Institute of Economic Development and Research

SCHOOL OF ECONOMICS
University of the Philippines System

Discussion Paper No. 74-11 August, 1974

INCOME DISTRIBUTION IN THE PHILIPPINES: THE EMPLOYED AND THE SELF-EMPLOYED

by

José Encarnación, Jr.

Abstract

Using cross-section data from the 1968 National Demographic Survey, this paper gives quantitative estimates of the relative contributions of various factors (education, occupation, etc.) to income inequality among heads of families, measuring inequality by the variance of income logarithms. Particular attention is given to the characteristics of the self-employed (more than 40% of the total) as distinguished from the employed. Not unexpectedly, mean income is lower but variance significantly larger for the self-employed. For both groups, education is the single most important variable explaining income variation, and excepting college graduates, the self-employed earn less at every education level.

For the aggregate, other factors following education in order of importance are: occupation, geographical region of residence, sex, sector of employment, amount of working time, age and class of worker (whether employed or self-employed). Considering that little can be done over the medium-term in regard to most of these factors, the suggestion made is that it might be useful to effect institutional changes that combine self-employed workers into more efficient groupings in order (among other things) to reduce income inequality.

Income Distribution in the Philippines: The Employed and the Self-employed[1/]

by J. Encarnación

1. Introduction

Among the characteristics that distinguish LDCs from the more developed economies would seem to be the larger proportion of the self-employed in the labor force. One conjectures that incomes of the self-employed are less than those of the employed, as the self-employed are likely to be working with less capital per worker at least at the lower educational levels. One then expects that most of the self-employed would prefer to be in the "employed" category but for the lack of employment opportunities.

This paper reports the results of an empirical inquiry into the determinants of income among the employed and the self-employed, using cross-section results from the 1968 National Demographic Survey. We consider only heads of families with positive incomes, and we calculate the quantitative effects of the more obvious factors (education, age, occupation, sector of employment, etc.) in the determination of income. Measuring income inequality by the variance of income logarithms, we thus obtain the relative contributions of these factors to income inequality. Most of the results are not unexpected: e.g., education turns out to be the single most important variable in explaining income variation, and female workers earn less than their male counterparts.

2. Theoretical considerations

A person's income from work depends on his productivity, and this depends partly on his skill level (proxied by years of schooling), experience (age) and occupation. The sector of employment--whether agriculture, manufacturing, or other--also matters as different sectors have different capital-labor ratios. Of course, the amount of time spent at work is also a determinant.

In addition, one might expect the geographical region of residence to help explain a person's income, for different regions vary as to land-man ratios as well as proximity to distribution centers and cost of living. Location of residence--whether urban or rural--would be a similar variable. To take account of income from property, we use home-ownership as a proxy variable. We expect urban home-owners to have higher incomes than rural home-owners *ceteris paribus*. Sex differences (and discrimination) need also to be considered--we would expect their effect on income to be stronger among the employed than among the self-employed. Finally, for obvious reasons migrants might be expected to have higher incomes than non-migrants.

As is well known, the lognormal curve is a fairly good approximation to the size distribution of income particularly at the lower and middle ranges; the lognormal is also the most convenient to work with statistically among all skew, unimodal distributions (Aitchison and Brown 1957; Cramer 1971). If the income distribution is at all approximately lognormal, therefore, the simplest assumption to make--which we will follow--is that the effects of the variables cited above are additive on (the logarithm of) income.

3. The statistical model

Consider the explanation of a variable y (the logarithm of income in the present instance) in terms of two classificatory variables A and B. A could be, say an industrial sector classification with different categories such as manufacturing, agriculture, etc.. Each classification consists of categories which are exhaustive and mutually exclusive, and we wish to explain an individual's y as a result of his belonging to some category of A and some category of B. We also want to estimate the relative contributions of those variables to the explanation of y variation. It will be apparent how the discussion would proceed if there were three or more classificatory variables.

Let the categories of A be indexed by k $(k = 0, 1, \ldots, K)$, those of B by j $(j = 0, 1, \ldots, J)$; and let y_{kji} be the ith observation in the (k, j)-cell, $i = 1, \ldots, n_{kj}$. There are n observations, so

$$n = \sum_{k=0}^{K} \sum_{j=0}^{J} n_{kj} .$$

Write

$$n_{k.} = \sum_{j=0}^{J} n_{kj}$$

for the number of observations in the kth category of A; similarly, we write $n_{.j}$ for the number in the jth category of B. The mean y in category k of A is given by

$$\bar{y}_{k.} = \sum_{j=0}^{J} \sum_{i=1}^{n_{kj}} y_{kji}/n_{k.}$$

and for the mean y we write $\bar{y}$.

Let

$$x_k = \begin{cases} 1 \text{ if an observation belongs to } k \text{ of } A \\ 0 \text{ otherwise} \end{cases}$$

and define z_j similarly for category j of B. With these dummy variables we can calculate a regression equation

$$(1) \quad y' = c + a_1'x_1 + \ldots + a_K'x_K + b_1'z_1 + \ldots + b_J'z_J$$

Note that the variables x_0 and z_0 are omitted from the regression in order to get determinate estimates of the coefficients; see (Suits 1957). (Cf. the fact that a sex classification has two categories--male and female--but only one dummy variable would be included in a regression equation, i.e. one dummy variable is omitted.) For a given individual, at most one of the x_k, $k = 1, \ldots, K$, and at most one of the z_j, $j = 1, \ldots, J$, can be nonzero. Eq. (1) thus shows y', the predicted y, to depend on an individual's membership in the different categories of A and of B.

We wish to write (1) in the form

$$(2) \quad y' = \bar{y} + a_k + b_j$$

where $k = 0, 1, \ldots, K$ and $j = 0, 1, \ldots, J$; see e.g. (Anderson and Bancroft 1952) or (Hogg and Craig 1965) for general background. Then a_k would measure the effect on y resulting from belonging to category k of A, which effect is measured from $\bar{y}$. Similarly for b_j in regard to j of B. (We will later call the a_k and the b_j category effects.)

In order to get the a_k, from least squares regression properties we know that in (1),

$$c = \bar{y} - \sum_1^K a'_k \bar{x}_k - \sum_1^J b'_j \bar{z}_j$$

$$= \bar{y} - \sum_1^K a'_k n_{k.}/n - \sum_1^J b'_j n_{.j}/n$$

But c is the predicted y when one belongs to category 0 of A (in which case $x_1 = \ldots = x_K = 0$) and 0 of B. From this it follows that

$$(3) \qquad a_0 = -\sum_1^K a'_k n_{k.}/n$$

$$b_0 = -\sum_1^J b'_j n_{.j}/n$$

On the other hand, if one belongs to category k $(k = 1, \ldots, K)$ of A and 0 of B, the predicted y is $c + a'_k$. Accordingly we have

$$(4) \qquad a_k = a_0 + a'_k \qquad (k = 1, \ldots, K)$$

Eqs. (3) and (4) thus determine the a_k, and a similar procedure gives the b_j.

Multiplying (4) by $n_{k.}$, summing both sides and then adding $n_{0.}a_0$ to the results, one gets

$$\sum_0^K n_{k.}a_k = na_0 + \sum_1^K n_{k.}a'_k$$

which, in view of (3), implies that the weighted sum of the a_k is zero:

(5) $$\sum_{0}^{K} a_k n_{k.}/n = 0$$

In order now to get (2) from (1), write

$$\begin{aligned} y' &= \bar{y} + a_0 + b_0 + \sum_{1}^{K} a'_k x_k + \sum_{1}^{J} b'_j z_j \\ &= \bar{y} + a_0 + b_0 + \sum_{1}^{K} (a_k - a_0)x_k + \sum_{1}^{J} (b_j - b_0)z_j \\ &= \bar{y} + a_0(1 - \sum_{1}^{K} x_k) + \sum_{1}^{K} a_k x_k + b_0(1 - \sum_{1}^{J} z_j) + \sum_{1}^{J} b_j z_j \end{aligned}$$

But $1 - \sum_{1}^{K} x_k = x_0$ and similarly for the other set of dummy variables, and we obtain (2). Hence (1) and (2) yield identical predicted values (and therefore identical error terms).

The beta coefficients discussed by Morgan et al. (1962) serve to measure the relative contributions of the classificatory variables to the explanation of y variation. In the case of A,

(6) $$\beta_A^2 = \frac{\sum_{k=0}^{K} n_{k.} a_k^2/(n - 1)}{s_y^2}$$

where s_y is the standard deviation of y. A similar definition applies to B with $\sum n_{.j} b_j^2$ in the formula instead. If, say, $\beta_A > \beta_B$, that would indicate that more of the variation in y is due to the classificatory variable A. As Morgan et al. point out, these coefficients are analogous to the partial beta coefficients of standard multiple regression.[2/]

Following Meesook (1972), an approximate F-test can be used to test the significance of a classificatory variable in explaining y. For from (2),

$$e_{kji} = y_{kji} - \bar{y} - a_k - b_j \tag{7}$$

where e is the error term, and under certain assumptions, $(n - 1)s_e^2/\sigma_y^2$ would be a chi-square variable with $n - K - J - 1$ degrees of freedom. Accordingly in the case of A,

$$F_A = \frac{\sum_{k=0}^{K} n_{k.} a_k^2 / K}{\sum_{k=0}^{K} \sum_{j=0}^{J} \sum_{i=1}^{n_{kj}} e_{kji}^2 / (n - K - J - 1)} \tag{8}$$

If the sum of squares in the numerator is small relative to that of the denominator, that would indicate that the use of A does not add much to the explanation of y.

4. Data base and notation

The 1968 National Demographic Survey (NDS), a joint undertaking of the University of the Philippines Population Institute and the then Bureau of the Census and Statistics (now the National Census and Statistics Office), is a nation-wide stratified random sample of over 7,000 households. It contains a valuable body of data which, though not flawless, gives both economic and demographic information for individuals and households. A post-enumeration survey indicates that some completed questionnaires were apparently merely filled in by the interviewers, so it is difficult to check precisely the general accuracy of the NDS data (Raymundo 1974).

In regard to the income data in particular, it appears from a comparison with the 1957, 1961 and 1965 Family Incomes and Expenditures Surveys of the Census Bureau and the GNP estimates for the corresponding years, that family income data of the NDS could have been underreported by perhaps 12 percent on the average (Encarnación et al. 1974). Undercoverage of income in kind appears to be the major cause, so that income data from rural areas is probably more seriously understated than those for urban areas. Despite the data defects of the NDS, it is the only one of its kind yet available in the Philippines, and we have no other recourse. Since our chief interest is the explanation of income variation rather than absolute income levels, and since we can use a dummy variable for urban-rural differences, hopefully the NDS defects are less serious for present purposes than they would be otherwise.

For estimation purposes we use a subsample of heads of families from the NDS by checking for non-responses to certain information (see Appendix B). The resulting data file, which we will call the IDS (for income distribution), consists of 2092 urban and 1588 rural observations. Since the NDS rural sampling fraction is 1/3 that of the urban, we have in effect an IDS sample of 2092 + 3 x 1588 = 6856 "observations."

We use the following notation (all pertaining to heads of families):

Y annual income, in thousand pesos

LY natural logarithm of Y

EH education level:

0 = no schooling

1 = grades 1 to 4

2 = grades 5 to 7

3 = one to three years of high school

4 = high school graduate

5 = one to three years of college

6 = college graduate

AG age:

0 = age 15 to 24

1 = 25-34

2 = 35-44

3 = 45-54

4 = 55-64

5 = 65 and over

OC occupational code:

0 = farm tenants and owner-tenants, farm laborers, fishermen and loggers

1 = farmers-owners

2 = service and unskilled (non-farm)

3 = skilled workers and transportation and communications workers

4 = clerical and sales

5 = professional, administrator, management

SE sector of employment:

0 = agriculture, forestry and fishing

1 = manufacturing and mining

2 = construction

3 = transport, communications and utilities

4 = commerce

5 = services

HH hours worked during survey week:

0 = 1 to 19 hours

1 = 20-39

2 = 40-49

3 = 50 and over

GR geographical region:

0 = greater Manila

1 = rest of Luzon

2 = Visayas

3 = Mindanao

HO home ownership:

0 = no

1 = yes, rural

2 = yes, urban

LR location of residence = 0 if urban, 1 if rural

SX sex = 0 if male, 1 if female

MS migrant status = 0 if non-migrant, 1 if migrant

CW class of worker = 0 if self-employed, 1 if employed.

Some self-employed persons (less than 0.5 per cent of the total NDS) have employees; the employed are employees in the government as well as the private sector.

5. The results

Logarithmic probability graphs of the distribution of incomes in our IDS, employed and self-employed sub-samples show roughly straight lines, even though (as might be expected) the self-employed group indicates a slight curvature at the higher income ranges. Computations give the following figures:

	IDS	Employed	Self-employed
Mean LY[3/]	-0.02810	0.13710	-0.24184
Standard deviation	1.00313	0.94754	1.03167
Sample size	6856	3868	2988

The mean LY of the employed is significantly higher than that for the self-employed, and an F-test rejects the hypothesis that the variances are equal (the calculated F = 1.185 is significant for the relevant d.f. at the .01 level).

5.1 Sample proportions

Tables 1 through 11 show the relative distributions of the categories of the various classificatory variables for the IDS, the employed and the self-employed. Corresponding figures for the NDS where available are also shown as a check on the IDS, though in some tables (as noted) the figures for the NDS pertain to all workers instead of family heads only.

Table 1 shows that relatively more of the self-employed are at the lower education levels compared to the employed, and relatively more of the employed have more education. At EH2--corresponding to grades 5 to 7--we have the same proportion (28 percent) in both groups.

Table 2 gives the age distributions in the two groups, and it is interesting to note that relatively more of the self-employed belong to older age cohorts. This might be due to a shift towards self-employment as one gets older after gaining experience as an employee. But it seems more likely, and quantitatively more important, that the age distributions in the two groups relate to the distribution by educational level, as relatively more of the younger age cohorts would have had access to formal schooling considering the time pattern of the growth of the educational system.

As expected, Table 3 shows more than half (54 percent) of the self-employed as "farmers-owners" (OC1), and also more than half (51 percent) of the employed as "farm tenants and owner-tenants, farm laborers, fishermen and loggers" (OC0).[4/] Among the employed, 0.01 percent are reported as farmers-owners; presumably these persons have part-time employment off the farm, although it is also possible that these are cases of misreporting. Excepting clerical and sales workers (OC4) and farmers-owners already mentioned, there are more of the employed than self-employed in the other occupations. Sales workers, it might be noted, include market vendors, street and sidewalk vendors, peddlers, and proprietors of small neighborhood stores.

In Table 4, we see relatively more of the self-employed in the agricultural (SE0)[5/] and commerce (SE4) sectors. In manufacturing and mining (SE1), construction (SE2), transport, communications and utilities (SE3) and services (SE5), there are more of the employed.

It is apparent from Table 5 that relatively more of the self-employed are underemployed, and one also infers that more under-employment is to be found among non-family heads than among heads of families.

Table 6 gives a geographical breakdown. We note that relatively more of the self-employed are in Mindanao (GR3), a region with a high land-man ratio and a focus of migration from the Visayas (GR2) and parts of Luzon (GR1). Greater Manila (GR0) has more employed than self-employed, as was to be expected.

Table 7 gives a breakdown by home-ownership. Almost two-thirds of the employed do not own their homes, while one half of the self-employed (mostly farmers-owners obviously) own rural homes.

Tables 8 through 11 give the proportions by location of residence (urban or rural), sex, migrant status[6/] and class of work (employed or self-employed).

5.2 EH, AG, OC and SE category effects

Beginning with Table 12 we report the results of the procedures worked out in Section 3. Table 12 gives the results using the specification LY: (EH), meaning LY regressed on the variables following the colon, (EH) being short for EH1, EH2, ..., EH6.[7/] Six columns of figures are given in the table, the first three of which are the regression coefficients obtained (with t-values underneath) using the IDS, the employed and the self-employed subsamples respectively. R^2 and the standard error of estimate (s.e.) are also given for each equation. The last three columns give the effects of belonging to the various categories of EH--the a_k of Section 3--which we may call category effects, the first row of which corresponds to EH0. For example, in the self-employed group, the effect (relative to the mean) of belonging to category EH0 is -0.4718. The relative magnitudes of the category effects conform to expectation, with effects greater for higher education levels.

It should be noted that the category effects calculated for the three groups (IDS, employed and self-employed) are not directly comparable since their means are different. In order to have comparable figures, one has to add category effects to the appropriate means and then get the antilogarithms of the results, which are then in money terms. This is done in Table 12A where we find that the self-employed earn less than the employed at every education level except the last (college graduates). Finally, in Table 12, we also give the F-values discussed in Section 3. (With only one classificatory variable, $\beta^2 = R^2$.) While less of the variation is explained in the self-employed group than in the employed--this is so in all that follows--it is apparent from the F-values that EH is a significant variable.

Table 13, using LY: (AG), gives income distributions by age-cohorts. We expect here a unimodal distribution with peak income at AG4 (age 55-64), which is the case with the self-employed. The employed group shows the curious result, however, of the peak occurring at AG2 (age 35-44). This fact might be due to a shift of higher-earning individuals towards self-employment after reaching AG2--a possibility mentioned earlier--but it may simply be that an age classification alone is misleading. Indeed, it will be seen that in the specification LY: (EH), (AG), (OC), (SE) of Table 16, the category effects of AG exhibit the usual pattern. AG explains very little of income variation, but nonetheless it is significant (at the .01 level) as a classificatory variable. Table 13A, in money terms, shows the self-employed earning less in every age group except the last.

There are no surprises in Table 14 based on an occupational classification, except possibly the relative magnitudes of the OC0 and OC1 effects for the self-employed. One would have expected OC1 to have a higher category effect than OC0, but the straightforward explanation here would be that the large mass of small subsistence farmers-owners has simply pulled down the figure for OC1.[8/] In any event, Table 14A shows the self-employed earning less in every occupational group.

In Tables 15 and 15A, using a sectoral classification, the possibly surprising item is the higher earnings of the self-employed in SE3 (transport, communications and utilities). Outside of sampling variations, the explanation here would be the presence of bus operators among the self-employed in this sector.

Table 16 uses EH, AG, OC and SE to explain LY. β^2s and F-values are given at the bottom of the table, and also $\sum\beta^2$ on the last line. Some regression coefficients are not significant but the important point of interest is that all the classificatory variables considered are significant at the .01 level.

Recall that β^2 is the ratio of the variance associated with a classificatory variable to the total variance. While R^2 is of course higher with the specification using all four classificatory variables than with single classificatory variables, it is not necessarily the case that $\sum\beta$ is higher--it could be lower--if more classificatory variables are used. Table 16A, reproduced from Tables 12-16 for easier reference, shows this for IDS and the employed. Excepting AG, $\sum\beta^2$ there is lower than β^2 from the other variables used singly.

Numerically, the reduction in the magnitudes of the β^2s is due to the decrease in the absolute values of the a_k--the category effects--when more variables are included in the specification. This happens when variables correlated with the old variables are added to the specification, because then part of the effects associated with the old variables get appropriately assigned to the new ones. It is also possible, under the opposite conditions, for the β^2 of a variable to increase with the addition of more variables (see that of AG for the self-employed, which increases from .011 to .013).[9/]

The numerical results show a sharper reduction in the values of the β^2s for the employed than for the self-employed, and the apparent reason is that there is stronger correlation among the variables in the case of the employed. One rather expects this intuitively, since the self-employed are a more heterogeneous group whose incomes are not dependent on employers' decisions. Thus, even though $\sum\beta^2$ is about the same for both groups, R^2 is substantially lower for the self-employed.[10/]

For both groups, education has the largest effect on income, followed by occupation, sector of employment, and then age. Though not unusual with cross-section data on individual observations, the fraction of the variance explained is rather small in both cases, though less so for the employed as might be expected.

What appears surprising about Table 16 is the apparently substantial change in the ordering of the OC categories by size of effect, as compared with the results in Table 14. We postpone discussion of this to Section 5.4.

5.3 Results with additional variables

In Table 17, using the specification LY: (EH), (AG), (OC), (SE), (HH), we find regression coefficients for (HH) whose relative magnitudes conform to expectation. Those for the self-employed group are not statistically significant however--this is so in all the regressions tried--suggesting much variability in time spent on work from week to week among the self-employed. Still, the classificatory variable HH itself is significant (at the 0.1 level) as shown by its F-value.

If instead of HH we were to add GR, HO, LR, SX, MS and CW separately to our four basic classificatory variables (EH, AG, OC and SE) for explaining LY variation, we would get tables of results like those of Table 17. For reasons of space those results are not presented here, but we briefly summarize the main points.

With GR added, it is interesting to observe that GR3 (Mindanao) has a larger effect than GR1 (Luzon) for the self-employed while the opposite holds for the employed.[11/] This fact seems related to the higher land-man ratio and the higher proportion of (enterprising) migrants (33.7 percent) in Mindanao, which has a large proportion of self-employed (66.5 percent).

The results with HO show urban home-ownership to have a positive effect but rural home-ownership a negative one (relative to the mean) for both the employed and the self-employed, while that of non-ownership is positive even if small. These relationships are probably to be expected considering that the great mass of rural homes are small farm-huts.

Other results using urban-rural residence, sex, migrant status and class of worker (for IDS only) conform to expectations. For example, being female has a negative effect in both groups, but the effect is less for the self-employed. However, because the employed earn more, ceteris paribus the self-employed female still earns less than the employed female.

5.4 All variables

The last set of results is Table 18 where all variables that we have considered are included in the specification. Of course, the employed and self-employed regressions exclude CW.

Looking at the education variable, the IDS shows progressively higher effects with more education. Separately, however, the results for the employed and the self-employed show contrary relative magnitudes between EH2 and EH3 for the employed and between EH1 and EH2 for the self-employed.[12/] Instead of adding to income, some years of high school apparently detract from the employed and some years of intermediate schooling from the self-employed. However, looking at the regression coefficients, the differences involved are quite small and well within standard errors, and we conclude that at such low education levels, a few more years of schooling do not make any real difference.

The age variable behaves as expected, declining in effect at retirement level for the employed but still increasing for the self-employed.

The occupation variable might deserve a closer look. From Table 14 (where only this variable is used) and Table 18, we have

Table 18A which gives the ordering of the occupations by size of effect. The employed OC1 group, we recall, is a very small fraction of the total (.01 percent) and apparently involves some misreporting. In any event, the standard error of its regression coefficient is such that it could well fall below OC2 in size of effect. As for the self-employed, the ordering of occupations implied by Table 18 is rather unreliable because of the very low t-values of the regression coefficients.

A similar table in regard to sector of employment is given in Table 18B. The ordering for the employed appears fairly coherent, the only noteworthy aspect of it being that manufacturing (SE1) now tops the list, which seems intuitively to be expected. In case of the self-employed, three sectors (SE1, SE4 and SE5) merely reverse their order. Low t-values of the regression coefficients should be noted, however. Most of the results pertaining to the other variables contain no surprises, all having expected relative magnitudes except in the case of HH for the self-employed. Low t-values make the results here unreliable.

R^2 for the employed equation is .423, that for the self-employed is .233. It is interesting that $\sum\beta^2$ is about the same for both groups (.162 and .169).[13/] The classificatory variables as a whole, therefore, explain about the same amount of income variability in the two groups, but the self-employed simply have more income variability.

Comparing the β^2s of each variable for the employed and the self-employed, we find those for EH, GR and SX about the same for both groups. SE and AG have relatively larger β^2s for the self-employed; OC is an opposite case. Ordering the variables by size of β^2 gives Table 18C. By the F-test, all variables are significant at the .01 level except where noted. Home-ownership, migrant status and location of residence are relatively weak factors in explaining income differences. It is interesting that age ranks high for the self-employed but quite low for the employed.

By far the most important variable is education. Sector of employment is high on the list for the self-employed because of the agricultural sector. Occupation is high on the employed because different occupations have fairly standard wage rates so far as employers are concerned; among the self-employed, this consideration does not apply.

For both groups, geographical region explains more of income differences than does sex-classification, and for the employed, sex-classification explains even more than age-classification.

Finally, it is of some interest to compare the R^2s and $\sum\beta^2$s obtained from the different specifications that we have considered. Recall that Table 16 uses our four basic variables: education, age, occupation and sector of employment. Table 17 adds HH to the basic four, while Table 18 includes all variables. These statistics are summarized in Table 18D where the line-entries come from different tables (some not reported here) seriatim. The third line, for example, derives from using GR in addition to the four basic variables. Comparing the rankings of the non-basic variables according to $\sum\beta^2$ in Table 18D with the rankings in Table 18C, it is apparent that pair-wise comparisons can be quite misleading as to the relative sizes of the

β^2s in the (correct) simultaneous determination context of Table 18C. For instance, one might infer (wrongly) from Table 18D that GR for the employed is among the least important of the non-basic variables. Table 18C shows, however, that GR stands fairly high on the list, exceeded only by HH among the non-basic variables. If one is to use Table 18D at all as a guide to the relative magnitudes of the β^2s, it is R^2 that appears to be the better (though clearly not infallible) index.

[illegible]. Concluding remarks

While there are data defects in our sample and there is the possibility that the self-employed may understate their incomes more so than would the employed--and we have not considered interaction effects at all--the following broad conclusions appear warranted.

In our IDS sample of family heads with employment, over 40 per cent are self-employed. On the average they have less education, are older and earn less than the employed. They are relatively more numerous in the agricultural and commerce sectors, and there is more underemployment and greater income inequality among them.[14/]

Different factors vary in their relative contributions to the explanation of income inequality in the two groups, but education is most important for both. Considering the IDS sample as a whole, in order of importance these factors are: education, occupation, geographical region, sex, sector of employment, amount of working time, age, class of worker, location of residence and migrant status. The home-ownership variable is not significant.

Only in the case of college graduates do we find the self-employed earning more. At all lower education levels they earn less than their employed counterparts, and we expect that they would prefer employment if this were available. Not only would employment income be higher, it would also be more regular because short-run fluctuations are usually absorbed by the firm. As development proceeds, we expect the self-employed to decrease in proportion to the total--a factor additional to those discussed by Kuznets (1963) that make for narrowing income inequality over time. Meanwhile the question for the immediate term is how might incomes be raised and inequalities reduced.

We conjecture that (physical) capital per worker is lower among the self-employed compared to the employed. In addition, however, lower productivity among the self-employed may be due in part to less discipline on the part of the worker when there is no employer-relationship involved, and surely also to the absence of the division-of-labor economies inherent in the firm. This suggests that an institutional change that collects self-employed workers into more efficient groupings should help the situation. Other social scientists in addition to economists need to work on this problem.

Table 1. Sample Proportions in Categories of EH

	NDS	IDS	Employed	Self-employed
EH0	.195	.160	.131	.196
EH1	.323	.324	.293	.364
EH2	.256	.282	.282	.282
EH3	.078	.080	.085	.074
EH4	.071	.078	.103	.046
EH5	.026	.027	.036	.015
EH6	.051	.049	.070	.023
	1.000	1.000	1.000	1.000

Table 2. Sample Proportions in Categories of AG

	NDS	IDS	Employed	Self-employed
AG0	.030	.029	.033	.025
AG1	.232	.258	.295	.211
AG2	.281	.316	.319	.312
AG3	.215	.223	.217	.231
AG4	.155	.131	.105	.163
AG5	.087	.043	.031	.058
	1.000	1.000	1.000	1.000

Table 3. Sample Proportions in Categories of OC

	NDS	IDS	Employed	Self-employed
OC0	.365	.383	.514	.214
OC1	.269	.236	.001	.540
OC2	.076	.081	.127	.020
OC3	.156	.172	.229	.098
OC4	.085	.084	.070	.103
OC5	.049	.044	.059	.025
	1.000	1.000	1.000	1.000

Table 4. Sample Proportions in Categories of SE

	NDS*	IDS	Employed	Self-employed
SE0	.576	.622	.525	.749
SE1	.116	.103	.123	.078
SE2	.031	.048	.073	.016
SE3	.030	.053	.080	.018
SE4	.113	.070	.044	.102
SE5	.134	.104	.155	.037
	1.000	1.000	1.000	1.000

*All workers

Table 5. Sample Proportions in Categories of HH

	NDS*	IDS	Employed	Self-employed
HH0	.096	.010	.011	.009
HH1	.239	.119	.097	.149
HH2	.340	.391	.426	.345
HH3	.325	.480	.466	.497
	1.000	1.000	1.000	1.000

*All workers

Table 6. Sample Proportions in Categories of GR

	NDS	IDS	Employed	Self-employed
GR0	.096	.111	.168	.037
GR1	.436	.461	.476	.441
GR2	.279	.259	.255	.264
GR3	.189	.169	.101	.258
	1.000	1.000	1.000	1.000

Table 7. Sample Proportions in Categories of HO

	NDS	IDS	Employed	Self-employed
HO0		.508	.645	.329
HO1		.378	.241	.556
HO2		.114	.114	.115
		1.000	1.000	1.000

Table 8. Sample Proportions in Categories of LR

	NDS	IDS	Employed	Self-employed
LR0	.302	.305	.369	.223
LR1	.698	.695	.631	.777
	1.000	1.000	1.000	1.000

Table 9. Sample Proportions in Categories of SX

	NDS	IDS	Employed	Self-employed
SX0	.901	.948	.973	.916
SX1	.099	.052	.027	.084
	1.000	1.000	1.000	1.000

Table 10. Sample Proportions in Categories of MS

	NDS	IDS	Employed	Self-employed
MS0	.758	.763	.721	.818
MS1	.242	.237	.279	.182
	1.000	1.000	1.000	1.000

Table 11. Sample Proportions in Categories of CW

	NDS	IDS	Employed	Self-employed
CW0	.464	.436	.000	1.000
CW1	.536	.564	1.000	.000
	1.000	1.000	1.000	1.000

Table 12. LY: (EH)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-0.6166	-0.5011	-0.7136	-0.5885	-0.6382	-0.4718
EH1	0.3880 (11.9)	0.3039 (7.09)	0.4517 (9.14)	-0.2005	-0.3343	-0.0201
EH2	0.5798 (17.3)	0.6214 (14.4)	0.4742 (9.14)	-0.0087	-0.0168	0.0024
EH3	0.8290 (17.9)	0.8458 (14.9)	0.7267 (9.54)	0.2406	0.2076	0.2549
EH4	1.1613 (24.9)	1.1078 (20.6)	1.0788 (11.9)	0.5728	0.4697	0.6071
EH5	1.5180 (21.6)	1.5014 (19.6)	1.3028 (8.75)	0.9295	0.8632	0.8310
EH6	2.0386 (37.1)	1.9204 (31.7)	2.1374 (17.6)	1.4501	1.2822	1.6657
R^2	.224	.283	.132			
s.e.	0.884	0.803	0.962			
F				330.4	254.1	75.7

Table 13. LY: (AG)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-0.3743	-0.1587	-0.7371	-0.3462	-0.2958	-0.4953
AG1	0.3136 (4.23)	0.2408 (2.74)	0.4173 (3.31)	-0.0326	-0.0550	-0.0779
AG2	0.3964 (5.39)	0.4042 (4.61)	0.4636 (3.74)	0.0502	0.1084	-0.0317
AG3	0.3797 (5.08)	0.3334 (3.72)	0.5363 (4.27)	0.0335	0.0376	0.0410
AG4	0.3823 (4.91)	0.2954 (3.09)	0.6377 (4.98)	0.0361	-0.0004	0.1424
AG5	0.1263 (1.38)	-0.2527 (-2.10)	0.5984 (4.21)	-0.2199	-0.5485	0.1031
R^2	.007	.019	.011			
s.e.	1.000	0.939	1.027			
F				9.7	14.9	6.8

Table 12A. Calculated Incomes Based on EH

	IDS	Employed	Self-employed
EH0	0.5460	0.6059	0.4899
EH1	0.7956	0.8210	0.7696
EH2	0.9639	1.1278	0.7871
EH3	1.2367	1.4116	1.0132
EH4	1.7241	1.8345	1.4410
EH5	2.4631	2.7190	1.8026
EH6	4.1455	4.1342	4.1533

Table 13A. Calculated Incomes Based on AG

	IDS	Employed	Self-employed
AG0	0.6878	0.8533	0.4785
AG1	0.9412	1.0856	0.7264
AG2	1.0223	1.2783	0.7607
AG3	1.0054	1.1909	0.8181
AG4	1.0080	1.1466	0.9054
AG5	0.7803	0.6627	0.8705

Table 14. LY: (OC)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-0.3377	-0.3280	-0.3710	-0.3096	-0.4651	-0.1291
OC1	-0.0549 (-1.95)	0.5455 (1.20)	-0.0225 (-0.49)	-0.3645	0.0804	-0.1516
OC2	0.5942 (14.3)	0.6288 (15.9)	0.2839 (2.15)	0.2846	0.1637	0.1548
OC3	0.7240 (23.2)	0.8673 (27.3)	0.2936 (4.23)	0.4144	0.4022	0.1645
OC4	0.9014 (22.1)	1.2060 (23.7)	0.6603 (9.67)	0.5918	0.7409	0.5312
OC5	1.6677 (31.0)	1.7059 (31.2)	1.5511 (12.9)	1.3581	1.2408	1.4220
R^2	.214	.314	.093			
s.e.	0.889	0.786	0.984			
F				374.2	352.8	61.0

Table 15. LY: (SE)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-0.3576	-0.3219	-0.3902	-0.3295	-0.4590	-0.1484
SE1	0.7852 (21.3)	0.9770 (23.7)	0.3528 (5.16)	0.4557	0.5180	0.2044
SE2	0.6823 (13.2)	0.7392 (14.4)	0.1568 (1.07)	0.3528	0.2802	0.0085
SE3	0.8490 (17.2)	0.7729 (15.7)	1.1083 (8.18)	0.5196	0.3139	0.9599
SE4	0.8885 (19.2)	1.1838 (18.4)	0.6613 (10.9)	0.5090	0.7249	0.5129
SE5	1.0748 (29.3)	1.0956 (29.1)	0.8073 (8.39)	0.7453	0.6367	0.6589
R^2	.184	.271	.075			
s.e.	0.906	0.810	0.993			
F				309.3	286.9	48.4

Table 14A. Calculated Incomes Based on OC

	IDS	Employed	Self-employed
OC0	0.7134	0.7204	0.6901
OC1	0.6753	1.2430	0.6748
OC2	1.2924	1.3510	0.9167
OC3	1.4716	1.7148	0.9256
OC4	1.7572	2.4060	1.3356
OC5	3.7821	3.9665	3.2550

Table 15A. Calculated Incomes Based on SE

	IDS	Employed	Self-employed
SE0	0.6994	0.7248	0.6769
SE1	1.5336	1.9253	0.9633
SE2	1.3836	1.5179	0.7919
SE3	1.6347	1.5699	2.0505
SE4	1.6176	2.3679	1.3114
SE5	2.0486	2.1680	1.5176

Table 16. LY: (EH), (AG), (OC), (SE)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-1.0002	-0.8347	-1.2028	-0.4735	-0.4126	-0.4412
EH1	0.3692 (11.6)	0.2460 (5.94)	0.4452 (9.18)	-0.1043	-0.1666	0.0039
EH2	0.4913 (14.7)	0.4498 (10.4)	0.4514 (8.78)	0.0178	0.0371	0.0102
EH3	0.6224 (13.2)	0.4441 (7.55)	0.7035 (9.28)	0.1489	0.0314	0.2622
EH4	0.8105 (16.6)	0.6028 (10.4)	0.9146 (10.1)	0.3370	0.1902	0.4734
EH5	1.1388 (15.9)	0.9091 (11.2)	1.1762 (8.06)	0.6653	0.4964	0.7350
EH6	1.4892 (21.4)	1.1647 (15.0)	1.6892 (12.1)	0.9656	0.7521	1.2480
AG0				-0.3065	-0.2313	-0.4367
AG1	0.2490 (3.95)	0.1885 (2.67)	0.3302 (2.86)	-0.0575	-0.0429	-0.1065
AG2	0.2839 (4.53)	0.2309 (3.27)	0.3867 (3.41)	-0.0226	-0.0004	-0.0500
AG3	0.3727 (5.82)	0.3009 (4.15)	0.4979 (4.31)	0.0662	0.0696	0.0612
AG4	0.4217 (6.31)	0.3192 (4.10)	0.5940 (5.05)	0.1152	0.0879	0.1573
AG5	0.3350 (4.26)	0.1011 (1.03)	0.5942 (4.53)	0.0286	-0.1302	0.1574
OC0				-0.0955	-0.1754	-0.0252
OC1	-0.0685 (-2.54)	0.6177 (1.42)	-0.0183 (-0.42)	-0.1640	0.4423	-0.0435

Table 16. (continued)

	IDS	Employed	Self-employed	IDS	Employed	Self-employed
OC2	0.1884 (2.66)	0.1734 (2.49)	0.1121 (0.50)	0.0929	-0.0021	0.0869
OC3	0.2563 (3.66)	0.3877 (5.40)	-0.1925 (-1.21)	0.1608	0.2123	-0.2177
OC4	0.2981 (3.58)	0.3628 (4.21)	0.3410 (1.62)	0.2026	0.1873	0.3158
OC5	0.6124 (6.83)	0.6497 (7.00)	0.6588 (3.17)	0.5169	0.4742	0.6336
SE0				-0.0967	-0.1523	-0.0578
SE1	0.2726 (3.81)	0.4042 (5.46)	0.3118 (1.96)	0.1759	0.2519	0.2540
SE2	0.2536 (3.22)	0.2410 (3.01)	0.1644 (0.90)	0.1569	0.0887	0.1066
SE3	0.3493 (4.43)	0.2302 (2.92)	0.9932 (5.04)	0.2527	0.0779	0.9354
SE4	0.2004 (2.49)	0.4066 (4.51)	0.1152 (0.55)	0.1128	0.2543	0.0575
SE5	0.2216 (3.00)	0.3123 (4.21)	0.0211 (0.10)	0.1250	0.1599	-0.0366
R^2	.289	.374	.187			
s.e.	0.847	0.752	0.933			

	IDS		Employed		Self-employed	
	β^2	F	β^2	F	β^2	F
EH	.107	171.9	.092	94.5	.092	55.9
AG	.006	12.5	.005	6.4	.013	9.8
OC	.030	57.9	.047	57.5	.025	18.0
SE	.016	30.8	.031	38.0	.023	16.5
$\sum\beta^2$	.160		.176		.153	

Table 16A. β^2s from Tables 12-16 using variables singly and simultaneously

	IDS		Employed		Self-employed	
	Singly	Simul-taneously	Singly	Simul-taneously	Singly	Simul-taneously
EH	.225	.107	.283	.092	.132	.092
AG	.007	.006	.019	.005	.011	.013
OC	.214	.030	.314	.047	.093	.025
SE	.184	.016	.271	.031	.075	.023
$\sum\beta^2$		.160		.176		.153

Table 17. LY: (EH), (AG), (OC), (SE), (HH)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-1.4289	-1.5166	-1.2766	-0.4605	-0.4044	-0.4210
EH1	0.3511 (11.1)	0.2332 (5.72)	0.4166 (8.59)	-0.1094	-0.1712	-0.0044
EH2	0.4784 (14.4)	0.4459 (10.4)	0.4266 (8.32)	0.0179	0.0415	0.0056
EH3	0.6141 (13.2)	0.4282 (7.39)	0.7046 (9.34)	0.1536	0.0238	0.2836
EH4	0.7973 (16.5)	0.6004 (10.6)	0.8788 (9.71)	0.3368	0.1960	0.4578
EH5	1.1178 (15.8)	0.8885 (11.1)	1.1698 (8.06)	0.6573	0.4842	0.7488
EH6	1.4138 (21.1)	1.1492 (15.0)	1.6410 (11.8)	0.9533	0.7448	1.2200
AG0				-0.3069	-0.2325	-0.4497
AG1	0.2420 (3.87)	0.1883 (2.71)	0.3269 (2.84)	-0.0649	-0.0442	-0.1227
AG2	0.2747 (4.42)	0.2208 (3.17)	0.3890 (3.43)	-0.0322	-0.0117	-0.0607
AG3	0.3685 (5.81)	0.2925 (4.10)	0.5140 (4.46)	0.0616	0.0600	0.0643
AG4	0.4370 (6.59)	0.3394 (4.43)	0.6290 (5.34)	0.1302	0.1070	0.1793
AG5	0.4292 (5.47)	0.2366 (2.42)	0.6540 (4.98)	0.1223	0.0041	0.2043
OC0				-0.0887	-0.1708	-0.0212
OC1	-0.0668 (-2.50)	0.5394 (1.26)	-0.0027 (-0.06)	-0.1555	0.3686	-0.0239

Table 17. (continued)

	IDS	Employed	Self-employed	IDS	Employed	Self-employed
OC2	0.1645 (2.34)	0.1603 (2.34)	0.0227 (0.10)	0.0758	-0.0105	0.0015
OC3	0.2424 (3.49)	0.3801 (5.38)	-0.2078 (-1.31)	0.1537	0.2093	-0.2290
OC4	0.2688 (3.26)	0.3419 (4.02)	0.2692 (1.28)	0.1801	0.1710	0.2480
OC5	0.6040 (6.79)	0.6550 (7.16)	0.5904 (2.85)	0.5153	0.4842	0.5692
SE0				-0.0998	-0.1431	-0.0684
SE1	0.2816 (3.96)	0.3830 (5.26)	0.3299 (2.08)	0.1818	0.2400	0.2615
SE2	0.2705 (3.46)	0.2335 (2.96)	0.2070 (1.14)	0.1706	0.0904	0.1386
SE3	0.3394 (4.84)	0.2058 (2.65)	0.9810 (5.00)	0.2395	0.0627	0.9125
SE4	0.2218 (2.66)	0.3879 (4.38)	0.1756 (0.84)	0.1220	0.2448	0.1072
SE5	0.2322 (3.17)	0.2909 (3.98)	0.0906 (0.45)	0.1324	0.1478	0.0221
HH0				-0.4449	-0.7030	-0.0745
HH1	0.1944 (1.83)	0.3569 (2.88)	-0.0503 (-0.27)	-0.2505	-0.3461	-0.1248
HH2	0.4333 (4.18)	0.7217 (6.03)	-0.0194 (-0.11)	-0.0116	0.0188	-0.0939
HH3	0.5259 (5.07)	0.7733 (6.46)	0.1785 (0.98)	0.0811	0.0704	0.1039
R^2	.301	.394	.197			
s.e.	0.840	0.740	0.928			

Table 17. (continued)

	IDS		Employed		Self-employed	
	β^2	F	β^2	F	β^2	F
EH	.104	170.0	.091	95.9	.088	53.8
AG	.008	15.3	.005	6.2	.017	12.4
OC	.028	53.9	.046	58.0	.019	13.8
SE	.017	32.9	.028	34.9	.024	17.7
HH	.013	41.0	.021	45.1	.010	12.4
$\sum\beta^2$	.169		.191		.158	

Table 18. LY: (EH), (AG), (OC), (SE), (HH), (GR), (HO), (LR), (CW), (SX), (MS)

	Regression Equations			Category Effects		
	IDS	Employed	Self-employed	IDS	Employed	Self-employed
const.	-1.0878	-0.9741	-1.0259	-0.3876	-0.3253	-03980
EH1	0.3115 (10.0)	0.1780 (4.43)	0.4159 (8.60)	-0.0761	-0.1473	0.0179
EH2	0.4147 (12.6)	0.3727 (8.81)	0.4020 (7.74)	0.0271	0.0474	0.0040
EH3	0.5193 (11.2)	0.3418 (5.96)	0.6495 (8.60)	0.1316	0.0165	0.2515
EH4	0.6234 (12.9)	0.4662 (8.23)	0.7583 (8.37)	0.2358	0.1409	0.3603
EH5	0.8819 (12.5)	0.7070 (8.86)	1.0531 (7.29)	0.4943	0.3817	0.6551
EH6	1.1212 (16.5)	0.9402 (12.2)	1.4533 (10.5)	0.7335	0.6149	1.0553
AG0				-0.3157	-0.2206	-0.4954
AG1	0.2427 (3.98)	0.1805 (2.65)	0.3644 (3.23)	-0.0730	-0.0401	-0.1310
AG2	0.2974 (4.90)	0.2196 (3.22)	0.4492 (4.04)	-0.0183	-0.0010	-0.0462
AG3	0.3699 (5.95)	0.2739 (3.90)	0.5522 (4.86)	0.0542	0.0533	0.0568
AG4	0.4428 (6.80)	0.2977 (3.94)	0.6770 (5.83)	0.1271	0.0771	0.1816
AG5	0.4384 (5.69)	0.2113 (2.20)	0.6957 (5.35)	0.1227	-0.0093	0.2002
OC0				-0.1036	-0.1549	0.0077
OC1	0.0494 (1.43)	0.4690 (1.12)	-0.0207 (-0.47)	-0.0542	0.3141	-0.0130

Table 18. (continued)

	IDS	Employed	Self-employed	IDS	Employed	Self-employed
OC2	0.1165 (1.70)	0.1681 (2.49)	-0.1325 (-0.59)	0.0130	0.0132	-0.1248
OC3	0.1915 (2.82)	0.3175 (4.55)	-0.2596 (-1.66)	0.0880	0.1626	-0.2519
OC4	0.2800 (3.47)	0.3292 (3.93)	0.1943 (0.94)	0.1765	0.1743	0.2020
OC5	0.5842 (6.75)	0.6277 (6.98)	0.4600 (2.26)	0.4807	0.4728	0.4678
SE0				-0.0777	-0.0902	-0.0767
SE1	0.2361 (3.39)	0.2571 (3.54)	0.3878 (2.49)	0.1584	0.1670	0.3111
SE2	0.1656 (2.17)	0.1390 (1.78)	0.1481 (0.83)	0.0879	0.0488	0.0714
SE3	0.2089 (2.72)	0.0829 (1.08)	0.8861 (4.60)	0.1312	-0.0073	0.8094
SE4	0.2242 (2.71)	0.2294 (2.60)	0.2042 (1.00)	0.1464	0.1392	0.1275
SE5	0.1789 (2.50)	0.2033 (2.80)	0.1844 (0.93)	0.1012	0.1131	0.1077
HH0				-0.3742	-0.5855	-0.0940
HH1	0.1920 (1.86)	0.2676 (2.20)	0.0263 (0.15)	-0.1822	-0.3179	-0.0677
HH2	0.3373 (3.34)	0.5944 (5.03)	-0.0089 (-0.05)	-0.0370	0.0089	-0.1029
HH3	0.4575 (4.54)	0.6564 (5.56)	0.1875 (1.06)	0.0833	0.0710	0.0934
GR0				0.1836	0.1707	0.1597
GR1	-0.1412 (-3.38)	-0.1370 (-3.17)	-0.1137 (-1.12)	0.0423	0.0337	0.0460
GR2	-0.3561 (-7.94)	-0.3199 (-6.72)	-0.3612 (-3.45)	-0.1725	-0.1492	-0.2015

Table 18. (continued)

	IDS	Employed	Self-employed	IDS	Employed	Self-employed
GR3	-0.1558 (-3.33)	-0.2386 (-4.52)	-0.0553 (-0.52)	0.0278	-0.0679	0.1044
HO0				-0.0106	-0.0042	-0.0109
HO1	0.0042 (0.16)	-0.0020 (-0.06)	-0.0050 (-0.11)	-0.0064	-0.0062	-0.0158
HO2	0.0840 (2.16)	0.0411 (0.95)	0.1186 (1.60)	0.0734	0.0368	0.1078
LR0				0.1157	0.0981	0.1195
LR1	-0.1665 (-4.57)	-0.1554 (-3.86)	-0.1538 (-2.19)	-0.0508	-0.0573	-0.0343
CW0				-0.0917		
CW1	0.1626 (5.7[illegible])			0.0709		
SX0				0.0247	0.0163	0.0316
SX1	-0.4725 (-9.93)	-0.5948 (-8.10)	-0.3752 (-5.64)	-0.4479	-0.5785	-0.3436
MS0				-0.0218	-0.0156	-0.0276
MS1	0.0920 (3.53)	0.0558 (1.87)	0.1516 (3.26)	0.0702	0.0402	0.1239
R^2	.340	.423	.233			
s.e.	0.817	0.723	0.908			

Table 18. (continued)

	IDS		Employed		Self-employed	
	β^2	F	β^2	F	β^2	F
EH	.064	110.8	.061	67.0	.070	44.5
AG	.008	16.0	.004	4.9	.018	13.6
OC	.019	39.0	.038	50.0	.015	11.7
SE	.010	20.9	.012	15.8	.025	18.8
HH	.009	31.4	.018	38.6	.008	10.5
GR	.012	42.3	.013	28.5	.014	18.4
HO	.001	3.5	.000	0.6	.001	2.7
LR	.006	60.2	.006	41.4	.004	14.7
CW	.006	66.5				
SX	.011	113.1	.011	69.5	.010	39.1
MS	.002	15.7	.001	4.6	.003	12.3
$\sum\beta^2$	.148		.162		.169	

Table 18A. Ordering of occupations by size of effect

Employed		Self-employed	
From Table 14	From Table 18	From Table 14	From Table 18
OC5	OC5	OC5	OC5
OC4	OC1	OC4	OC4
OC3	OC4	OC3	OC0
OC2	OC3	OC2	OC1
OC1	OC2	OC0	OC2
OC0	OC0	OC1	OC3

Table 18B. Ordering of sectors by size of effect

Employed		Self-employed	
From Table 15	From Table 18	From Table 15	From Table 18
SE4	SE1	SE3	SE3
SE5	SE4	SE5	SE1
SE1	SE5	SE4	SE4
SE3	SE2	SE1	SE5
SE2	SE3	SE2	SE2
SE0	SE0	SE0	SE0

Table 18C. Ordering of variables by size of β^2

IDS	Employed	Self-employed
EH	EH	EH
OC	OC	SE
GR	HH	AG
SX	GR	OC
SE	SE	GR
HH	SX	SX
AG	LR	HH
CW	AG	LR
LR	MS*	MS
MS	HO**	HO**
HO**		

*Significant at .05 level.
**Not significant at .05 level.

Table 18D. R^2 and $\sum\beta^2$ using different specifications

	IDS		Employed		Self-employed	
	R^2	$\sum\beta^2$	R^2	$\sum\beta^2$	R^2	$\sum\beta^2$
Basic	.289	.160	.374	.176	.187	.153
+ HH	.301	.169	.394	.191	.197	.158
+ GR	.306	.157	.389	.163	.205	.167
+ HO	.293	.150	.376	.169	.194	.147
+ LR	.300	.144	.382	.154	.196	.149
+ SX	.304	.169	.386	.184	.199	.164
+ MS	.295	.156	.377	.170	.194	.156
+ CW	.295	.160				
All*	.340	.148	.423	.162	.233	.169

*Excepting CW for employed and self-employed

Appendix A. Relationship of the β^2s to R^2

Consider the interesting (but false) conjecture that

$$R^2 = \beta_A^2 + \beta_B^2$$

where $R^2 = 1 - s_e^2/s_y^2$ is the coefficient of determination given by (1). One might argue in the following way. We have

$$y_{kji} - \bar{y} = a_k + b_j + (y_{kji} - \bar{y} - a_k - b_j)$$

$$\sum_{kji}\sum\sum (y - \bar{y})^2 = \sum\sum\sum [a_k + b_j + (y_{kji} - \bar{y} - a_k - b_j)]^2$$

Suppose cross-product terms on the right-hand side all vanish. Then

$$\sum_{kji}\sum\sum (y - \bar{y})^2 = \sum\sum\sum a_k^2 + \sum\sum\sum b_j^2 + \sum\sum\sum (y_{kji} - \bar{y} - a_k - b_j)^2$$

in which case

$$\frac{(n-1)s_y^2}{\sigma_y^2} = \frac{(n-1)s_a^2}{\sigma_y^2} + \frac{(n-1)s_b^2}{\sigma_y^2} + \frac{(n-1)s_e^2}{\sigma_y^2}$$

where $s_a^2 = \sum_k n_{k.} a_k^2/(n-1)$, $s_e^2 = \sum\sum\sum (y_{kji} - \bar{y} - a_k - b_j)^2/(n-1)$, etc. Assuming that y is normally distributed, it is known that $(n-1)s_y^2/\sigma_y^2$ is a chi-square variable with $n - 1$ degrees of freedom, so that the terms on the right-hand side must then be independent chi-square variables whose degrees of freedom add up to $n - 1$. That is,

$$\chi^2_{y,n-1} = \chi^2_{a,K} + \chi^2_{b,J} + \chi^2_{e,n-K-J-1}$$

and one then gets the F-statistic (8). Moreover,

$$1 = \frac{s_a^2}{s_y^2} + \frac{s_b^2}{s_y^2} + \frac{s_e^2}{s_y^2} = \beta_A^2 + \beta_B^2 + \frac{s_e^2}{s_y^2}$$

which says that R^2 is the sum of the β^2s. This statement is wrong, however, as cross-product terms do not all vanish. For instance,

$$\sum\sum\sum a_k (y_{kji} - \bar{y} - a_k - b_j) = \sum_k a_k n_{k.} \bar{y}_{k.} - \bar{y} \sum_k n_{k.} a_k - \sum_k n_{k.} a_k^2 - \sum\sum\sum a_k b_j$$

The second and fourth terms on the right-hand side are easily seen to vanish, but

$$\sum_k a_k n_{k.} \bar{y}_{k.} - \sum_k n_{k.} a_k^2 = \sum_k a_k n_{k.} (\bar{y}_{k.} - a_k) \quad .$$

is zero only if $\bar{y}_{k.} - a_k$ is a constant, as would be the case if $a_k = \bar{y}_{k.} - \bar{y}$. The last clearly does not hold in general, however. If it did, the calculated a_k would not change as one adds more classificatory variables in the explanation of y, but in fact the a_k--which derive from the regression coefficients of (1)--would change with a different specification.

Appendix B. Selection of IDS Sample

The following criteria were used in obtaining our sample from the NDS master file. First, only families whose household data type were "normal" (i.e. those which had all seven records of the NDS) were selected. Second, from this group, the following items were checked for NR codes ("no response" or "not reported"). If one or more of these items turned out NR, the corresponding family was excluded from the sample:

Farm type
House ownership
Lot ownership
Sex of head
Age of head
Migration code of head
Education of head
Occupation of head
Class of worker of head
Hours worked by head
Major industry
Language most often spoken by head
Total number of children born alive in all marriages
Income of head
Total family income.

The resulting proportions of urban and rural families in our IDS sample may be compared with those of the NDS:

	N D S	I D S
Urban	.302	.305
Rural	.698	.695
	1.000	1.000

Footnotes

1. Thanks are due the World Employment Programme of the International Labour Office for research support. Writing this paper has benefited from discussions with Hakchung Choo, Oey Meesook, Felix Paukert and Richard Wada, and especially Lydia H. Flores particularly in regard to Section 3. Benjamin Nolasco and Ruben de la Paz provided research assistance, and Porfirio Sazon, Jr. did the computations at the University of the Philippines Computer Center.

2. We note that the sum of the β^2s does not necessarily equal the coefficient of determination R^2 given by (1); see Appendix A.

3. These figures imply median incomes of 0.972 for IDS, 1.147 for the employed and 0.785 for the self-employed.

4. "Farmer-owners" work their own farms; "owner-tenants" work land owned by somebody else in addition to their own.

5. The figure for SE0 self-employed (.749) cannot be less than the sum of OC0 and OC1 self-employed in Table 3 (.214 + .540), so there must have been some misreporting here.

6. We define a person as a non-migrant if his 1960 residence, 1968 residence and father's residence are all the same.

7. It will be noticed that we use these symbols as dummy variables (with values 0 or 1) and also as categories, but this should cause no confusion. Note that (EH) does not include EH0. We will also refer to EH as a classificatory variable and will follow the same

convention in regard to the other classificatory variables so that (SX), for example, is simply SX1.

8. Underreporting might have played a role here, though why any possible underreporting should be more severe for farmers-owners than for farm-tenants and fishermen, etc., is not obvious.

9. A pictorial explanation might be helpful here. If we think of EH and AG as ordinary numerically-scaled variables, then the regression coefficient of AG in the specification LY: AG would be lower than in LY: AG, EH. The reason is that relatively more observations at higher values of AG have low EH, thus pulling down the average LY of higher AG.

10. Following the analogy of the preceding footnote, the regression planes for the two groups are not too different (the β^2s and $\sum\beta^2$ are about the same) but the dispersion of the self-employed is greater (so that R^2 is lower).

11. This asymmetry remains in Table 18 where more variables are included in the specification.

12. This pattern appears with some of the earlier specifications also.

13. As noted previously, equality of β^2s or of $\sum\beta^2$ for the two groups does not mean the same monetary magnitudes, as the mean LY of the self-employed is lower.

14. The degree of inequality may be somewhat less with respect to <u>family</u> incomes, however, as wives are more likely to be in

the labor force if their husbands earn less. Regression results from a subsample (size 3629) of the NDS give the following:

$$\begin{array}{rl} LPW = & 0.3202 - 0.0093\ FYH - 0.1184\ EWN + 0.0656\ EWX \\ & \qquad\quad (-3.78) \qquad\quad (-9.79) \qquad\quad (7.06) \end{array}$$

$$R^2 = .030, \qquad F = 38.1$$

where LPW = 1 if the wife is in the labor force and 0 otherwise; FYH is the husband's annual income (in thousand pesos); and

$$EWN = \min(EW - 2.5,\ 0)$$

$$EWX = \max(EW - 2.5,\ 0)$$

where EW is a numerically scaled variable representing the wife's education level. The coding for EW follows that of EH so that, e.g., EW = 2 means five to seven years of grade school. As might have been expected, FYH and EWN have negative coefficients while EWX has a positive one. The lower is the husband's income and the lower is EW (as proxy for earning power) below some education threshold, the more likely is the wife's participation in the labor force in order to reach some subsistence level of family income. Above the threshold, more EW means higher participation because of greater opportunity costs. The effect of education on wives' participation rates thus appears to be essentially nonlinear; because of this, the net result on family income distribution is unclear.

References

Aitchison, J. and Brown, J.A.C. The Lognormal Distribution, Cambridge University Press, 1957.

Anderson, R.L. and Bancroft, T.A. Statistical Theory in Research, McGraw-Hill, 1952.

Cramer, J.S. Empirical Econometrics, North-Holland, 1971.

Encarnación, J. et al. Appendix 4 of "An Economic-Demographic Model of the Philippines," in Studies in Philippine Economic-Demographic Relationships by A. Kintanar et al., School of Economics, IEDR, University of the Philippines System, 1974.

Hogg, R.V. and Craig, A.T. Introduction to Mathematical Statistics, 2nd ed., Macmillan, 1965.

Kuznets, S. "Quantitative Aspects of the Economic Growth of Nations: Distribution of Income by Size," Economic Development and Cultural Change, January 1963 (Part II).

Meesook, O.A. "Income Distribution in Brazil," unpublished Ph.D. dissertation, University of California at Berkeley, 1972.

Morgan, J.N. et al. Income and Welfare in the United States, McGraw-Hill, 1962. Appendix E.

Raymundo, C. "The Methodology of the 1968 NDS," Appendix C in A Demographic Path to Modernity: Patterns of Early Transition in the Philippines, ed. W. Flieger and P.C. Smith, University of the Philippines Press, 1974.

Suits, D.B. "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, December 1957, pp. 548-51.