

Self Selection or Labour Market Segmentation? Evidence from Pakistan

Nadeem Ul Haque

Research Report Series # 4
1992

All rights reserved. No part of this paper may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or information storage and retrieval system, without prior written permission of the publisher.

A publication of the Sustainable Development Policy Institute (SDPI).

The opinions expressed in the papers are solely those of the authors, and publishing them does not in any way constitute an endorsement of the opinion by the SDPI.

Sustainable Development Policy Institute is an independent, non-profit research institute on sustainable development.



RR- 004- 002- 065- 1992- 024

© 1992 by the Sustainable Development Policy Institute

Mailing Address: PO Box 2342, Islamabad, Pakistan.
Telephone ++ (92-51) 278134, 278136, 277146, 270674-76
Fax ++(92-51) 278135, URL:www.sdpi.org

Table of Contents

Abstract	1
Introduction	1
Theoretical Framework	2
Estimation Issues	3
Some Hypothesis	6
Results	8
Conclusions	12
References	13
Appendix	15

The Sustainable Development Policy Institute is an independent, non-profit, non-government policy research institute, meant to provide expert advice to the government (at all levels), public interest and political organizations, and the mass media. It is administered by an independent Board of Governors.

Board of Governors:

Mr V. A. Jafarey
Chairman of the Board

Khalid Ahmed
Editor, AAJ KAL

Syed Babar Ali
Advisor, Packages Ltd.

Dr Tariq Banuri
Executive Director, SDPI

Dr Parvez Hassan
Senior Partner, Hassan & Hassan Advocates

Aban Marker Kabraji
Country Representative, IUCN-Pakistan

Sohaib Sultan Khan
Senior Advisor, NRSP

Dr G. M. Khattak
Co-ordinator, Sarhad Provincial Conservation Strategy

Abdul Rahim Mahsud
Regional Representative for West Asia, Nippon Gieken Associates

Dr Atta-ur-Rahman
Director, HEJ Research Institute, Karachi University

Dr Arshad Zaman
Former Chief Economist to the Government of Pakistan

In the Research Report Series, the SDPI publishes finished research papers written either by the regular staff of the Institute or affiliated researchers. These papers present research findings directly or indirectly related to sustainable and just development. These findings are meant to provide insights into economic, political, social and other issues in the context of sustainable development.

Self-Selection or Labour Market Segmentation? Evidence from Pakistan

Nadeem Ul Haque

Abstract

The paper examines the issue of labor market imperfections as reflected in wages in urban Pakistan. The empirical analysis does not find any market imperfections in the determination of wages as support for the hypothesis of self-selection according to comparative advantage. It is also determined that individual self-selection decisions are based on the criterion of expected gain in monthly income which is the product of total labor supplied and the wage rate, and not just relative wage gain. Consequently, both relative wage gain and relative labor supply gains are important in self-selection decisions. This finding would be consistent with behavior in a poorer society where the effort is to maintain a certain minimum level of income.

Introduction

Traditionally analysis of the distribution of earnings have been concerned primarily with the measurement and explanation of observed inequality in the personal earnings of individuals. Notable examples are Mincer (1974), Psacharopoulos (1977), and Psacharopoulos and Layard (1979). Such analysis have generally sought to estimate the returns to human capital as measured by schooling and work experience. Other individual characteristics, where available, were used mainly as control variables. As a consequence, issues of individual self-selection into occupations or into higher levels of education (see Willis and Rosen 1979) which in themselves affect the eventual distribution of earnings have not been explicitly analyzed. Furthermore, observations of the self-employed segment of the labor force have often been excluded from the analysis because of the difficulty of separating the wage component from that of the return to capital. Such a practice in the context of industrial economies where the self-employed are a very small proportion of the labor force may be considered to be somewhat defensible. However, in developing economies where, as has been noted in earlier studies such as Chiswick (1977), Haque (1977), and Wong (1981), the self-employed can constitute as much as up to a third of the labor force, basing the analysis of the distribution of earnings solely on the employed segment of the labor force could lead to serious biases. This would be especially true if the assignment of individuals to the employed or the self-employed work status was not random but according to some underlying choice mechanism. In that case, the work-status decision would be a choice variable in the analysis and estimation which does not account for this choice is likely to be mis-specified.

An explicit analysis of the endogeneity of the work-status decision, i.e., that between employment and self-employment, would not only lead to better model specification for education purposes, but could also help provide insights into the functioning of labor markets in Pakistan. In terms of the analysis, it would need to be taken into account that whereas individual choices would be based on intimate knowledge of expectations of returns in each market based on individual characteristics and market circumstances, the analyst would observe only the choice that has been made and some of the individual characteristics. In terms of possible hypotheses, the one most consistent with economic

theory would be that which is in keeping with maximizing behavior. Such behavior would imply that individuals are observed to choose the particular work status that yields the higher return. In the event that this hypothesis is not supported by the data, some institutional rigidities or labor market imperfections would be indicated. Thus, it could be that for equity considerations institutional wages have been depressed leading to the observation that the employed are unable to self-select according to comparative advantage. Alternatively, whereas the employed earn a wage that is consistent with comparative advantage the self-employed may be unable to gain fully from self-selection as relative to the employed sector, costs of entry for the self-employed sector are lower. In this case we would observe higher wages in the employed sector and lower wages in the self-employed sector which would tend to support the "segmentation" hypothesis of Mazumdar (1981). According to this hypothesis, insofar as the self-employed are representative of the "informal" sector, their earning should be, among other things, lower and characterized by greater inequality. Because of these lower and more unequally distributed earnings, individuals in the "informal" or self-employed sector would be seeking to enter the employed sector, but would be prevented from doing so.

An examination of labor markets for any existing imperfection is obviously of considerable importance to policy, as the existence of imperfections may require policy intervention. It is therefore important to both test for the existence of imperfections and how the imperfection manifests itself and affects individual welfare. The analysis here attempts to do this. Further detailed analysis to determine the factors that cause the imperfection will probably be required for policy purposes.

The rest of the paper is organized as follows: the next section sets out the theoretical framework for analyzing the problem of occupational choice; section 3 discusses the estimation issues involved; section 4 sets out some of the hypotheses that will be tested; section 5 discusses the results of the estimations; and finally a conclusion summarizes the main results of the paper.

Theoretical Framework

Haque (1986) estimated the rates of return to human capital across employment categories in the standard Becker-Mincer model. Those estimations, however, did not account for the differing individual abilities and backgrounds which may be important in determining the amount of education that is required. The standard assumptions on which our earlier discussion was based and on which much of the literature in the area is also based, are that individuals, given their innate ability, seek to maximize lifetime earnings. Since the skill requirements across occupations differ, an individual is faced with an array of expected lifetime earnings at the point of choosing his career. The choice that is made and one that is observed at the time of data evaluation is the one that maximized that individual's expected lifetime earnings given his skills and abilities. This line of reasoning, therefore, implies that occupational choice is itself an endogenous variable and that the estimation of earnings functions should take into account this endogeneity.

To illustrate this approach, consider that the costs and benefits of an occupation are randomly distributed among people according to their tastes, perceptions, expectations and an array of talents that affect performance in work activities. Some of these characteristics are observed while others are not. Individual sorting into occupational categories takes place according to the interaction of a selection criterion, such as maximum present value and the underlying joint distribution of individual characteristics. The selection rule partitions the underlying joint density into a corresponding realized occupation distribution.

If we represent potential lifetime earnings of person in occupation j by Y_{ij} and note that these earnings are a function of, on the one hand, vectors of observed and unobserved talent and ability variables represented by X_i and t_i , respectively, and on the other, vector of observed and unobserved background variables represented by B_i and b_i , respectively, then lifetime earnings can be expressed as follows:

$$Y_{ij} = Y_j(X_i, B_i, t_i, b_i) \quad j = 1, \dots, n \quad (1)$$

This expression summarizes for us how given the requirements of each occupation, different skill and talent combinations indicate a value for each occupation. The joint distribution of the unobserved components can be written as follows:

$$F(t, b) \quad (2)$$

The specification of a selection rate closes the model. Thus, if we maintain the selection rule of maximization of present value, the selection rule is:

$$i \text{ belongs to } j \text{ if } Y_{ij} = \max(Y_{i1}, \dots, Y_{in}) \quad (3)$$

Since observed assignments of individuals to occupations are selected on (X, B, t, b) , earnings observed in each class may be non-random samples of population potential earnings, because those with larger net benefits in a particular occupation have a higher probability of being observed in it.¹ Estimation should therefore take into account the self-selection at the individual level that according to this reasoning may be expected. It is to that issue that we now turn.

Estimation Issues

Operationalizing the model presented in the earlier section in the context of our two-choice (employment/self-employment) model involves taking explicit account of the endogeneity of the work-status choice variable. Alternatively, the basic idea can be stated as that of individual self-selection into either employment or self-employment on the basis of the marginal gain to the concerned individual that the individual has foregone as a result of his choice and foregoing the alternative. Writing wages or earnings Y_i ($i = E, S$) as the sum of the observed fixed effect M_i ($i = E, S$) and the unobserved random effect U_i ($i = E, S$), the earnings equation for employment denoted by E and self-employment denoted by S can be written as

$$Y = M_i + U_i \quad i = E, S \quad (4)$$

Constructing a new variable I , which measures relative gain of employment over self-employment, we have

$$I = Y_E - Y_S = M_E - M_S + U_E - U_S \quad (5)$$

1 The basic framework for such a model of occupational choice was first discussed by Roy (1951). An important application can be found in Willis and Rosen (1979) where a structural model of the demand for college education is derived from the theory of comparative advantage and estimated using statistical models of self-selection.

Individuals who gain from employment (i.e., $I > 0$) are likely to be observed in employment, while those who benefit by self-employment activity (i.e., $I < 0$) have a greater probability of being observed in that category. Consequently, our observation of individuals takes place after their self-selection into a particular work status choice based on their relative gain. The expected earnings of an individual in employment are thus conditioned on I being positive and are not equal to M_E .

$$E(Y_E | I > 0) = M_E + E(U_E | I > 0) \quad (6)$$

Similarly, the expected earnings of an individual in self-employment are conditioned on I being negative and are again not equal to M_S

$$E(Y_S | I < 0) = M_S + E(U_S | I < 0) \quad (7)$$

The expressions (6) and (7) serve to illustrate the difficulty with earnings function estimates from the two subsamples by standard least squares methods. The error terms no longer satisfy the classical assumptions since their distribution are truncated because of the selection criterion. The correct estimation procedure, therefore, is to take into account the qualitative dependent variable of occupational choice in a simultaneous equation framework along with the earnings functions. Assuming U_E and U_S to be jointly normally distributed with marginal probability distribution functions, $N(0, \sigma^2_E)$ and $N(0, \sigma^2_S)$, respectively, we can define a new variable $U = U_S - U_E$ and distributed as $N(0, \sigma^2)$, (6) and (7) can now be rewritten as:²

$$E(Y_E | I \geq 0) = M_E + \sigma_{EU}/\sigma^2 E(U | U \geq M_E - M_S/\sigma) \quad (8)$$

$$E(Y_S | I < 0) = M_S + \sigma_{SU}/\sigma^2 E(U | U > M_E - M_S/\sigma) \quad (9)$$

Expressions (8) and (9) point to an approach to consistent estimation of the structural parameters. It can be seen from these equations that the problem we face is one of specification bias which could be resolved if we knew the conditional means of U . As suggested by Heckman (1979) and as also discussed in Lee (1978), using our knowledge of the criterion function, instruments can be constructed for the conditional means that we need.³ Since individuals in either category of employment are identified, it is possible to estimate (5) by, say, the probit method. Thus, denoting M_E and M_S linear functions of individual characteristics X_E and X_S , respectively, we have:

$$M_E = X_E b_E; M_S = X_S b_S \quad (10)$$

Equation (5) can now be rewritten as

$$I = X_E b_E - X_S b_S + U_E - U_S = Z\alpha - U \quad (11)$$

Estimation of (11) by probit maximum likelihood methods gives consistent estimates of α (denoted by $\hat{\alpha}$). Consistent estimation of the conditional means of U in (8) and (9) with the functional forms specified in (10), can then be constructed using these estimates.⁴

2 σ_{iu} is the covariance of U_i ($i = E, S$) and U .

3 A similar statistical model and estimation technique was used by Lee (1978) to estimate union/non-union wage differentials, and Lee and Trost (1978) applied it to the problem of housing demand with choices of owning and renting. The model of self-selectivity was also used by Willis and Rosen (1979) in studying the decision to go to college.

4 The circumflex " $\hat{\alpha}$ " indicates an estimate.

$$E(U | \geq M_E - M_S/\sigma) = \lambda_E - f(Z\hat{a})/F(Z\hat{a}) \quad (12)$$

$$E(U | < M_E - M_S/\sigma) = \lambda_S = f(Z\hat{a})/1 - F(Z\hat{a}) \quad (13)$$

Here the $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions of the standard normal variable, respectively.

Equation (8) and (9) can now be written in estimable form as:

$$\ln Y_i = X_i b_i + c_i \lambda_i + V_i; c_i = \sigma_{iu}/\sigma^2; i = E, S \quad (14)$$

With the introduction of λ_i in each of the two equations, the V_i are now normally distributed with zero mean. Application of ordinary least squares would give us consistent estimates of the structural coefficients as well as the coefficients of λ_i . However, since the V_i are heteroscedastic, for efficient estimation a procedure such as generalized least squares should be used (Heckman 1979).

For estimation, the remaining issue is the appropriate definition of the relative gain, and the variables that define it in (5). If the individual criterion when choosing a particular employment status is to maximize the relative wage gain, then the Y_i ($i = E, S$) in the criterion function (5) is W_i ($i = E, S$). Another valid criterion, however, may be the maximization of relative monthly or yearly gain, especially if the individual choice to enter a particular sector enables him, based on his characteristics, to command not only a certain wage but also choose his hours such that gross income is affected. Self-employment probably allows a greater flexibility in the choice of hours since in employment an increase in the choice of hours probably depends on the availability of either a job that allows overtime or multiple jobs. If the criterion is relative gain in monthly incomes, an estimable form of the criterion function would be

$$I_j = \alpha (\ln W_{E_j} - \ln W_{S_j} - \ln W_{S_j}) + \beta (\ln h_{E_j} - \ln h_{S_j}) + \alpha Z_j + U_j \quad (15)$$

Where the W_j ($j = E, S$) is the wage rate and H_{ij} ($i = E, S$) is the hours worked in a year and Z_j is the set of exogenous variables. Using this approach at the second stage we would have a set of two simultaneous equations. A wage equation for each category as specified earlier and repeated below:⁵

$$\ln W_i = X_i^W \beta_i^W + \zeta_i \quad i = E, S \quad (16)$$

and an hours worked or labor supply equation for each category

$$\ln h_i = \eta_i \ln W_i + X_i^h \beta_i^h + \xi_i \quad i = E, S \quad (17)$$

Here the X_i^W and X_i^h and β^W and β^h are the set of individual characteristics and their respective coefficient vectors, in wage and labor supply equations, respectively.

Substituting (16) and (17) into (15) a reduced form probit equation can be derived for the estimation of the criterion function for the choice of employment. By constructing the selectivity variables as described above, the specification bias can be corrected for in (16) and (17). Since with the introduction of these instruments the error terms in (16) and (17) are heteroscedastic, generalized

5 The superscripts "w" and "h" indicate variables and coefficients particular to the wage and labor supply equation.

least squares will be used to estimate the two equations sequentially. An estimate of $\ln W_i$ from (16) will be used as an instrument in (17), to account for simultaneity between wages and labor supply.

With these estimated values of the parameters of (16) and (17), predicted values of the two wage rates W_E and W_S and the labor supply of the two employment categories can be used to estimate the structural probit equation (15). The advantage to doing this is that positive and significant values for both α and β would constitute an additional test of the joint test of comparative advantage and total income maximization. The hypothesis of self-selection according to comparative advantages would be borne out if only α were positive and significant or if both α and β were positive and significant. If α alone were positive and significant the hypothesis that individual choices of self-selection are based only on relative wage maximization, would appear to be supported by the data. In the event that both α and β were positive and significant self-selection would be based on relative income maximization.

Some Hypothesis

Using the selectivity model also allows us to test the following additional hypothesis relating to labor markets in Rawalpindi.

- a. The significance of the self-selection hypothesis can be tested by the level of significance of the coefficients (c_j) of the constructed selection variables (λ_j) in equation (14) above.
- b. The comparative advantage hypothesis, which states that individuals positively self-select would be validated by $c_E < 0$ and $c_S > 0$, since $\lambda_E < 0$ and $\lambda_S > 0$. Matching signs on the coefficients of the constructed selectivity variables imply that observed earnings after individual self-selection has taken place in each category is higher than say a notional average that the individual would have obtained had the assignment been random.⁶
- c. The labor market segmentation hypothesis which has been discussed in the development literature (Mazumdar 1981, 1983) can also be tested with appropriate signs for λ_E and λ_S . According to this hypothesis, it is possible that because of barriers to entry, a more desirable sector (from the point of view of the workers), and a less desirable sector can exist side by side. The protective barriers could be institutional such as unionization and government regulation, while benefits could take the form of fringe benefits, job security and paid vacations among other things. Among the major problems with empirically testing this hypothesis is that it is hard to identify the two sectors without prejudging the conclusion. Among the low income population, the employed/self-employed dichotomy is sometimes regarded as a close approximation to the dual sectors of the segmentation hypothesis.⁷ To the extent that our definition of these sectors approximates the rational one, we would be able to test for the hypothesis only as far as earnings are concerned by looking for $c_E < 0$ and $c_S \leq 0$.

6 This hypothesis was successfully tested by Willis and Rosen (1979) for individual self-selection into college education based on comparative advantage. As also noted by these authors, this approach allows for a rather eclectic view of talent and abilities, placing no restrictions on the observable and unobservable components of talent. This is in direct contrast to the one-factor ability-as-IQ specification which implies a hierarchical sorting of individuals, with an individual with a higher hierarchical rank dominating the one with a lower ranking in terms of the former's superior ability in all occupations. The comparative advantage noncomparable "absolute" ability ranking.

7 As we have already seen, the self-employed earn more per month than the employed with the former's incomes, as measured by the log variable, being more equally distributed. However, the employed take paid vacations, work shorter hours and therefore earn a higher hourly wage rate. The initial evidence on segmentation is therefore already mixed.

⁸The existence of barriers to entry to the employed sector implies that those in that sector can positively self-select, while those excluded from it are prevented from doing so. In the event, the self-employed may actually be doing worse than they would have done in random assignment because of, say, excessive competition in that sector, or because of the impediments in the way of individuals to truly exploit their comparative advantage.⁹

As mentioned in the previous section, estimation of the structural probit with estimated relative wage and hourly gains as independent variables would also enable us to see if the hypothesis of comparative advantage is supported by the data or not. To the extent that this hypothesis remains unverifiable, there would be an indication of some labor market rigidities. In that case, the segmentation hypothesis, though not confirmed, cannot be ruled out. If, on the other hand, self-selection is borne out by the data, barriers to entry though ruled out in terms of protection of incomes cannot be ruled for the purposes of job security, fringe benefits, etc.

An additional advantage of the estimation procedure that is used here is that at the first stage information is obtained on the determinants of the choice of employment. Some hypothesis of interest related to this choice are:

- a. Although the marginal rate of return to schooling is expected to be positive in both sectors, the estimated rate may vary by sector. It may be recalled that the comparison being made here is between employment in large enterprises or institutions such as the federal government on the one hand and small owner-operated enterprises on the other. It could be conjectured that the larger institutions might value education more than the small owner-operated outfits since the skill-requirements of the latter is perhaps rudimentary and/or best obtained in on-the-job training. For the larger establishments which employ most of the workers in the sample, size consideration and the resultant requirements for managerial control may necessitate a certain minimum amount of formal school education for employees with increasing returns for schooling beyond the minimum. According to this line of reasoning greater rates of return amongst the employed than the self-employed could be expected. If this were the case then individual expectations would take this into account leading those with larger amounts of education to seek employment rather than self-employment. Such reasoning leads to the hypothesis that the probability of employment would increase with additional years of schooling.
- b. As argued in point (a), education is likely to be positively associated with the probability of employment. Consequently, both own education and educational family background which enhances own educational achievements would also positively affect the choice towards employment.
- c. Experience of a particular sector would provide sector-specific skills and knowledge increasing thereby the earnings potential of the individual concerned. The costs of a move into the alternative sector would, thus, be increased as experience of a sector increases by at least the opportunity cost of doing so. This would imply that the probability of staying in a particular category would increase with the length of tenure in that category. On the other

8 Note that definitions of λE and λS rule out $cE > 0$ and $cS < 0$, i.e., negative self-selection for both sectors. Moreover, it is also implied that $cS - cE > 0$. These can be used as consistency checks for the model.

9 A word of caution is perhaps in order here. The test proposed here principally tests for self-selection. Segmentation can only in a very narrow sense of earnings be ratified here. However, since earnings may only be one of the set of attributes that make the employed sector more desirable, segmentation or the evidence of barriers between the two sectors cannot be entirely ruled out by the test proposed here. However, since incomes are a significant consideration in individual choice it is interesting to see if the segmentation hypothesis is borne out in its stronger form of higher incomes in the preferred sector.

- hand, as employment income peaks and stabilized, after a certain age and as an employee gets closer to his retirement age, there could be an incentive to invest in some form of self-employment activity. Pension income could then be combined with income from self-employed activity. Consequently, as work experience or age increases beyond a certain point, the probability of being observed in self-employment increases. In other words, it is quite likely that a concave age-probability of employment profile will be observed.¹⁰
- d. *Ceteris paribus*, increased availability of capital will increase the reservation wage for employment according to standard analysis. On the other hand, greater individual wealth makes available the fixed costs of entering self-employment. It could, therefore, be conjectured that great wealth would be positively correlated with the probability of employment.
 - e. Women might find employment to be the preferable course especially if a large enterprise or the governments were to provide a legal or institutional umbrella in an otherwise male-dominated society.

Results

The data upon which the analysis is based is derived from a survey of a random sample of households in Rawalpindi, the fifth largest city in Pakistan, with a population of 716,761 in 1977 at the time of the survey¹¹ and historically an important regional metropolis and administrative center. It was the nation's capital during the 1960s and today is only ten miles from the current capital, Islamabad. Though there is manufacturing activity in the city, the development of Islamabad on the outskirts of Rawalpindi has allowed an expansion of wholesale trade and construction. For a description of the data and some of its salient features as well as summary statistics of some of the variables used, see Appendix B.

The results of the probit which are presented in Table 1 reveal that most of the hypothesis that were identified in the last section are borne out.¹² Education has a significant and positive effect on the decision to seek employment.¹³ The same is true for education background as measured by father's education. Although in the early years of labor market experience individuals may prefer employment, with increased market experience chances of moving to self-employment grow larger. As captured by the convex probability of earnings work-experience profile, after eight years of work experience, the probability of moving to self-employment increases with each succeeding year.¹⁴ The length of tenure in the current job variable also tends to confirm this tendency among the older workers to move to the self-employed sector. The coefficient of the RENT variable, a proxy for wealth and capital, is significant and negative confirming our hypothesis. The FEMALE coefficient,

10 It may, however, be noted that an alternative and perhaps indistinguishable reason, as far as the data at hand is concerned, may be the fact that as opportunities expanded with economic growth, more employment opportunities opened up for the young. Maximizing behavior in each cohort was limited to a particular opportunity set, leading us to observe older workers in self-employment and younger workers in employment. Vintage effects in the skills of older workers could also lead to a similar observation.

11 The population figure is a projection of the 1972 census estimate at a 3.2 percent annual growth rate.

12 For the convenience of the reader all the variables used have been listed in Appendix A.

13 An alternative model is to use education as an endogenous variable that is determined by the exogenous variables in the model. The predicted value of education is then used in the regressions that have been presented. As the results of this procedure do not yield results that are substantially different, they are not presented here.

14 Using age instead of work experience in the probit indicated that the probability of being in self-employment increased with each year beyond age 31.

however, is negative but is insignificant indicating no structural differences in the two work-status choices in terms of opportunities for women.

Table 1: Maximum Likelihood Estimates of the Probability of Employment for Male Heads of Household

Variable	Coefficient	Marginal Probability at the Means	Asymptotic t ratio
Human Capital			
GRAD	.090	.034	12.102
YRSWKED	.007	.003	.706
YRSQ	-.0004	-.0002	-2.426
YRSATJOB	-.008	-.003	-2.436
Educational Background			
FCLAS	.017	.007	2.060
Sex			
FEMALE	-.075	-.029	-.616
Background			
URBAN	-.003	-.001	1.963
RENT	-.005	-.002	-2.408
Constant	-.021	-.008	-.248
-2 log L	-1283.9		

The result of the probit were used to construct the selectivity variables (λ_E and λ_S) for the estimation of the selectivity corrected wage and earnings equations.¹⁵ These results are presented in Table 2.¹⁶ In both the wage equations, selectivity, as judged by the significance of the selectivity variables, appears to be important. Moreover, since $C_E < 0$ and $C_S > 0$, the hypothesis of comparative advantage seems to be supported by the data. Apparently, individuals in Rawalpindi as a result of self-selection earn more, on an hourly basis, than a hypothetical mean which would have obtained if selection was random. Alternatively, both the observed self-employed and the employed are earning a higher wage rate than they would have received if it were not possible to self-select.¹⁷

Looking beyond the selectivity variables, it can be seen from the regression results that most of the coefficients are significant and of the right sign. Schooling continues to remain important for the employed category with the estimated returns to schooling indicated to be larger after the selectivity corrections. For the self-employed, however, despite the selectivity corrections, the coefficient of schooling remains small and insignificant. As expected female wage discrimination is stronger among the self-employed. The joint observation of an insignificant coefficient for the female variable in the wage equation and the negative and significant coefficient for the female dummy in the earnings equation for the employed is probably indicative of female labor supply decisions. Surprisingly, the

15 As indicated above the estimation procedure corrects for the heteroscedasticity that results from the use of the instrument for λ_i .

16 While no clear theory is available for determining and identifying restrictions, the nonlinearity of the selectivity variable provides us with an additional variable at the second stage of estimation. In our estimations, exogenous variables were also experimented with to enable an identifying variable to be determined which turned out to be URBAN.

17 In a similar analysis of heads of household only, both self-selection according to comparative advantage and segmentation were rejected (see Haque, (1984)).

coefficient of the job tenure variable, YRSATJOB, is insignificant in the employed category although in keeping with our hypothesis, it retains a significantly positive value in the self-employed and are hopefully isolating some of the effect of capital. As expected, the reporting of a second job is indicative of the individual being on the lower end of the wage scale.

As mentioned earlier, the greater flexibility that the self-employed have with respect to the choice of labor supply could help to reinforce the selectivity that has been observed in wage behavior when total earnings are being considered. However, in the results presented in Table 2, it can be observed that in the earnings equation estimations, the employed appear to positively self-select whereas the self-employed do not as the coefficients of the selectivity variables though of the right sign are not significant. It is possible that the larger observed variability in wages and hours worked among the self-employed is contributing to increasing the standard error of this estimate of the coefficient of the selectivity variable, λ_S . Moreover, the effects of selectivity on total earnings can be rewritten as:

$$dwh/d\lambda = dw/d\lambda * h + (dh/d\lambda + dh/dw dw/d\lambda)w \quad (18)$$

If the observed labor supply is characterized by a negative elasticity (i.e., a backward bending labor supply curve), and as observed there is positive self-selection in wage behavior the second term in the parentheses would be negative. Depending on how individual self-selection affects labor supply choices, the first term in the parentheses could be positive or negative. Consequently, the observed positive self-selection in wage behavior could be swamped out by the second term in (18), leading to an insignificant selectivity effect on earnings.

Table 2: Selectivity Corrected Wage and Earnings GLS Regressions for all Employed and Self-Employed Earners

Variable/Dependent Variable	Employed		Self-Employed	
	Log Wage	Log Earnings	Log Wage	Log Earnings
GRAD	.165 (10.797)	.131 (9.639)	.029 (.969)	.012 (.391)
FEMALE	-.004 (-.006)	-.171 (-3.410)	-.308 (-3.578)	-.756 (-8.586)
YRSWKED	.057 (12.660)	.054 (13.405)	.032 (6.048)	.040 (7.504)
YRSQ	-.001 (-9.750)	-.001 (10.244)	-.0005 (-3.800)	-.0007 (-5.684)
YRSATJOB	-.003 (-1.216)	-.003 (-1.52)	.012 (4.098)	.008 (2.838)
NWI	.0001 (1.583)	.0002 (2.051)	.0005 (2.972)	.0004 (2.478)
RENT	-.0002 (-1.516)	-.0001 (-.896)	.0009 (4.379)	.0008 (3.634)
SECJOB	-.045	-.117	-.106	-.276
λ_E	-1.045 (-5.762)	-.889 (-4.870)		
λ_S			.605 (1.979)	.379 (1.151)
INTERCEPT	-1.4274 (-5.762)	4.099 (18.678)	-.548 (-1.455)	5.161 (13.390)
R ²	.50	.44	.15	.21

Note: t ratios in parentheses

As can be seen from the labor supply estimations that have been presented in Table 3, the self-employed labor supply is backward bending and thus could be counteracting the observed positive wage self-selection in the earnings equation. Selectivity appears not to be an issue when it comes to labor supply. The greater flexibility with hours of work among the self-employed is reflected in the concave hours-experience profile that is observed in this group. The employed labor supply behavior probably because of institutional constraints on labor supply shows no sensitivity to either wage or experience. Females in both categories are strongly indicated to work fewer hours in the market.

Table 4 presents the results of the structural probit equation (17) above that was estimated using the predicted relative wage and labor supply gains from the earlier estimates of the wage and labor response functions. The most interesting result here is the joint verification of the hypothesis of self-selection based on individual comparative advantage and the hypothesis of income maximization as a criterion for individual choice. This can be seen from the positive and significant coefficients of GAINEMP (defined as the predicted value of the log of hourly wage rate for the concerned individual in employment divided by the predicted value of the log of the hourly wage rate for the concerned individual in employment divided by the predicted value of the log

of the hourly wage rate for the same individual in self-employment, both of which are obtained from the estimations in Table 2) and GAINHRS (defined as the predicted value of the log of annual hours worked for the concerned individual in employment divided by the predicted value of the log of annual hours worked for the same individual in self-employment which are obtained from the regressions presented in Table 3). Consequently, the data strongly support the conclusion that individual self-selection is according to the comparative advantage hypothesis. The segmentation hypothesis is, at the same time, not supported by the data.¹⁸ Additionally, we observe that it is on the basis of relative gain in total earnings that individual decisions are made, i.e. while relative wage gain is important to individuals in the population, the possibility of increased labor supply to a particular work status relative to the other is another important consideration. Such behavior would be consistent with both an individual desire for maintenance of a minimum level of income and a relatively low marginal valuation for leisure. In a poor LDC labor market both this desire for a minimum level of income and low marginal value for leisure may be expected, thus providing some intuitive basis for our result.

Table 4: Structural Probit Estimates of the Choice of Employment

Variable	Coefficient	Asymptotic t ratio
FEMALE	-.871	-.483
NWI	-.262	-.121
RENT	.0009	3.494
FCLAS	.012	1.318
MCLAS	.603	.165
GAINEMP	.358	7.231
GAINHRS	3.124	4.809
Constant	1.927	6.004
-2log likelihood	-1479.7	

Conclusions

In view of the empirical differences between the employed and the self-employed that were discovered in the analysis of Haque (1986a), and the possibility that the assignment of individuals to either work status may not be a random phenomenon, the analysis in this paper has rigorously tested for the endogeneity of the work-status decision. A model similar to Roy (1951), Willis and Rosen (1979), and Lee (1978) was adapted to the problem of the choice of work status. The model allowed the testing of two important hypotheses pertaining to urban labor markets in less developed countries. These were (1) the hypothesis of self-selection of individuals according to their comparative advantage which is associated with the three studies mentioned above and, (2) the hypothesis of labor market segmentation which is associated with Mazumdar (1981, 1983). The data provided strong support for the former hypothesis. It was therefore concluded that the choice of a work status was made on the basis of the relative gain that

18 It is perhaps worth emphasizing that though segmentation for the purposes of observed earnings has been ruled out, it has not been ruled out for other purposes such as job security and fringe benefits.

individual expected from the chosen work status. As a consequence, there was no evidence to suggest that there any barriers to entry or market imperfections affected earnings in the Rawalpindi labor market.

While in selecting their preferred work status, individuals gravitate towards the sector where they can command a higher wage, it was also indicated by the data that selection on the basis of choice of labor supply was also important. Alternatively, individuals prefer to be in the work status that allows them to maximize total monthly earnings and not just hourly wages. This finding would be quite consistent with individual behavior that seeks some level of subsistence income or a minimum level of income. This behavior, in turn, is quite likely to be observable in the kind of market that is under consideration where the relative poverty of the population might result in a relatively low individual marginal valuation for leisure. The conclusion that the individual criterion function for choice of work status is to maximize total earnings and not just the hourly wage rate is, therefore, justifiable.

In addition to these main conclusions that have just been noted, the analysis also provided evidence on many other issues. For Example, interesting evidence was also found on female labor market behavior. No evidence was found to indicate any significant work-status preferences for females in the sample. Alternatively, women showed no clear preference for whatever, if any, institutional safeguards that employment might provide. In terms of wage rates, however, women do tend to earn less than men though only in self-employment. Total earnings of women in both sectors were lower than for men in the respective work statuses. However, this was because of labor supply behavior as women were found to work significantly fewer hours than men.

Among the factors that positively affected the decision to choose employment over self-employment, the most important was increasing levels of schooling. It may be noted that this result is in keeping with the estimations of the rate of return to schooling where higher rates of return were estimated for the employed (see Haque, 1986a). Evidence was also found to indicate that with age the probability is increased of an individual being found in a state of self-employment as opposed to employment. Alternatively, the choice of employment-work experience profile is concave. The estimates indicated that after eight years of work experience, the probability of moving to self-employment is increased with each succeeding year.

References

- Chiswick, C. U. 1977. On Estimating Earnings Functions for LDCs. *Journal of Development Economics* 3 (March): 67-78.
- Haque, N. U. 1977. An Economic Analysis of Personal Earnings in Rawalpindi City. *Pakistan Development Review* 46 (Winter): 353-382.
- Haque, N. U. 1984. Work Status Choice and the Distribution of Family Earning. *Rand Paper Series* (November):7307.
- Haque, N. U. 1986. *The Determinants of the Distribution of Family Earnings: A Case Study of Rawalpindi, Pakistan*. Ph.D. dissertation. Chicago: University of Chicago.
- Haque, N. U. 1986a. The Rates of Return to Education in Rawalpindi, Pakistan. Chicago: University of Chicago.
- Heckman, J. J. 1979. Sample Selection Bias as a Specification Error." In *Female Labor Supply: Theory and Estimation*, ed. J. P. Smith. Princeton: Princeton University Press.
- Lee, L. F. 1987. Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *International Economic Review* 19 (June): 415-433.

- Lee, L. F., and Trost, R. P. 1978. Estimation of Limited Dependent Variables with Application to Housing Demand. *Journal of Econometrics* 8 (December): 357-382.
- Mazumdar, D. 1981. *The Urban Labor Market and Income Distribution: A Study of Malaysia*. London: Oxford University Press.
- Mincer, J. 1974. *Schooling Experience and Earnings*. New York: Columbia University Press.
- Psacharopoulos, G. 1977. Schooling Experience and Earnings: the case of an LDCs. *Journal of Development Economics* 4 (March): 39-48.
- Psacharopoulos, G. and Layard, R. 1979. Human Capital and Earnings: British Evidence and a Critique. *Review of Economic Studies* 46 (July): 485-503.
- Roy, A. D. 1954. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* n.s. 3 (June): 135-46.
- Willis, R. J. and Rosen, S. 1979. Education and Self-Selection. *Journal of Political Economy* 87 (October): S7-S36.
- Wong, Y. C. 1981. Family, Work Status and Earnings Distribution in Hong Kong. Ph.D. dissertation, Chicago: University of Chicago.

Appendix A

Definition of Variables

Variable	Definition
FCLAS	Number of completed years of head of household's father's education
FEMALE	1 if female; 0 otherwise
GAINEMP	Predicted value of the log of the hourly wage rate for the employed divided by the predicted value of the log of the hourly wage rate for the self-employed
GAINHRS	Predicted value of the log of annual hours worked for the employed divided by predicted value of the log of annual hours for the self-employed
GRAD	Years of completed education
NWI	Non-wage income
PREDLWAGE	Value of the log of wage rates predicted from the wage equation
RENT	Rental value of the house the family resides in, equals the actual rent if rented and the imputed rent if owned
SECJOB	1 if individual holds another job apart from principal employment; 0 otherwise
URBAN	1 if head of household spent childhood in urban environment; 0 otherwise
YRSATJOB	Total number of years in the current job
YRSWKED	Total number of years of experience
YRSQ	$(YRSWKED)^2$

Appendix B

A Description of the Data and Rawalpindi City

In August and September of 1977, the Pakistan Institute of Development Economics undertook a survey to collect socio-economic information on 2000 Rawalpindi households. The sample size and the restriction to one city are the result of a limited budget. In the two-stage sampling procedure, the household clusters first were randomly selected from a 400 cluster sampling frame (a cluster being a group of approximately 263 households). At the second stage a fixed proportion of households per cluster ($2000/16 = x$) were selected from the clusterwise address list of all structures and semi-structured dwelling. A fairly detailed questionnaire was developed, pretested, and revised.¹⁹ The revised questionnaire was completed for each household through an interview with the head of household or the oldest household member. Numerous checks were devised to ensure a proper implementation of the survey and accuracy of the information.

A description of the labor force and family characteristics of the data are set out in Table 1. Among the more striking features of the data are the extremely low level of female labor force participation (6.81 percent), and the large proportion of self-employed in the labor force (40 percent) and the significant proportion of non-nuclear families in the sample (36.7 percent). The first finding probably is explained by the population's cultural and religious norms. The proportion of self-employed seems high even when one takes into consideration that the sample is in an urban area of a developing economy. Compared with other samples from similar economies, the reported proportion of self-employed in Rawalpindi may be high. Wong (1981), for example, reports 16 percent self-employed in the total labor force in Hong Kong.²⁰

The sample reports a feature which has been observed quite commonly among developing economies--a large proportion of observed families reporting a non-nuclear or an extended structure. This information can be helpful to the analysis of the choice of family structure and the interaction of this choice and other forms of household decisionmaking, such as individual labor supply. In this view, valuable insights into family decisionmaking, an area of continuing interest (see Becker (1981)) in economics, can be gained. Given the low level of female labor force participation, about 38.5 percent of all earners are either children of the household, who have chosen gainful employment over the alternative of furthering their schooling, if any, or extended family members.

Summary statistics of some of the more important variables in the analysis that follows are presented in Table 5.²¹ As has already been observed, the data indicate that the labor market is characterized by a large proportion of the self-employed. As the employed/self-employed dichotomy is to play a considerable role in then analysis, it is being retained in this table. The figures in the table present a picture of the labor markets where in terms of gross monthly income the self-employed earn more than the employed. However, this gross income reflects the labor supply choices of the self-employed as they work

19 A major pretest for the 1977 Rawalpindi survey was the smaller 1975 survey of 1000 households (see Haque (1977)). Attempts were made in the 1977 survey to go back to the same 1000 households of the earlier survey and at the same time collect information on 1000 more.

20 A probable explanation is the large rural-urban migration into Rawalpindi from a population hinterland. Hong Kong does not have such an immigration.

A complete lost of all definitions of the variable that are used in the analysis below is presented in the appendix.

21 A complete lost of all definitions of the variable that are used in the analysis below is presented in the appendix.

considerably longer hours and earn a lower hourly wage rate. Individuals in this group take no holidays whereas those in the employed group take about 12 holidays a year. The self-employed appear to be an older group of people, indicating perhaps a cohort effect in work-status selection as a result say of greater employment opportunities available to the young as the economy expanded. Both the "wealth" proxy variables NWI and RENT indicate that the employed are a relatively richer group. Interestingly enough, female labor force participation is similar across the two groups.

Table 5: Summary Statistics of Some Important Variables

	Employed		Self-Employed	
	Mean	Standard Deviation	Mean	Standard Deviation
MGINC	527.78	345.89	579.84	485.41
ANHRS	2075.91	568.43	2830.05	662.58
WAGE	3.34	2.58	2.51	2.05
NWI	56.68	156.44	40.17	121.23
RENT	164.71	162.26	145.44	135.43
GRAD	7.53	4.74	3.84	4.18
YRSWKED	15.63	11.96	22.62	14.40
YRSATJOB	10.29	9.11	15.00	13.08
HOLI	41.97	31.80	2.19	9.07
SECJOB	.06	.24	.03	.18
FEMALE	.06	.24	.06	.24

Note: For definitions see Appendix A