World Bank Reprint Series: Number Forty-four

Carmel Ullman Chiswick

On Estimating Earnings Functions for LDCs

ON ESTIMATING EARNINGS FUNCTIONS FOR LDCs

Carmel Ullman CHISWICK*

World Bank, Washington, DC 20433, U.S.A.

Received April 1976, revised version received July 1976

A technique is developed for estimating earnings functions from data in which both wage-earners and self-employed persons are included; it is argued that the return to schooling is overestimated if only wage-earners are in the sample. A procedure for simultaneously imputing labor incomes to the self-employed is also developed and carried out with 1971 data for Bangkok, Thailand. The evidence suggests that ceteris paribus the self-employed have labor earnings at least as high as their wage-earning counterparts.

1. Introduction

It is generally agreed that the development and deployment of human resources is a key aspect of economic growth and increasingly, as the data become available, economists have been estimating human capital earnings functions for a wide variety of LDCs. Although each new study seems to confirm the predicted relationship between schooling, experience (or age) and earnings, the coefficient of schooling tends to be somewhat higher for LDCs than has been found for the United States.1 This paper explores the possibility that such high coefficients are due at least in part to a systematic bias in the selection of the sample population and develops a procedure to overcome this problem.

Estimates of the coefficients of earnings functions are biased when the samples are restricted to wage-earners, or 'employees,' often further restricted to those working in 'modern' manufacturing establishments. Limiting the analysis to

*The author is an Economist on the staff of the World Bank's Development Research Center. Neither the World Bank nor any of its affiliates, however, are responsible for the views expressed in this paper. The material presented here is part of a research project on Income Distribution in Thailand; a more detailed reporting of equations is given in Working Paper Series B-1 of that project.

1For example, the coefficient of schooling in a regression of the natural log of earnings on schooling for males over age 25 is 0.08 for the U.S. in 1960, 0.08 for Canada in 1960, and 0.14 for Mexico in 1963 [Chiswick (1974, pp. 83-88)]. When the regression is run for urban men of all ages, with experience and weeks worked held constant, the coefficient of schooling is 0.07 for the U.S. in 1970 and 0.18 for Mexico in 1963 [Chiswick (1974, p. 87)]. Although other published estimates for LDCs are not sufficiently compatible to permit strict comparisons, it is not unusual for them to find coefficients well over 0.10.
wage-earners poses little problem for the United States or other developed countries where these constitute a very large proportion of the labor force. It might also be appropriate in cases of extreme rigidity where schooling is available only to those who will later become wage-earners in the modern sector. For many LDCs, however, a substantial fraction of the employed labor force is self-employed. Although they tend to be disproportionately concentrated among those with little schooling, they are not necessarily less well-educated than wage-earners in comparable occupations. Moreover, it is reasonable to suppose that among those with little or no schooling the self-employed as a group compare more favorably to wage-earners with respect to enterprise and other earnings-related ability characteristics. If so, the expected wage (hence opportunity cost) of a self-employed person with little schooling would tend to be higher than the average wage actually received by employees at the same schooling level. Since this discrepancy is not nearly so pronounced for higher schooling levels, restricting the sample to wage-earners would result in an overstatement of the expected gain from an additional year of schooling, especially for low schooling levels.

Restricting the sample to wage-earners has seemed necessary in part because of the data limitations but mainly because employees are the only category of workers for which reported income can be interpreted as the earnings of labor. In a competitive labor market, however, where workers are mobile between firms and have reasonable access to information about job opportunities, the opportunity cost of a self-employed person would be the wage commanded by employees with similar skills (or other productivity characteristics). An extension of the model deriving the earnings function for wage-earners may be used to derive the potential earnings of a self-employed person, and hence the opportunity cost of that person's labor. The procedure developed and tested in this paper yields an estimate of this opportunity cost as well as an unbiased estimate of the schooling coefficient in the earnings function.

The data for testing the model and estimating labor's share of self-employment income are from a household survey of Bangkok in 1971. This particular survey is well-suited to the task: it covers the entire labor force; it contains information on most of the desired variables for each income recipient; it pertains to a city with many small businesses and highly competitive markets in which dualism is not an obvious feature; and slightly more than a third of all

---

2 In a dualistic labor market the opportunity cost of a self-employed person's labor would depend on his potential earnings in whichever sector he might expect to find employment. An earnings function for the modern sector can be used to estimate the opportunity cost of only those persons who know they could find modern sector jobs. It may be possible in some cases to estimate a separate earnings function for the traditional sector; the opportunity cost of self-employment could then be calculated as the expected value of potential earnings, e.g. average predicted earnings in each sector weighted by the probability that the individual would find employment in that sector.
income earners are self-employed. The model is developed in section 2 and applied to these data in section 3. Section 4 is a summary and conclusion.

2. The model

Let \( W \) denote the wage or salary income of employees, let \( P^* \) denote the income from self-employment (profits), and let \( Y \) denote all earnings; for each individual \( i \),

\[
Y_i = W_i + P_i^*.
\]

(1)

Self-employment income may be attributed to two components, one a return to the labor of the self-employed person and one a return to the owner's non-labor inputs into the business. Let \( \alpha_i \) represent labor's share of self-employment income for the \( i \)th individual, \( \alpha_i \geq 0. \)

Assuming that all wages are returns to labor, total labor income \( Y_{Li} \) would be

\[
Y_{Li} = W_i + \alpha_i P_i^*.
\]

(2)

From (1) and (2) it is clear that

\[
Y_i = Y_{Li} + (1 - \alpha_i)P_i^*. \]

(3)

Now define \( p_i \) as the fraction of earnings derived from self-employment, \( P_i^* = p_i Y_i \); substitute this into eq. (3) and collect terms to obtain

\[
Y_{Li} = [1 - (1 - \alpha_i) p_i] Y_i. \]

(4)

Solving for \( Y_i \) and taking logarithms yields the basic result:

\[
\log Y_i = \log Y_{Li} - \log [1 - (1 - \alpha_i) p_i]. \]

(5)

Since the \( \alpha_i \) are unknown and the \( p_i \) are fairly easy to estimate, it would be especially convenient to be able to factor out \( p_i \) from the last term in eq. (5) and assume that it is distributed independently of its coefficient. This would permit estimation of the equation

\[
\log Y_i = \log Y_{Li} + \bar{F} p_i, \]  

(5')

where

\[
\bar{F} = E \left[ -\log \left[ \frac{1 - (1 - \alpha)p}{p} \right] \right],
\]

In general, \( \alpha_i \leq 1 \). However, an \( \alpha > 1 \) may arise when unreported income is more important for self-employment income than for wages. Suppose, for example, that wages are fully reported but that the money value of income from business is actually \( P_i = (1 + \pi_i)P_i^* \), \( \pi_i > 0 \). Then the \( \alpha_i \) of eq. (2) is labor's share of \( P_i^* \) inflated by a factor of \( (1 + \pi_i) \), a bias which should be kept in mind when interpreting empirical estimates of \( \alpha \).
by simply including the variable \( p \) in an ordinary least-squares estimate of the earnings function. Although \( F \) is clearly not independent of \( p \), it will be shown below that it is bounded from both sides by limits which are determined by \( \alpha \) alone, and further that these bounds are sufficiently close together so as to permit under certain conditions the estimation of \( F \) as a constant whose value is independent of \( p \).

For simplicity of exposition, define a new function:

\[
\chi(p, \alpha) = \frac{1}{1-(1-\alpha)p},
\]

which is positive throughout the interval \( 0 \leq p \leq 1 \) for all values of \( \alpha \).\(^4\) The coefficient of \( p \) in eq. (5') may now be written:

\[
F(p, \alpha) = \begin{cases} 
\log \frac{x}{p}, & p \neq 0, \\
1-\alpha, & p = 0.
\end{cases}
\]

For values of \( \alpha \) less than one, \( x \) is always greater than unity and \( F \) is therefore positive; for values of \( \alpha \) greater than one, \( x \) is less than unity and \( F \) is negative. \( F \) is a continuous nondecreasing function of \( p \) throughout the interval \( 0 \leq p \leq 1 \), implying that it is bounded by \( F(0, \alpha) \) and \( F(1, \alpha) \). For a proof, take the first partial derivative of eq. (7) with respect to \( p \):

\[
\frac{\partial F}{\partial p} = \frac{1}{p^2} \left[ \frac{p \partial x}{x \partial p} - \log x \right], \quad p \neq 0.
\]

From eq. (6),

\[
\frac{\partial x}{\partial p} = (1-\alpha)x^2
\]

and

\[
(x-1) = (1-\alpha)px.
\]

Substituting (9a) and then (9b) into eq. (8) yields

\[
\frac{\partial F}{\partial p} = \frac{1}{p^2} \left[ (x-1)-\log x \right].
\]

\(^4\)The function \( x \) is not defined for the case where \( p = 1 \) and \( \alpha = 0 \) — the case of the non-working 'capitalist.'

\(^5\)Using L'Hopital's rule,

\[
\lim_{p \to 0} F(p, \alpha) = \lim_{p \to 0} \frac{1}{x} \frac{\partial x}{\partial p} / 1 = \lim_{p \to 0} (1-\alpha)x = (1-\alpha).
\]
But \( x - 1 \geq \log x \) for any positive \( x \), with equality holding only for \( x = 1 \) (which occurs when \( \alpha = 1 \) or \( p = 0 \)). Since \( \partial F/\partial p \) is nonnegative, \( F \) is bounded from below by \( F(0, \alpha) = 1 - \alpha \) and from above by \( F(1, \alpha) = -\log \alpha \).

Thus, even though individual values of \( F \) are not distributed independently of \( p \), its range is determined only by the individual values of \( \alpha \). This is an especially powerful result when \( \alpha \) is near unity, for in that case \( 1 - \alpha \approx -\log \alpha \) and \( F \) is virtually constant\(^6\) throughout the interval \( 0 \leq p \leq 1 \). If the interdependence of \( F \) and \( p \) biases upward the value of \( \bar{F} \), it will be closer to \( -\log \alpha \); if the bias is downward, \( \bar{F} \) will be closer to \( (1 - \alpha) \). As long as we know that \( F \) is bounded by these values, however, we can infer maximum and minimum values of \( \alpha \) from the estimated \( \bar{F} \).

The full estimating equation used here is

\[
\log Y = b_0 + b_1 S + b_2 T - b_3 T^2 + b_4 p + \varepsilon, \tag{10}
\]

where:

- \( Y \) = annual income from wages, salaries, and self-employment;
- \( S \) = years of formal schooling;
- \( T \) = years of post-schooling experience;
- \( P \) = self-employment income (profits) as a fraction of annual income \( (0 \leq p \leq 1) \);
- \( \varepsilon \) = stochastic term.

This model takes the usual human capital earnings function\(^7\) and adds a new variable, \( p \), the coefficient of which is an estimate of \( F(p, \alpha) \). If this estimate is based on a pooled sample of persons with only wage and salary income and those with only self-employment income so that \( p \) is a dummy variable which takes the value of unity for profit-earners and zero for wage-earners, the coefficient \( b_4 \) may be interpreted as \( F(1, \alpha) = -\log \alpha \). An unconstrained least-squares regression estimate of eq. (10) will thus always give a positive estimated value of \( \alpha \).

3. Estimates of labor income

The Thailand Socio-Economic Survey 1971 is a household survey which includes information on education, age and income by type for each household member separately. The sample of 5,739 earners comprises individuals age 15 and over in the Bangkok region who reported nonzero incomes from wages, salaries or ‘own business’ during the previous year. Wage and salary income

\(^6\)The approximation holds to one significant digit for \( \alpha \geq 0.75 \) and to two significant digits for \( \alpha \geq 0.90 \).

\(^7\)The equation is developed fully in Mincer (1974), to which the interested reader is referred.
during the year was used for the variable $W_i$ and income from 'own business' was used for $P_i^*$. The variables $Y$ and $p$ were computed directly from this data.

The survey reports educational attainment by level and type of program. This information was used to construct the variable $S$ by attributing an approximate number of years spent in school to each educational group. Years of work experience ($T_i$) were then computed as the difference between present age and the age at leaving school (estimated as $S_i + 6$).

The first two columns of table 1 show the distribution of earners in Bangkok by sex and type of income; the remaining columns show the distribution of each category by education level and the proportion with incomes from farm occupations. Wage-earners with no self-employment income account for 65 percent of the men and 62 percent of the women; another 4 percent of the men (and a few women) are wage-earners with some self-employment income as well. The rest are self-employed persons with no wage income at all (i.e. persons for whom $p = 1$).

The concentration of persons with low levels of education among the self-employed is consistent with an underlying structure which is similar for all earners. If wages are positively related to schooling and if the expected gain from own-business activities relative to the opportunity cost is higher for labor at low wage levels, the self-employed should represent a larger proportion of all earners among people with low levels of schooling. This is in fact the case in Bangkok, where people with only profits account for 54 percent of all nonfarm earners with no formal education, 33 percent of those with no more than primary schooling, and only 10 percent of those with more than a primary education.

It has been mentioned earlier that if the self-employed have higher rates of under-reporting of income than wage-earners, the estimated value of $x$ will be biased upward. This is especially troublesome for the case of farmers: men with income from farm occupations account for fully 33 percent of the men with incomes from both wages and self-employment in Bangkok. The exclusion of income in kind and home-production from the measured value of $P^*$ means that their incomes may well be substantially under-reported, and interpretation of $x$ is accordingly weakened. Persons in farm occupations have therefore been excluded from the sample.

Earnings functions have been estimated separately for men and for women and the results are presented in table 2. The first two equations pool wage-earners with the self-employed and the third is for wage-earners alone; the second equation differs from the first in that it excludes persons with both types of income from the sample. The coefficient of schooling is lowered by 1.5 percentage.

---

8These values of $S$ are necessarily somewhat arbitrary. In the general academic program, people with primary education (1–9 years of schooling) were assigned $S = 6$, for lower secondary $S = 10.5$, for upper secondary $S = 13$, and for higher education $S = 17$. For the vocational program, middle- and upper-level categories were assigned $S = 10$ and $S = 14$ respectively.
Table 1

Distribution of earners by type of earnings, education level and sex (persons age 15 and over, Bangkok, 1971).*

<table>
<thead>
<tr>
<th></th>
<th>Numberb</th>
<th>Percent of total</th>
<th>Distribution by education level (percent)</th>
<th></th>
<th>All occupations</th>
<th>Percent in farm occupations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All men</strong></td>
<td>599,580</td>
<td>100</td>
<td>16</td>
<td>54</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(3,416)</td>
<td></td>
<td>(2,299)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages only (p = 0)</td>
<td>389,702</td>
<td>65</td>
<td>10</td>
<td>51</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(2,299)</td>
<td></td>
<td>(125)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages and self-employment income (0 &lt; p &lt; 1)</td>
<td>24,058</td>
<td>4</td>
<td>8c</td>
<td>57</td>
<td>35</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(125)</td>
<td></td>
<td>(20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-employment income only (p = 1)</td>
<td>185,821</td>
<td>31</td>
<td>30</td>
<td>59</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(992)</td>
<td></td>
<td>(20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All women</strong></td>
<td>398,980</td>
<td>100</td>
<td>18</td>
<td>58</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(2,291)</td>
<td></td>
<td>(1,447)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages only (p = 0)</td>
<td>248,079</td>
<td>62</td>
<td>11</td>
<td>54</td>
<td>35</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(1,447)</td>
<td></td>
<td>(20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages and self-employment income (0 &lt; p &lt; 1)</td>
<td>3,701</td>
<td>1</td>
<td>10c</td>
<td>68</td>
<td>22c</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td></td>
<td>(20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-employment income only (p = 1)</td>
<td>147,199</td>
<td>37</td>
<td>30</td>
<td>64</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(824)</td>
<td></td>
<td>(824)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


bSample sizes are given in parentheses.

cSample size ≤ 10.
Table 2
Earnings functions by sex and importance of self-employment (persons age 15 and over in nonfarm occupations, Bangkok, 1971).

<table>
<thead>
<tr>
<th>Sample pools wage-earners with</th>
<th>Estimated coefficients of: (^a)</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(S)</td>
<td>(T)</td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All earners ((0 \leq p \leq 1))</td>
<td>0.091</td>
<td>0.071</td>
</tr>
<tr>
<td>((p = 0) or (p = 1))</td>
<td>(0.0036)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>Self-employed with no wages</td>
<td>0.089</td>
<td>0.072</td>
</tr>
<tr>
<td>((p = 0))</td>
<td>(0.0037)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>No self-employed ((p = 0))</td>
<td>0.104</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.0036)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All earners ((0 \leq p \leq 1))</td>
<td>0.130</td>
<td>0.038</td>
</tr>
<tr>
<td>((p = 0) or (p = 1))</td>
<td>(0.0047)</td>
<td>(0.0042)</td>
</tr>
<tr>
<td>Self-employed with no wages</td>
<td>0.129</td>
<td>0.038</td>
</tr>
<tr>
<td>((p = 0))</td>
<td>(0.0047)</td>
<td>(0.0042)</td>
</tr>
<tr>
<td>No self-employed ((p = 0))</td>
<td>0.145</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0042)</td>
</tr>
</tbody>
</table>

\(^a\)Dependent variable is \(\ln Y\) for all equations.
points (for both men and women) when the self-employed are included in the
sample along with wage-earners. Estimated values of $P(1, x) = -\log x$ are given
by the coefficients of $p$ in the second equation; the 'wage' imputed to persons
who depend entirely on self-employment is thus 57 percent of their income for
men and 76 percent for women. It is possible that the lower values of $x$ for men
as compared to women reflect more highly capitalized business for the former,
although it should be noted that this finding may also reflect a systematic
tendency to attribute to the household head the nonlabor component of profits
from a family business.

3.1. Interaction between $p$ and $x$

The value of $F$ (estimated as the coefficient of $p$) is somewhat higher when
persons with both wages and self-employment income are included. Since these
are people with low values of $p$ (and since $F$ is an increasing function of $p$ and a
decreasing function of $x$), this suggests a positive relationship between $x$ and $p$.
That is, men whose earnings derive mostly from self-employment (high $p$) tend
to report incomes that are close to the wages that constitute their opportunity
cost (high $x$), while men for whom 'own business' is not the main source of
income (low $p$) have a smaller share of income from that business attributable
to their own labor (low $x$). (No comparable statement is made for women since
there are too few observations.)

To provide evidence for an underlying relationship between $x$ and $p$ and at
the same time give some perspective on the robustness of the estimates, the
equation was estimated for several samples of wage-earning men differing only
by the proportion of their wages in total earnings. The coefficients of $S$, $T$ and
$T^2$ are the same for all samples of wage-earners regardless of whether or not
men with self-employment income are included, and the coefficient of $p$ is
always highly significant (see table 3). The value of $x$ is very low for the samples
in which self-employment income is supplemented by wages (or vice versa). It
would seem therefore that the opportunity cost of labor is a much smaller
component of self-employment income when the business from which it derives
is not the main source of income.

3.2. Interaction between schooling and $x$

About 90 percent of the people whose earnings are entirely composed of
own-business income are in the two lowest schooling categories. Since the
variable $S$ in the earnings function is least satisfactory for people with primary
or no education, separate estimates were made for persons with no education
at all, those with primary education only, and those with more than a primary
education. The results are summarized in table 4. The explanatory power of
the equation ($R^2$) increases with education level and is somewhat better for men
Table 3
Estimated values of $\alpha$ for different levels of $p$ (men aged 15 and over, in nonfarm occupations, Bangkok, 1971).

<table>
<thead>
<tr>
<th>Sample pools wage-earners with:</th>
<th>Number of cases with $p \neq 0$</th>
<th>$R^2$ of estimated equation</th>
<th>Coefficient$^a$ of $p$</th>
<th>Estimated $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All earners ($0 \leq p \leq 1$)</td>
<td>892</td>
<td>0.307</td>
<td>0.590</td>
<td>0.41 (0.0337) 0.55</td>
</tr>
<tr>
<td>Earners with no wages ($p = 0$ or $p = 1$)</td>
<td>797</td>
<td>0.303</td>
<td>0.567</td>
<td>0.43 (0.0340) 0.57</td>
</tr>
<tr>
<td>Earners with some wages ($p &lt; 1$)</td>
<td>95</td>
<td>0.374</td>
<td>1.608</td>
<td>-0.61 (0.1588) 0.20</td>
</tr>
<tr>
<td>Earners with mostly wages ($p \leq 0.6$)</td>
<td>79</td>
<td>0.362</td>
<td>1.409</td>
<td>-0.41 (0.2206) 0.25</td>
</tr>
</tbody>
</table>

$^a$Standard error is reported in parentheses.

Table 4
Summary of findings: Estimated values of $\alpha$ by education level and sex (persons aged 15 and over in nonfarm occupations, Bangkok, 1971).

<table>
<thead>
<tr>
<th>Education level</th>
<th>Percent of self-employed with some wages</th>
<th>$R^2$ of estimated equation</th>
<th>Coefficient$^a$ of $p$</th>
<th>Estimated $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All levels of education</td>
<td>11</td>
<td>0.307</td>
<td>0.590</td>
<td>0.41 (0.0337) 0.55</td>
</tr>
<tr>
<td>No formal education</td>
<td>3</td>
<td>0.221</td>
<td>0.648</td>
<td>0.35 (0.0863) 0.52</td>
</tr>
<tr>
<td>Primary education or less</td>
<td>8</td>
<td>0.316</td>
<td>0.601</td>
<td>0.40 (0.0434) 0.55</td>
</tr>
<tr>
<td>More than primary education</td>
<td>30</td>
<td>0.369</td>
<td>0.446</td>
<td>0.55 (0.0645) 0.64</td>
</tr>
<tr>
<td>($^b$)</td>
<td></td>
<td>0.373</td>
<td>0.343</td>
<td>0.66 (0.0652) 0.71</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All levels of education</td>
<td>2</td>
<td>0.282</td>
<td>0.285</td>
<td>0.71 (0.0423) 0.75</td>
</tr>
<tr>
<td>No formal education</td>
<td>0</td>
<td>0.142</td>
<td>0.602</td>
<td>0.40 (0.1101) 0.55</td>
</tr>
<tr>
<td>Primary education or less</td>
<td>2</td>
<td>0.161</td>
<td>0.268</td>
<td>0.73 (0.0531) 0.76</td>
</tr>
<tr>
<td>More than primary education</td>
<td>8</td>
<td>0.330</td>
<td>0.169</td>
<td>0.83 (0.0886$^c$) 0.84</td>
</tr>
</tbody>
</table>

$^a$Standard error is reported in parentheses.
$^b$Equation excludes men with both profits and wages.
$^c$Not significant at the 5% level.
than for women. It isrespectably high for men even when those with no schooling are considered separately. The estimated value of \( \alpha \) is slightly over 0.50 for men and women with no formal schooling. This is about the same as for men with a primary education and for all men together. The value of \( \alpha \) for men with more than primary schooling is about 0.60, and when men with both types of earnings are dropped from the sample it increases to 0.71. The increase in \( \alpha \) by education level is even more pronounced for women: it is 0.75 for women with some primary schooling and 0.84 for those with more than primary education.

Earlier it was noted that the expected gain from entrepreneurial activity (over and above the value of labor per se) may be higher relative to the value of labor for people at low schooling levels. Since the parameter \( \alpha \) represents the imputed share of labor in self-employment income, the finding that \( \alpha \) is higher for better-educated entrepreneurs is consistent with this hypothesis. In contrast, the positive association between education level and \( \alpha \) would contradict the notion that the more educated self-employed have a higher proportion of capital or other nonlabor incomes. Although this conclusion may be weakened if under-reporting is proportionately greater for people at high education levels, there is no a priori evidence that this is the case.

4. Conclusions

The technique developed in this paper is for analysing labor earnings in a competitive market where self-employment is an important alternative to wage-employment. This involves imputing the labor income of self-employed people as the wages they could otherwise earn as employees. The earnings function from which this imputation is made should be estimated from a pooled sample of both types of workers since a sample of wage-earners alone may not be random with respect to earnings-related characteristics. This may be done by including in the earnings function as an independent variable the share of self-employment income in total earnings. Moreover, the coefficient of this new variable provides a basis for comparing the imputed wage of self-employed persons with their total earnings.

The procedure was applied to 1971 data for the Bangkok region of Thailand. The coefficient of schooling in the earnings function is 0.091 for all nonfarm male earners. If only wage-earners were included in the sample, however, the estimate for men in Bangkok would have been 0.104. For women the estimated coefficient of schooling is 0.130 for all earners and 0.145 for wage-earners alone. These findings are consistent with the hypothesis that restricting the sample to employees results in an overestimate of the partial effect of schooling on earnings.

Estimates of labor’s share of self-employment income were made and some sensitivity tests were conducted. Persons whose earnings depend entirely on their own business were shown to have a much closer correspondence between self-employment income and the opportunity cost of their labor than is the case
for persons with both wage and business incomes. The procedure was also found to be satisfactory for explaining the earnings of self-employed persons with relatively low educational attainment. In short, the proposed technique is most useful for studying an important part of the labor force hitherto considered least amenable to empirical analysis.

Although the level of individual earnings is well explained by years of formal schooling and post-school experience, the analysis suggests that managerial skill may be an important explanatory variable, especially if earnings attributable to this type of skill are more likely to be a component of wages for high-level white-collar and technical employees than for ordinary hired labor. There is a close correspondence between wages and self employment-income for people at high education levels, whereas at lower education levels the self-employed earn more than wage-earners. The disproportionate concentration of low-education groups among the self-employed is further evidence of this relationship.

Estimates of labor's imputed share of self-employment income for various subgroups of Bangkok's labor force provide insights into the structure of labor markets in that city. Self-employment and wage-employment are alternative modes competing for the same labor; workers choose one or the other according to their relative earnings potential, and the market functions so as to equalize those potential earnings at the margin. It does not appear that differences in ownership of capital contribute much to an explanation of the income of self-employed persons. Indeed, the people with the highest imputed nonlabor incomes (mostly men with high educational attainment) also have wage incomes which are well explained by their labor characteristics and which are large even in comparison with their high nonlabor incomes.

References

World Bank reprints


No. 27. Efrain Friedmann, "Financing Energy in Developing Countries," *Energy Policy*


No. 29. V. V. Bhatt, "On Technology Policy and its Institutional Frame," *World Development*

No. 30. Bela Balassa and Ardy Stoutjesdijk, "Economic Integration among Developing Countries," *Journal of Common Market Studies* (also available in Spanish as published in *El Trimestre Económico*)


No. 32. Marcelo Selowsky, "A Note on Preschool-Age Investment in Human Capital in Developing Countries," *Economic Development and Cultural Change*


No. 34. Shlomo Reutlinger, "A Simulation Model for Evaluating Worldwide Buffer Stocks of Wheat," *American Journal of Agricultural Economics*

No. 35. John Simmons, "Retention of Cognitive Skills Acquired in Primary School," *Comparative Education Review*

No. 36. Montek S. Ahluwalia, "Inequality, Poverty, and Development," *Journal of Development Economics*


No. 43. Dipak Mazumdar, "The Urban Informal Sector," *World Development*

No. 44. Carmel Ullman Chiswick, "On Estimating Earnings Functions for LDCs," *Journal of Development Economics*