AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON, D. C. 20523

BIBLIOGRAPHIC INPUT SHEET

FOR AID USE ONLY
Batch 55

1. SUBJECT CLASSIFICATION	A. PRIMARY: TEMPORARY
	B. SECONDARY

2. TITLE AND SUBTITLE
The determinants of income and its distribution in four villages in India

3. AUTHOR(S)
Lopez, Michael

4. DOCUMENT DATE	5. NUMBER OF PAGES	6. ARC NUMBER
1977	~~78p.~~ 81p.	ARC

7. REFERENCE ORGANIZATION NAME AND ADDRESS
Mich.

8. SUPPLEMENTARY NOTES *(Sponsoring Organization, Publishers, Availability)*
(In Discussion paper no.64)

9. ABSTRACT
(ECONOMICS R&D)

This report examines the determinants of the distribution of income among households by explaining the determinants of the income of each household. A model is presented in which income is a function of the resources or assets of each household. The distribution of income is an indirect function of the distribution and values of these assets. The article contains two major components: an analysis of the determination of each household's income and an analysis of the distribution of household incomes within the villages. Four villages were surveyed. The common characteristics of the villages were: 1. they were developing economically, 2. the irrigated area increased substantially in each of the villages, 3. all were the objects of government development programs, and 4. three of them were near enough to a city or town that some residents could commute to urban jobs. Simple linear regressions on variables representing land, labor, education, and milchstock account for 36% to 87% of the variance in household incomes per capita. The influence of caste is found to be primarily indirect. The role of economic factors varied greatly with occupation. In the second part of the paper, an original method, based on correction for the systematic accumulation of errors, refines the estimated income distribution obtained from the set of regression equation predictions for individual households. Inequality of land distribution is shown to be the only important factor in explaining income inequalities. The income distribution effects of a hypothetical land reform are simulated.

10. CONTROL NUMBER PN-AAD-391	11. PRICE OF DOCUMENT
12. DESCRIPTORS	13. PROJECT NUMBER
	14. CONTRACT NUMBER CSD-2547 211(d)
	15. TYPE OF DOCUMENT

AID 590-1 (4-74)

The Determinants of Income and its Distribution in Four Villages in India

Michael Lopez

CENTER FOR RESEARCH ON ECONOMIC DEVELOPMENT
The University of Michigan
Ann Arbor, Michigan 48108

Discussion Paper No. 64 February 1977

ABSTRACT

The Determinants of Income and its Distribution in Four Villages in India

The personal distribution of income is assumed to be a function of the values of factors of production, and their distribution among households.

In the paper's first part, simple linear regressions on variables representing land, labor, education, and milchstock account for 36% to 87% of the variance in household incomes per capita, for data from nine surveys. The influence of caste is found to be primarily indirect, through different caste groups possessing different amounts of resources.

Closer inspection reveals that the role of economic factors varied greatly with occupation.

In the paper's second part, an original method, based on correction for the systematic accumulation of errors, refines the estimated income distribution obtained from the set of regression equation predictions for individual households. Inequality of land distribution is shown to be the only important factor in explaining income inequalities.

The income distribution effects of a hypothetical land reform are simulated.

† † †

On assume que la distribution personnelle du revenu est une fonction des valeurs des facteurs de production et de leur répartition entre les ménages.

Dans la première partie de ce document, pour des données provenant de neuf enquêtes, les régressions linéaires simples sur les variables représentant la terre, le travail, l'éducation et les procuits laitiers représentent 36 à 87% des variations des revenus familiaux par personne. On a trouvé que l'influence de caste est essentiellement indirecte et ceci à travers des groupes de caste différents possédant différentes quantités de ressources.

Un examen plus attentif révèle que le rôle des facteurs économiques varie beaucoup avec l'occupation.

Dans la seconde partie du document, une méthode originale, basée sur la correction de l'accumulation systématique d'erreurs, épure la distribution estimée du revenu obtenue à partir de l'ensemble des prédictions des équations régressives sur les ménages individuels. L'inégalité de la répartition des terres s'avère être le seul facteur important expliquant les inégalités de revenu.

Les effets d'un programme de réforme agraire hypothétique sur la répartition de revenu sont simulés.

The Determinants of Income and its Distribution in Four Villages in India*

Michael Lopez**

Introduction

The welfare of a people is clearly dependent not only upon the sum total of national income, but also on how it is distributed. For most countries, however, there is little knowledge of how the present distribution came about, or how it might be affected by future policies. This is especially true for less developed countries, where statistics are scant. Most of the literature on income distribution has analyzed the distribution among factors of production. The problem with allocating income to the abstract entities of "labor", "land", "capital" and "human capital" is that individual households possess varying combinations of these resources, especially in rural areas. For example, since the large landlords usually have above-average education, without information

---

* I would like to thank Professors Robert Evenson and Guy Orcutt, then both of Yale University, for their extensive comments on an earlier version of this work. In India, where I went to gather data in 1972, the staffs of the Agro-Economic Research Centres were extraordinarily hospitable. At Delhi, I received help from the Deputy Director, Dr. H. Laxminarayan, as well as from V. P. Bahl and S. S. Tyagim among others. In Vallabh Vidyanagar the Deputy Director, Dr. M. D. Desai and D. M. Brahmbhatt, R. M. Patel, and others were exceptionally kind and patient. Several Indian graduate students helped me transcribe data: Manjit Gandhi and Ravi Verma in Delhi, and K. D. Vankar in Vallabh Vidyanagar, worked diligently and thoughtfully. Back in the United States, Howard Gilbert, Terry O'Conner, and Joseph Vitale of the User Services Staff at the Yale Computer Center provided programming advice. Financial assistance for the research and computer work was received from the Yale Economics Department and the Yale Economic Growth Center, from the Agency for International Development, U.S. Department of State (Grant AID-CSD-2492), and from the National Science Foundation (Grant NSF FS-36863). It should also go without saying that the opinions expressed here are mine, and not the viewpoints of the various agencies which funded the research; and that mine, too, is the responsibility for the errors which remain despite the help I have received.

** Maitre Assistant Associé, Centre des Recherches Economiques Appliquées, Faculté des Sciences Juridiques et Economiques, University of Dakar, Senegal, and Assistant Research Scientist, Center for Research on Economic Development, the University of Michigan.

on the exact combinations of factors it is impossible to disentangle the effect of education from the effect of land.

My belief is that the best approach to an explanation of the determinants of the distribution of income among households is to start by explaining determinants of the income of each household. Specifically, I shall present a model in which each household possesses certain resources or assets (such as land and education), and in which income is a function of these assets. The distribution of income is an indirect function of the distribution and values of these assets. Therefore this article contains two major components: first, an analysis of the determination of each household's income, and second, an analysis of the distribution of household incomes within the villages.

Review of the Literature

I have been able to find only four empirical analyses of income distributions which adopt an approach similar to mine: one for the United States by F. Gerard Adams, one for Great Britain by T. P. Hill, one for Denmark by Kjeld H. Bjerke, and one for Bombay by V. R. and P. R. Panchamukhi. In "The Size of Individual Incomes: Socio-Economic Variables and Chance Variations"[1], Adams took U.S. data from 1949, limited his sample to white males, and regressed an income-determining equation using dummy variables for age, age squared, geographical location (South or non-South; and country, town and city), type of job (e.g., blue-collar and white-collar), and extent of employment (more or less than eleven months). Hill's study of April 1953 to March 1954 data, "An Analysis of the Distribution of Wages and Salaries in Great Britain"[2], is similar, with dummy variables for age, geographical location, and education, plus an elaborate scheme of dummy variables to represent occupations. Similar to these two articles, except for the conspicuous absence of education as an explanatory variable, is Bjerke's "An Analysis of the Personal Income Distribution for Wage and Salary Earners in 1955"[3], concerning the wage and salary incomes received by the heads of a sample of urban Danish households. The article "Socio-Economic Variables and Urban Incomes"[4] by the Panchamukhi brothers includes regressions of the incomes of 22,859 residents of Bombay City on no less than 47 dummy variables, representing categories of education, age, industry (e.g., construction), occupation (e.g., managerial), employment status (e.g., self-employed), family income and sex. The data appear to date to 1954, though the year is not explicitly stated. The purpose of the family income dummy variables was to measure the role of "connections" in obtaining jobs. This variable makes sense, but only if it is family income exclusive of the income of the person in question (for example, his parents' income). Obviously -- particularly for households with only one earner -- the results of a regression of individual income on total household income are meaningless. Ominously, the article does not mention subtraction of individual earnings from household earnings. An R^2 of 17% is reported for the linear regression, with the set of dummy variables for family income categories far more important than any other group of dummy variables (which would not be surprising if household income in fact included the dependent variable). Finally, it should be noted that the distribution of a factor like labor, through occupations,(and industries) may be thought of as proxies for the varying qualities of labor which they employ.

Source of data

The statistics which enable this analysis to bridge the usual gap between the factoral distribution of income and the personal distribution of income are a result of the Continuous Village Surveys project organized by the Indian Ministry of Food and Agriculture. Nine Agro-Economic Research Centres were organized, each affiliated with a university. The centres were instructed to conduct intensive socio-economic surveys of each household in selected villages in their regions. These villages were to be canvassed periodically, at (in practice) intervals ranging from four to eleven years. The basic goal was to obtain a continuing picture of the long-range effects of development, rather than the usual one-shot survey after a major change. For this reason the usual criterion for selection of a village was that an important change was expected to occur in the near future, but subsequent to the first survey: for example, a village in the path of an irrigation project.

From the available studies I selected four villages for analysis. There were several requirements for selection. The village had to have been surveyed at least twice (one was surveyed three times). These surveys had to be in years of reasonably ordinary weather. The quality of the data had to be good. Study of the original filled-in questionnaires and conversations with the people who had conducted the interviews gave otherwise unobtainable insights into the quality of the data. It was obvious that some of the surveys were conducted by a dedicated and well-supervised field staff, who accomplished the difficult task of convincing farmers that they were not spies for the tax bureau or the land reform agency; whereas for other surveys, these ideal conditions did not prevail. And of course, permission had to be secured from the Agro-Economic Center to transcribe and use their raw data.[5]

Description of the Four Villages[6]

ANKODIA was surveyed for the "crop-year" July 1960 to June 1961, when the sample included 269 households containing 1,515 individuals (an additional fifteen households were absent from Ankodia when the interviews were conducted and another household refused to answer questions about its income).[7] Ankodia was surveyed again for the crop-year 1968-69, when the sample included all 298 households in the village, with their population of 1,718. This was a relatively prosperous village, growing a mixture of cash and food crops; it is located in Gujarat state, about 250 miles north of Bombay. Between the two surveys, a milk cooperative was established in the village, improving prices received by milk producers. The road connecting the village to the city of Baroda nine miles away was paved and bus service was established. Electricity was brought to Ankodia. The cheapness of running electric pumps compared to diesel pumps contributed to an increase in the gross area irrigated, from 40% to 63% of the gross crop area.[8]

BHATIAN was surveyed for three crop years: 1955-56, when there were 481 inhabitants in 94 households; 1960-61, when the sample included 80 households with 526 inhabitants (one other household with seven members was exlcuded due to missing income data); and 1971-72, when there were 893 inhabitants in 151 households. The farmers in the village were refugees from Pakistan, they

had been resettled in Bhatian because the village farmland had been abandoned by its former Muslim owners, who had fled to Pakistan. The refugees were all members of the Sikh religion, a sect whose relationship to Hinduism is roughly analogous to the relationship of Protestantism to Catholicism, both because Sikhism had its genesis in a reform movement and because Sikhs have a belief in the intrinsic morality of hard work not unlike the "Protestant Ethic"[9]. Both before and after India's independence, the landless laborers and artisans in the village were Hindus.

In 1955-56, the time of the first survey, the refugees had arrived recently enough so that they had not yet brought all of their land allotments under cultivation. However, even then Bhatian was economically advanced. The village had been electrified and already contained twenty tubewells. Bus service was available to the city of Ludhiana, just four miles away, on the highway which passes by Bhatian. By 1960-1961, the time of the second survey, all of the land had been brought under cultivation and seven more tubewells had been put into operation. Nine landowners had moved their households to nearby cities. Most of these still retained control over their land, farming with hired labor instead of renting out the land or selling it. But because these households did not live in Bhatian they were not included in the second survey. Aside from the lure of "city lights" for some landowners, the effects of the growth of Ludhiana were limited to an induced change in the crop pattern. In the eleven-year interval between the second and third surveys further growth of Ludhiana -- it became a burgeoning manufacturing center of bicycles and agricultural machinery -- provided employment opportunities for landless laborers, more than off-setting the downwards pressure on wages caused by the immigration of workers from other parts of India. Many of the 74 households which moved into Bhatian between the second and third surveys had workers commuting by bicycle to jobs in Ludhiana. Since 1969 the state of Punjab, in which Bhatian is located, was in the center of the so-called "Green Revolution" in wheat. By 1971-72, the number of tubewells in the village had swollen to 37, and Bhatian's agriculture was immune from the current severe drought because 99% of its land was irrigated.

Finally, it should be noted that Bhatian has been the focus of numerous government development programs. The district of Ludhiana had been chosen as one of the Intensive Agricultural Development Program targets. Within the district, the "taluka" and "block" subdivisions including Bhatian had each been singled out to receive special attention, and Bhatian itself had been selected to be a Model Village. Perhaps more important than this series of programs was the fact that Bhatian is a few miles away from the Punjab Agricultural University. The University holds a Farmers Fair twice a year at which (for example) sample packages of the latest hybrid seeds are distributed at low prices. Although it should be pointed out that the Intensive Agricultural Development Program has been criticised as ineffective, for what it is worth Bhatian has had available far more government aid and information programs than an average Indian village, and is in that sense a prototype.

NAURANGDESHAR was surveyed for the crop-year 1961-62, when 192 households were included in the sample (twenty-five households were missing -- twenty

were absent from the village at the time the survey was conducted, and five refused to answer questions); and again for the crop-year 1968-69, when all 298 households then resident in the village were included in the sample. The village is located in an extremely arid part of the Rajasthen desert. At the time of the first survey, drinking water was literally more difficult to obtain than milk. On the other hand, land was relatively abundant, and a majority of the households either owned land outright or had received allotments from the state of Rajasthen, which in practical terms was selected for study because it was near the headwaters of a massive canal project. In 1962, the first trickles of water had arrived from the canal, irrigating 6% of the land area for only two weeks; by 1968-69 three-fifths of the land area was under irrigation, more local feeder channels were being built, and the village had been completely transformed. The increased requirements for labor on the newly irrigated land and the availability of a small amount of additional state land allotments had attracted a large number of immigrants, and the population of the village had increased from 1,133 (including the 125 members of the households omitted from the first survey) to 1,807. The village was located on a road which had been paved before the time of the first survey, with even then bus service to the town of Hanumangargh, 14 miles distant. Although electrification of the village began between the two surveys, it had neglibible effect on agricultural production, since the irrigation was not dependent on pumps.

SOHALPUR GARA was surveyed for the crop-year 1954-55, when there were 98 households with 443 inhabitants, and for the crop-year 1958-59, when there were 99 households with a population of 499. Virtually all of Sohalpur Gara's inhabitants were Muslims. Nevertheless, as frequently happens even in non-Hindu India, the society was differentiated into castes. Between the two surveys a sugar mill went into operation a mile and a half from the village (actually, it had been in operation even during the first-point survey, but it had not yet had an impact). Although the report prepared by the Agro-Economic Research Centre is not explicit on this point, it appears that the mill itself did not employ any villager. It did change the cropping pattern toward sugar-cane, but there were only small increases in the production of cane, both in terms of the area (from 8.5% to 11.1% of the gross area sown to all crops) and in terms of the value of output (from 29.2% to 31.6% of the value of all crops). Also associated with the construction of the mill was the paving of the road which went from Sohalpur Gara to the town of Rourkee; this was one factor behind the greater number of households having jobs outside the village during the second survey. The most important factor of change between the two surveys was probably not the mill, but the construction of state-owned tubewells and the associated irrigation ditches, which caused the portion of land irrigated to rise from zero to 18% of the gross sown area. Despite the improvements in the economic environment, productivity and incomes remained stagnant.

Some important features of the villages are summarized in Table 1. Certain characteristics were common across villages. All were developing economically, though in Sohalpur Gara the improvements in the economic infrastructure had not yet been translated into changes in income. The area irrigated increased substantially in each of the villages, but the irrigation technology varied: private tubewells in Ankodia and Bhatian, state tubewells in Sohalpur Gara, and a state canal in Naurangdeshar. To varying degrees (relatively more for Bhatian and Naurangdeshar, relatively less for Ankodia and Sohalpur Gara), all were the

object of government development programs. All except Naurangdeshar were near enough to a town or city so that at least a few residents commuted to urban jobs.

The fact that all these villages were progressive -- or at least the target of government development projects -- has to do with the policy of the Agro-Economic Research Centres to select villages where changes were expected to occur. Rarely, a stagnant village not the focus of government aid was chosen for comparative purposes. I was frustrated in my effort to include such a case in the data I brought back from India: either I was unable to obtain permission to copy the material, or the quality of the statistics was too poor to be usable.

Definition of Income

Before analysis of income determinants and its distribution, it is useful to look at what it actually was. Income, as defined here, included not only cash receipts but also receipts in kind, such as the free meals received by an agricultural laborer or the value of crops retained for home consumption by a farmer. Income was net of cash and kind expenses, such as the value of fodder crops grown by a farmer but fed to his livestock instead of sold. Only actual expenses were netted out: for example, with owner-operated farms no attempt was made to calculate a "shadow value" for family labor in order to subtract it from family earnings. Expenditures for capital improvement (like the purchase of land or new farm machinery) were not subtracted from current earnings. By the same token, receipts from the sale of assets (mainly livestock) were not included in the current income. Also excluded were transfer payments (usually remittances to the family from sons working elsewhere in India). The villagers were not subject to taxation, and the other taxes they paid were counted as expenses. Total household income was divided by the number of members resident in the household. In short, the concept used was current net income per capita after taxes and before transfer payments.

Precision of Income Data

Are these figures accurate? After observing the interviews for the third survey at Bhatian, speaking at length with staff members at Delhi and at Vallabh Vidyanagar who had prepared that and other surveys, and examining the filled-in questionnaires, my conclusion was that the four villages included here, the income statistics are in fact accurate. The first surveys are to the job, and because the villagers had to be convinced that these strangers were not evil government agents. (In some places -- not included here -- the village residents were not convinced on the first round, though by the second survey everyone realized that the previous visit of the Agro-Economic Research Centre staff had been harmless.) For all the villages, the least reliable figures are for "traders" ("merchants" in American usage), for whom imprecise expense estimates had to be subtracted from imprecise sales estimates; however, each village contained only a few merchants. In the first survey of Ankodia, home-grown fodder fed to the household's livestock was probably not subtracted as an expense, causing minor exaggeration to the income of cultivators.

Household Resources and Incomes

How is income determined? The approach used here is that each household possesses certain resources or assets, and that expected income streams flow from these assets.

The most important resources available to a household in a rural Indian context are land, labor, caste or social status, farm capital, and education or "human capital". Several measures of each of these resources were calculated from the available data, then one measure (or set of measures) was selected from each category by three criteria: a search for the most significant variables by a computer regression program, a desire to adopt common measures across all surveys when practicable, and the imposition of constraints based upon economic considerations.

Assuming that a household makes use of some or all of its resources in the production of earnings, the next question is -- How do the resources combine? Here, the simplest answer was chosen: that resources combine linearly. Algebraically:

$$(1) \quad \text{Income} = a + \sum_i (b_i \text{ Resource}_i) + \text{Error}$$

in which the "a" and "b_i" are estimated coefficients, and in which the usual econometric assumptions were made about the error term (essentially, that it averages out to zero and is random).[10] Since the dependent variable was household income per capita -- more relevant from a welfare point of view than total household income -- for consistency the explanatory variables were also expressed in per capita terms. For example, land was measured as hectares of land owned per capita. To avoid excess repetition, in the rest of this article the variables entering the regression equations will not always be called "per capita", but the reader should remember that they always are per capita.

Aside from the reasoning that some simple arrangement is the most logical way to start, the specified form of Equation (1) has two arguments in its favor. As a first approximation, it does seem plausible to say that resources contributed linearly to income. Since it was possible to rent land in or out (though land reform legislation had the effect of reducing the amount of land available for rental), and since there was a wage market for laborers, the rental rate and the wage rate at least tended to put a floor under the marginal revenue from a household's land and labor. Also, the estimate for income derived from Equation (1) can be decomposed into the sum of the separate effects of each explanatory variable, plus the constant term. This property of additivity will be useful in the subsequent analysis of income distribution. Complementing the advantages of a linear specification are the disadvantages of non-linear specification. Logarithms could not be taken -- as for the estimation of a Cobb-Douglas function -- because the independent variables often had a value of zero (e.g., no land owned), and because the dependent variable, net income per capita, was sometimes negative. The introduction of cross-product terms among independent variables introduced enough problems of multicollinearity so that it became difficult to interpret what the estimated coefficients meant.

Regressions of Income on Land

Land, not surprisingly, turned out to be by far the most important economic variable. The land area owned was chosen as the standard land variable for inter-survey comparisons.[11]

Table 2 shows the results of regressions of household income on the land area owned by each household. As with all regressions reported in this article, income was measured as household income per capita in current rupees, while the land area was measured in hectares per capita.

In order to allow unbiased comparisons between regression equations containing different numbers of explanatory variables or different numbers of observations, an adjustment must be made to the magnitude of the R^2 statistic, essentially to subtract out the component probably due to coincidence.[12] The unadjusted R^2 has also been reported in the table, despite its bias, because it is by far the more commonly used statistic, the one printed by most computer programs and reproduced in most publications.

The high R^2 statistics reported for a majority of the village surveys in Table 2 are somewhat surprising, since most of the households in them did not own any land. The low correlation between income and land during the 1961-62 survey of Naurangdeshar was due to the presence of highly-paid but landless staff members from the canal project together with the infertile quality of the still unirrigated land.

When considering the two-tailed significance statistics reported in the tables of this article, one must remember that the lower the magnitude of the significance statistic, the smaller the probability that the estimated coefficient was entirely the product of happenstance, and the greater its significance in the ordinary sense of the word.[13]

Regressions of Income on Land, Education and Labor

The next set of regressions, reported in Table 3, included variables representing land, education and labor.

"Human capital" in its most general sense included skills acquired outside school (for example, experience as a craftsman), as well as those acquired in school. However, the only components of "human capital" for which measurements were available were education and literacy. A problem with using education as an explanatory variable for income is that education is a consumption good as well as an investment good, and richer families within the village tend to obtain more education for their children. To reduce as much as possible the effect of more income causing more education, in regressions designed to estimate whether more education caused more income, only the education of working males not currently attending school was considered. Thus the variable chosen to represent "human capital" was the total years of schooling completed by male workers not currently attending school. Naturally, in regressions for income per capita the total education variable was also divided by the number of household members.[14]

A regression of household income on the number of workers in the household is a special case of the classic and difficult set of econometric problems which arise when supply is confused with demand (in this case, the supply and demand of wage services). As household income rises, the family can afford to have more of its children attend or stay longer in school, and to have its women stay at home. By itself, this would lead to a negative correlation between income and the number of household workers. On the other hand, the more family members work, the more income they can earn collectively, implying a positive correlation. The net effect of these opposing factors on the coefficient of labor in a multiple regression of income on labor and other variables is ambiguous, and any significance test on the labor coefficient has no certain interpretation. To get around this problem, the concept of each household's potential labor force was implemented: that is, the number of males within a certain age bracket, whether or not they were actually working.[15] Typically about 90% in the age group were listed as working. Furthermore, the concept of "potential workers" fits in well with the basis of the analytical approach used here: examining the determination of income in light of the resources available to each household. For simplicity and to avoid severe problems of multicollinearity, female and child potential workers were not included in the regressions, even though large numbers of women and young teen-aged children did work -- always at lower pay and usually for a lesser number of days per year than men.

Unfortunately, it was not possible to select a standard form of the potential labor variable for use in all surveys, because in some surveys the significance of the labor variable was quite sensitive to the age limits chosen, and because there was no pattern of the most significant ages across villages, or even within the same village across time. Non-per-capita regressions did not demonstrate any consistently superior age range, either. Three starting ages were assayed (15, 19 and 21 years) with four ending ages (50, 55, 60 and 65), yielding twelve combinations. The most significant version of the labor variable was chosen for each survey's regression.[16]

All but one of the variable coefficients reported in Table 3 were significant at under 1%; the weakest significance level was 2.2%. Only two explanatory variables caused problems: the number of working-age males in Bhatian in 1955-56 and again in 1971-72. Although these variables attained reasonably high significance levels (3.4% for 1955-56, 6.1% for 1971-72), their coefficients (-410 and -700 respectively) had the wrong sign. Since labor was measured as the potential number of workers, the negative sign cannot be explained away as due to richer households having fewer of their members at work. Whatever the reasons for the negative coefficients, it is absurd to believe that each additional male of working-age caused a large reduction in household income. It made more sense to exclude the labor variable from the regression equations, particularly since doing so made very little change to the coefficients for the remaining variables.

In fact, it was reassuring to observe during the preparation of Table 3 that in every survey the coefficient for each category -- land, labor, "human capital" -- varied little as changes were made in the other categories. Thus changing the measure of education from total schooling to maximum schooling, or switching from schooling to literacy, or even omitting the "human capital" variable altogether, produced only small changes in the coefficients for land and

labor (the changes usually were considered smaller than the coefficients' standard errors). The stability of the coefficients was greatest when the variables were entered into the regression equation in order of their significance (most significant first).

Income Regressions with Dairy Animals

The number of dairy animals owned per capita, measured in "cow-equivalents",[17] was assayed as an additional explanatory variable. Table 4 shows the regression results for the three surveys in which this variable had a significant role.[18] It can be noted that in two of these surveys (Naurangdeshar in 1961-62 and Sohalpur Gara in 1954-55), inclusion of dairy animals lowered the land coefficient by about one-fourth from the levels reported in Table 3; otherwise, inclusion of dairy animals caused little change in the estimations of the coefficients for other resources. In five other surveys, inclusion of the dairy animals variable added almost nothing to the R^2 of the regressions reported in Table 3; in these surveys the coefficient for dairy animals was usually quite insignificant and in most cases even negative. For the remaining survey (Ankodia in 1960-61), the animal census information was missing.

Conditions of dairying were strikingly dissimilar in the two villages for which the dairy animals variable was significant. In Sohalpur Gara, most of the milk (59% in the first survey, 66% in the second survey) was sold, with almost all sales to customers outside the village.[17] Naurangdeshar was subject to unusual conditions before the canal was built. Drinking water was extremely scarce and the little available was of poor quality. There was also at the time a local taboo against selling milk. For these reasons, nearly four out of five households produced their own milk (which was counted as income in kind), and only 3% of the milk was sold.[18]

Income Regressions with Caste

It is a truism that caste plays an extremely important role in India, especially in rural areas. To give one example, people outside the higher castes may have a more difficult time obtaining services such as connections to irrigation canals. In the past, members of the "untouchable" castes took it for granted that sending their children to school would be an unthinkable social outrage -- fortunately this is no longer the case for the present generation of children. The number of castes and sub-castes in India is in the thousands; in a given village, there may be a score of different sub-castes. To simplify this excessively detailed classification, I grouped the castes into four categories. The first was priestly castes (such as Brahmins), traditionally the highest caste. Next came cultivator castes (Jats, Rajputs, and so on). The third group, called "intermediate" castes here, contained the remaining "clean" castes -- mostly artisans. The last group consisted of the "untouchable" castes (Harijans, Naik, and so forth -- referred to as "scheduled castes" in Indian documents, terminology which will henceforth be adopted here). One must realize that for the last several decades castes have not been limited to their traditional roles. Land reform legislation and the secularization which usually accompanies development have moved most Brahmins away from their traditional combined occupation of

priest and non-cultivating landlord, and into various white-collar careers. Many members of artisan castes, such as weavers, have been unable to continue their traditional occupations in the face of competition from factory-made goods.[19] And in all of the villages studied here except Sohalpur Gara, even some of the scheduled castes owned land. Also, even though Muslims and Sikhs are not supposed to observe caste distinctions, in practice every Muslim and Sikh in the villages knew to which caste his family belonged, and inter-caste marriages did not occur.

Three dummy variables representing membership in the priestly castes, in the intermediate castes, and in the scheduled castes were added to the regression equations reported in Table 3. (The caste dummy variables were not divided by the number of household members.[20]) Under this specification of the equation, the intercept term in effect applied to members of the cultivator castes only, and the coefficients of the dummy variables were the deviations of the other caste groups' intercept terms from the value of the intercept term for cultivator castes. Thus the coefficient and the significance statistic of each caste dummy variable indicated the degree to which per capita income received by members of that caste group was different from the per capita income received by a household in the cultivator castes and the significance of this difference, after controlling for landholdings, etc. Obviously, any caste group could have been chosen as the basis for comparison.

Table 5 shows the coefficients of the caste dummy variables. The regression equations included all caste dummy variables, regardless of their significance; obviously, those coefficients with very weak significance statistics (greater than 25%, for example) are suspect. Since differences from the second-ranked cultivator caste group are being measured, the downtrodden scheduled castes should have a negative coefficient, the intermediate castes should have a coefficient somewhat larger, though still negative, while the top-ranking priestly castes should have a positive coefficient. Asterisks in the table indicate coefficients failing to meet these expectations. Data on the education variable is also reproduced, since inclusion of the caste variables sometimes altered the size and significance of the education coefficient. The coefficients for labor and land are not reported in the present table, because they changed only slightly from their levels and significances as reported in Table 3.

The priestly castes variable was definitely significant in both surveys of Ankodia, and membership in the scheduled castes had a significantly negative effect in Naurangdeshar during 1961-62. Even though the other caste coefficients reported in Table 5 had weak or very weak significance levels, often in combination with the wrong ranking, and even though the increase in the adjusted R^2 induced by the inclusion of the caste variables was quite small, one still cannot conclude that it made little economic difference to which caste group a household belonged. For the coefficients in Table 5 measure the effects of caste membership for households possessing the same quantities of land, education, and males of working age. But caste could be the main factor which determines how much land and education a household has. After statistically controlling for these factors, the additional effects of the caste variables might be neglible despite an important, though indirect, role of caste.

A test of the effect of caste on the determination of earnings which avoids these problems is to compute regressions including only caste variables. Table 6 shows the percent of variance in income per capita which can be explained by predictions derived solely from knowledge of each household's caste (the R^2 statistic). That table also shows the average per capita incomes of the caste groups, which usually were in the expected order. The coefficients with the wrong ranking were with one exception for caste groups containing ten or fewer households, and the difference in per capita incomes between wrongly ranked castes were never significant at less than 25% and mostly not significant at 50%.

Three conclusions about caste can be made. First, after statistical adjustment for the number of households and the number of explanatory variables, caste explained less than 10% of the variance in income per capita in two-thirds of the surveys, a moderate part of the variance in two surveys, and a large portion of the variance (35.5%) in only one survey, even using a form of the regression equation which maximizes the amount of influence attributed to caste variables Second, to the extent that caste played a role in the determination of income it was through different castes having control of different amounts of economic resources, and not through discrimination among households having the same amounts of land, education, and available family labor. Third, as an exception to the point just made, in two surveys members of the priestly castes earned significantly more, and in one survey members of the scheduled castes earned significantly less, than could be expected from their endowments of land, education, and family labor.

Disaggregating Income

Up till this point, all types of income have been lumped together in the dependent variable. But the importance of the explanatory variables may be systematically different for different types of occupation -- for example, one would expect that education would be of little use to unskilled laborers. Not only is it of intrinsic interest to see how the factors are related to specific kinds of income, but failing to make distinctions can reduce the accuracy of the prediction for total income.

I classified the many sources of income listed in the questionnaire forms into six categories: cultivation, diary, unskilled labor, skilled occupations, land rent receipts, and house rent receipts. Though it was clear where most of the income belonged, there remained many instances where the classification was arbitrary, especially as between unskilled labor and skilled occupations. "Skilled occupations" was a hodge-podge of employments not elsewhere classified.[21] The category of cultivation income incorporated all receipts stemming from the use of farm equipment and draft animals, thus including income from the rental of bullocks, the rental of rice hulling equipment, and the sale of surplus water from irrigation pumps. Very few households received income from house rent.

Many households pursued more than one occupation. It was impossible to allocate the labor of these households among their jobs, because of the impracticability of measuring the time spent at work by the self-employed (e.g., farmers and

artisans). Therefore the following tactic was adopted. It was assumed that only three sources of income -- unskilled labor, skilled occupations, and cultivation -- used significant inputs of labor. (Dairy often did involve labor, but in smaller amounts, and cattle often were grazed by children.) Regressions for each type of income were prepared including only households which derived all of their "labor-using" income from within one category. Rental and dairy receipts were ignored, but a household deriving income simultaneously from cultivation and unskilled labor, for example, would not be included in either the cultivation regression nor the unskilled labor regression. In each occupational regression, the dependent variable was the net income per capita received by households from the specified source (e.g., net income from cultivation per capita).

Regressions for Income from Unskilled Labor

Table 7 shows the regressions for income from unskilled labor. Since households in this occupational category contained the largest proportion of working women, the number of potential female workers was included in the regressions along with the number of potential male workers. The number of females of working age (assumed to be 19 to 50 years) (alternative age specifications were not calculated for women) always had a coefficient smaller and less significant than that for males; in many cases, the female labor variable had to be excluded from the regression equation because its coefficient was grossly insignificant or even negative.

When the education variable was added to the unskilled labor regressions, it was positive and significant for only one survey: Bhatian in 1955-56. Though significant at 1.3%, that coefficient was small, 23 rupees per year of schooling (with a standard error of 9). Otherwise, the education coefficient was either negative with significance level of 29% or even weaker, or it was positive but with a significance level of 49% or worse. For these reasons, it was concluded that the small amount of education possessed by workers in unskilled labor households had no significant economic effect, and the education variable was left out of the regressions reported in Table 7.

Regressions for Income from Skilled Occupations

Regressions for income per capita from skilled occupations are displayed in Table 8. In a few surveys the R^2 statistic was reduced by several percentage points through adjustment for the fact that only a small number of households was included in each regression. The significance of the coefficients for workers' education and for males of working age varied from less than 1% (highly significant) to over 50% (practically meaningless). The weak significance levels occured when the extremely diverse nature of skilled occupations precluded a uniform relationship between labor and income or between education and income. When significant, the coefficient per year of schooling usually was high.

Regressions for Income from Cultivation

The results of regressions of net income from cultivation on the land area cultivated, the number of potential male workers, and the education of male workers are shown in Table 9. The net land area cultivated was chosen because during the first survey of Bhatian the refugee families had not yet brought all of their land allotments into cultivation. In other surveys it made little difference whether one chose the net land area cultivated, the land area owned, the land area operated (i.e., after rentals), or the adjusted land area operated (the adjustment consisting of subtracting half the land area rented in, on the grounds that rental terms were usually 50% cropsharing).

Special treatment had to be given Naurangdeshar in 1961-62. For this survey statistics were available showing the total expenses for cultivation and dairy, but the expenditure was not broken down between the two categories. (This data was compiled directly from the raw questionnaire forms by Gary Y. Burtless, then preparing his Senior Essay for Yale University.[22] Although his expenditure data is less accurate than that prepared by the Sardar Patel University Agro-Economic Research Centre for its later surveys, an approximate account of costs is better than none.) A computer program which did not include an intercept term was used to regress net income from cultivation-plus-dairy on gross income from cultivation and gross income from dairy. The sample was limited to the 129 cultivator households in Table 9, and as usual all of the variables entering the regression were expressed in per capita terms. The R^2 was 97% (with or without adjustments). The regression equation was estimated to be as follows (standard errors are in parentheses):

$$\text{(2)}\quad \underset{\text{CULTIV. + DAIRY)}}{\text{(NET INCOME FROM}} = \underset{(\ .028)}{0.645}\ \text{(CROSS CULTIVATION INCOME)} + \underset{(\ .052)}{0.802}\ \text{(GROSS DAIRY INCOME)}$$

Therefore while the regression for the 1961-62 Naurangdeshar survey reported in Table 9 was based on gross income from cultivation per capita as the dependent variable, all the coefficients and standard errors appearing in the table have been multiplied by 0.645 to convert them into approximate coefficients for net income. It can be mentioned in this connection that there were problems in all surveys with regard to the allocation of costs between cultivation and dairy (especially with the value of unmarketed home-grown fodder, which was eaten by draft and milch cattle), so that data for the sum of net income from cultivation plus net income from dairy is more accurate than either of its components.

There are problems with the coefficients estimated for labor and education in Table 9: six of the labor coefficients turned out to be negative, as did five of the education coefficients. Although most of the negative coefficients had very weak significance levels, five were significant at 8% or better, and the most absurd coefficient of all (-6142 rupees per potential worker during the 1971-72 survey of Bhatian) happened to have the strongest significance level (under 0.1%0. Another problem is that, unlike the coefficients reported in previous tables, the education and labor coefficients were often clearly

interrelated. Thus when an education variable with a negative coefficient was removed from a regression, it usually caused a drop in the magnitude of the labor coefficient (for Sohalpur Gara, a reduction of the land coefficient instead).[23] This kind of phenomenon indicates that the coefficients reported in Table 9 measure not the separate contributions of labor and education to income from cultivation, but the interactions of labor and education.

Still unexplained is why education or labor might have a negative interaction. Three lines of argument may be advanced. The simplest is bad luck. In these fairly small samples, for which most of the variance ever to be explained was explained by land, the negative coefficients for some of the remaining variables may have been the result of coincidence. While coincidence can account for the presence of some negative coefficients, it is an unsatisfactory explanation for coefficients which are significant as well as negative.

The second line of argument is based on the specification of the cultivation regressions. If the true relationship between income and education is nonlinear, then the estimated education coefficient in a linear regression including a constant term will approximate the marginal differences in income associated with moderate differences of education about its average level. Now since education is a consumption good as well as an investment, with high prestige attached to schooling separate from its economic benefits, it is probable that families which could afford to do so -- landowners, for example -- invested in education beyond the point where its marginal product in rupees and paisa was significantly positive. This does not mean that if they had been illiterate they would have been able to manage their farms just as profitably: instead, the implication is that additional years of schooling beyond the prevailing level (or a few years less then the prevailing level) did not significantly alter farm income. Another feature of the specification of these regressions which could have led to an understatement of the role of education was the method of selecting households for inclusion in the cultivation sample. Whenever a well-educated farmer used his acquired skills to branch out into a non-cultivation activity (for example, acting as a wholesale dealer in farm produce), then his household was excluded from the cultivation regression because of its supplementary income from a skilled occupation. Misspecification of the form of the regression equation or aspects of the criteria for including households in the regression could explain why education might not have a significant and positive coefficient, even if education did on average contribute to income, but again this fails to provide an explanation of why the coefficient might be both negative and significant.

The third line of argument is that educated Indian villagers have an extreme aversion to manual labor. This observation was emphasized in conversations I had at the Sardar Patel Agro-Economic Research Centre with several staff members who had participated in the studies of Ankodia and Naurangdeshar. "Once they are educated, they do not want to work," one of them said. The same point was accentuated by the Indian sociologist Kusum Nair in her study Blossoms in the Dust: The Human Factor in Indian Development.[24] According to this reasoning, in some surveys a costly tendency of educated farmers to limit their efforts to supervising hired hands more than offset the better managerial abilities their education presumably affords them.

Because the labor and education coefficients reported in Table 9 often had weak significance, often had the wrong sign, and because in any case the meaning of the coefficients was unclear, it was decided to exclude these variables from the regressions. The results of the abbreviated regressions of cultivation income on the land area cultivated are shown in Table 10. The adjusted R^2 statistics for these regressions were in most cases close to (in two surveys, a bit higher than) the adjusted R^2 statistics for the cultivation regressions including education and labor as well as land.

Comparison of Resource Coefficients Over Time

Up till this point all coefficients have been expressed in current rupees. It may be of interest to observe how the magnitudes of the coefficients changed over time in real terms. To do this, the coefficients were deflated using the series of consumer price index numbers listed in Table 11. The consumer price index for Sohalpur Gara was derived from the Rural Consumer Price Index for Western Uttar Pradesh; the indices for the other villages were derived from the Consumer Price Index for Agricultural Labourers in the appropriate region.[25]

When comparing the selected "constant rupee" coefficients which are shown in Table 12, severe limitations must be kept in mind:

(a) The standard errors in the coefficients often were sufficiently large to preclude the differences between estimated coefficients from being statistically significant. In such cases the differences should be regarded just as tentative indicators.

(b) The consumer price indices, based on regions containing about twenty million inhabitants apiece, may not have been accurate reflections of the budgets of the various classes in the villages studied here.

(c) The consumer price indices did not have a common all-India base and therefore do not permit comparisons between villages.

(d) Because the regression equations included constant terms and other variables, the coefficients are estimates of the marginal contributions (of an additional male of working age to household income, for example). Thus when one coefficient in an equation changed, the effect may have been offset for most households by changes in the constant term or in other coefficients.

(e) The potential male workers variable did not include the same age range for all surveys.

The main features of Table 12 can be summarized as follows. The marginal value of land -- particularly as estimated by the coefficient for the net land area cultivated in regressions for income from cultivation -- increased markedly in all villages except Sohalpur Gara. In Sohalpur Gara the decline in this coefficient was due in part to the fact that by the time of the second survey

lower quality land had been brought into cultivation, and due in part to the fact that prices for crops grown in the village did not rise quite as fast as the overall cost of living index.

The real level of the education coefficient showed no particular trend, which is not surprising considering the heterogenous and changing collection of "skilled occupations" in which education played an important economic role.

The coefficient for the number of males of working age also fluctuated erratically. The comparisons shown in Table 12 are practically meaningless, because the range of working ages was not uniform and because other coefficients in the regression equations changed along with that for potential male workers. Unfortunately, no coherent pattern emerged even when females were excluded from the regression equations, the age range for potential male workers was standardized (to 15 through 55 years), and the regressions were limited to the per capita wage earnings in constant rupees of unskilled labor households containing at least one man aged 15 to 55. In particular, as shown in Table 13, the coefficient for potential male workers did not follow the same trends as the average earnings per capita or the average earnings per potential worker.[26] The averages of course include the earnings of women, children, and men more than 55 years old. Although in theory the regression coefficients measure the contribution only of males in the prime age range, while the earnings of others are subsumed in the intercept (constant) terms, in practice the small sample sizes and large standard errors, among other problems, indicate that the evidence is just too uncertain to allow conclusions about changes in the real labor coefficient over time.

Actual and Predicted Income Distributions

The set of predictions for individual households' incomes implies a prediction for the distribution of income among households -- obviously, the distribution of the predicted incomes. Graph Set 1 shows for each survey the distribution of household per capita incomes predicted by the regression equations listed in Tables 3 and 4. As benchmarks, each plot also shows the actual distributions of per capita income and the average per capita income. Table 4 (containing regressions of household per capita incomes on their per capita resources of land owned, males of working age, male workers' schooling, and dairy animals) was used to calculate the predicted income distribution for both surveys of Sohalpur Gara and for the first survey of Naurangdeshar, and Table 3 (containing regressions like those in Table 4, but excluding dairy animals) was used to calculate the predicted distributions for the remaining six surveys. Predicted incomes and actual incomes were separately ranked from highest to lowest.

Although the predicted distributions are reasonably good approximations to the actual distributions (in some cases, very close approximations), inspection of Graph Set 1 reveals a common trend: the predicted distributions always overestimate the lowest incomes in the village and always underestimate the highest incomes.

The explanation of this trend is simplified if one discusses incomes predicted by a singly explanatory variable -- for example, regressions of per capita incomes on per capita landholdings. The upper part of Graph Set 2 demonstrates, for a hypothetical village containing ten households of equal size, observations of income per capita plotted against landholdings per capita along with the line showing the relationship between the two as estimated by a linear regression. The predicted income per household is the point on the line corresponding to the amount of land owned per member of that household. Thus the vertical distance between each point and the line is the difference between predicted income and actual income -- in other words, the error of prediction, or residual.

The mathematics of the regression equation guarantee that there will be no linear trend in the residuals as one goes from the households owning the least land to those owning the most. But there will be a trend in the residuals after the households have been reordered from the poorest to the richest in terms of their per capita income. The reason is that households earning less than their predicted income will tend to fall into the poorer part of the income distribution, while households earning more than their predicted income will tend to rise into the upper part of the income distribution. In the first graph of Graph Set 2, for example, the three households with the lowest incomes ("A", "B", and "C") all earn less than the smallest predicted income in the village, while the household receiving the highest village income ("D") earns more than the greatest predicted income. Therefore in the second graph of Graph Set 2, where both the set of actual incomes and the set of predicted incomes have been ordered into distributions, the three lowest points on the distribution of actual incomes are lower than the three lowest points on the distribution of predicted incomes, while the highest point on the distribution of actual incomes is higher than the highest point on the distribution of predicted incomes. In general, the larger the number of households, the more smooth will be the trend of the predicted distribution to overstate the lower range of the actual distribution and to understate the uppper range; and the larger the average absolute value of the residuals, the greater will be the magnitude of this systematic divergence between the actual distribution and the predicted distribution.

One can obtain a closer fit to the actual income distribution by taking this phenomenon explicitly into account. The procedure is to add to the predicted household per capita incomes a series of random numbers whose distribution is similar to the distribution of the error terms in the regressions. The modified predictions are then reordered into what I shall call the "simulated distribution" of income (in order to distinguish it from the "predicted distribution" without the random comment). The simplest appropriate distribution to simulate the expected effect of errors of prediction would be a series of random numbers following a normal distribution with an average value of zero and standard deviation equal to the standard error of the constant term in the regression equation. However, it was clear that the magnitude of the error terms varied systematically with the magnitude of incomes,[27] and allowing for this greatly improved the fit of the simulated distributions. The absolute values of the error terms were regressed on the per capita incomes predicted by the equations in Table 3 & 4, yielding results summarized in Table 14. For example, the absolute value of the error term for the 1958-69 survey of Naurangdeshar was estimated to be 27 rupees plus 39.4% of the estimated per capita income.

(Of course, leaving in positive and negative signs, the average value of the error terms was zero, with no linear trend.)

The simulated error terms for each survey were generated by the equation below:

(3) $E_i = N_i \ (a + bY_i)$

where "E_i" is the simulated error term, "N_i" is the ith entry in a table of random numbers following a normal distribution with a mean of zero and variance of one, "a" is the intercept and "b" is the coefficient for estimated income as reported in Table 14, and "Y_i: is the estimated per capita income of the ith household. To avoid repetition, only the plot of the distribution of simulated incomes for the second survey of Naurangdeshar is shown as Graph 3. That survey was chosen because it had the poorest fit among the predicted distributions: after the incorporation of the random element, the fit of the simulated distribution was very close.[28]

There is, however, an argument that the more fundamentally valid picture is obtained from the original predicted distribution not adjusted for the systematic influence of the errors in prediction. If the residual between an individual household's income and the estimation of that household's income is a transitory phenomenon, a random fluke which cannot be expected to recur year after year, then the expected income which the household will receive on average over the years is the predicted income, and the distribution of predicted incomes is equivalent to the distribution of long-run expected incomes. In Milton Friedman's terminology, the residuals would be transient incomes and predictions would be permanent incomes.[29] The hypothesis that the predicted distribution corresponds to the distribution of "permanent incomes" is unlikely to be true for two reasons. First, not all residuals are due to transient factors. When the residual is due to a factor not taken into account in the regressing equation which is likely to be a continuing influence -- for example, because some household workers are unusually intelligent or stupid, strong or unhealthy, or because the household owns land which is unusually fertile or poor -- then the residual cannot be regarded as a "transient" component of income. Second, not all transient factors are manifested in the residuals. In a rural society the chief cause of fluctuation in earnings is weather, but in one year in one village the weather is the same for everybody. In conclusion, the adjustment from predicted distribution to simulated distribution is appropriate; and failing to make the adjustment would not leave one with the distribution of "permanent incomes" [30]

Analysis of the Income Distribution by Means of Simulated Deviational Effects

If every household in a village possessed the same quantity of each resource per capita, then according to the regression equation all would have the same predicted per capita income -- equal to the average per capita income in the village. Thus the deviations of income from an egalitarian distribution can be analyzed in terms of the deviations of each resource from an egalitarian distribution.

This concept can be implemented in a number of ways. No one approach will be the best for all conceivable applications. The following procedure was used as the operational definition of the "simulated deviational effect" of a resource. A set of predicted incomes was obtained using a regression equation as before. Then a new set of predicted incomes was calculated, assuming that the given resource (for example land) had been evenly redistributed so that all households owned the average amount per capita, while the distribution of the other resources remained unchanged. A more equal distribution of predicted incomes resulting from the egalitarian redistribution of one resource could be presumed to produce a more homogenous distribution of the errors or prediction. Thus it is plausible to suppose that if land were distributed more evenly, then unpredicted fluctuations in land income would be more evenly distributed too. Equation (3) was used with the same entries "N_i" from the table of random numbers in the same order to calculate both the original and the modified series of simulated residuals. Therefore the differences between the two sets of simulated residuals were due exclusively to differences in the magnitudes of the predicted incomes (the "Y_i" terms). Each series of simulated residuals was added to the corresponding series of predicted incomes, and the two series of simulated incomes were ordered into two simulated distributions. The simulated deviational effect of the resource was equal to the original simulated distribution minus the simulated distribution calculated with the assumption that every household owned an average per capita amount of the resource in question. Thus the simulated deviational effect of land is a measure of the extent to which the inequality in the distribution of landholdings contributed to the inequality in the distribution of total incomes: it shows where and how much the distribution of income would have changed if only land had been distributed equally.

Graph Set 4 shows the simulated deviational effects attributed to each of the variables entering the regression equations used to predict the distributions of village income. In the preparation of these graphs, the simulated deviational effects were added to the average per capita incomes, since without deviational effects all households would be expected to earn the same income plus or minus a usually small random component. For reference, the actual distributions of income were also plotted.

The most striking feature of these graphs is the extent to which land dominated the deviational effects. In a majority of the surveys, the simulated deviational effect of land by itself accounted for most of the variance in incomes. Although many households derived the bulk of their income from their resources of labor, because this factor was fairly evenly distributed across households, its deviational effects were small.

Application of the Model to Simulate the Distributional Effects of a Land Reform Program

The models of the income distribution developed in this article can also be applied to predict the distributional effects of possible policies. A discussion of land reform policy is presented here purely as an illustration of methodology. A more realistic appraisal would require estimates of the impact of the land reform on farmers' efficiency and changes in employment within the

agricultural sector.[31] Here it is assumed that the marginal value of all economic assets -- land, labor, education, and dairy animals -- would not be significantly changed by the land redistribution. And needless to say, conditions in four villages surveyed over a span of twenty years cannot be projected to India as a whole.

Even a rather simple model of a land reform program requires a detailed specification of who gains and who loses and how much. These assumptions are used:[32]

(1) One-third of the total land area owned by households in each survey was redistributed, all within the village. (Alternatively, the land area received from absentee landlords resident elsewhere was equal to the land area lost elsewhere by absentee landlords resident in the village.)

(2) The compensation for expropriated land was equal to 50% of the land coefficient listed in Table 3 or Table 4. This payment was received by the former owners (possibly in the form of interest on non-redeemable government bonds), and was given up by the new owners (possibly in the form of 25% crop-share tax payments to the government), so that the government neither subsidized nor financially exploited the reform program. Obviously, if the expropriated landlords had received full compensation for their land, and the new landlords had been required to pay taxes and fees sufficient to finance this reimbursement, the distribution of income would have been just as before.

(3) To decide who would give up land, a ceiling was calculated just low enough to cause one-third of the village land area to be expropriated. The ceiling was determined in "cultivator units": a household was allowed a certain area for each male aged 18 or over.[33] To avoid confiscating all the land of widows and orphans, households containing no male aged 18 or over were assigned one cultivator unit.

(4) It was assumed that the government did not evenly distribute the land among all remaining households in the village, partially out of concern that a small allotment to each household would not be of an economically viable size. (In point of fact, many of the smallest landowners rented out their tiny plots, indicating that very small holdings were economical.) A "floor" area was set, initially equal to one-fourth of the ceiling area, and like the ceiling area defined in terms of the cultivator units in a household. As a first priority, land was distributed to small landowners to bring them up to the floor, and also to households renting land though owning none. After small holders and landless tenants had been brought up to the minimum, the remaining reformed land was distributed in plots of the floor size until it ran out. If there was more than enough reformed land area to bring all small landholders, landless tenants, and landless unskilled labor households up to one-fourth of the ceiling area, the floor area was increased as much as possible consistent with provision of equal minimums to eligible recipients (including the small landowners who became eligible after the floor size was increased).

Table 15 shows what the ceiling and floor areas would be, how many households would lose land, and how many would gain, if such a policy were applied to each of the village surveys. The first survey of Bhatian was excluded because

land tenure conditions then had obviously not yet reached equilibrium: for example, several households rented in land which they did not cultivate. A procedure analogous to that used for calculating simulated deviational effects was used to simulate the difference in each quintile's average per capita income caused by the land reforms. (The households present in a quintile before the reforms were not always the same as the households in the quintile after the reforms.) The results of this exercise are presented in Table 16.

Except for the first survey of Sohalpur Gara, where a small land regression coefficient caused the simulated effects of the land redistribution to be neglibibly small, the transfer of one-third of the village land area with 50% compensation to its former owners had a substantial impact on the distribution of per capita income. The income of the top quintile was always reduced (by an unweighted average of 8% among the eight surveys). Most of this income was transferred to the middle three quintiles, since not all landless labor households received land allotments. But due to the low base level of average per capita income in the poorest quintile, the percentage increase in income for this quintile was usually greater than that of any other quintile, despite a smaller gain in rupees. In Naurangdeshar, where the land was widely distributed already (largely as a consequence of state government policy), redistribution of one-third of the village area would have resulted in virtual equality of ownership, though of course much of this land would be subject to special taxes to finance the reform program.

This section illustrates how the methodology developed in this article may be applied to evaluate the distributional implications of almost any policy. In fact, the simulated deviational effects derived in the previous section are special cases. For example, the simulated deviational effects of education indicate that a policy of equalizing educational attainment in the villages, though probably quite desirable for other social reasons, would have usually caused little change to the income distribution. More complete policy appraisals would require estimations of the changes in the values of resources, from projections of agricultural productivity, rural employment, and wages.

Conclusions

I. The Determinants of Household Income

The first conclusion is that the approach proposed here works. With a simple linear regression containing two to four explanatory variables plus a constant term, it was possible to explain from 36% to 87% of the variance in household per capita income, after discounting the R^2 statistic for the number of degrees of freedom used up by the explanatory terms. The coefficients for the households' resources of land, labor, and education as estimated in the village-wide regressions were almost without exception highly significant and of the right sign. The coefficients of these variables were also stable, in the sense that if the definition of one of the variables was modified, the magnitudes of the other variables' coefficients changed only slightly.

The second general conclusion is that the role of a family's resources depended a great deal upon their occupation. For example, in the same village

survey the "human capital" variable might have a significantly positive correlation with income from skilled occupations, virtually no income from unskilled labor, and a significantly negative correlation with income from cultivation.

The remaining conclusions concern particular variables.

Land, as expected, was the most important factor. The only surprise was the extent to which it dominated the regressions. In two-thirds of the surveys, land by itself accounted for most of the variance which was ever explained. Except for the first survey of Bhatian, during which there was the unusual circumstance that the recently resettled refugee families had not yet brought all of their land allotments under cultivation, it made little difference how land was measured -- owned land, net cultivated land, gross cultivated land, or other alternatives. Over time, the size of the land coefficient in constant rupees increased dramatically (especially when the regression was restricted to cultivator households and cultivation income), except in Sohalpur Gara, where physical productivity stagnated. Causes of the increased value of land elsewhere were major increases in the area irrigated combined with more double-cropping and greater use of chemical fertilizers. In Bhatian, another factor in the recent upsurge in land productivity was the new seed varieties from the "Green Revolution". Because cereals dominate the budgets used in the construction of rural Indian consumer price indices, cultivators to a large extent were automatically compensated for changes in the cost of living by changes in the evaluation of their crops.

To measure the contribution of labour to income it is necessary to carefully specify the way labour is measured. Counting the number of workers could give misleading results, because households receiving substantial amounts of income anyway -- as a result of their ownership of land, for example -- tend to cut back on the number of household members working, especially women and children. Developing a model to predict how households decide to allocate the time of their available family members between gainful employment and ungainful household work or leisure, in light of their alternative earnings, was beyond the scope of this article, particularly since no information was available on the time spent at work by the self-employed. Instead, a simpler approach was used. The labour variable was defined as the pool of family labour available (usually specified as the number of males within a certain age range), whether or not they were actually employed. Nearly everyone within the age brackets chosen did in fact work.

The importance of family labour varied greatly with the household's "labour-using" occupation -- cultivation, unskilled labour, or the set of skilled occupations. For households deriving all of their "labour-using" income from cultivation, the estimated contribution of family labour often was insignificant or even negative (especially when no education variable was included in the regression equation). This lends support to the "labor surplus" viewpoint: the marginal contribution of family labour to income appeared to be negligible as a rule for cultivator households.[34] This does not mean that the marginal product of labour for the village economy as a whole was near zero. The great bulk of unskilled labour was farm work, and there is no reason to believe that cultivators would

hire farm hands when the product of their labour did not repay its cost. Instead, the mechanism appears to have been that the members of households owning enough land to derive all of their "labour-using" income from cultivation (supplemented usually by dairy income and sometimes by rental income) largely limited their efforts to supervision and management, employing hired hands for the physical work. Thus the "labour surplus" was latent in the leisure of the cultivator households.

The best indicators of the role of labour were the wage incomes received by households which derived all of their "labour-using" income from unskilled labour. The topic was confused by the fact that the marginal product of labour behaved quite differently from the average product of labour. If one is investigating the welfare of the poorest occupational group in the villages, the average income is clearly the superior measure. If one is seeking a measure of the marginal product of "raw labour" for each village economy, the answer in theory is the coefficient for males of working age estimated by a regression on the wage earnings of households which had unskilled labour as their only "labour-using" employment. In practice, the small sample sizes, changes in the working age ranges, and other problems all cast doubt upon the accuracy of this measure. The evidence is too uncertain to allow comparisons of estimated marginal products of labour at different points in time.

The average of total household wage income per male aged 15 to 55 among unskilled labour households declined in Sohalpur Gara and rose elsewhere, measured in constant rupees. The percentage increases in average wage incomes were lower than the percentage increases in the land coefficient among cultivators (in Sohalpur Gara, the relative decline in the wage average was greater than the relative decline in the land coefficient). One should note that the standard deviations about the averages were large enough to make all these comparisons tentative. Thus in the villages where the development of agriculture made it profitable to hire more labour, particularly to carry out increased double-cropping, population growth and immigration from the rest of India supplied enough unskilled labourers to prevent a rapid increase in average real wages.

The most important aspect of education was that it was of economic benefit only for certain occupations. It is not surprising that schooling did not make any difference to an unskilled labourer's income. It is surprising that the coefficient for education was often negative or insignificant for cultivators. Apparently the costly effect of education causing its recipient to feel above manual labour often more than offset the benefit of increased managerial ability. These findings are somewhat uncertain: a more definite assessment would require more careful specification of the regression equation for income from cultivation (gross cultivation income might be specified to follow a Cobb-Douglas function, for example), would require information on farm capital to form part of the improved equation, and for reliable estimation would require larger sample sizes.[35] The data available from the village studies examined here were sufficient to constitute a clear warning against regressing income on education without including other variables in the equation, lest education as the product of high incomes be mistaken for education as the cause of high incomes.

The main economic role of caste was through different castes possessing different amounts of resources (such as land). For households having the same amounts of resources, normally caste made little economic difference. Thus when they had land, members of the lower castes farmed about as well as members of the traditional cultivator castes. By classifying households into four caste groups, one could account for about 10% of the variance in income per capita for five surveys, 15% to 20% for two surveys, and a little under 40% for one survey. In all cases, this was considerably less than the variance in per capita income which was accounted for by the single most significant non-caste variable (usually land).

II. The Distribution of Income

A profile of the "predicted income distribution" can be obtained by ranking the expected incomes for individual households derived from a regression equation designed to optimize the prediction of household incomes considered one at a time. But the "predicted income distribution" for the households considered as a group systematically overestimates the poorest end of the income distribution, where there is a concentration of households which received less income than would be predicted by the regression equation, and systematically underestimates the richest end of the income distribution, where there is a concentration of households which received more income than would be predicted by the regression equation. Thus the estimation of the income distribution can be improved by adding to the elements of the "predicted income distribution" a random component having a distribucion similar to that of the error term of the regression equation, producing what has been called here the "simulated income distribution".

Up till now most analyses of the size distribution of income have been in terms of combinations of distributions, arbitrarily arrived at (e.g., an unexplained urban distribution combined with an unexplained rural distribution.[36]). In rare studies, the size distribution has been analyzed by regressions estimated for individual recipients. The "simulated income distribution", combining regression estimates with simulated distributions of errors in the prediction equations, is an example of a class of models which can be developed to accurately portray the distribution of income, while at the same time permitting evaluation of its component causes. The "simulated income distribution" also provides a very flexible technique to estimate the distributional influence of almost any policy.

FOOTNOTES

(1) Review of Economics and Statistics (August 1958)

(2) Econometrica (August 1959)

(3) Denmark, Statistical Department, "Statistical Inquiries", No. 12; Copenhagen, Statistical Department, 1964. As far as I was able to determine, this 40-page monograph is available in the United States only at the Library of Congress.

(4) Chapter 24 of Measurement of Cost Productivity and Efficiency of Education, N.N. Pandit, editor (Madras: National Council of Educational Research and Training, 1969).

(5) Permission was usually refused if the Center had not itself finished analyzing the data and published a report. Because a number of village survey reports have been published since my return from India, other researchers should now be able to have access to material unavailable to me. Interested parties are warned that because of inadequate storage facilities at some of the Centres, after completion of the report the data is often thrown away or fed to termites.

(6) The reports prepared by the Agro-Economic Research Centres were indispensable to the preparation of this work, both to give a general view of each village unobtainable from a collection of raw numbers, and at times to resolve ambiguities in the interpretation of the statistics. The reports were as follows:

(a) R.M. Patel, Ankodia (Baroda District, Gujarat): Economic Life in a Tobacco Village (Vallabh Vidyanagar, District Kaira, Gujarat, India: Agro-Economic Research Centre, Sardar Patel University, 1964). This work will henceforth be referred to as the Ankodia Survey Report.

(b) [R.M. Patel], Ankodia (Baroda District, Gujarat): Change in Economic Life of a Tobacco Village (ibid., 1970). Henceforth referred to as the Ankodia Resurvey Report.

(c) M.L. Bhat, Naurangdeshar (Sriganganagar District, Rajasthan): A Village in the Rajasthan Canal Zone (ibid., 1964). Henceforth referred to as the Naurangdeshar Survey Report.

(d) D.M. Brahmbhatt, Naurangdeshar (Sriganganagar District, Rajasthan): Impact of Irrigation on a Rajasthan Canal Village (ibid., 1974). Henceforth referred to as the Naurangdeshar Resurvey Report.

(e) Sulekh Chandra Gupta, Sohalpur Gara: District Saharanpur: Socio-Economic Changes in Rural India 1954-55 -- 1958-59: Case Study of a Village in Western U[ttar] P[radesh] (Delhi: Agricultural Economics Research Centre, University of Delhi, n.d.). Henceforth referred to as the Sohalpur Gara Resurvey Report. There also exists a report for the first survey; but that was produced as a typescript with a few carbon copies, all of which are in libraries in Delhi.

(f) Bhatian Resurvey Report. A friend visiting Delhi obtained a copy of this mimeographed report for me in 1974. The pages are tied together with a piece of string and there is a handwritten sheet where the title page would normally appear. I surmise that the document -- completed in 1973 or 1974 -- is to be bound with a printed title as its cover. I

apologize for the lack of an official-type footnote, but that is the best that I can do, and I am grateful for what I got. The Bhatian Resurvey Report covers the two surveys 1955-56 and 1960-61. A typescript report (and several carbon copies) covering only the first survey are available in Delhi libraries. No report has yet been prepared for the third survey.

Those reports still in print (fortunately, the resurvey reports repeat most of the information contained in the first survey reports) can be obtained at cost from the Deputy Director of the Agro-Economic Research Centre which prepared them. All of the reports (including those issued only in typescript form) are on file in the Economics and Statistics Library of the Ministry of Agriculture in New Delhi.

(7) This forms a minor discrepancy with the Ankodia Survey Report, p. 40 and elsewhere, which states that there were 268 households in the village plus the 15 households temporarily absent, and that these households contained a population of 1,533. My data was copied directly from the questionnaire forms filled out in the village.

(8) In the measurement of "gross" areas, land irrigated (cropped) in both of India's growing seasons was counted double.

(9) For a discussion of the psychological characteristics of the refugee Sikhs, see Kusum Nair, Blossoms in the Dust: The Human Factor in Indian Development (New York: Praeger, 1961), pp. 102-115.

It should be noted that she believes that in explaining their development, the work-oriented ethic of these Sikhs may have been a less important factor than their refugee status: uprooted from their ancestral villages, people are more willing to innovate.

(10) More formally, the assumptions are that the expected values of the error terms are zero, and that the variance-covariance matrix of the error terms is a scalar multiple of the identity matrix. Also being assumed is that the relationship between income and its explanatory variables is actually linear, and that these variables cause income, not the other way round. See J. Johnston, Econometric Methods, (2nd ed.; New York: McGraw Hill Book Co., 1972), pp. 121-22.

(11) "Owned land" included the categories "land mortgaged with possession" and "land rented from state". The status of land rented from a state is for practical purposes like owned land: rental payments are nominal, especially when compared to private rental rates, and the state very rarely changes tenant families; the main difference is that the tenants may not sell the land. In Naurangdeshar, most of the area "rented" from the state of Rajasthan consisted of formerly uncultivated arid tracts allocated to farmers; otherwise, the land "rented" from states was transferred from its former non-cultivating owners to the control of its former share-cropping tenants by means of land reform legislation. For a further discussion of land reform legislation, see Daniel Thorner, The Agrarian Prospect in India: Five Lectures on Land Reform Delivered in 1955 at the Delhi School of Economics (Delhi: University Press, 1956), a good concise survey.

Statistics on the value of owned land were available for only one-third of the surveys, the most recent ones. Alternative land area statistics included

land operated (that is, land owned plus land rented in minus land rented out), net land area cultivated (operated land minus waste and fallow land), and gross land area cultivated (the same as net land area cultivated, but with areas sown in both of India's growing seasons counted double). The land area owned was chosen as the standard land variable for inter-survey comparisons because it can be argued that it is conceptually the best alternative and because it was the land area measure showing the highest correlation with income per capita in seven of the nine surveys. Nevertheless, it generally made little difference which alternative land measure was chosen, either in simple or in multiple regressions.

(12) If "R_a^2" is the adjusted R^2, "R_u^2" is the unadjusted R^2, "n" is the number of households, and "c" is the number of estimated coefficients including the constant term ("c" equals two for Table 2), then the formula for the adjusted R^2 is:

$$R_a^2 = R_u^2 - (1-R_u^2)\ (c-1)\ /\ (n-c)$$

See Ronald J. Wonnacott and Thomas H. Wonnacott, Econometrics (New York: John Wiley Sons, 1970), footnote on page 311, where k = c - 1.

(13) The two-tailed significance test is formally defined as the probability that a series of numbers drawn from a random normal distribution and inserted into the regression equation in place of the variable being tested would produce an estimated coefficient with a t-statistic at least as large in absolute value as the t-statistic for the variable being tested. For a two-tailed test, it does not matter whether the spurious coefficient for the random number series is positive or negative.

(14) The literacy variable, a crude three-way classification (illiterate/semi-literate/able to read and write), generally did not do as well as the education variable; it can be noted in passing that many people with no formal education were listed as literate. One might believe that education helps a farmer or other head of household to make better decisions, but otherwise had no economic significance. For example, an educated son might read the instructions on a package of fertilizer, but there is no need for two children to recite them as a duet. To test this view, the maximum schooling level achieved among each household's male workers (and the maximum literacy) were also assayed, but these variables generally had lower partial correlations with income in multiple regressions than did total education.

(15) It is true that even the number of potential workers may be in part determined by income, because more children may survive to working age in relatively well-to-do families. But the concatenation of correlations involved, from present income to past income to past demographic events to the present labour force, is far more tenuous than the causal connection from the present number of household members of working age to the present income.

(16) Two reasons can be advanced for the discrepancies in the most significant age ranges. One is that patterns of school-going and retirement for families following a given occupation changed over time. Probably the more important reason is that the ages at which people began and stopped working differed widely from one occupation to another while the occupational mix varied a great deal between villages and within the same village over time.

Where there is no theoretical reason for choosing one among a set of alternative measures, it is appropriate to use the maximum likelihood method employed here. However, the significance level should be viewed with skepticism: when trying twelve similar variables, it might be preferable to cite the significance statistic of the sixth or seventh ranked alternative (though that method is not used in the text). That the labour coefficients were "legitimate" is indicated not only by the very high significance levels of the best alternatives, but also by the fact that for most surveys all other alternatives also had positive coefficients, while the few negative labour coefficients had very weak significance levels. The two exceptions to this rule -- that is, labour coefficients with negative signs and reasonable significance levels -- are discussed in the text.

(17) Based on relative milk productivities, the following scale was used: a she-buffalo counted as 1 1/2 cows and a she-goat counted as 1/6 of a cow.

(18) It is interesting to note that conditions of dairying were strikingly dissimilar in the two villages for which the dairy animals were significant. Naurangdeshar was subject to unusual conditions before the canal was built. Drinking water was extremely scarce and the little available was of poor quality. There was also at the time a local taboo against selling milk. For these reasons, nearly four out of five households produced their own milk (which was counted as income in kind) and only 3% of the milk was sold. (Naurangdeshar Survey Report, pp. 12, 149-50, 177-79.) By contrast, in Sohalpur Gara most of the milk was sold (59% in the first survey, 66% in the second survey), with almost all sales to customers outside the village. (Sohalpur Gara Resurvey Report, pp. 95-97, and Tables 6.1, 6.2, and 6.3 in the unpaginated appendix. The figures cited on page 96 for the proportions of milk sold in the two surveys -- 68.5% and 75.4% -- differ from mine for two reasons: because the report's ratios are based only on households which sold some of their milk, while mine include households which produced milk exclusively for home consumption; and because the report's statistics refer to fluid milk only, while mine include milk converted into ghee, clarified butter.)

(19) This gradual erosion of traditional caste roles, which is in progress throughout India, is commented upon in the Ankodia Survey Report, pp. 35-37, 196-97, and in the Naurangdeshar Survey Report, pp. 28, 30.

(20) Since the entire household was obviously of the same caste, this can be viewed as the result of multiplying the caste variables by the number of household members, then dividing the "total caste" in the household by the number of household members, with the division cancelling the multiplication.

(21) "Skilled occupations" included the Agro-Economic Research Centre's categories of "arts and crafts", "professions", "service" (i.e., salaried positions), and "trade" (i.e., commerce). The correspondence is not exact, because I reclassified some jobs within these categories as unskilled labor -- for example, people employed by the railway (hence in "service") as sweepers.

(22) "The Relationship of Education and Income Among Agriculturalists in Naurang-Deshar Village, India", Yale Senior Essay (typescript, 1972).

(23) The interrelations between labor and education were more easily visible in regressions of total household cultivation income (not per capita) on total household resources (not per capita).

(24) Pp. 145-65. In this connection, one should avoid the common confusion of tacitly assuming that "lazy" is synonymous with "irrational".

(25) The Rural Consumer Price Index for Western Uttar Pradesh is to be found in the Monthly Bulletin of Statistics published by the Economics and Statistics Department of the Government of Uttar Pradesh. This publication is available in the New York City Library, but to find it one has to know that it is catalogued under the colonial name Uttar Pradesh spelled sideways (Agra and Oudh, United Provinces of). The Consumer Price Index for Agricultural Labourers in various regions is published by the Labour Bureau of the Government of India in the Indian Labour Journal (formerly called the Indian Labour Gazette), and in the Indian Labour Annual. This index goes back only to the calendar year 1957; for Bhatian it was extended back to the crop year July 1955 to June 1956 by linking it to the Food component of the Consumer Price Index for the Working Class in the nearby city of Ludhiana, given in the Indian Labour Gazette. A detailed discussion of the price indices and their appropriateness is to be found in Michael Lopez, "The Determinants of Income and its Distribution in Four Villages in India" (unpublished Ph.D. dissertation, Department of Economics, Yale University), pp. 32-34, 45-47, 51-52.

(26) The standard deviation about the averages are comparable to the standard errors about the regression coefficients.

(27) This relationship, technically called heteroscedasticity, is a minor violation of the standard set of assumptions mentioned in the first footnote of the previous chapter. The presence of heteroscedasticity does not bias the estimates of the coefficients. While it is possible to reduce heteroscedasticity by weighting the observations, doing so would destroy the equality between the average size of the predicted per capita incomes and the average size of the actual per capita incomes.

(28) Lovers of paradox will appreciate that the household predictions combined with random components form a better prediction of the village distribution than the household predictions alone. But while the individual random components are unpredictable, the variance of the set from which these random elements are drawn is carefully controlled.

(29) Milton Friedman, A Theory of the Consumption Function, (Princeton: Princeton University Press, 1957), p. 209.

(30) For a further discussion of problems related to the residuals, see Jørgen S. Dich, "On the Possibility of Measuring the Distribution of Personal Income", Review of Income and Wealth (September 1970). See also Richard Ruggles' criticism of that article on page 216 of the same issue.

(31) R. Albert Berry presents a model of the relationship between land reform and rural employment and wages, in "Land Reform and the Agricultural Income Distribution", Pakistan Development Review (Spring 1971). His article does not apply the model to empirical evidence.

(32) The hypothetical land reform policy described here may be compared to the rough policy model presented in B.S. Minhas, "Rural Poverty, Land Redistribution, and Development", Indian Economic Review (April 1970). Minhas does not mention compensation for expropriated land.

(33) No distinction was made between irrigated and unirrigated land, mainly because the regression equations did not make this distinction.

(34) The seminal article on the "labor surplus" theory is W.A. Lewis, "Development with Unlimited Supplies of Labour", The Manchester School of Economic and Social Studies (May 1954); see also his "Unlimited Supplies of Labour: Further Notes", ibid. (January 1958). A formal model of the "labour surplus" theory is presented by John C.H. Fei and Gustav Ranis in Development of the Labour Surplus Economy: Theory and Policy (Homewood, Illinois: Richard D. Irwin, inc., 1964), Chapter Two.

(35) An attempt was made to estimate a regression for net income from cultivation, assuming that gross income was created by a Cobb-Douglas function in land and labour (but without requiring constant returns to scale), and assuming that the ratio between land and non-labour expenses were constant. The attempt had to be abandoned when it was discovered that for the numbers of observations available, the non-linear-regression computer program did not converge accurately enough to produce reliable solutions. (The program was BMDP3R, version of May 8, 1972, in the Bio-Medical Data Program series prepared by the Health Sciences Computing Facility of the University of California at Los Angeles.)

(36) For example, Subramanian Swamy, "Structural Changes and the Distribution of Income by Size: The Case of India", Review of Income and Wealth (June 1967)

APPENDIX I

Table 1: Main Characteristics of Surveyed Villages

Village	Crop year	No. of households included in survey	Pct. of households in village included	Pct. land irrigated	Paved road?
Ankodia	1960-1961	269	95%	40%	No
Ankodia	1967-1968	298	100%	63%	Yes
Bhatian	1955-1956	94	100%	13%	Yes
Bhatian	1960-1961	80	99%	18%	Yes
Bhatian	1971-1972	151	100%	99%	Yes
Naurangdeshar	1961-1962	192	88%	6%	Yes
Naurangdeshar	1968-1969	291	100%	59%	Yes
Sohalpur Gara	1954-1955	98	100%	0%	No
Sohalpur Gara	1958-1959	99	100%	18%	Yes

Table 2: Regressions of Income Per Capita on Land Area Owned Per Capita

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	32.3%	64.0%	79.5%	58.8%	83.7%
(before adjustment	32.6%	64.1%	79.7%	59.3%	83.8%
Number of households	269	298	94	80	151
Intercept	188	437	208	182	628
(Standard error)	(331)	(486)	(547)	(207)	(824)
Coefficient for land area owned	439	1364	102	395	2091
(Standard error)	(39)	(59)	(5)	(37)	(75)
t-statistic	11.	23.	19.	11.	28.
Significance	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%

Village:	Naurangdeshar	Naurangdeshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	2.7%	19.5%	17.0%	58.5%
(before adjustment	3.2%	19.8%	17.9%	58.9%
Number of households	192	291	98	99
Intercept	560	550	138	146
(Standard error)	(625)	(544)	(103)	(145)
Coefficient for land area owned	55	231	82	295
(Standard error)	(22)	(27)	(18)	(25)
t-statistic	2.5	8.4	4.6	12.
Significance	1.4%	<0.1%	<0.1%	<0.1%

Table 3: Regressions of Income Per Capita on Land Owned, Workers' Education, and Number of Working-Age Males, Per Capita

Village: Year:	Ankodia 1960-61	Ankodia 1967-68	Bhatian 1955-56	Bhatian 1960-61	Bhatian 1971-72
R^2 (after adjustment	41.6%	68.1%	84.4%	66.9%	87.0%
(before adjustment	42.2%	68.4%	87.4%	68.2%	87.2%
Number of households	269	298	94	80	151
Intercept	29	180	110	66	464
(Standard error)	(308)	(458)	(477)	(186)	(734)
Coefficient for land area owned	449	1376	83	355	1810
(Standard error)	(38)	(58)	(6)	(37)	(81)
t-statistic	12.	24.	14	9.7	22.
Significance	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
Coefficient for workers' education	60	76	123	53	245
(Standard error)	(22)	(24)	(23)	(22)	(39)
t-statistic	2.8	3.1	5.5	2.3	6.3
Significance	0.7%	0.3%	<0.1%	2.2%	<0.1%
Coefficient for number of working-age males	561	546	(omitted since nega-tive)	334	(omitted since nega-tive)
(Standard error)	(119)	(178)		(123)	
t-statistic	4.7	3.1		2.7	
Significance	<0.1%	0.3%		0.9%	
Range of working age	21-50	15-60		15-65	

-- Table continued on next page --

Table 3: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	41.8%	36.5%	40.0%	65.7%
(before adjustment	42.7%	37.2%	42.8%	66.8%
Number of households	192	291	98	99
Intercept	157	252	42	85
(Standard error)	(483)	(483)	(87)	(132)
Coefficient for land area owned	105	232	51	198
(Standard error)	(18)	(24)	(17)	(35)
t-statistic	6.0	9.6	2.9	5.7
Significance	<0.1%	<0.1%	0.5%	<0.1%
Coefficient for workers' education	114	121	40	87
(Standard error)	(15)	(25)	(12)	(26)
t-statistic	7.6	4.7	3.3	3.4
Significance	<0.1%	<0.1%	0.2%	0.2%
Coefficient for number of working-age males	592	1058	267	251
(Standard error)	(173)	(232)	(51)	(76)
t-statistic	3.4	4.6	5.3	3.3
Significance	0.1%	<0.1%	<0.1%	0.2%
Range of working age	21-65	19-50	15-55	19-60

Table 4: Regressions of Income Per Capita on Number of Dairy Animals, Land Owned, Workers' Education, and Number of Working-Age Males, Per Capita

Village:	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1954-55	1958-59
R^2 (after adjustment	47.5%	43.3%	66.0%
(before adjustment	48.6%	45.6%	69.3%
Number of households	192	98	99
Intercept	39	33	55
(Standard error)	(459)	(86)	(127)
Coefficient for number of dairy animals*	458	85	171
(Standard error)	(99)	(39)	(61)
t-statistic	4.6	2.2	2.8
Significance	<0.1%	3.4%	0.7%
Coefficient for land area owned	78	39	184
(Standard error)	(18)	(18)	(34)
t-statistic	4.4	2.2	5.4
Significance	<0.1%	3.1%	<0.1%
Coefficient for workers' education	125	42	83
(Standard error)	(15)	(12)	(25)
t-statistic	8.6	3.4	3.3
Significance	<0.1%	0.1%	0.2%
Coefficient for number of working-age males	578	248	248
(Standard error)	(164)	(51)	(74)
t-statistic	3.5	4.9	3.4
Significance	0.1%	<0.1%	0.2%
Range of working age	21-65	15-55	19-60

* Dairy animals were measured in "cow-equivalents"

Table 5: Regressions of Income Per Capita on Land Owned, Workers' Education, and Number of Working-Age Males, Per Capita, and Caste Dummy Variables

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
Adjusted R^2	43.5%	68.7%	84.7%	66.8%	87.0%
(Same, without castes)	(41.6%)	(68.1%)	(84.4%)	(66.9%)	(87.0%)
Number of households	269	298	94	80	151
Coefficient for priestly castes	399	526	181	130	15
(Standard error)	(115)	(179)	(254)	(151)	(341)
t-statistic	3.5	2.9	0.71	0.86	0.04
Significance	0.1%	0.4%	48%	39%	50%
Coefficient for intermediate castes	-1	-17	-236	-131	-181
(Standard error)	(58)	(84)	(214)	(119)	(406)
t-statistic	0.03	0.20	1.1	1.1	0.45
Significance	>50%	>50%	27%	27%	>50%
Coefficient for scheduled castes	16*	-1*	147*	-36*	148*
(Standard error)	(51)	(68)	(130)	(70)	(171)
t-statistic	0.31	0.01	1.1	0.51	0.86
Significance	>50%	>50%	26%	>50%	39%
Coefficient for workers' education	39	60	138	59	255
(Standard error)	(23)	(25)	(28)	(24)	(40)
t-statistic	1.7	2.4	5.0	2.4	6.3
Significance	8.7%	1.9%	<0.1%	1.9%	<0.1%

* Indicates coefficients with "wrong" sign, or that coefficient for the scheduled castes was greater than coefficient for the intermediate castes. "Correct" signs were positive for priestly castes, negative for intermediate castes and scheduled castes.

-- Table continued on next page --

Table 5: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
Adjusted R^2	42.9%	36.1%	40.7%	67.1%
(Same, without castes)	(41.8%)	(36.5%)	(40.0%)	(66.7%)
Number of households	192	291	98	99
Coefficient for priestly castes	169	59	None in village	None in village
(Standard error)	(180)	(141)		
t-statistic	0.94	0.42		
Significance	35%	>50%		
Coefficient for intermediate castes	-115	107*	-14	-22
(Standard error)	(117)	(121)	(22)	(33)
t-statistic	0.99	0.89	0.63	0.68
Significance	33%	38%	>50%	50%
Coefficient for scheduled castes	-198	28*	-41	-28
(Standard error)	(91)	(77)	(33)	(54)
t-statistic	2.2	0.37	1.2	0.52
Significance	3.2%	>50%	22%	>50%
Coefficient for workers' education	108	120	40	90
(Standard error)	(16)	(29)	(12)	(26)
t-statistic	6.9	4.1	3.2	3.4
Significance	<0.1%	<0.1%	0.2%	0.2%

* Indicates coefficient with "wrong" sign, or that coefficient for the scheduled castes was greater than the coefficient for the intermediate castes. "Correct" signs were positive for priestly castes, negative for intermediate castes and scheduled castes.

Table 6: Regressions of Income Per Capita on Castes and Average Income Per Capita by Castes

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	9.5%	8.9%	9.0%	36.5%	19.3%
(before adjustment	10.2%	9.8%	11.9%	38.9%	20.9%
Average income per capita:					
Priestly castes	957	1358	1605	242	1407*
Cultivating castes	327	913	930	619	2861
Intermediate castes	249	507	217	267	864
Scheduled castes	177	399	185	165	597

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	8.7%	7.9%	4.3%	15.9%
(before adjustment	10.1%	8.9%	6.3%	17.6%
Average income per capita:				
Priestly castes	1133	954	---	---
Cultivating castes	724	949	195	340
Intermediate castes	834*	681	136	151
Scheduled castes	330	551	147*	157*

* Indicates income with wrong ranking

Table 7: Regressions of Unskilled Labor Income Per Capita on the Numbers of Working-Age Males and Females, Per Capita

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	41.5%	35.8%	78.4%	51.5%	63.1%
(before adjustment	42.8%	37.2%	81.0%	54.4%	64.3%
Number of households	88	91	27	21	62
Intercept	43	121	-39	94	64
(Standard error)	(59)	(162)	(79)	(48)	(105)
Coefficient for number of working-age males	217	565	522	239	765
(Standard error)	(31)	(82)	(54)	(50)	(86)
t-statistic	7.1	6.9	9.7	4.8	8.9
Significance	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
Range of working age	21-65	19-60	15-60	15-65	15-60
Coefficient for number of females aged 19-50	148	317	236	(Omitted	537
(Standard error)	(30)	(124)	(108)	since not	(126)
t-statistic	5.0	2.6	2.2	signif.	4.3
Significance	<0.1%	1.3%	3.9%	at 50%)	<0.1%

-- Table continued on next page --

Table 7: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	74.7%	57.5%	56.9%	9.9%
(before adjustment	81.0%	58.5%	58.6%	13.8%
Number of households	5	45	26	24
Intercept	-68	177	28	71
(Standard error)	(61)	(163)	(32)	(42)
Coefficient for number of working-age males	1326	1629	202	81
(Standard error)	(371)	(209)	(35)	(43)
t-statistic	3.6	7.8	5.8	1.9
Significance	3.8%	<0.1%	<0.1%	7.4%
Range of working age	21-55	21-60	15-55	19-60
Coefficient for number of females aged 19-50 omitted because:	(Negative)	(Negative)	Not signif. at 48%)	Not signif. at 50%)

Table 8: Regressions of Skilled Occupation Income Per Capita on Workers' Education and the Number of Working-Age Males, Per Capita

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	53.9%	38.5%	76.4%	46.3%	67.5%
(before adjustment	57.3%	41.4%	79.8%	53.0%	68.6%
Number of households	28	43	8	9	31
Intercept	-409	212	283	185	353
(Standard error)	(509)	(438)	(162)	(238)	(421)
Coefficient for workers' education	149	111	79	(Omitted	236
(Standard error)	(67)	(54)	(16)	since not	(30)
t-statistic	2.2	2.0	4.9	signif.	8.0
Significance	3.5%	4.9%	0.3%	at 50%)	0.1%
Coefficient for number of working-age males	1841	627	(Omitted	633	(Omitted
(Standard error)	(385)	(515)	since	(225)	since
t-statistic	4.8	1.2	negative)	2.8	negative,
Significance	0.1%	23%		2.7%	not
					signif.
Range of working age	19-50	15-65		19-50	at 50%)

-- Table continued on next page --

Table 8: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	40.7%	57.6%	75.2%	11.1%
(before adjustment	44.9%	60.4%	77.8%	15.1%
Number of households	29	24	20	23
Intercept	106	162	31	102
(Standard error)	(881)	(514)	(75)	(73)
Coefficient for workers' education	77	160	25	(Omitted
(Standard error)	(42)	(47)	(19)	since not
t-statistic	1.8	3.4	1.3	signif.
Significance	7.7%	0.3%	20%	at 50%)
Coefficient for number of working-age males	1240	919	491	138
(Standard error)	(507)	(526)	(79)	(72)
t-statistic	2.4	1.7	6.2	1.9
Significance	2.2%	9.6%	0.1%	618%
Range of working age	21-50	19-65	15-50	19-50

Table 9: Regressions of Cultivation Income Per Capita on Net Land Area Cultivated, Workers' Education and the Number of Workers, Per Capita

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	51.8%	79.5%	95.7%	72.6%	91.4%
(before adjustment	53.8%	80.3%	95.1%	76.7%	92.2%
Number of households	74	74	34	21	35
Intercept	24	-414	196	-11	602
(Standard error)	(288)	(476)	(273)	(261)	(1053)
Coefficient for net land area cultivated	526	1969	175	512	1811
(Standard error)	(62)	(117)	(9)	(113)	(217)
t-statistic	8.5	17.	20.	4.5	8.3
Significance	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
Coefficient for workers' education	-14	-91	-83	139	593
(Standard error)	(50)	(65)	(34)	(84)	(134)
t-statistic	0.29	1.4	2.4	1.7	4.4
Significance	>50%	17%	2.1%	12%	<0.1%
Coefficient for number of working-age males	-119	1000	-661	-832	-6143
(Standard error)	(271)	(503)	(362)	(715)	(1083)
t-statistic	0.44	2.0	1.8	1.2	5.7
Significance	>50%	5.1%	7.8%	26%	<0.1%
Range of working age	15-50	15-60	21-60	19-60	21-65

-- Table continued on next page --

Table 9: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62*	1968-69	1954-55	1958-59
R^2 (after adjustment	47.3%	33.6%	61.6%	70.4%
(before adjustment	48.5%	34.9%	64.8%	79.2%
Number of households	129	156	37	31
Intercept	143	-35	51	-3
(Standard error)	(208)	(471)	(78)	(114)
Coefficient for net land area cultivated	91	287	237	317
(Standard error)	(10)	(37)	(33)	(65)
t-statistic	9.4	7.8	7.1	4.9
Significance	<0.1%	<0.1%	<0.1%	<0.1%
Coefficient for workers' education	136	67	-43	-63
(Standard error)	(34)	(53)	(18)	(46)
t-statistic	3.9	1.3	2.3	1.4
Significance	<0.1%	21%	2.6%	18%
Coefficient for number of working-age males	-198	365	-184	353
(Standard error)	(104)	(357)	(138)	(186)
t-statistic	1.9	1.0	1.3	1.9
Significance	6.0%	31%	19%	6.9%
Range of working age	15-50	19-50	21-55	21-60

* For the 1961-62 survey of Naurangdeshar, the dependent variable was gross income from cultivation, but all coefficients and their standard errors have been multiplied by 0.645 to convert approximately to net income. See text for details.

Table 10: Regressions of Cultivation Income Per Capita on Net Land Area Cultivated Per Capita

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	52.7%	78.8%	92.5%	71.2%	80.7%
(before adjustment	53.3%	79.1%	92.7%	72.6%	81.3%
Number of households	74	74	34	21	35
Intercept	-20	-232	-55	-191	-544
(Standard error)	(285)	(483)	(362)	(268)	(1580)
Coefficient for net land area cultivated	524	1920	142	628	2088
(Standard error)	(58)	(116)	(7)	(89)	(174)
t-statistic	9.1	17.	20.	7.1	12.
Significance	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62*	1968-69	1954-55	1958-59
R^2 (after adjustment	41.1%	33.3%	57.3%	71.4%
(before adjustment	41.6%	33.7%	58.5%	72.4%
Number of households	129	156	37	31
Intercept	123	39	15	102
(Standard error)	(219)	(472)	(82)	(127)
Coefficient for net land area cultivated	95	308	182	272
(Standard error)	(10)	(35)	(26)	(31)
t-statistic	9.5	8.9	7.0	8.7
Significance	<0.1%	<0.1%	<0.1%	<0.1%

* For the 1961-62 survey of Naurangdeshar, the dependent variable was gross income from cultivation, but all coefficients and their standard errors have been multiplied by 0.645 to convert approximately to net income. See text for details.

Table 11: Consumer Price Indices

Village	Year	Index	Village	Year	Index
Ankodia	1960-61	1.00	Naurangdeshar	1961-62	1.00
	1967-68	1.60		1968-69	1.96
Bhatian	1955-56	0.89	Sohalpur Gara	1954-55	1.00
	1960-61	1.00		1958-59	1.68
	1971-72	2.05			

Table 12: Selected Coefficients (and their Standard Errors), in Constant Rupees

Variable	Regression sample	Table	Ankodia 1960-61	Ankodia 1967-68	Bhatian 1955-56	Bhatian 1960-61	Bhatian 1971-72
Land area owned	General	3	449 (38)	860 (36)	93 (7)	355 (37)	883 (39)
Net land cultivated	Cultivation	10	524 (58)	1200 (73)	160 (8)	628 (89)	1019 (85)
# working-age males	General	3	561 (119)	341 (111)	*	334 (123)	*
# working-age males	Unskilled labour	7	217 (31)	353 (51)	586 (60)	239 (50)	373 (42)
Workers' education	General	3	60 (22)	47 (15)	139 (25)	53 (22)	111 (19)
Workers' education	Skilled occupations	8	149 (67)	69 (34)	89 (18)	*	115 (14)

Variable	Regression sample	Table	Naurang-deshar 1961-62	Naurang-deshar 1968-69	Sohalpur Gara 1954-55	Sohalpur Gara 1958-59
Land area owned	General	3	105 (18)	118 (12)	51 (17)	118 (21)
Net land cultivated	Cultivation	10	95 (10)	157 (18)	182 (26)	162 (19)
# working-age males	General	3	592 (173)	540 (119)	267 (51)	149 (45)
# working-age males	Unskilled labour	7	1326 (371)	830 (107)	202 (35)	48 (26)
Workers' education	General	3	114 (15)	62 (13)	40 (12)	52 (15)
Workers' education	Skilled occupations	8	77 (42)	82 (24)	25 (19)	*

* Indicates that coefficient was not estimated

Table 13: Average and Estimated Marginal Wage Incomes, in Constant Rupees, for Males Aged 15-55 in Unskilled Labour Households

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
R^2 (after adjustment	8.6%	38.3%	85.9%	13.8%	50.6%
(before adjustment	9.9%	39.0%	85.5%	18.6%	51.5%
Number of households*	72	84	26	19	59
Intercept	95	86	10	108	74
(Standard error)	(70)	(99)	(72)	(53)	(59)
Coefficient for males aged 15-55	114	400	596	191	442
(Standard error)	(41)	(55)	(48)	(97)	(57)
t-statistic	2.8	7.2	12.	2.	7.8
Significance	0.7%	<0.1%	<0.1%	6.6%	<0.1%
Average wage income per male aged 15-55	473	734	612	651	758
(Standard deviation)	(238)	(321)	(169)	(268)	(226)
Average wage income per capita	133	220	276	159	198
(Standard deviation)	(73)	(126)	(188)	(55)	(83)

* The number of households in this table is slightly lower than the number in Table 7 because the households containing no males aged 15 through 55 were excluded from the calculations for this table.

-- Table continued on next page --

Table 13: (continued)

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
R^2 (after adjustment	74.1%	55.9%	60.2%	-4.5%
(before adjustment	80.6%	57.0%	61.9%	0.7%
Number of households*	%	42	25	21
Intercept	-16	61	17	(See
(Standard error)	(62)	(81)	(31)	below)
Coefficient for males aged 15-55	875	764	226	(Not
(Standard error)	(248)	(105)	(37)	signif.
t-statistic	3.5	7.3	6.1	at 50%)
Significance	3.9%	<0.1%	<0.1%	
Average wage income per male aged 15-55	782	1037	279	182
(Standard deviation)	(215)	(280)	(90)	(89)
Average wage income per capita	188	260	107	60
(Standard deviation)	(109)	(120)	(48)	(25)

* The number of households in this table is slightly lower than the number in Table 7 because households containing no males aged 15 through 55 were excluded from the calculations for this table.

Table 14: Regressions of the Absolute Values of Residuals on Estimated Incomes

Village:	Ankodia	Ankodia	Bhatian	Bhatian	Bhatian
Year:	1960-61	1967-68	1955-56	1960-61	1971-72
Unadjusted R^2	26.8%	43.8%	19.3%	29.0%	27.1%
Intercept	3	6	165	21	216
(Standard error)	(222)	(271)	(398)	(124)	(520)
Coefficient for estimated income	0.514	0.358	0.164	0.295	0.166
(Standard error)	(0.052)	(0.024)	(0.035)	(0.052)	(0.022)
t-test	9.9	15.	4.7	5.6	7.4

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
Unadjusted R^2	13.9%	19.6%	18.3%	21.9%
Intercept	120	27	-4	32
(Standard error)	(383)	(296)	(61)	(89)
Coefficient for estimated income	0.360	0.394	0.415	0.271
(Standard error)	(0.065)	(0.047)	(0.089)	(0.052)
t-test	5.5	8.4	4.6	5.2

All coefficients for estimated income were significant at <0.1%

Table 15: Characteristics of Hypothetical Land Reform Programs

Village:	Ankodia	Ankodia	Bhatian	Bhatian
Year:	1960-61	1967-68	1960-61	1971-72
Areas in hectares:				
Ceiling	2.09	1.77	2.75	3.24
Floor	0.85	0.76	0.69	0.81
Floor/Ceiling	41%	43%	25%	25%
Numbers of households:				
Landowners	131	136	25	32
" , losing land	46	51	11	12
" , gaining land	54	53	5	5
Landless tenants	14	11	22	11
" , gaining land	14	11	22	11
Landless labor	82	90	21	62
" , gaining land	82	90	6	27

Village:	Naurang-deshar	Naurang-deshar	Sohalpur Gara	Sohalpur Gara
Year:	1961-62	1968-69	1954-55	1958-59
Areas in hectares:				
Ceiling	8.56	5.36	2.08	2.01
Floor	8.34	4.86	1.30	1.18
Floor/Ceiling	97%	91%	62%	59%
Numbers of households:				
Landowners	145	208	46	54
" , losing land	67	109	22	18
" , gaining land	78	81	10	21
Landless tenants	11	13	0	0
" , gaining land	11	13	0	0
Landless labor	5	43	25	22
" , gaining land	5	43	25	22

Table 16: Simulated Effects of Land Reform Programs on Quintile Average Per Capita Incomes

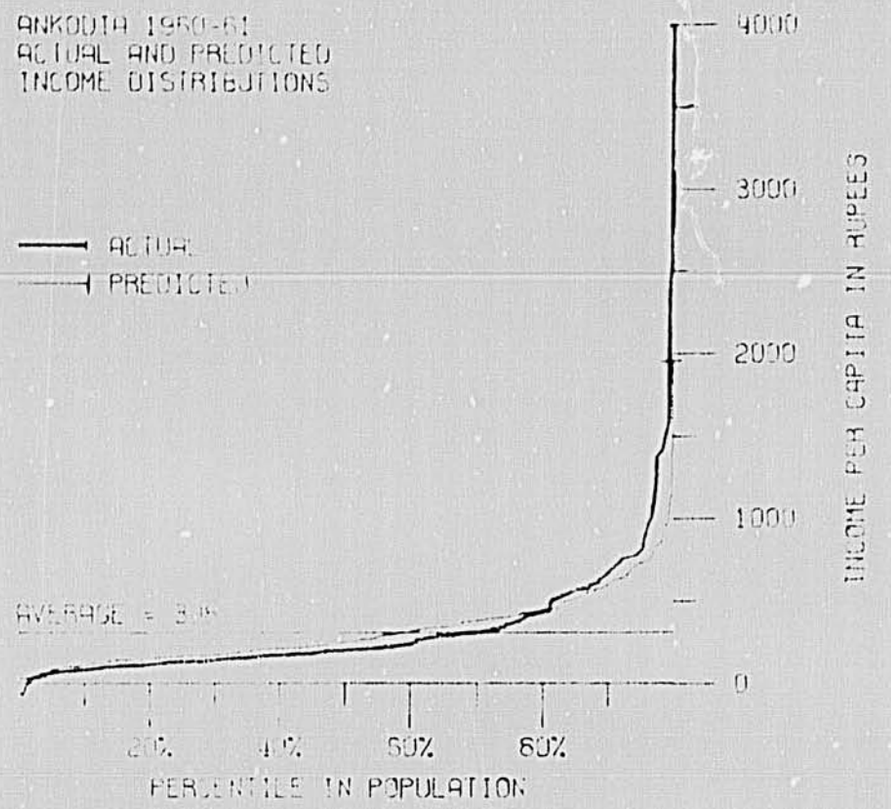
	Average Changes in Rupees					Changes as % of Pre-Reform Simulated Quintile Incomes				
Quintile:	I	II	III	IV	V	I	II	III	IV	V
Ankodia, 1960-61	22	35	28	27	-74	73%	32%	14%	8%	-9%
Ankodia, 1967-68	40	80	95	103	-259	20%	24%	19%	13%	-13%
Bhatian, 1960-61	7	9	20	16	-94	11%	6%	11%	6%	-11%
Bhatian, 1971-72	46	57	50	108	-453	37%	14%	7%	11%	-12%
Naurangdeshar, 1961-62	14	16	13	5	-28	22%	5%	3%	1%	-2%
Naurangdeshar, 1968-69	30	41	27	21	-95	13%	9%	4%	2%	-6%
Sohalpur Gara, 1954-55	1	2	2	0	-4	1.9%	2.1%	1.7%	-0.2%	-1.3%
Sohalpur Gara, 1958-59	8	8	15	13	-51	15%	6%	8%	5%	-9%

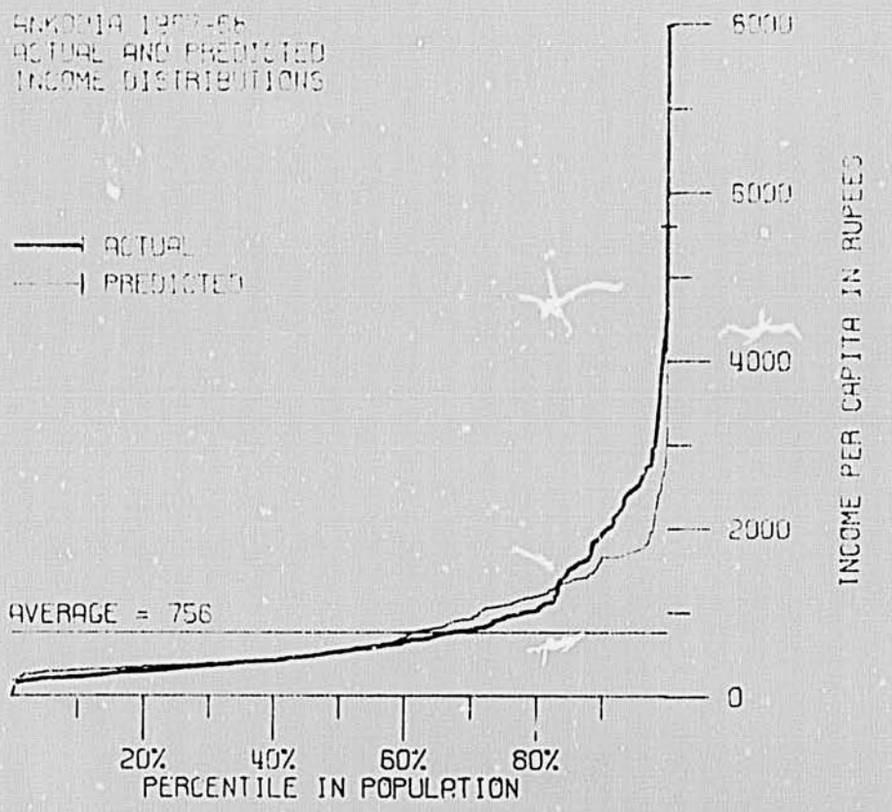
APPENDIX II

Graph Set 1	Actual and Predicted Income Distributions
Graph Set 2	Actual and Predicted Incomes and Distributions (Hypothetical Data)
Graph Set 3	Actual and Simulated Income Distributions for Naurangdeshar Village in 1968-69
Graph Set 4	Simulated Deviational Effects

Graph Set 1

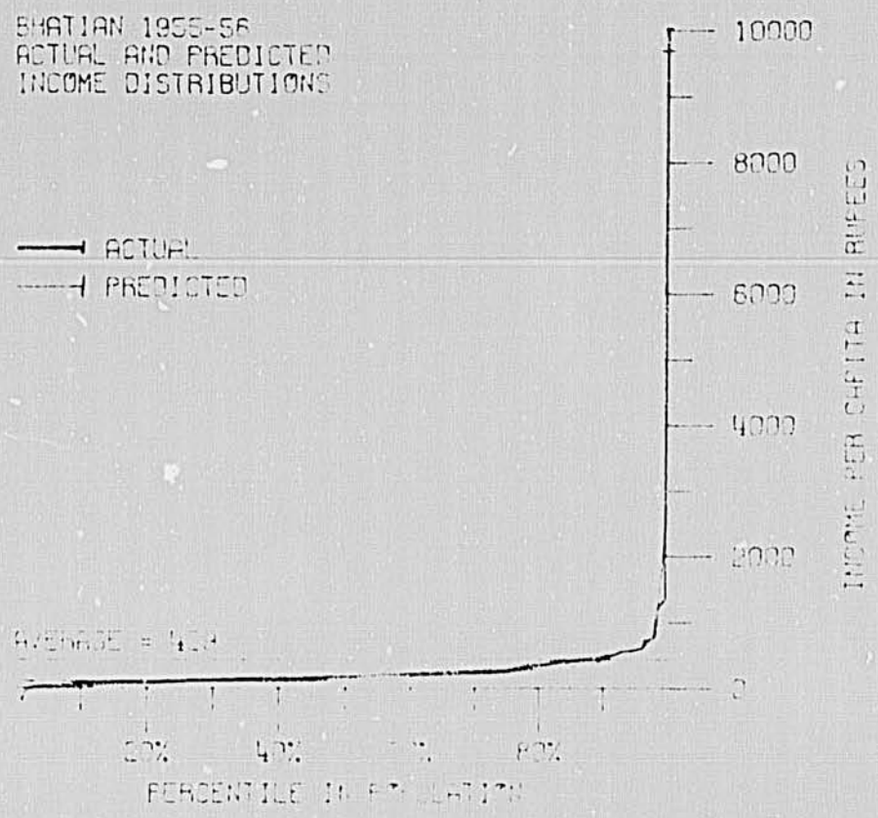
Actual and Predicted Income Distributions

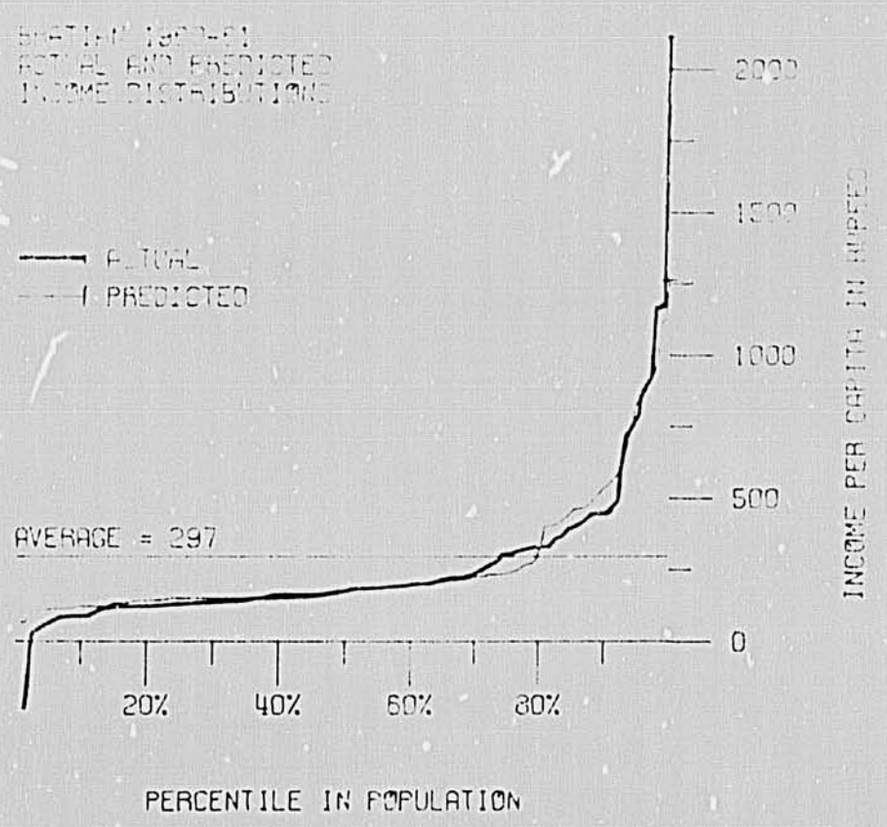




Graph Set 1

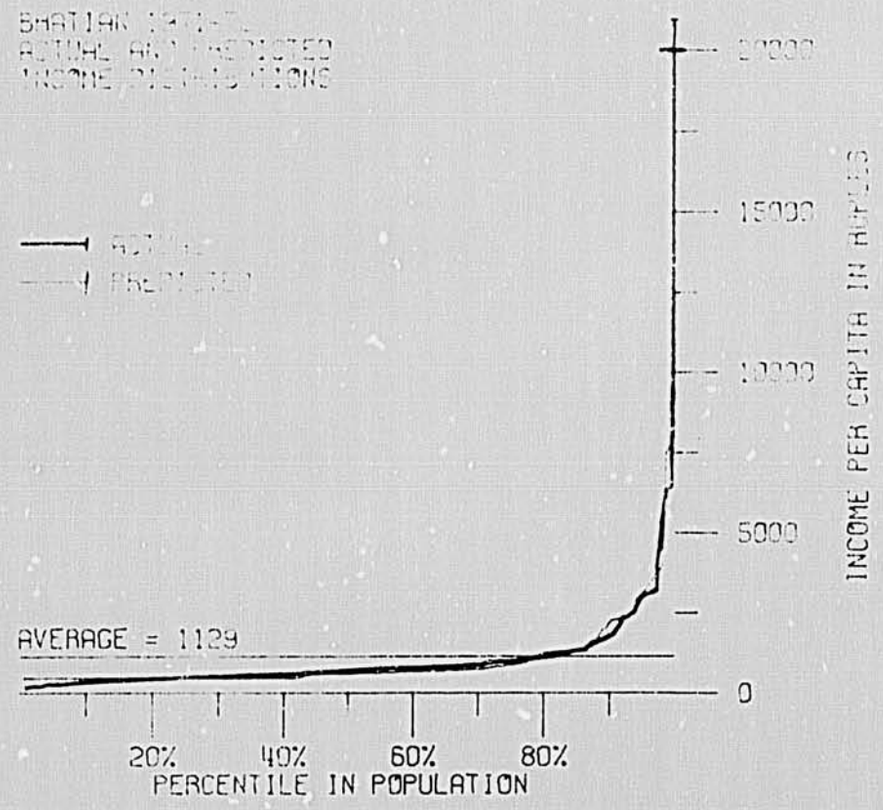
(continued)





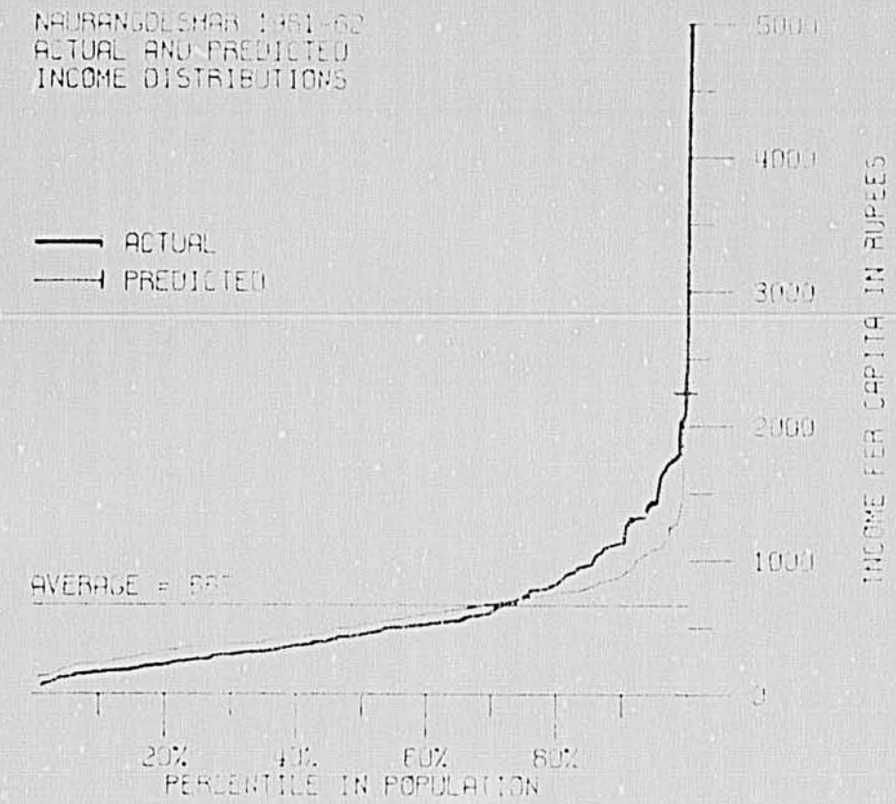
Graph Set 1

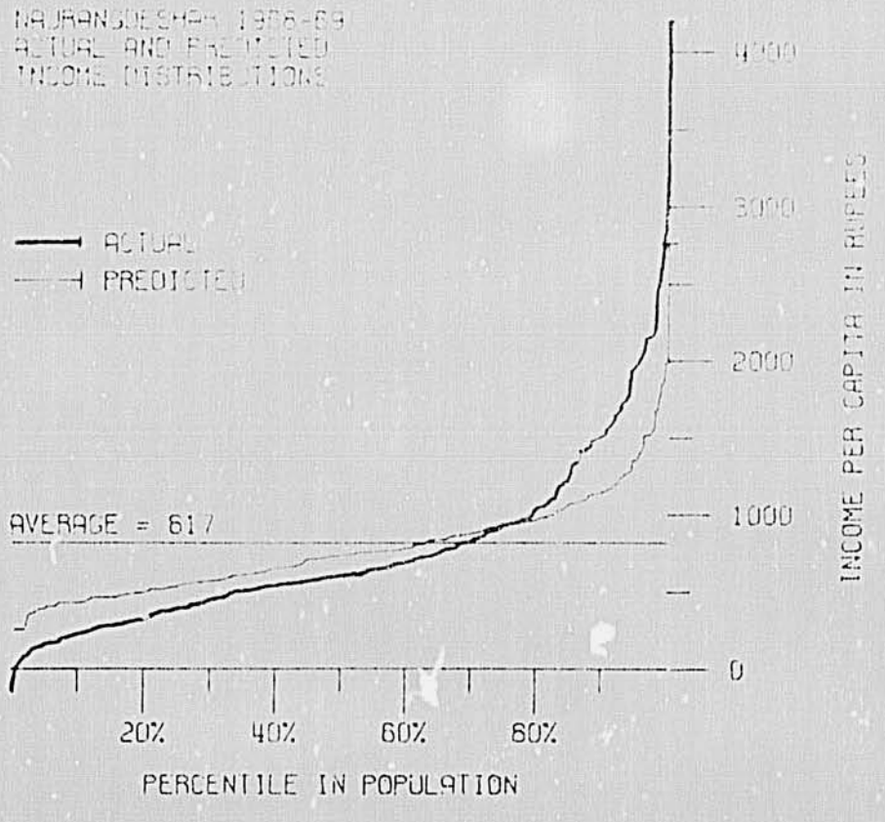
(continued)



Graph Set 1

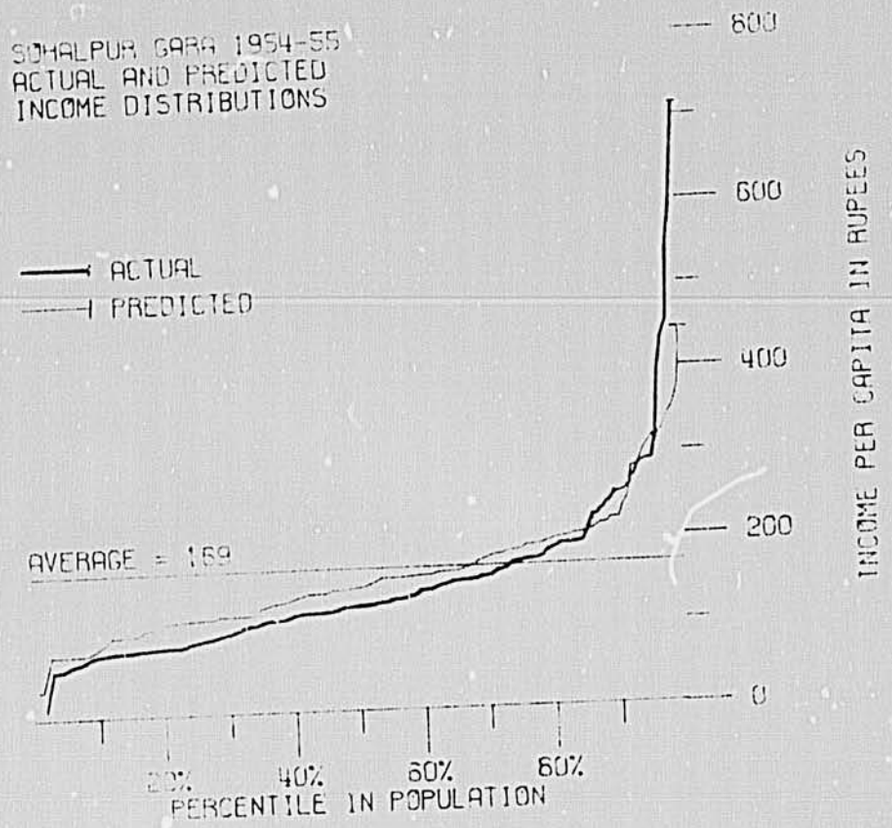
(continued)

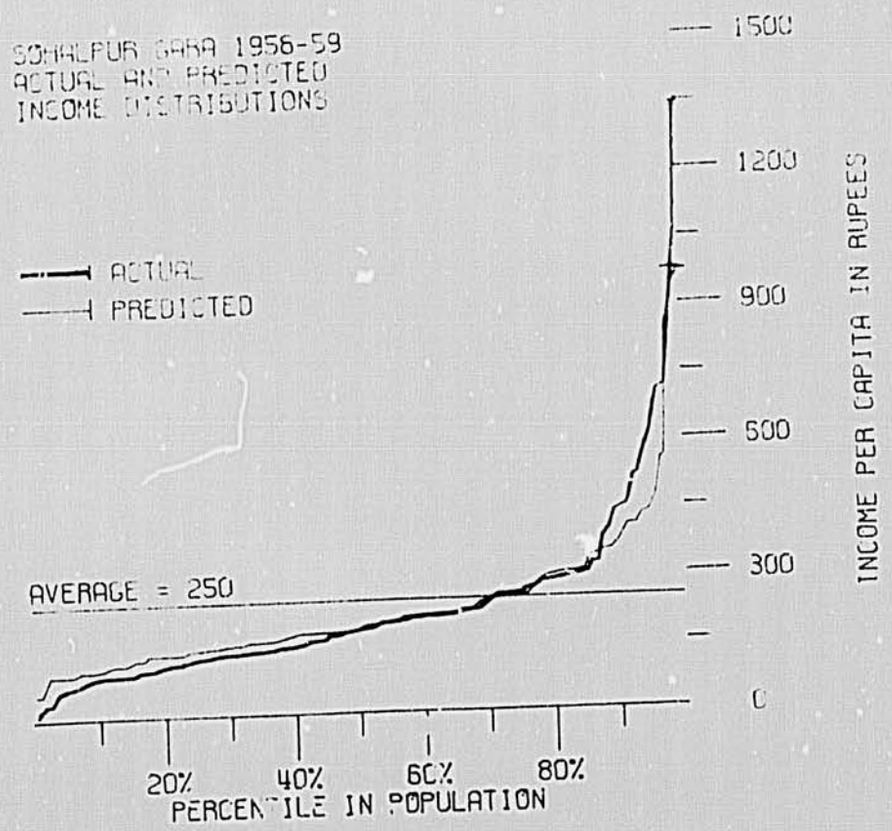




Graph Set 1

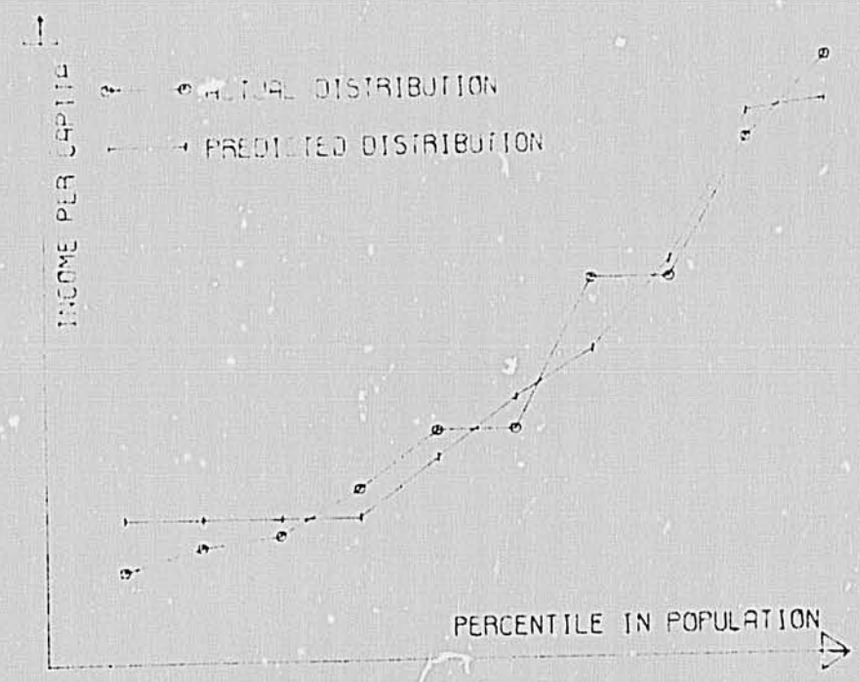
(continued)





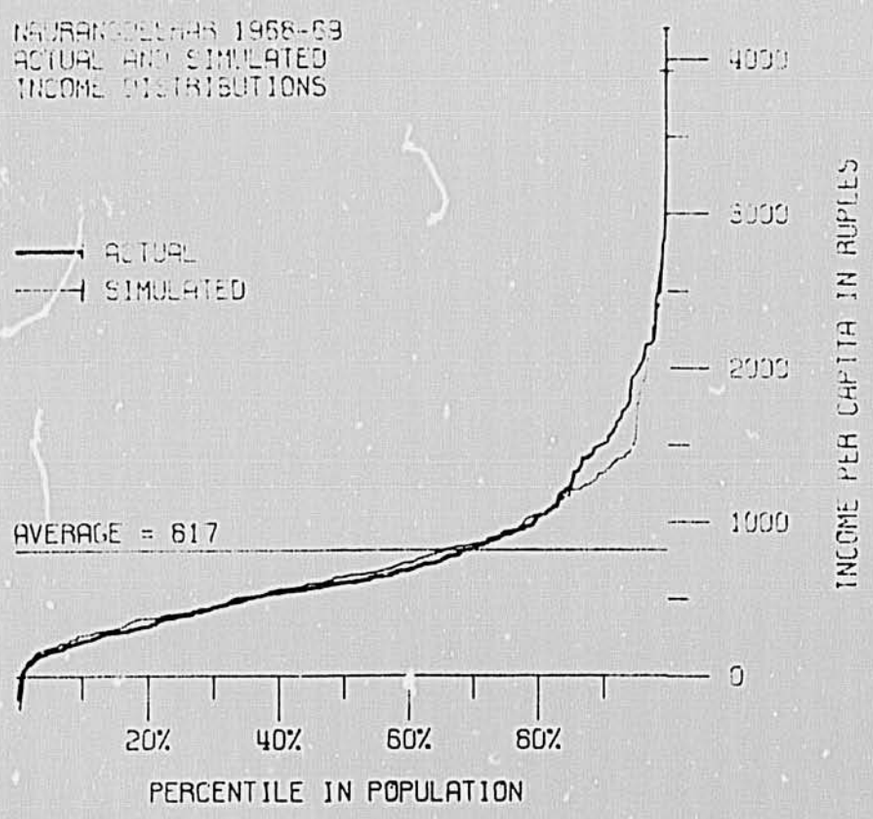
Graph Set 2

Actual and Predicted Incomes and Distributions (Hypothetical Data)



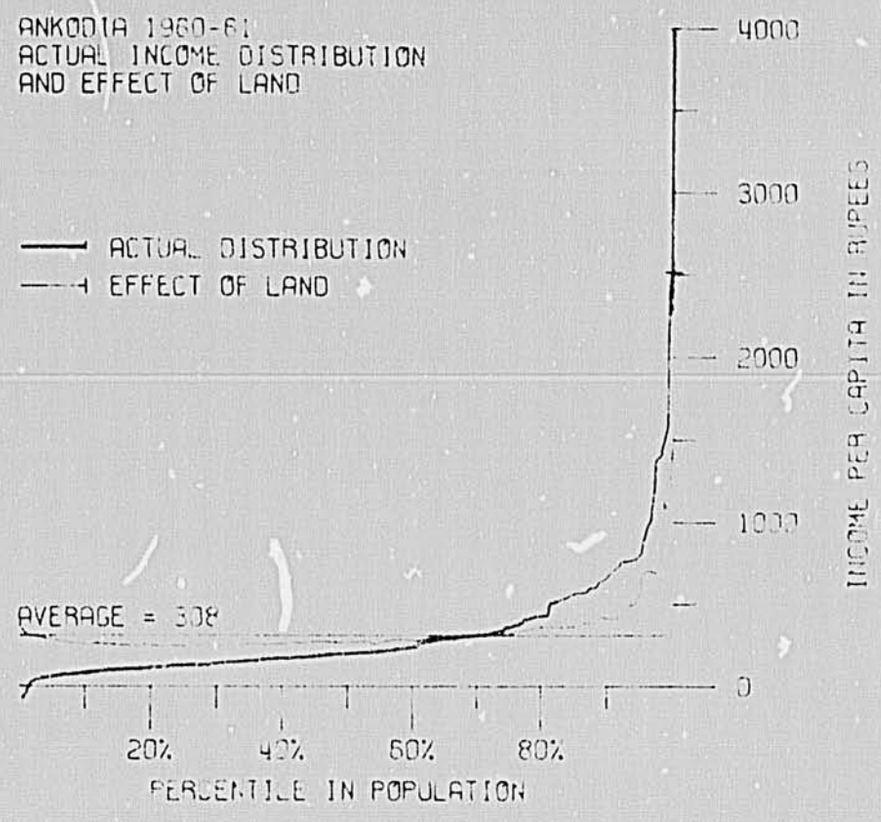
Graph Set 3

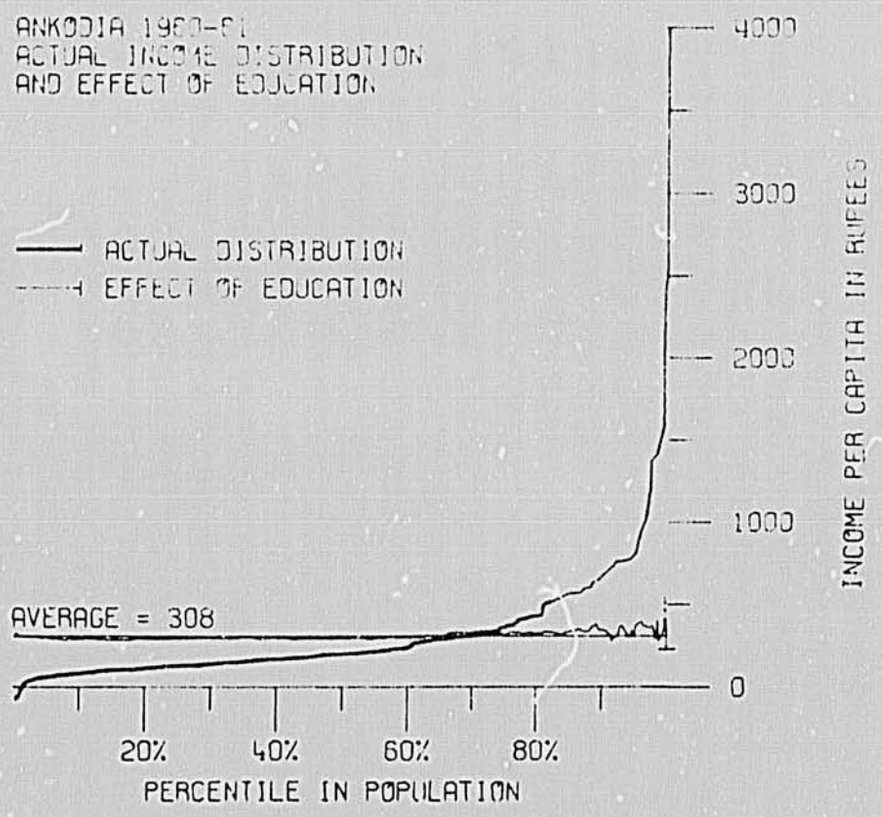
Actual and Simulated Income Distributions
for Naurangdeshar Village in 1968-69



Graph Set 4

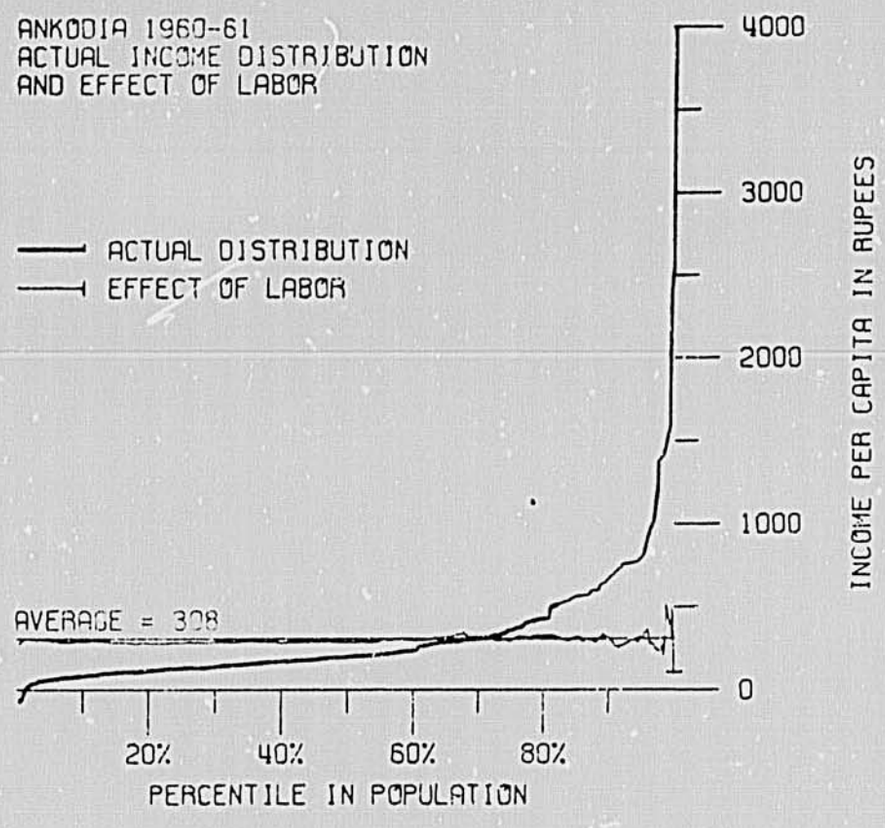
Simulated Deviational Effects

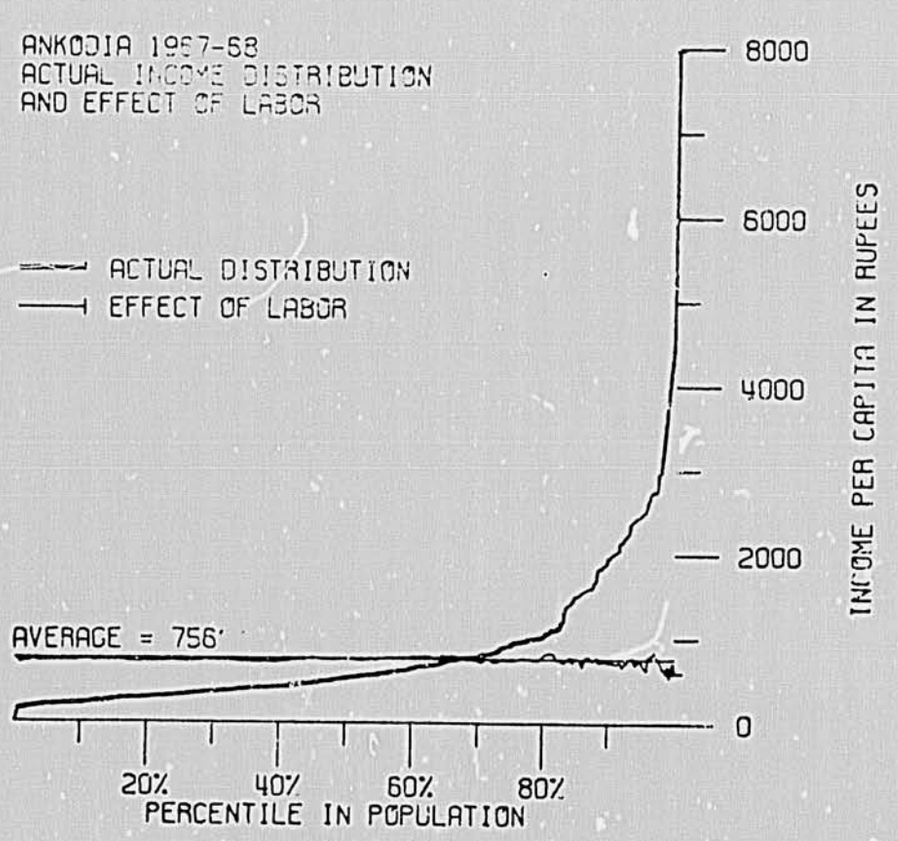




Graph Set 4

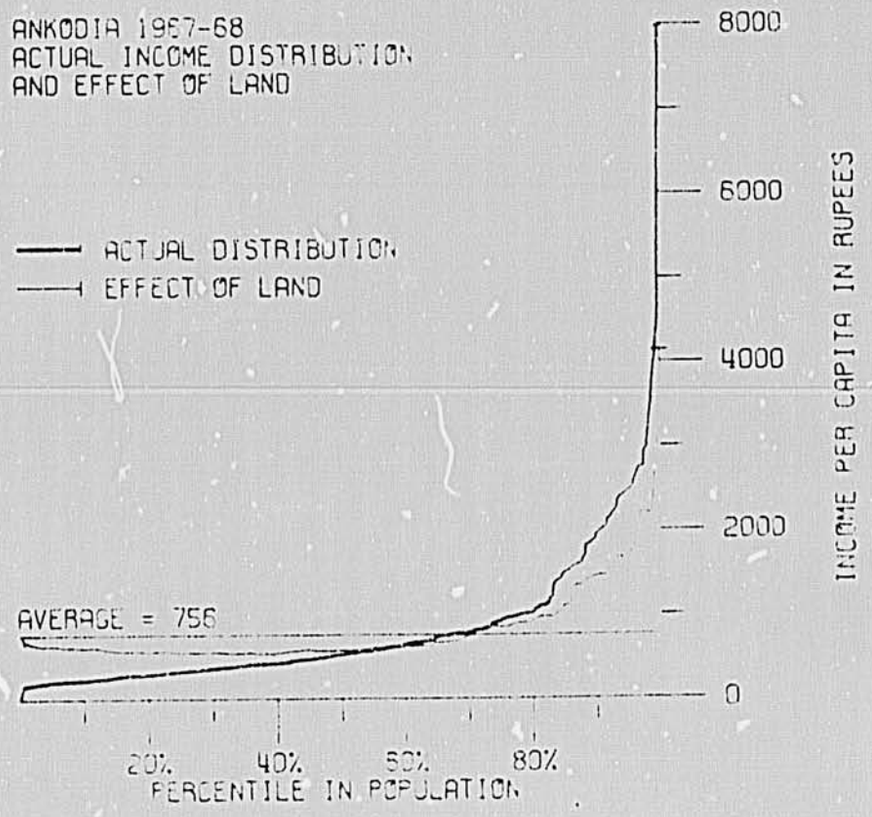
(continued)

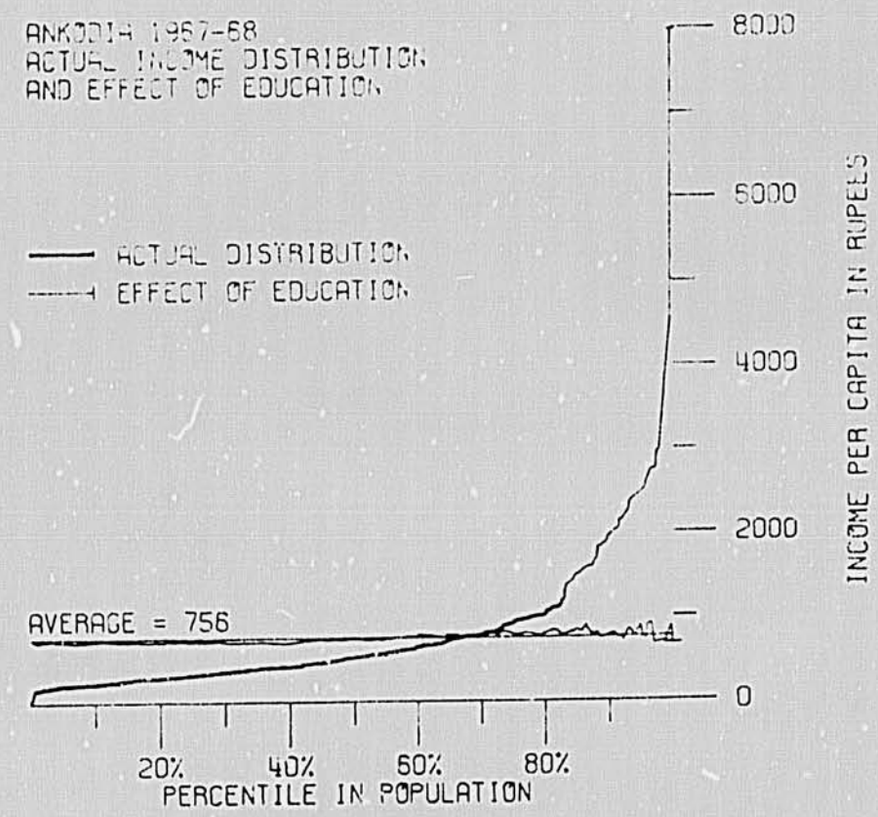




Graph Set 4

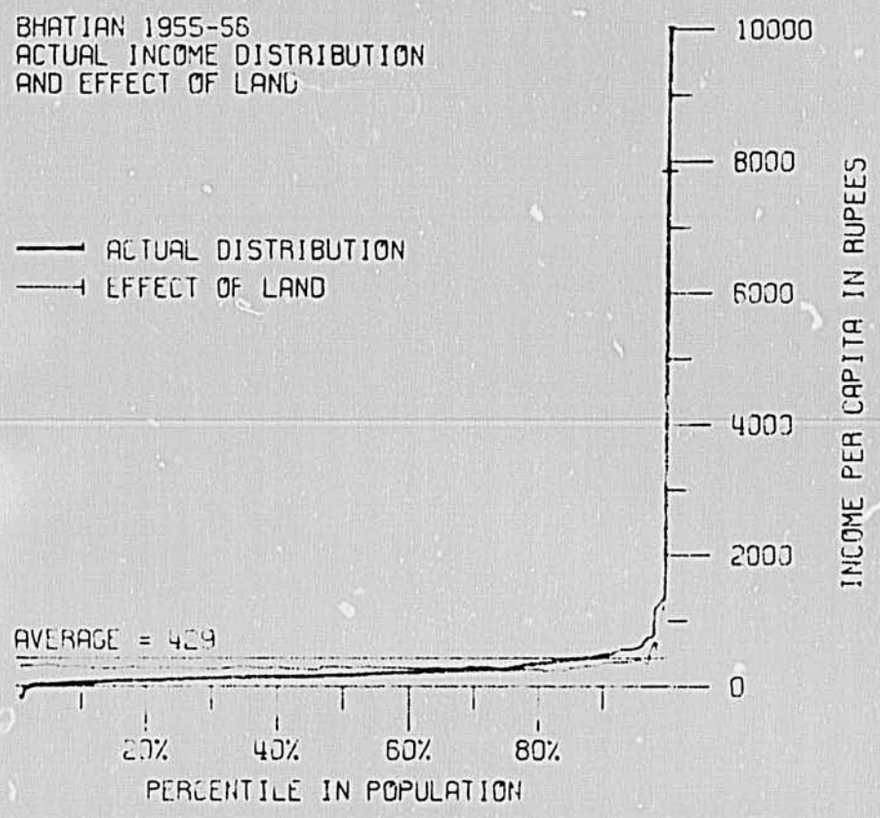
(continued)

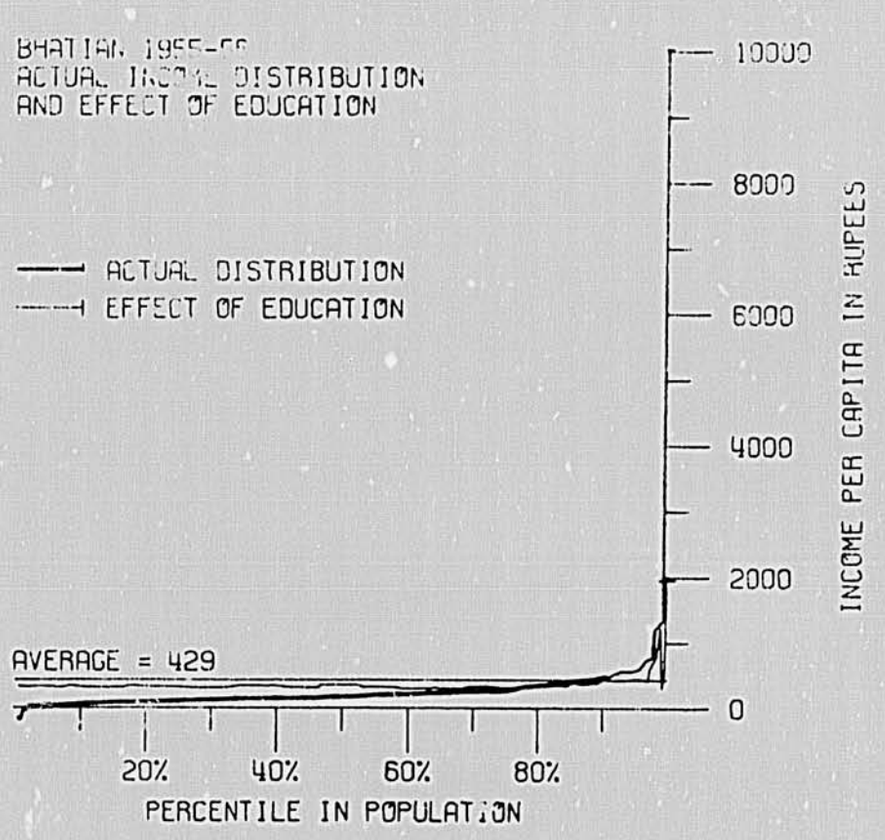




Graph Set 4

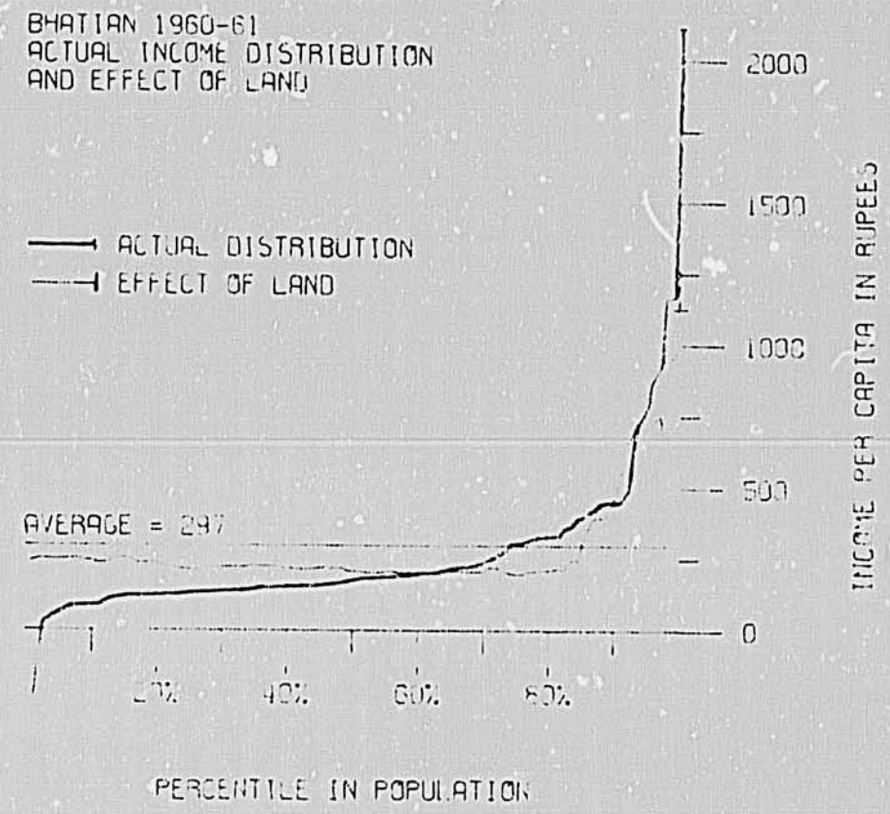
(continued)

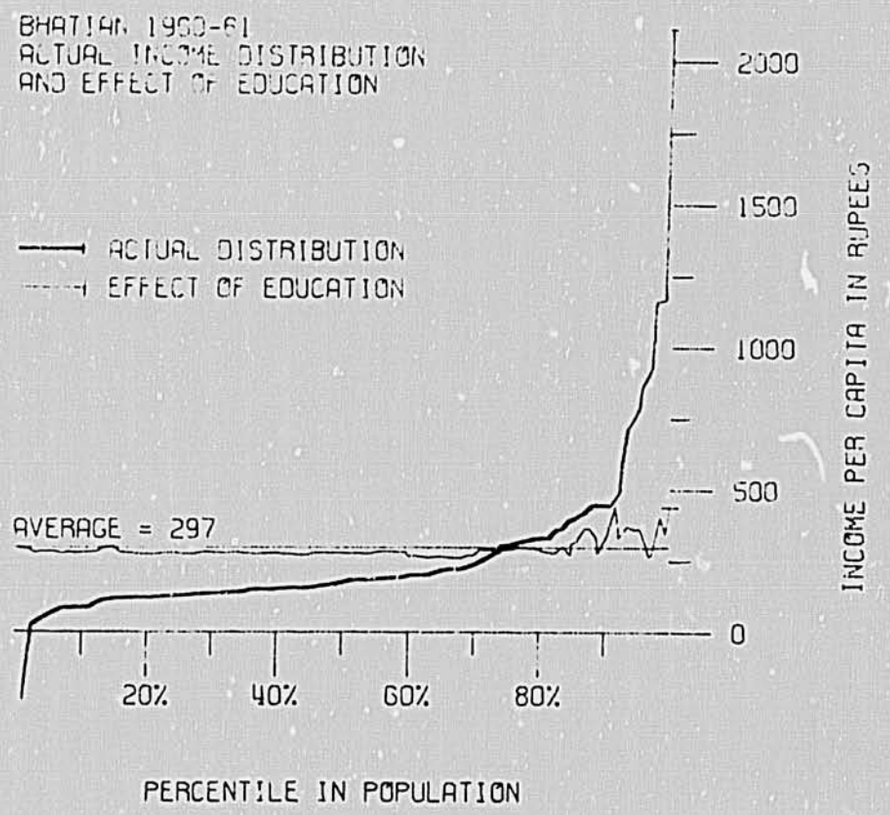




Graph Set 4

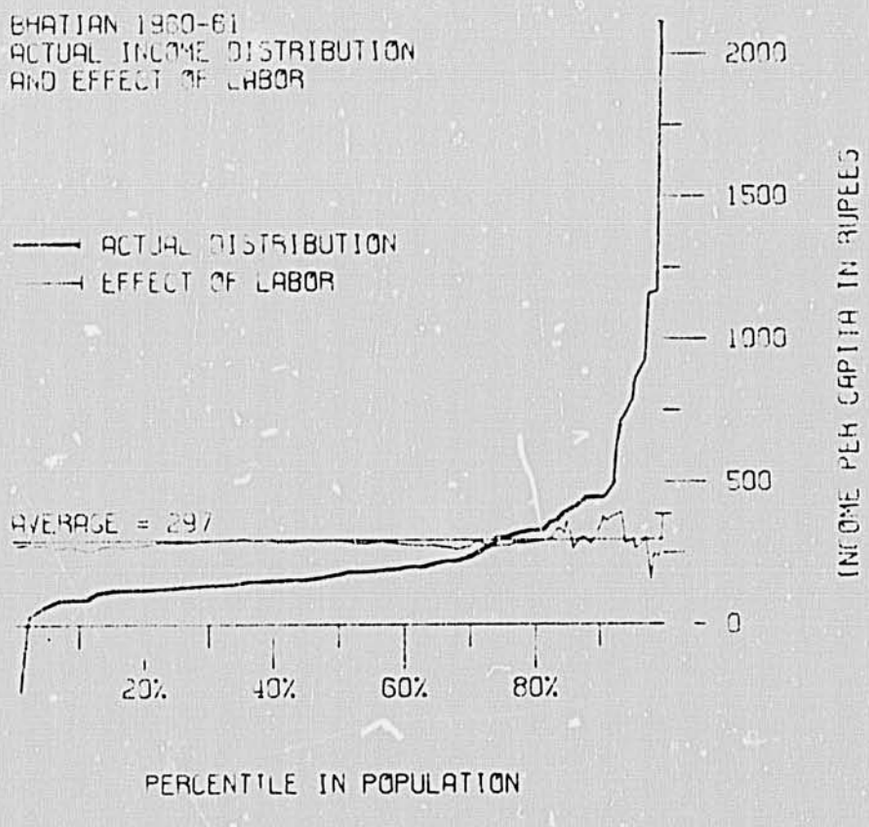
(continued)





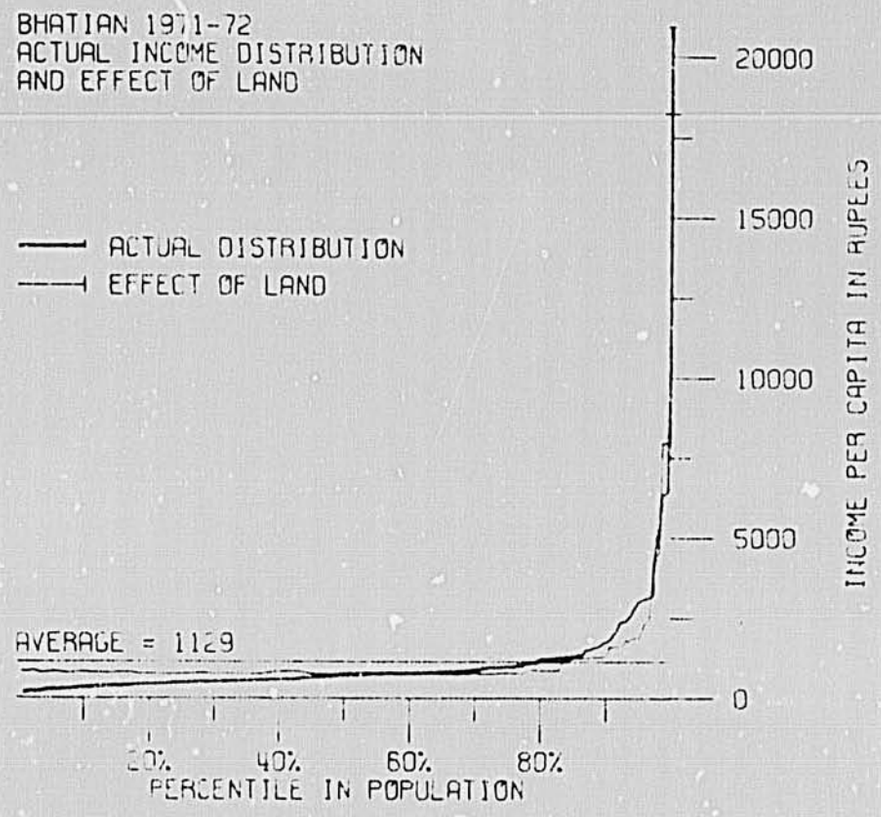
Graph Set 4

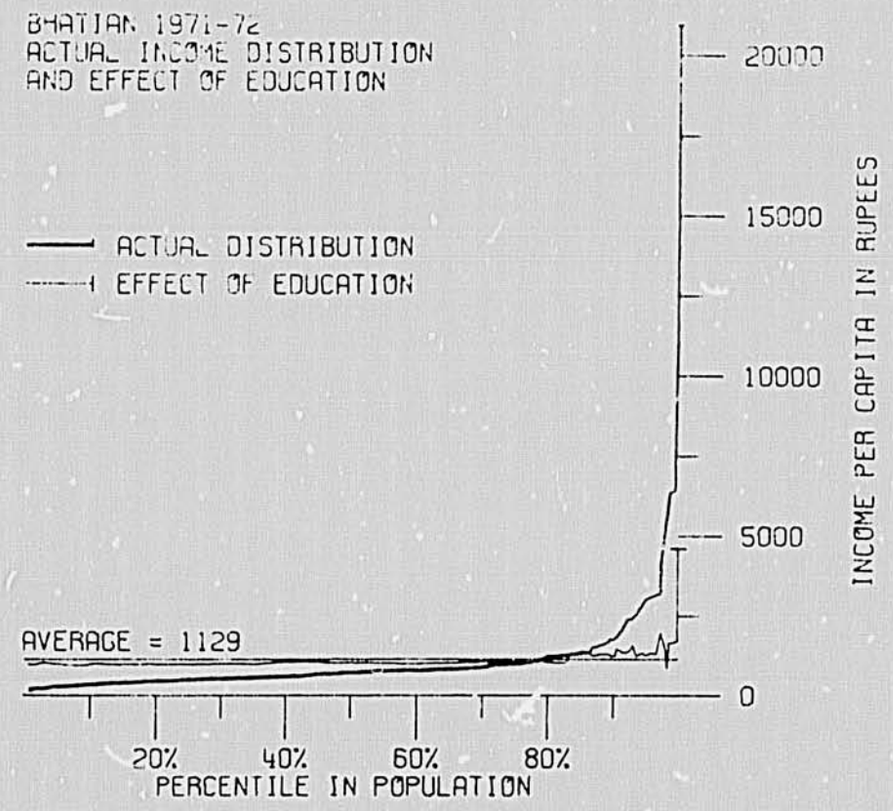
(continued)



Graph Set 4

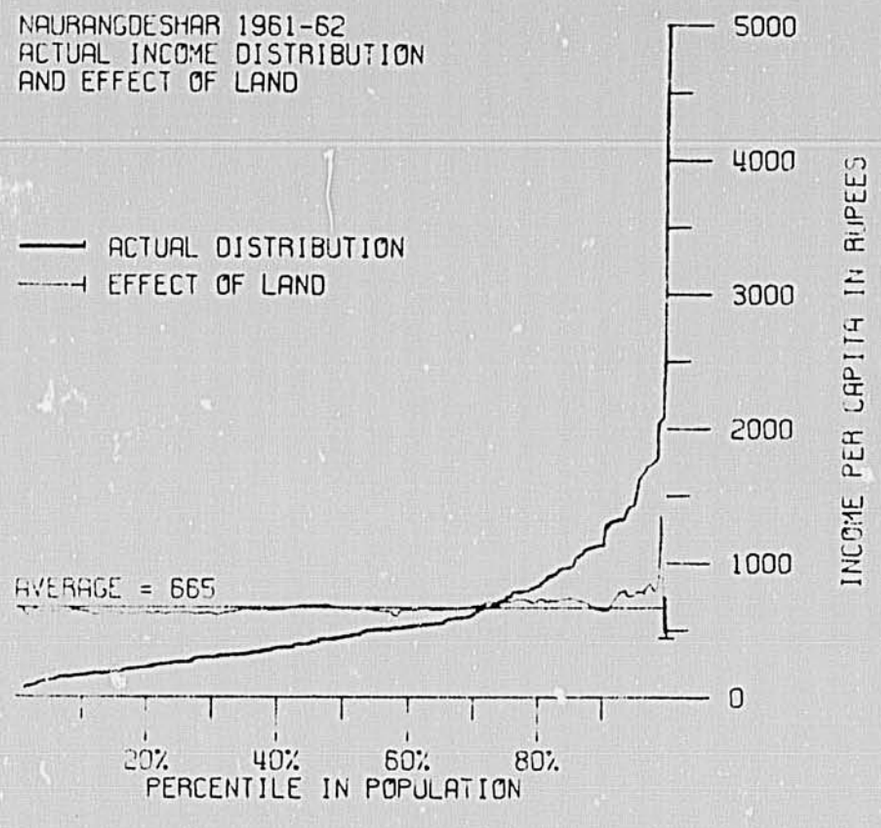
(continued)

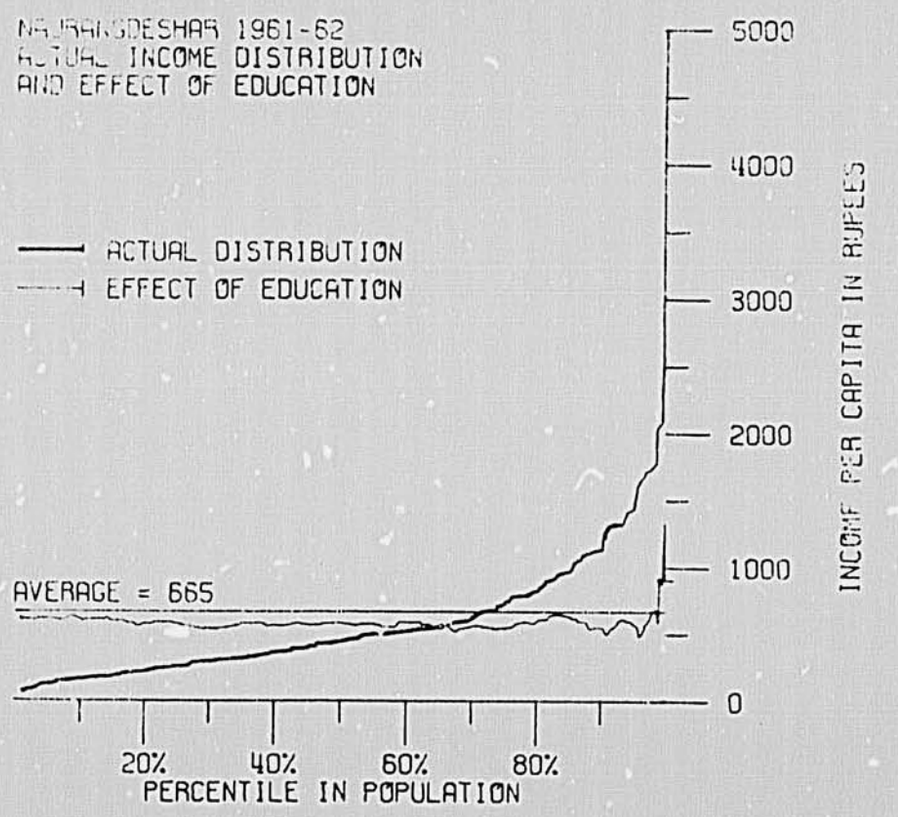




Graph Set 4

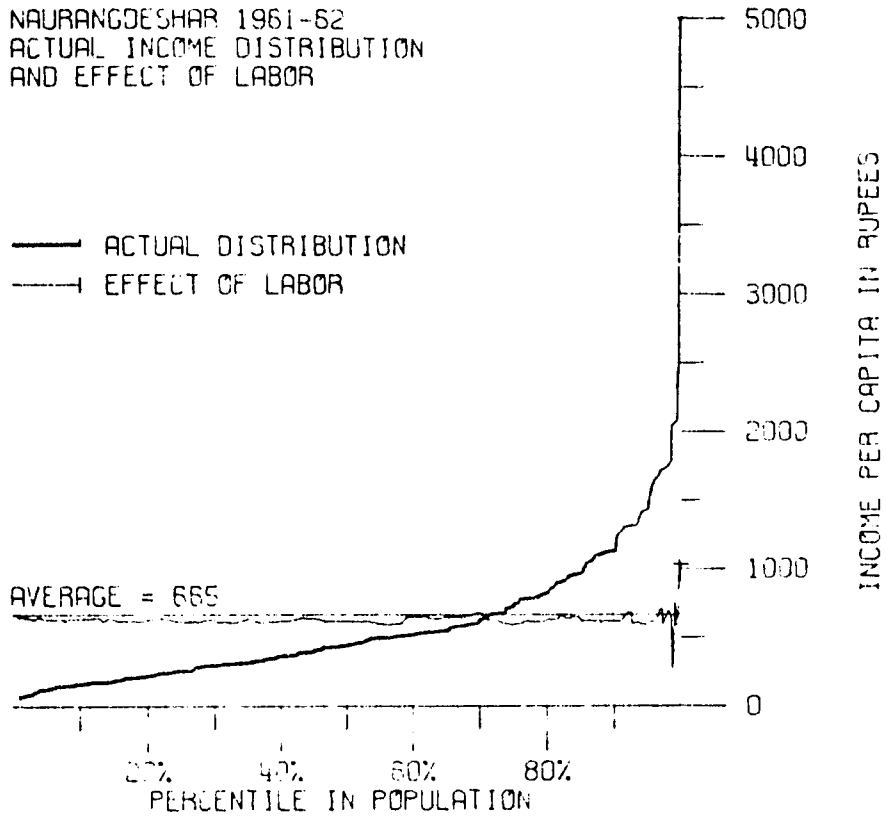
(continued)

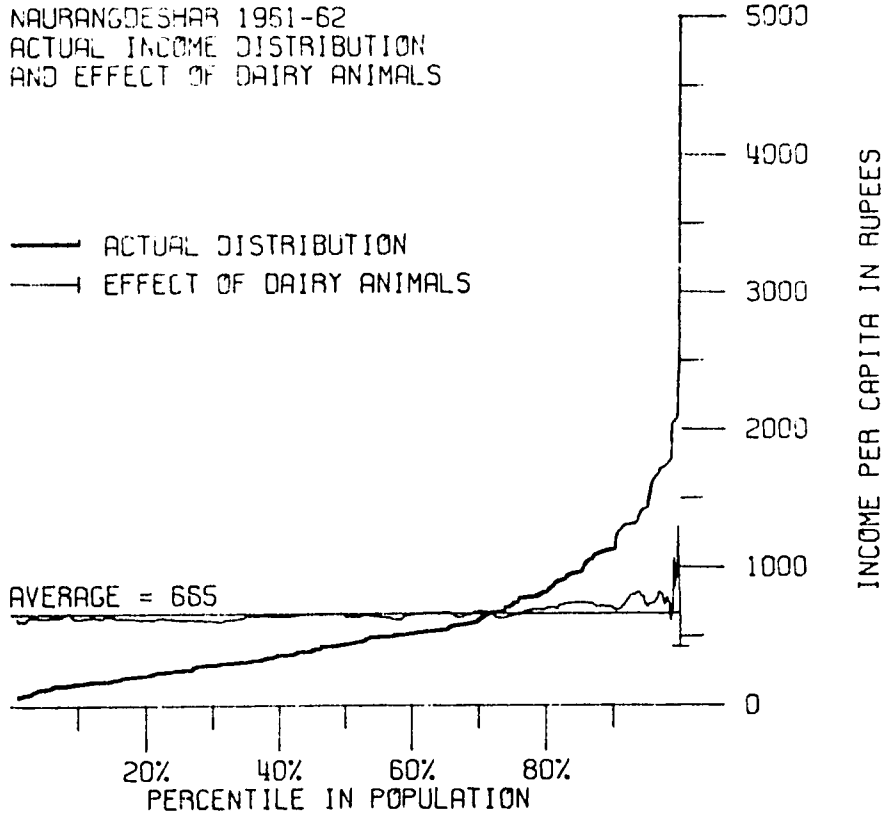




Graph Set 4

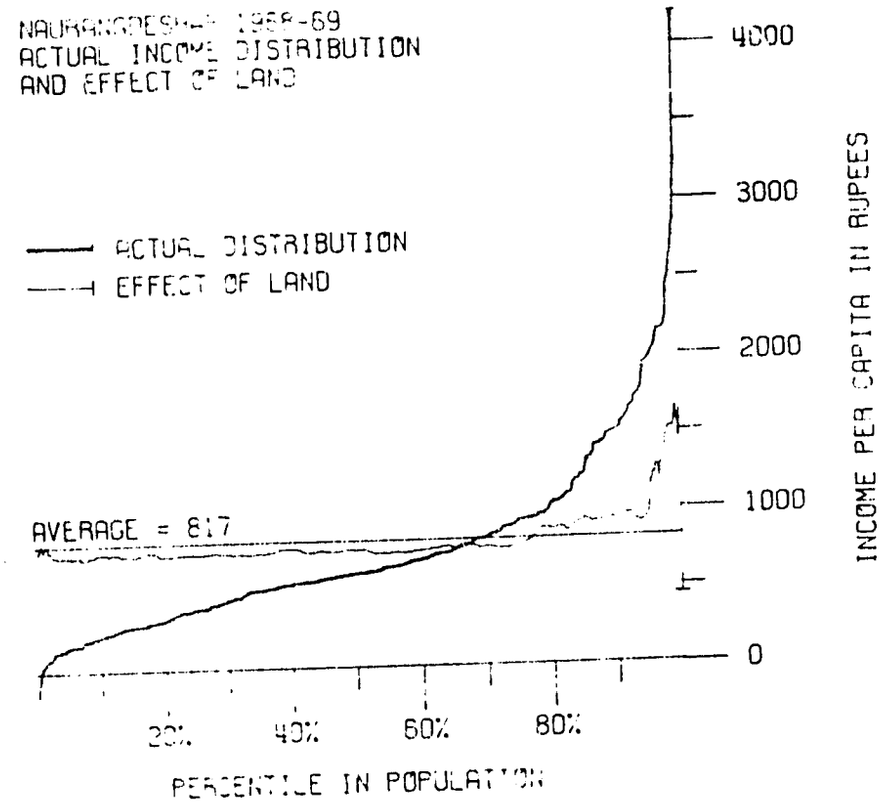
(continued)

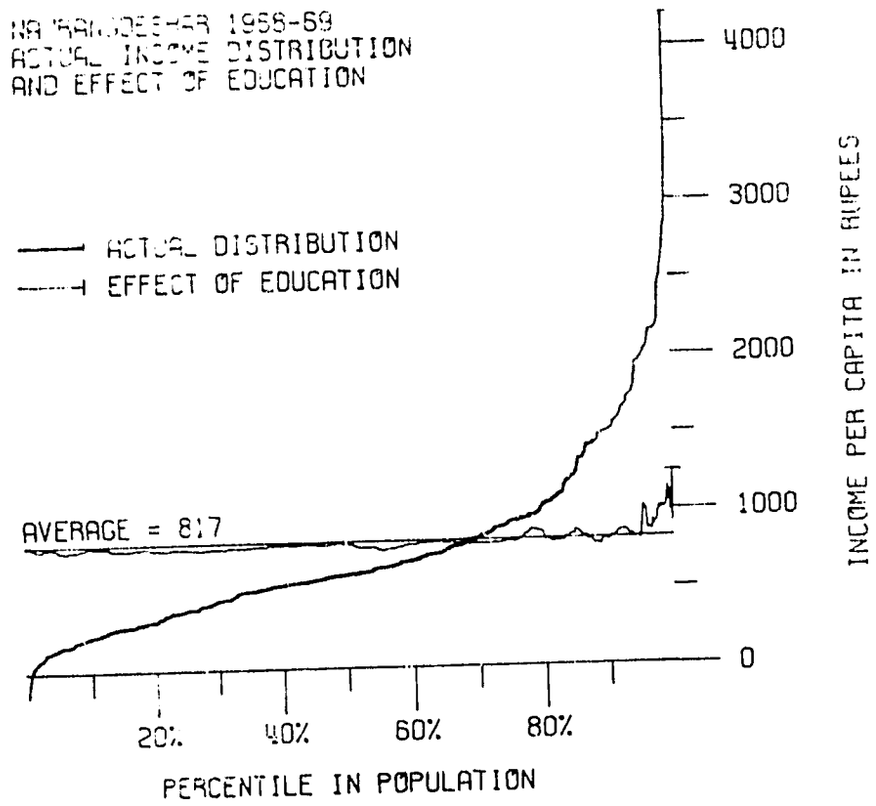




Graph Set 4

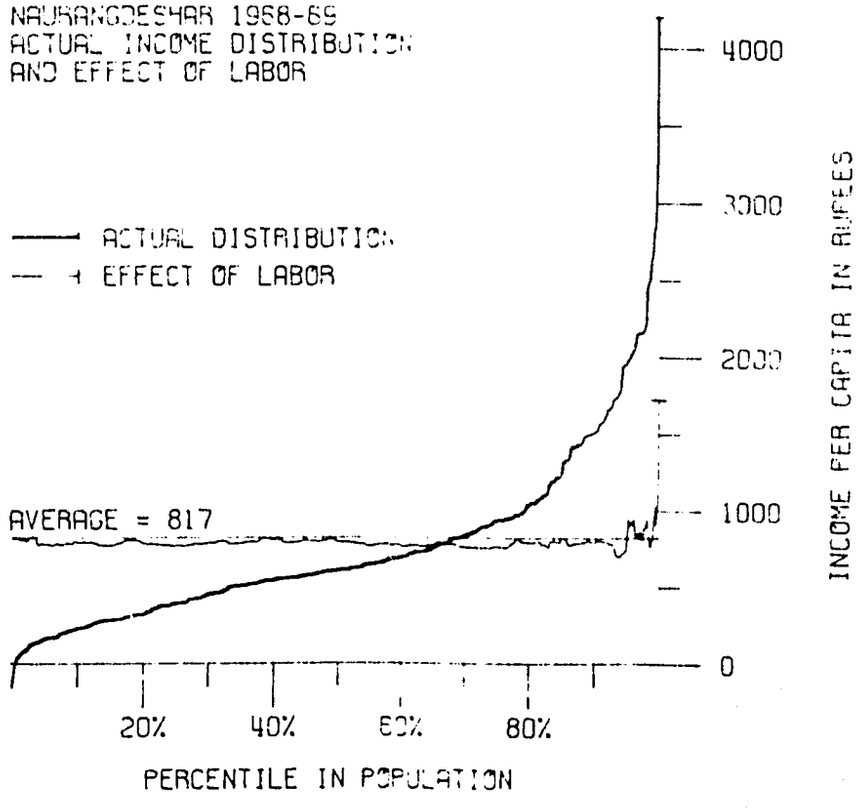
(continued)





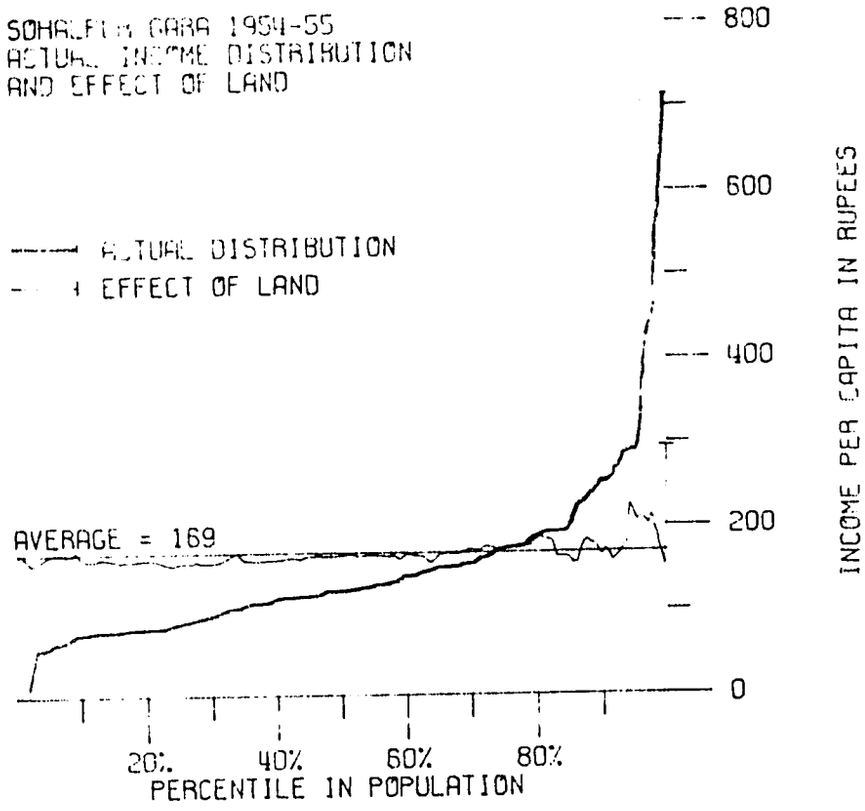
Graph Set 4

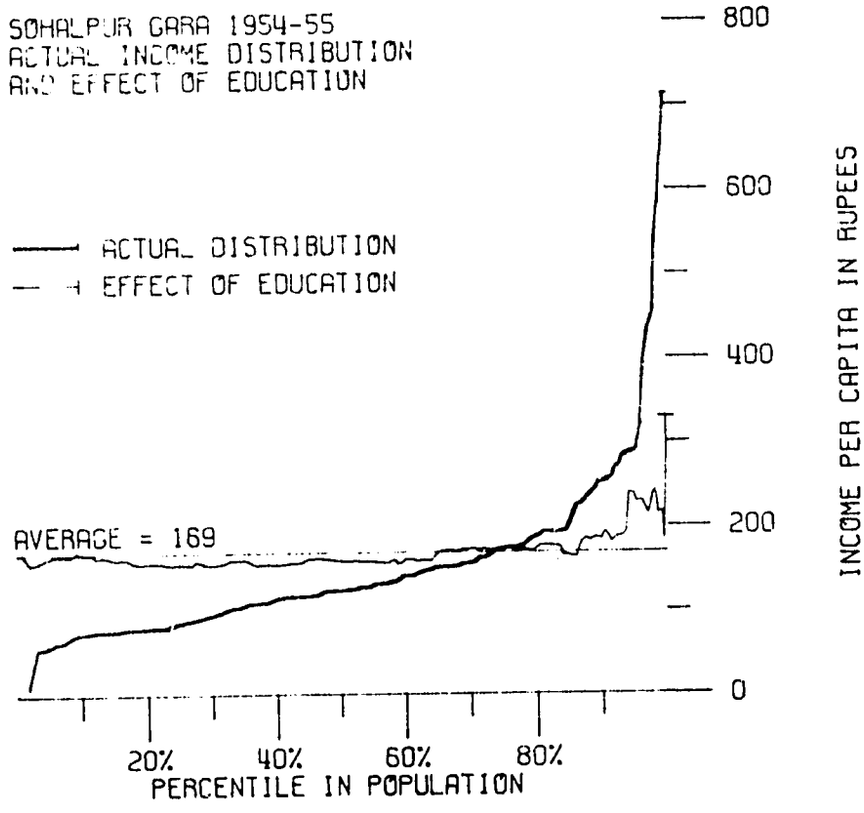
(continued)



Graph Set 4

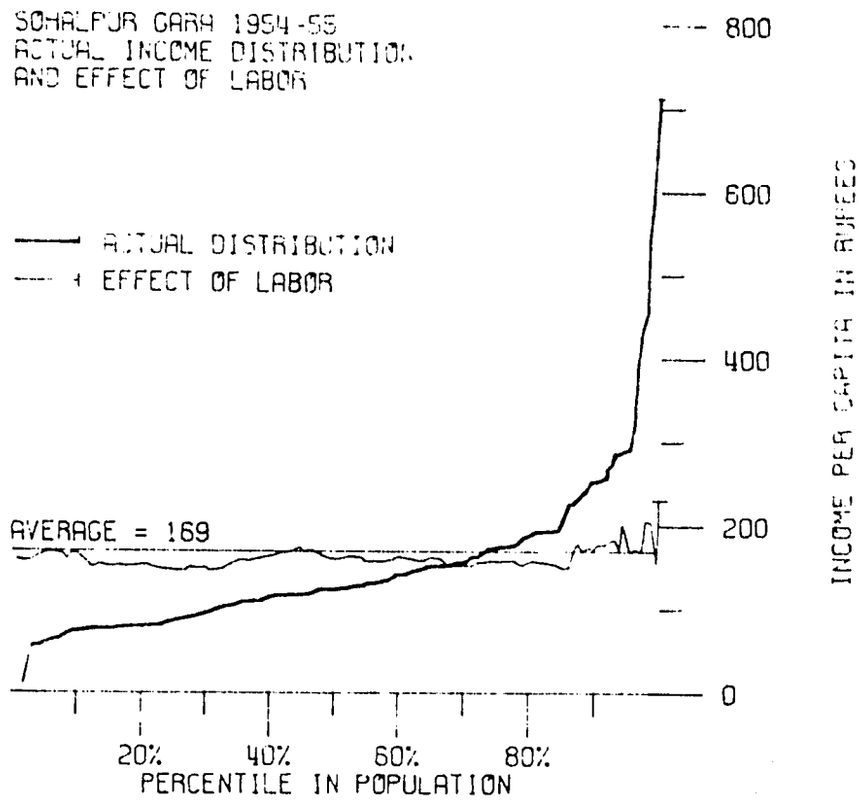
(continued)

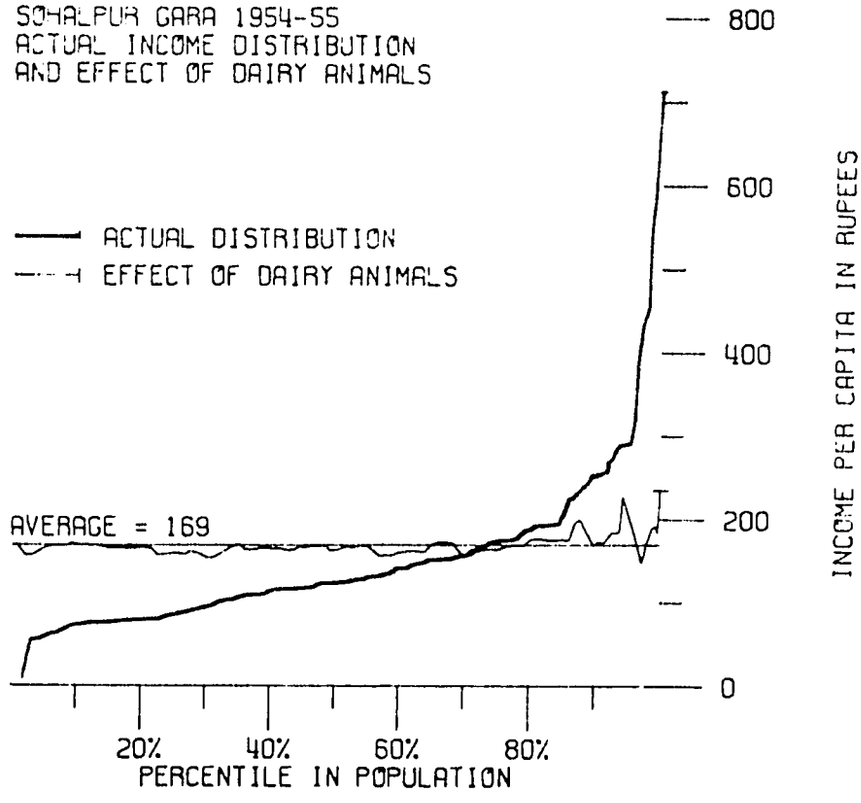




Graph Set 4

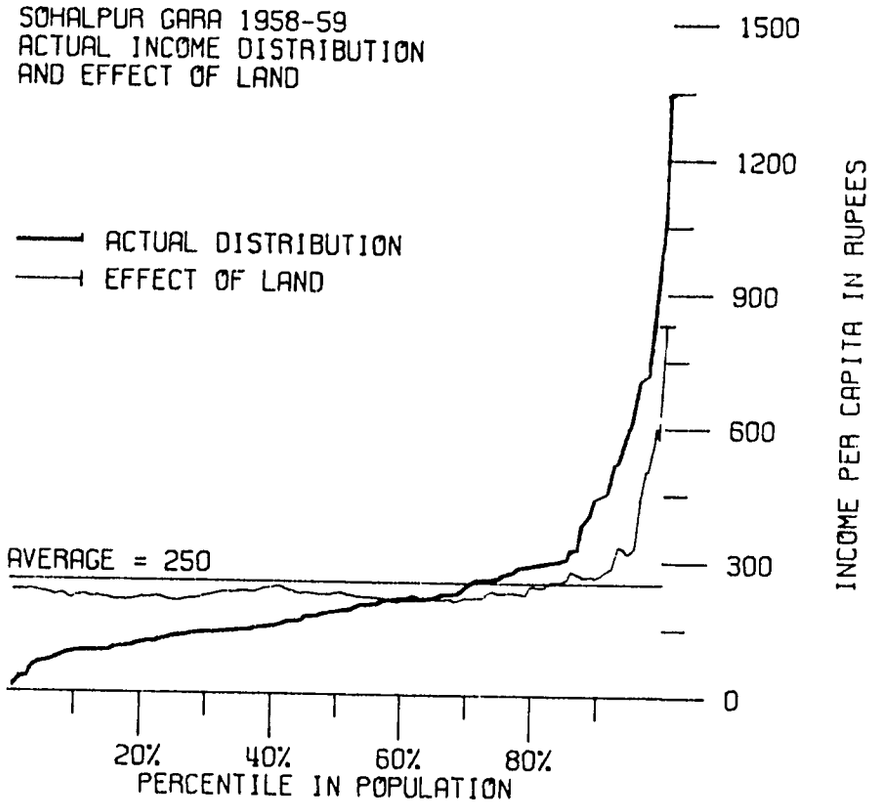
(continued)

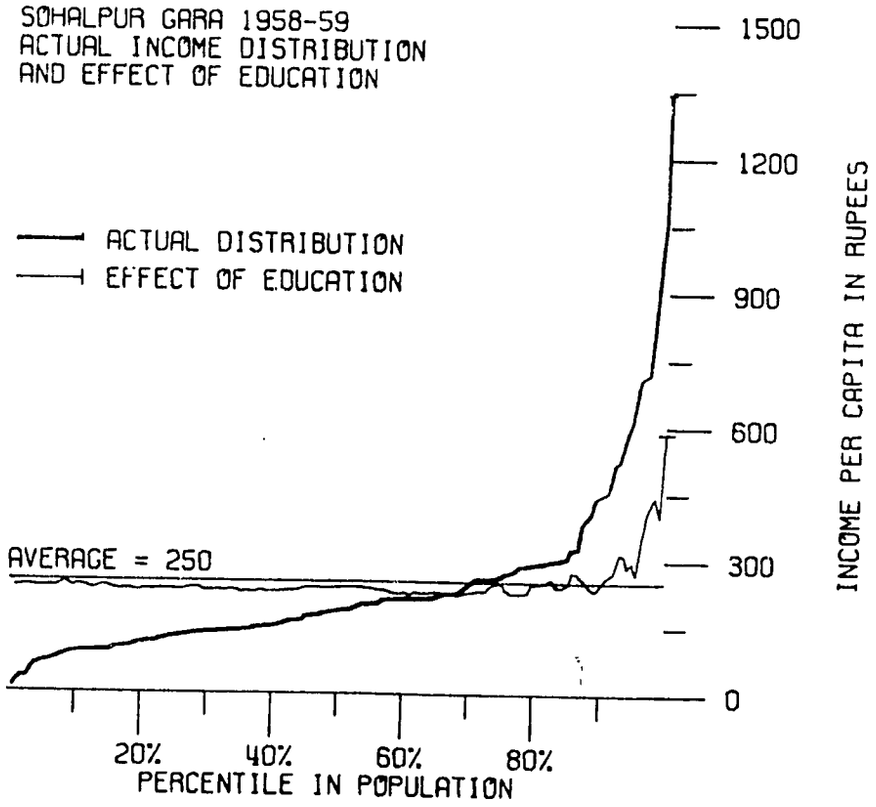


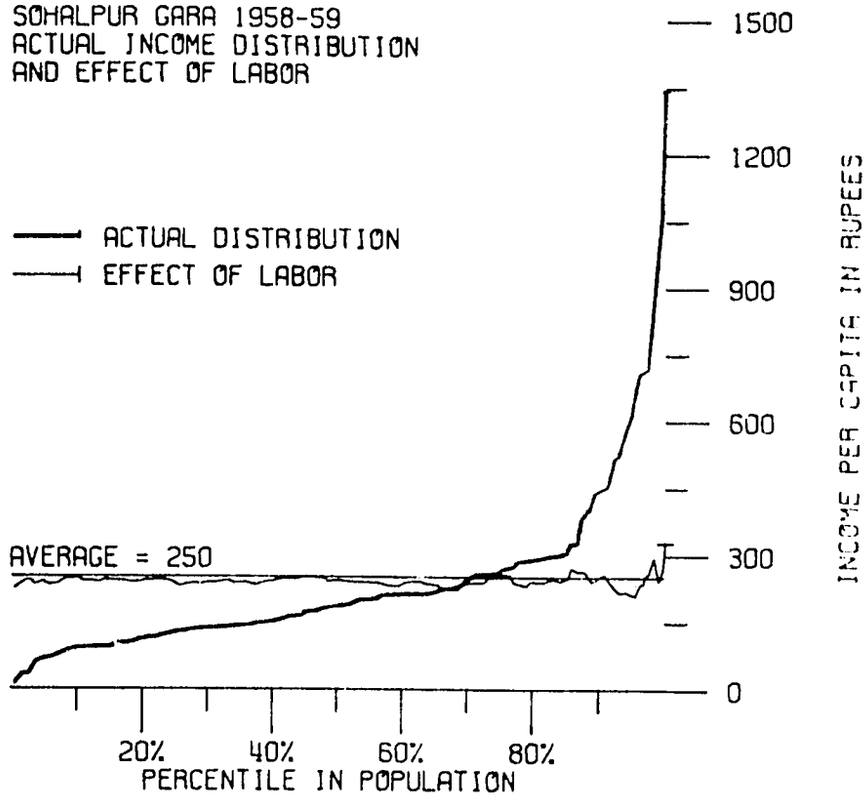
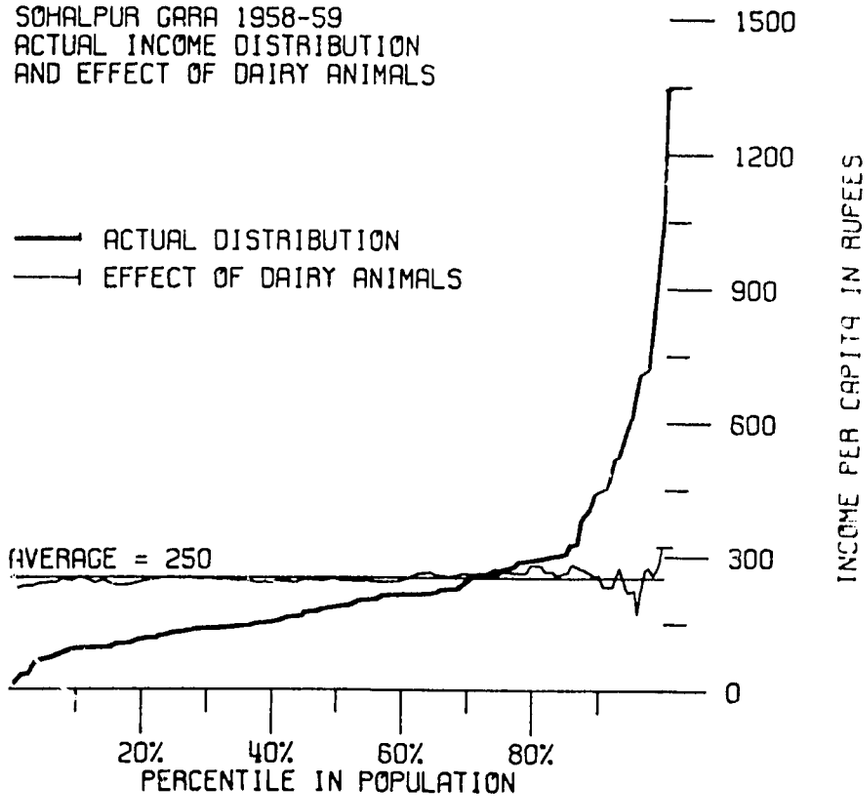


Graph Set 4

(continued)





Graph Set 4

(continued)