in 1999 while villages specializing in cultivation, agricultural labor or household industry in 1982 will continue to do so in 1999. Moreover, initial conditions with respect to the village labor endowment (such as literacy rates and household size), geographic endowment (such as rainfall and the relative performance of the *taluk*), and public facilities (such as health and irrigation infrastructure) as well as the asset and activity composition in the village are significant determinants of income premia and/or the VILS for cultivators, agricultural labor and/or household industry. In addition, the current level of labor endowment (with respect to literacy rates and household size), local governance, and indicators of the current vibrancy of the local village economy with respect to integration with the wider region and the presence of economic opportunities also significantly influence spatial inequality.

## 6. Conclusion

The analysis in this paper suggests that there is high degree of spatial variation in incomes among villages in India that is both persistent and rising during the period between 1982 and 1999. Thus, households resident in a village where the returns to location are high would benefit from the positive externalities and even poor households would have better long-term income prospects. Alternatively, poor households located in villages where negative externalities generate relatively persistently low returns to location would be more likely to stay chronically poor. Thus, such a phenomenon can produce the requisite preconditions for chronic poverty by reinforcing the inability of poor households to escape low-income status.

This paper posits that spatial inequality is a jointly determined outcome. Factors governing village income premia (a quality of life indicator) and clustering of activities (a result of specialization, skills, preferences, culture, and structures that prevent mobility) together determine spatial inequality. If households were free to move in response to differential returns to location we would expect spatial inequality in the welfare sense to disappear. The literature on village studies in general and mutual insurance in particular along with the limited data on migration from the village survey and the Census of India suggests that there is at best very limited mobility across villages even within the same state.

The foregoing discussion implies that designing policies to combat spatial inequality is unlikely to be straightforward. For instance, if the underlying driver of the persistence in spatial inequality as captured by low relative returns to location is low productivity, then the appropriate policy response could be to lower barriers to mobility and encourage out-migration. If, however, this low productivity is a function of low returns to assets due to asset constraints and imperfections in the asset markets, the appropriate policy could be to improve the functioning of asset and credit markets, encourage investment and reduce the dependency on less spatially transferable social networks in such areas. Thus, framing appropriate policies requires the identification of the underlying economic drivers of productivity differences across regions. This paper provides the first step in that direction by identifying several structural variables that influence the village income premia and activity specialization patterns. The importance of initial levels of labor, geographic and institutional endowments indicates that there are differential returns to assets across villages. It remains to be seen whether these are, in turn, driven by imperfections in asset markets. Our results suggest that government policies relating to land tenure and ownership patterns, local governance structures, rural literacy and rural non-farm employment opportunities have a significant role to play in reducing spatial inequality.

## References

Ahluwalia, M.S. 2002. 'State Level performance Under Economic Reforms in India', in Anne Krueger (ed.), *Economic Policy Reforms and the Indian* Economy. Chicago: University of Chicago Press.

Alesina, A. and Rodrik, D. 1994. 'Distributive Politics and Economic Growth', *Quarterly Journal of Economics,* 109(2), 465-490.

Andalón, L.M. and López-Calva, L.F. 2002. 'The Many Mexicos: Income Inequality and Polarization in Urban Mexico during the 90s', Paper presented at the Comell-LSE-WIDER Conference on Spatial Inequality, London.

Anderson, K. and Pomfret, R. 2003. 'Spatial Inequality and Development in Central Asia', Paper presented at the UNU/WIDER Project Conference on Spatial Inequality in Asia, Tokyo.

Arbache, Jorge Saba; Dickerson, Andy and Green, F. 2004. 'Trade Liberalisation and Wages in Developing Countries', *The Economic Journal,* 114(493), F73-F96.

Balisacan, A.M. and Fuwa, N. 2003. 'Is Spatial Income Inequality Increasing in the Philippines?', Paper presented at the UNU/WIDER Project Conference on Spatial Inequality in Asia, Tokyo.

Benabou, R. 1996. 'Inequality and growth', in B.S Bernanke and J.J. Rotemberg (ed.), *NBER Macroeconomics Annual.* Cambridge, Mass: MIT Press.

Borooah, Vani K. 2005. 'Caste, Inequality, and Poverty in India', *Review of Development Economics,* 9(3), 399-414.

Cashin, Paul and Sahay, Ratna 1996. 'International Migration, Center-State grants, and Economic Growth in the States of India', *International Monetary Fund Staff Papers*, 43(1).

Datt, Gaurav and Ravallion, Martin 2002. 'Is India's Economic Growth Leaving the Poor Behind?', (Unpublished Manuscript).

Datt, Gaurav and Ravallion, Martin 1998. 'Why Have Some Indian States Done Better Than Others at Reducing Rural Poverty?', *Economica,* 65(257), 17-38.

Davis, Donald R. and David E. Weinstein. 1999. 'Economic Geography and Regional Production Structure: An Empirical Investigation', *European Economic Review,* 43(2), 379-407.

Deininger, Klaus and P. Olinto. 2000. 'Asset distribution, Inequality, and Growth', Policy Research Working Paper, WPS 2375, World Bank.

Deininger, Klaus and Squire, Lynn 1998. 'New Ways of Looking at Old Issues: Inequality and Growth', *Journal of Development Economics,* 57, 259-287.

Eeckhout, Jan and Jovanovic, B. 2001. 'Knowledge Spillovers and Inequality', Working Paper, University of Pennsylvania.

Elbers, C., Lanjouw, P., Mistiaen, J., Ozler, B.and Simler, K.R. 2005. 'Are Neighbours Equal? Estimating Local Inequality in Three Developing Countries', in R. Kanbur and A.J. Venables (ed.), *Spatial Inequality and Development.* Oxford: Oxford University Press.

Engerman, Stanley L. and Sokoloff, K.L. 1998. 'Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View from Economic Historians of the United States?', in Stephen Haber (ed.), *How Latin American Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914*. Stanford, CA: Stanford University Press.

Escobal, J. and Torero, M. 1999. 'How to Face an Adverse Geography?: The role of Private and Public Assets', (Unpublished Manuscript).

Fan, C. and Casetti, E. 1994. 'The Spatial and Temporal Dynamics of U.S. Regional Income Inequality, 1950-1989', *Annals of Regional Science,* 28(2), 177-196.

Fields, Gary S. 1980. *Poverty, Inequality and Development*. Cambridge: Cambridge University Press.

Fleisher, M.B. and Chen, J. 1996. 'Regional Income Inequality and Economic Growth in China', *Journal of Comparative Economics,* 22, 141-164.

Foster, A.D. and Rosenweig, M.R. 2003. 'Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000', Paper presented at the D. Gale Johnson Memorial Conference, October 25.

Fujita, M., Krugman, P. and Venables, A.J. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, MA: MIT Press.

Galor, O. and Zeira, J. 1993. 'Income Distribution and Macroeconomics', *Review of Economic Studies,* 60, 35-52.

Gaston, N. and Trefler, D. 1994. 'Protection, Trade and Wages: Evidence from US Manufacturing', *Industrial and Labor Relations Review,* 47(4), 547-593.

Ghosh, B. and De, P. 2000. 'Linkage between Infrastructure and Income among Indian States: A Tale of Rising Disparity Since Independence', *Indian Journal of Applied Economics*, 391-431.

Glaeser, Edward L., Kallal, H., Scheinkman, J. and Shleifer, A. 1992. 'Growth in Cities', *Journal of Political Economy,* 100(6), 1126-1152.

Greene, William H. 1990. *Econometric Analysis*. Macmillan Publishing Company.

Haisken-DeNew, J.P. and Schmidt, C.M. 1997. 'Inter-Industry and Inter-Region Differentials: Mechanics and Interpretation', *Review of Economics and Statistics,* 79, 516-521.

Hare, D. and West, L.A. 1999. 'Spatial Patterns in China's Rural Industrial Growth and Prospects for the Alleviation of Regional Income Inequality', *Journal of Comparative Economics,* 27, 475-495.

Heltberg, R. 2003. 'Spatial Inequality in Vietnam: A Regression-based Decomposition', (Unpublished Manuscript).

Henderson, J.V., Shalizi, Z. and Venables, A.J. 2001. 'Geography and Development', *Journal of Economic Geography,* 1(1), 81-105.

Jalan, J. and Ravallion, M. 1998. 'Geographic Poverty Traps?', Institute for Economic Development Discussion Paper, 86, Boston University.

Kanbur, Ravi and Zhang, Xiaobo 2005. 'Fifty Years of Regional Inequality in China: A Journey through Revolution, Reform and Openness', *Review of Developmant Economics,* 9(1), 87-106.

Krueger, Alan B. and Summers, Lawrence H. 1988. 'Efficiency Wages and the Inter-Industry Wage Structure', *Econometrica,* 56(2), 259-293.

Miller, M., Gibson, L. and Wright, N.G. 1991. 'Location Quotient: A Basic Tool for Economic Development Analysis', *Economic Development Review,* 9(2), 65-68.

Mookherjee, Dilip and Ray, D. 2002. 'Is Equality Stable?', *American Economic Review* (Papers and Proceedings), 92(2), 253-259.

Munshi, Kaivan and Rosenzweig, Mark R. 2005. 'Why is Mobility in India so Low? Social Insurance, Inequality and Growth', CID Working Paper 121.

Nazara, Suahasil 2003. 'Spatial Inequality Analysis within a Regional Growth Decomposition Technique Indonesian case, 1976-1998', REAL Working Paper 03-T-11.

Nazara Suahasil and Hewings, G.D. 2003. 'Towards Regional Growth Decomposition with Neighbor's Effect: A New Perspective on Shift-Share Analysis', Working Paper.

Noorbakhsh, F. 2003. 'Spatial Inequality and Polarization in India', CREDIT Research Paper 03/16.

Özmucur, S. and Silber, J. 2002. 'Spatial Income Inequality in Turkey and the Impact of Internal Migration', (Unpublished Manuscript).

Perotti, R. 1993. 'Political Equilibrium, Income Distribution and Growth', *Review of Economic Studies*, 755-76.

Prakash, B.A. 1998. 'Gulf Migration and Its Economic Impact: The Kerala Experience', *Economic and Political Weekly,* December 12, 3210.

Rainwater, L. 1992. 'Social Inequality in Europe and the Challenge to Social Science', in M. Dierkes and B. Bievert (ed.), *European Social Science in Transition: Assessment and Outlook*. Frankfurt am Main: Campus Verlag, Boulder, CO: Westview Press.

Raiser, M. 1998. 'Subsidising Inequality: Economic Reforms, Fiscal Transfers and Convergence Across Chinese Provinces', *The Journal of Development Studies,* 34(3), 1-26.

Ravallion, Martin M. 2002. 'Externalities in Rural Development: Evidence for China', (Unpublished Manuscript).

Ravallion, Martin 2001. 'Growth, Inequality and Poverty: Looking Beyond Averages', *World Development,* 29(11), 1803-1815.

Sachs, Jeffrey 2003. 'Institutions Don't Rule: Direct Effects of Geography on Per Capita Income', Working Paper 9490, National Bureau of Economic Research.

Sachs, Jeffrey; Bajpai, Nirupam and Ramiah, Ananthi 2002a. 'Why Have Some Indian States Grown Faster than the Others?', Centre for International Development, Harvard University.

Sachs, Jeffrey; Bajpai, Nirupam and Ramiah, Ananthi 2002b. 'Why Have Some Indian States Grown Faster than the Others?', Rediff Budget 2002-2003, February 26.

Shand, Ric; Kalirajan, K. and Govinda Rao, M. 1999. 'Convergence of Incomes Across States: A Divergent View', *Economic and Political Weekly*, 34, 769-778.

Tadjoeddin, M.Z., Suharyo, W.L. and Mishra, S. 2003. 'Aspiration to Inequality: Regional Disparity and Centre-Regional Conflicts in Indonesia', Paper presented at the WIDER-UNU Conference on Spatial Inequality in Asia.

Wan, Guanghua and Zhou, Zhangyue 2005. 'Income Inequality in Rural China: Regression-based Decomposition using Household Data', *Review of Development Economics*, 9(1), 107-120.

Zhang, X. and Zhang, K.H. 2003. 'How does Globalisation Affect Regional Inequality within a Developing Country? Evidence from China', *The Journal of Development Studies*, 39(4), 47-67.

## Endnotes:

---

[1] The decompositions differ according to the inequality measure used because of the differences in the sensitivity to different parts of the income distribution and in the weights used to summarise within-group inequality.

[2] The other source for large survey data in India - National Sample Survey Organization (NSSO) – does not collect data that are representative at the village level, do not have a village questionnaire and do not permit the identification of villages by name.

[3] A *taluk* is a jurisdictional boundary within a district.

[4] Only 31 (16) villages have a sample size of less than 100 households in 1982 (1999).

[5] This is necessarily an *ad hoc* measure: some researchers prefer to trim the wage distribution using specific values (Krueger and Summers 1988) as adopted here, while others prefer to trim a percentage of the distribution at the tails (Arbache, Dickerson and Green 2004).

[6] The land data was reported in hectares in 1982 and in acres in 1999 as well as some local land units (e.g., bigha, cent, kanal, katha) in both years. These were converted to acres using the appropriate conversion factors. Note that the data on household land refers to land operated in 1982 and to land owned in 1999.

[7] These include four cultivator categories (marginal, small, medium and large farmers), agricultural labor (this the omitted reference category), fishing, animal husbandry, non-agricultural white-collar labor, non-agricultural blue-collar labor, non-agricultural business and transfer income. Marginal farmers cultivate land up to two acres, small farmers between two to four acres, medium farmers between four and ten acres and large farmers cultivate ten or more acres of land.

[8] It would have been desirable to repeat this analysis for other activities, such as manufacturing, trade and transport, construction, etc.. However, it is not possible to do so as the Census classifications have changed over time with the 1981 and 2001 Censuses combining all activities other than the three analysed here into a residual 'other' category.

[9] The Census defines main workers as those who have worked at least 183 days in a year in any economic activity.

[10] See Figure A1 in the Appendix.

[11] This represents the impact of the $(K+1)^{th}$ omitted village (assumed to be zero) so that the income differential for this village reduces to the (negative) size-weighted mean differential.

[12] The village population-to-basic employment ratios are defined as $P_v \Big/ \sum_{i=1}^{3} E_v^i$ ; where *P* represents the population of the village, the subscript *v* the village and superscript i the activities (cultivators, agricultural labor and household industry) and E the total number of main workers.

[13] Village premia were also computed using alternative weights (village population taken from the Census and village cell-sizes from the listing sheet data) as a robustness check. The overall picture with respect to dispersion and persistence over time is unchanged. The use of village area (from the Census) was preferred as these weights do not change over time so that no portion of the change in the income premia can be attributed to a change in the weights.

[14] Village premia for all villages are available on request.

[15] A Wald-type test for whether the income premia estimated in 1982 were significantly different from those in 1999 was set up as follows: $\chi_k^2$ statistic = $[\hat{\delta}_{t1}^* - \hat{\delta}_{t2}^*]'[\mathrm{var}(\hat{\delta}_{t1}^*) + \mathrm{var}(\hat{\delta}_{t2}^*)]^{-1}[\hat{\delta}_{t1}^* - \hat{\delta}_{t2}^*]$ where $\hat{\delta}_t^*$ is the (k+1x1) vector of the estimated village premia in year *t* (*t*=1,2) and *var*($\hat{\delta}_t^*$) is the (k+1 x k+1) variance-covariance matrix of these premia and *k* is the number of degrees of freedom. $\chi_k^2$ statistics are 30208.08 between 1982 and 1999.

[16] This is computed as follows: $SD(\hat{\delta}^*) = \sqrt{s'(Diag(\hat{\delta}^*))\hat{\delta}^* - s'Col(V(\hat{\delta}^*))}$; where *Diag(.)* transforms the (K+1x1) column vector into a (K+1xK+1) square matrix with the diagonal elements given by the column vector and *Col(.)* denotes the column vector formed by the diagonal elements of the matrix.

[17] OLS estimations of these equations yield very similar results. Given that the village income premia are themselves estimated from a first-stage household income regression model, we also estimated a weighted least squares model using the inverse of the variance of these premia as weights. The results were very similar to the separate OLS estimation and to the SUR estimates reported above. These additional results are available from the authors.

[18] The coefficient on the initial literacy rate is significant at the 16% level in the VILS for cultivators.

[19] This can almost certainly be attributed to low variation across our sample. These villages are old and have relatively well-developed infrastructure. For instance, 97% of the sample had at least one school within 2 km of the village.

[20] Anecdotal evidence also suggests that markets for certain high-value products such as turmeric are fragmented and very few in number. Hence, being located at a great distance from these select markets will not affect a household at the margin or may even have a positive impact.

[21] *Panchayats* are local village councils in rural India.

## APPENDIX

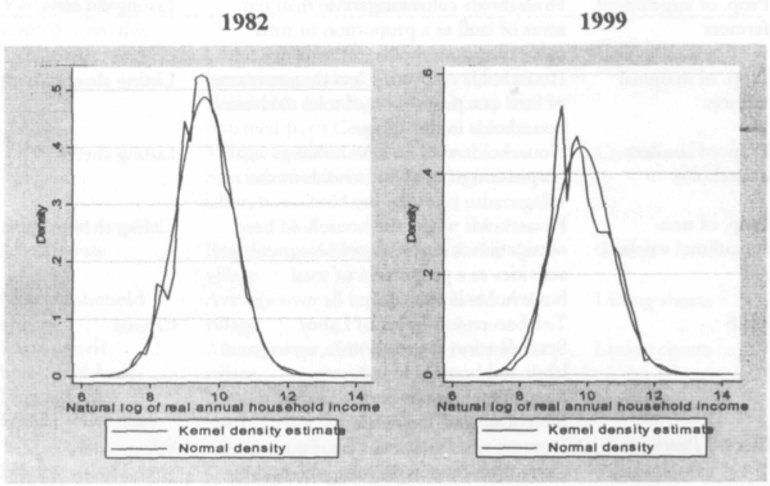
### Table A1: Variable definition and source for SUR estimation

| Variable name | Definition | Source |
|---|---|---|
| *Dependent variables:* | | |
| Village income premia | Obtained from first-stage household income regression<br>Weights (total village area in acres) obtained from Census | Listing sheets |
| VILS | Village-to-*taluk* Index of Labor Specialization for cultivators, agricultural labor, household industry and others | Census |
| *Independent variables:* | | |
| Literacy rate | Proportion of literate population in the village | Census |
| Mean household size | Average over all households listed in the village | Listing sheets |
| Mean age of household head | Average over all households listed in the village | Listing sheets |
| Population of nearest town (log) | Natural log of the population of the nearest town (as identified from the village survey and Census) | Census |
| School within 2 km | Coded one if there is a school of any type within 2 km of the village, zero otherwise | Village survey |
| Health facility within 2 km | Coded one if there is a health facility (hospital, clinic, health Centre, etc.) within 2 km of the village, zero otherwise | Village survey |
| Mean rainfall | Mean rainfall (mm) over the ten-years preceding the survey | NCAER - rain |
| Variability of rainfall | Standard deviation of rainfall over the ten-years preceding the survey | NCAER - rain |
| Prop. Of irrigated area | Irrigated area as proportion of cropped area | Village survey |
| Mean size of land holdings | Average size of land holdings over all land-owning households listed in the village | Listing sheets |
| Tenant cultivation | Dummy variable coded one if tenant cultivation is practiced in the village, zero otherwise | Village survey |
| Share cropping | Dummy variable coded one if share cropping is practiced in the village, zero otherwise | Village survey |
| Tenant cultivation x Mean hhd. Size | Interaction of tenant cultivation dummy with the mean household size | Village survey |
| Share cropping x Mean hhd. Size | Interaction of share cropping dummy with the mean household size | Village survey |

| | | |
|---|---|---|
| Prop. of cultivating households | Households where the household head occupation is reported as cultivation as a proportion of total households in the village | Listing sheets |
| Prop. of large farmers | Households cultivating more than ten acres of land as a proportion of total cultivating households in the village | Listing sheets |
| Prop. of marginal farmers | Households cultivating less than two acres of land as a proportion of total cultivating households in the village | Listing sheets |
| Prop. of landless households | Households with no land holdings as a proportion of total households in the village | Listing sheets |
| Prop. of non-agricultural workers | Households where the household head occupation is reported as non-agricultural activities as a proportion of total households in the village | Listing sheets |
| TILS | *Taluk*-to-region Index of Labor Specialization for cultivators, agricultural labor and household industry. Rainfall regions are as classified by the Meteorological Department | Census |
| Elected *Panchayat* | Dummy variable coded one if there is an elected *panchayat* in the village, zero otherwise | Village survey |
| Distance to district HQ | Distance (km) to district headquarters | Village survey |
| Distance to wholesale markets | Distance (km) to the nearest wholesale market (either *mandi* or weekly *haat*) | Village survey |
| Distance to retail markets | Distance (km) to the nearest retail market | Village survey |
| Distance to bus stop | Distance (km) to bus stop | Village survey |
| Distance to railway station | Distance (km) to railway station | Village survey |
| Mean yield rate | Average yield for major crops | Village survey |
| Mean yield rate x distance to markets | Interaction of average yield for major crops with the distance to wholesale markets | Village survey |
| Number of factories | Number of factories (including flour mill, lime kiln, *gur*-making unit, etc.) in the village | Village survey |
| Number of shops/service units | Number of shops/service units (including tea shops, grocery shops, eating houses, etc.) in the village | Village survey |

Note on data sources: Listing sheets and village survey refer to the ARIS-REDS data collected by NCAER (1981-82 and 1998-99); NCAER -rain to the NCAER dataset on annual actual and normal rainfall by meteorological region between 1970 and 2000; and Census refers to the Census of India (1971, 1981, 1991 and 2001).

**Fig. A1 : Kernel density plots for the natural log of real household income**




*Note: These kernel density plots are overlaid with the standard normal distribution*