REVIEW OF DEVELOPMENT ECONOMICS

Manuscript No: DE#3430, Acceptance date: November 21, 2004

**Human Capital, Productivity, and Stratification in Rural Pakistan***

*Takashi Kurosaki*

*Humayun Khan*

RRH: HUMAN CAPITAL AND PRODUCTIVITY IN PAKISTAN

LRH: Takashi Kurosaki and Humayun Khan

Abstract

This paper investigates the effects of human capital on productivity using micro panel data of rural households in the North-West Frontier Province, Pakistan, where a substantial job stratification is observed in terms of income and education. To clarify the mechanism underlying this stratification, the human capital effects are estimated for wages (individual level) and for self-employed activities (household level), and for farm and non-farm sectors. Estimation results show a clear contrast between farm and non-farm sectors — wages and productivity in non-farm activities rise with education at an increasing rate, whereas those in agriculture respond only to the primary education.

* Kurosaki: The Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8603 Japan. Tel: 81-42-580-8363, Fax: 81-42-580-8333, E-mail: kurosaki@ier.hit-u.ac.jp. Khan: The Institute of Development Studies, NWFP Agri-

cultural University, Peshawar, Pakistan. E-mail: humayun_ids@yahoo.com. We wish to thank an anonymous referee, Christopher Udry, T. N. Srinivasan, T. Paul Schultz, Martin Ravallion, Peter Lanjouw, Xavier Gine, Mir Kalan Shah, Anwar Hussain, Yasu Sawada, Nobu Fuwa, Kei Otsuka, and Katsuhide Isa for valuable comments. We acknowledge financial support from the Ministry of Foreign Affairs, Japan, and Seimei-Kai. The views reflected in this paper, however, are solely ours.

JEL Classification Numbers: O12, J24, Q12.

Number of Figures: 0. Number of Tables: 7.

Date: November 26, 2004

Address of Contact Author: Takashi Kurosaki, The Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8603 Japan. Tel: 81-42-580-8363, Fax: 81-42-580-8333, E-mail: kurosaki@ier.hit-u.ac.jp.

# 1 Introduction

In rural areas in contemporary developing countries, non-farm activities are becoming more important in determining the welfare of households (Lanjouw and Lanjouw, 2001). As a result, we often observe job stratification with a substantial income disparity between those who were successful in finding non-farm, lucrative jobs and those who were not. Underlying this stratification is a response of rural households in labor allocation to new economic opportunities, considering returns to human capital, which may differ from activity to activity. When farmers decide on their children's schooling, they are usually motivated by the desire of finding non-farm, lucrative jobs for their children. Therefore, investment in human capital in rural areas is more closely related with non-farm activities (Huffman, 1980; Yang, 1997; Fafchamps and Quisumbing, 1999; Lanjouw, 1999; Lanjouw and Lanjouw, 2001; Yang and An, 2002).

This paper is an empirical attempt to quantify the difference of returns to human capital across rural activities, using micro panel data of rural households in Pakistan's North-West Frontier Province (NWFP). The case of NWFP is particularly interesting because the weakness of economic development in South Asia is concentrated in this region — the incidence of income poverty is high and the deprivation in human development indicators is more serious than indicated by income growth. Another reason for studying NWFP economy is the general paucity of rigorous economic research on Pashtun society, which spreads over NWFP and Afghanistan.

The major contribution of this paper to the human capital literature in development economics is that a clear contrast between sectors and between employment statuses is shown through its comprehensive coverage of rural activities after controlling for endogenous selection. This paper is one of the few studies that apply the methodology of selection correction for polychotomous choice models to datasets from developing countries.[1] Since self-employment is important in developing countries (Newman and Gertler, 1994), this paper estimates comparable models of returns to human capital for four types of rural economic activities: non-agricultural wage/salary employment, agricultural wage employment, non-agricultural self-employment, and agricultural self-employment.[2] The empirical model is close to that of Yang (1997), who estimated non-linear production functions for farm value-added and linear wage functions for non-farm wage earnings. Unlike Yang (1997), however, this paper attempts to include non-farm enterprises and agricultural wages and to incorporate non-linear impacts corresponding to educational stages.

Another contribution of this paper is to give a clue to the controversy regarding the effects of education on farm productivity. Since Schultz (1961) emphasized the role of education in improving farm efficiency and in modernizing agriculture, microeconometric studies to test his hypothesis have been accumulated, showing mixed results from developing countries (Lockheed, Jamison, and Lau, 1980; Jamison and Lau, 1982; Yang, 1998). In the case of rural Pakistan, Fafchamps and Quisumbing (1999) found that private returns to education in farming are insignificant, whereas Kurosaki and Fafchamps (2002) demonstrated that the effects of schooling years on crop yields per acre are significantly positive. This paper gives one possible answer to this puzzle by allowing the effects of education to differ across different levels of education and at different aggregation levels of farm activities, in order to investigate whether education is important for efficient factor allocation within a farm.

# 2 Data and Key Features Identified in the Field

## 2.1 Data

This paper employs a panel dataset compiled from a sample household survey implemented in 1996 and 1999 in three villages in the Peshawar District of Pakistan's NWFP. NWFP is one of the four provinces of Pakistan. Compared with Punjab, which is the center of agriculture and related industries, and Sind, where the metropolitan city of Karachi is located, NWFP and Baluchistan are economically backward provinces. The incidence of income poverty (headcount index) in rural NWFP is estimated at 46.5% in 1998/99 (World Bank, 2002), which is the highest in Pakistan.

Since NWFP is a relatively land-scarce province with limited scope for agriculture-led sustained growth, human capital is expected to play an important role in poverty eradication. Yet, even in terms of human development, such as literacy rates and infant mortality rates, the province is lagging behind Sind and Punjab. This disparity, i.e., human development poverty being more serious than income/consumption poverty, is a notorious characteristic of South Asia as well as Pakistan, to which various issues of UNDP's *Human Development Reports* drew attention. This paper focuses on rural NWFP because this is a region where this disparity is stark.

Details of the dataset are given by Kurosaki and Khan (2004). In choosing sample villages in 1996, we controlled for village size, socio-historical background, and tenancy structure. At the same time, to ensure that the cross section data thus generated would provide dynamic implications, we carefully chose villages with different levels of economic development in terms of irrigation and transportation access. Table 1 summarizes characteristics of the sample villages. Village A is rainfed and is located some distance from main roads. This village serves as an example of the least developed villages. Village C is fully irrigated and is located close to a national highway, so serves as an example of the most developed villages. Village B is in between.

Out of 355 households surveyed in 1996, 304 were resurveyed in 1999, three of which had been divided into multiple households,[3] and 43 households were added as "replacement" samples. This paper, therefore, employs an unbalanced panel of 398 households, of which 301 are re-surveyed households without household division and 299 are those panel households with complete and comparable information.[4] Table 1 also shows characteristics of the panel households. Average household sizes are larger in Village A than in other two, reflecting the stronger prevalence of an extended family system. Average landholding sizes are also larger in Village A. Since the productivity of purely rainfed land is substantially lower than that of irrigated land, effective landholding sizes are comparable among the three villages. As is shown in the average household income or consumption per capita, the living standard is the lowest in Village A and the highest in Village C.[5]

## 2.2 Labor Force Allocation and Human Capital

Table 2 shows the distribution of working household members by their employment status. From those household members whose age is 15 years and above, students, retired people, and the unemployed are excluded, resulting in the total number of working members at 1,591 for the total 355 households in 1996 and 1,606 for the total 352 households in 1999.[6] Based on each individual's primary occupation, the table classifies the employment status into five categories: household work, non-agricultural wage/salary employment, agricultural wage employment, non-agricultural self-employment, and agricultural self-employment.

Agriculture is traditionally the most important source of employment in the study region. Because there are few large scale farms that are completely dependent on hired labor, most of those engaged in agriculture are self-employed. Their labor is sometimes supplemented by hired labor. Non-agricultural self-employment activities, or non-farm enterprises, are diverse: traditional, caste-based services such as carpenters, barbers, and blacksmiths (approximately 13% of the individuals self-employed in non-farm enterprises); low-capital, low-end jobs such as snack hawkers and shoe polishers (15%); those that require relatively large initial capital such as arms trading, general shops, wheat mills, and nursery shops (57%); and transportation service (15%), which includes all three types. Non-agricultural wage/salary employment are also diverse, including daily construction work, wage employment in those listed as non-agricultural self-employment activities, and office/shop work in the nearby towns. Since the size of establishments is universally small for those employees, we may classify them according to their contract duration — approximately 55% of the non-farm employees were hired casually on daily basis, while the rest were hired regularly.

Among males, employment in non-agriculture and self-employment in agriculture are more frequently found than the other two. The concentration of female workers on the category "household work" reflects the effects of *purdah*, the custom of social seclusion of women in South Asia, which is reinforced by Pashtun codes of maintaining family honor (Ahmed, 1980). Table 2 shows that there were only 15 cases of females employed by others for non-farm work, mostly in low-paid sectors. Because of this distortion, the following analysis focuses only on male labor allocation and the effects of human capital on it.

Panel B of Table 2 shows the level and composition of household income corresponding to the labor allocation in Panel A. The average household income declined more than consumption, indicating that households have *ex post* measures to cope with income risk (Kurosaki, forthcoming). The income composition shares show that the earning from non-agricultural employment is the most important one, followed by self-employment in agriculture and self-employment in non-agriculture. Therefore, the average income per worker in non-farm self-employment is highest, followed by that in wage employment in non-agriculture. The average income per worker in agriculture, whether it is self-employment or a wage job, is much lower than those in non-agriculture, suggesting a job stratification with a substantial income disparity. Then what determines the job stratification among these four activities?

This paper attributes the answer to a difference in returns to human capital in rural economic activities. Table 3 shows age and educational achievement for the same working males described in Table 2. The self-employed are older than employees and those working in agriculture are older than those in non-agriculture. The education level is indeed low — the average schooling years was 3.7 in 1996 and 4.0 in 1999; literacy rate was 43% in 1996 and 48% in 1999.[7] The difference in educational achievement is more significant between sectors than between employment statuses — those engaged in non-agriculture are generally more educated than those engaged in agriculture.

# 3 Empirical Specification

## 3.1 Labor Allocation

To investigate whether or not the association between higher earnings per worker and higher education levels in rural non-agricultural activities can be explained by a difference in returns to human capital, this section proposes empirical models that are comparable between the four rural activities and control for endogenous selection of the activities. Efficient allocation of household labor force requires that the factor be allocated based on a comparative advantage principle (Kurosaki and Khan, 2004). For example if the household's objective is to maximize expected income, when a household member can earn more as a non-farm employee than in self-employed farming or than in household work, the household allocates him/her to the non-farm employment even if he/she is a better farmer than other household members. If the household's objective is to maximize expected utility incorporating labor-leisure choice and risk aversion, the comparative advantage should be adjusted based on subjective equilibrium prices, which could diverge from the market returns to labor of each family member. With additional assumption that the household utility associated with allocating individual $i$ to activity $j$ has a non-stochastic component and a stochastic term with extreme-value distribution, the labor allocation can be characterized by a multinomial logit model (McFadden, 1974). We specify the multinomial logit model as

$$Prob(z_{it} = j) = \frac{\exp(X_{it}\gamma_{j1} + X_{ht}\gamma_{j2})}{\sum_{k=0,...,4} \exp(X_{it}\gamma_{k1} + X_{ht}\gamma_{k2})}, \quad j = 0, ..., 4, \tag{1}$$

and estimate it in the first stage of our empirical analysis, where $z_{it}$ is an indicator variable denoting the choice for individual $i$ in household $h$ with respect to $j$ in year $t$, $X_{it}$ is a vector of individual attributes such as education and age, $X_{ht}$ is a vector of household attributes such as household wealth and production assets, and $\gamma_{j1}$ and $\gamma_{j2}$ are vectors of coefficients to be estimated, associated with choice $j$ (household work = 0, non-agricultural wage employee = 1, agricultural wage employee = 2, non-agricultural self-employed = 3, and agricultural self-employed = 4).[8]

The multinomial logit model can be estimated by a maximum likelihood method. Then, the fitted probability of individual $i$ working in $j$, $\hat{Prob}(z_{it} = j)$ is given by expression (1) with $\gamma_{j1}$ and $\gamma_{j2}$ replaced by their estimates $\hat{\gamma}_{j1}$ and $\hat{\gamma}_{j2}$. Similarly, the fitted probability of household $h$ with its member(s) working in $j$ is given by

$$P\hat{r}ob(z_{ht} = j) = 1 - \prod_{i \in h} \frac{\sum_{k \neq j} \exp(X_{it}\hat{\gamma}_{k1} + X_{ht}\hat{\gamma}_{k2})}{\sum_{k} \exp(X_{it}\hat{\gamma}_{k1} + X_{ht}\hat{\gamma}_{k2})}. \qquad (2)$$

These fitted values are used to calculate selection terms in the second-stage estimation.

## 3.2 Determinants of Wage

Assuming wage labor markets to be exogenous to household decisions, the unit wage becomes a function of the human capital of the employee, $X_{it}$. To capture this idea, a standard Mincer equation is estimated in which $\ln W_{ijt}$ is regressed on $X_{it}$, where $W_{ijt}$ is the wage level of individual $i$ working in activity $j$ (=1, 2) in year $t$.

Two econometric issues are addressed in this paper. The first is sample selection. Because $W_{ijt}$ is observed only when individual $i$ works in $j = 1$ or 2, an error term to the Mincer equation conditional on this selection has non-zero mean. To control for this, a two-stage procedure is adopted in which a correction term $\hat{\lambda}_{ijt}$ compiled from estimation results of equation (1) is added as an additional regressor. Assuming that the error term to the wage equation is distributed normally, we adopt the correction term based on the general transformation of error terms to normality (Lee, 1983), because it facilitates a feasible computation of a selection term for the household-level regression in the next subsection.[9]

Another econometric issue is unobserved characteristics that affect wages received by those who work in the wage sector. An example is worker's ability that is known to the household but not observable to the econometrician. To minimize the bias from omitting these unobservables, a household specific effect, $\alpha_h$, is added to the wage regression.[10] It also controls for the possibility of segmented labor markets. The wage function is thus specified as

$$\ln W_{ijt} = X_{it}\beta_j + \rho_j \hat{\lambda}_{ijt} + \alpha_{hj} + \epsilon_{ijt}, \quad j = 1, 2, \tag{3}$$

where $\beta_j$ is a vector of coefficients, which represent returns to human capital for an activity $j$, $\rho_j$ controls for the selectivity bias, and $\epsilon_{ijt}$ is a zero mean error term.

## 3.3 Productivity in Self-Employment Activities

Unlike wage work, marginal returns to labor are unobservable for self-employment activities. We thus estimate production functions for value-added, as was adopted by Yang (1997).[11] Let $q_{hjt}$ denote the value-added from self-employment activity $j$ (= 3, 4) for household $h$ in year $t$. A Cobb-Douglas production function is assumed with two primary factors of production — the total labor input by household $h$ into activity $j$, denoted by $L_{hjt}$, and the total capital input (non-agriculture) or the total land input (agriculture) denoted by $H_{hjt}$. Each household is used as a unit of analysis and the natural log of value-added is used as a dependent variable.

Three econometric issues are addressed. The first is sample selection. To control for the fact that $q_{hjt}$ is observed only when household $h$ is involved in $j = 3$ or 4, Lee's (1983) general transformation of error terms to normality is adopted, as in the case of wage functions.[12] The second is a potential correlation between the error terms to the dependent variables on the one hand and right-hand-side variables on the other hand, which is likely to be serious for factor inputs, especially labor inputs. The correlation could occur when the right-hand-side variables are endogenous to household decisions even in the short run. Another reason for the potential correlation is measurement errors. To control for these problems, instruments are used for factor inputs and some other right-hand-side variables. The third is unobserved characteristics that affect the productivity of enterprises, such as land quality and inherent managerial ability of households in running enterprises. To minimize the bias from omitting these unobservables, a household specific effect, $\alpha_h$, is added to the value-added functions.

Therefore, the empirical model for self-employment is specified as

$$\ln q_{hjt} = b_{j0} + b_{j1} \ln L_{hjt} + b_{j2} \ln H_{hjt} + X_{hjt} c_j + \rho_j \hat{\lambda}_{hjt} + \alpha_{hj} + \epsilon_{hjt}, \quad j = 3, 4, \quad (4)$$

where $X_{hjt}$ is a vector of household $h$'s characteristics that affect productivity of activity $j$, such as household human capital (education, experience, etc.) and production/market environment, and $\epsilon_{hjt}$ is an i.i.d. error term. Parameters to be estimated are $b_0$, $b_1$, $b_2$, $\rho$, and vector $c$. It is assumed that various types of labor input are perfectly substitutable but the additive weights are different by its type, reflecting different productivity (Fafchamps and Quisumbing, 1999). Parameter vector $c$ is expected to capture these effects.

Theoretically, there are several routes through which human capital may affect productivity. The first route is its effects on the efficiency of labor input. If, for example, a literate laborer will be able to follow the instruction of a labor task more precisely, what matters to production is not the amount of hours of labor $L_{hjt}$ but the amount adjusted for its quality. Second, the accumulation of human capital might improve overall technical efficiency in production. Third, the accumulation of human capital might improve allocative efficiency at the household level.[13] We can investigate whether or not the third factor is important by estimating agricultural value-added functions at different aggregation levels. If the effects of education on the farm-level value-added are larger than those on value-added of individual crops, the difference could be attributable to educated farmers' superiority in allocating factors across crops.

# 4 Estimation Results

## 4.1 Determinants of Labor Allocation

Table 4 reports estimation results for the first-stage multinomial logit model (1). Variables in vector $X_{it}$ (individual characteristics that affect his/her productivity and market wage) include age, age squared, and educational achievement dummies, corresponding to five, eight, and more years of completed education.[14] Age and age squared are included to capture non-linear effects of experiences. The marginal effects of education dummies suggest a pattern with accelerating probability of joining non-farm wage markets at the cost of farm self-employment as the education level goes up. The effects of age show an inverted U shape for farm and non-farm wage employment and an U shape for farm self-employment.

The marginal effects of $X_{ht}$ show that households with more adult males and less dependent members are more likely to send their labor force to outside employment. Landed households are more likely to send their labor force to their own farms. These results imply that the necessity of family labor on family farms is an important determinant for a household to send its members to non-agricultural wage jobs.

## 4.2 Effects of Human Capital on Wages

With the sample selection term obtained from the results above, the second-stage wage equation (3) is estimated with an intercept dummy added for the second survey to control for macro shocks.[15] Estimation results in Table 5 show that there are significantly positive effects of education on the non-agricultural wage level.[16] A worker with primary education is expected to be paid 17% ($\approx e^{0.154} - 1$) higher than a non-literate worker (reference group); with middle school education, 31% higher;

and with high and higher school education, 64% higher. These parameters imply the following Mincerian rates of returns: 3.1% for education up to the primary level, 3.4% for education up to the middle level, and 4.4% for education up to the secondary and higher level; or 3.9% for additional middle education after primary education and 5.8% for additional higher education after middle education. This range is consistent with the estimates in earlier studies on the returns to schooling in rural non-farm activities in Pakistan (Fafchamps and Quisumbing, 1999; Alderman et al., 1996).

Since non-farm wage employment is diverse, distinguishing various types with more disaggregation could be important. Therefore, we extended the model in (1) by distinguishing those hired casually and regularly. The dummy variable for the regularly hired is significantly positive, indicating that their wages were on average 60% ($\approx e^{0.474} - 1$) higher than those for the casually hired.[17] The coefficients on education with the employment type dummy are much smaller than those without it. This is because more educated individuals are more likely to work regularly. Education thus not only increases the wage in non-agriculture but also increases the probability of working in non-agricultural activities with higher and stable payment.

Among other human capital variables, age as a proxy for job experience shows an inverted U-shape, suggesting that productivity in non-agricultural wage work responds positively with experience but at a diminishing rate. The selection term is significantly positive in all models, indicating a positive selection. Individuals whose propensity to be employed in non-agrucluture is high are expected to earn more even after controlling for the direct effects of their individual attributes on wages.

Table 5 also reports estimation results for agricultural wage earners. In sharp contrast to results for non-agriculture, only the coefficient on primary education dummy is significant with about 3.8% Mincerian returns. Education higher than the primary level does not seem to contribute to higher agricultural wages. Age and age squared show an inverted U-shape but the coefficients are smaller than those for non-agriculture. The non-response of farm wages to higher education reflects the nature of the farm labor market in the study region. Most of the workers are hired for unskilled, manual work on the farm such as weeding, harvesting, and transporting. It is no wonder that job experiences or education do not contribute much to improvement in productivity of such works. The selection term is not statistically significant.

## 4.3 Effects of Human Capital on Productivity of Enterprises

Production function (4) is estimated for both farm and non-farm self-employment (Table 6). Labor input is measured by the monetary sum of wages actually paid to hired workers and imputed wages for family workers using the same wages or village average wages imputed at daily basis. The second production factor, capital input for non-farm enterprises, is defined as the total capital used in production, approximated by the machinery/equipment depreciation and land rents for the non-farm enterprise. The second factor is land input for farm production, measured by the wheat-cropped area or the total farm area. Since the two factors of production are determined endogenously by the household and they are also likely to suffer from measurement errors, they are replaced by their fitted values using instrumental variables.

Estimation results for non-agricultural enterprises show that the coefficients on both of the production factors are statistically significant. Elasticities of production with respect to the two production factors are estimated in a reasonable range, with their sum around 0.83, indicating slightly decreasing returns to scale.

To capture the effects of education, several variables are available.[18] Because of the small sample size and high collinearity among these variables, simultaneous inclusion of these variables did not work well. Our results show that the average education among those household members who are engaged in the non-farm business performed marginally better than other specifications in terms of adjusted $R^2$. This could be due to the fact that the number of those engaged in non-farm business within a household is not large and they do not always include the household head and the individual with the highest education. The coefficients on educational stage dummies show significantly positive effects with higher reward for higher education. This is similar to the results for non-farm wages but the difference among educational stages is larger for non-farm enterprises than for wages. Coefficients on the age of the head and its quadratic term show an U-shape, but only the quadratic term is significant. This seems to suggest that experience improves productivity of non-farm enterprises at an increasing rate. The coefficient on the sample selection term is close to zero and not statistically significant, suggesting that errors in labor allocation decisions and those in value-added functions are not strongly correlated.

Since non-farm enterprises are diverse, distinguishing various types of non-farm activities with more disaggregation, e.g., low-end type jobs like hawkers and high-end type jobs like wheat mill owners, could be important (Lanjouw, 1999; Lanjouw and Lanjouw, 2001). Considering the limited number of observations, a dummy variable for those self-employed activities that are carried out in a permanent business space (for example, shop space or workshop space) is included. Since the dummy variable could be endogenous, the selection term was re-estimated by extending the model in (1) by distinguishing these two types of non-farm enterprises. The dummy variable has a significantly positive coefficient, indicating that those enterprises with business spaces are likely to belong to the high-end type jobs. The coefficients on education with the business type dummy are much smaller than those without it, implying that education increases not only productivity in non-farm self-employment but also the probability of having high-end type enterprises.

Table 6 also reports results for the farm value-added, either from wheat (the regions's staple food) or from all crops combined. The vector $X_{hjt}$ in equation (4) now includes production/market environment variables —irrigation (the share of farmland under irrigation) and sharecropping ratios (the share of sharecropped land in cultivating wheat or at the farm level). In the 2SLS estimation, the two production factors and the sharecropping ratio are replaced by their fitted values. The coefficients on production factors and irrigation are statistically significant with expected signs and reasonable magnitudes. Unexpectedly, the sharecropping ratio has a positive effect but it is significant only in Village A, after deleting insignificant cross terms with village dummies. This seems to suggest that sharecropping contracts are associated with superior access to capital for tenant farmers through landlords in Village A, where financial institutions are the least developed. The effects of irrigation and sharecropping ratio are stronger for the farm-level value-added than for wheat value-added, because crops competing with wheat are more irrigation sensitive and capital intensive than wheat. The coefficient on the sample selection term is not statistically significant.

Regarding the household education variables, the average education among those household members who are engaged in farming performed the best in terms of adjusted $R^2$.[19] In sharp contrast to results for non-farm enterprises, none of the coefficients on education dummies are significant for wheat and two of the education dummies have significant coefficients but with similar magnitudes for the farm-level value-added. The null hypothesis that the coefficients on the three dummies are the same was not rejected at 10% level. Therefore, acceleration of returns to education is not observed in agricultural self-employment. Having additional years of education beyond the primary or middle levels does not seem to contribute to higher farm productivity. When the three stages are merged, the impact of the average literacy of family farm labor is statistically significant at 1% for the farm-level output and its magnitude is much higher than the case for wheat (Kurosaki and Khan, 2004).

When value-added functions were estimated for individual crops other than wheat, the coefficients on education were not significant, possibly due to the small sample size. When they were estimated for non-wheat crops combined, the coefficients were similar to or smaller than those shown in Table 6. Our field observations suggest that gains in efficiency units of labor or in technical efficiency due to education in each cultivation cycle are small, if any. Therefore, we interpret that the larger effect of education at the farm level suggests that educated farmers are more able to allocate quasi-fixed inputs efficiently among different sub-sectors (Yang and An, 2002).

Our finding that the additional gain from education higher than the primary level is not large in farming seems to contradict the existing literature on technical efficiency in Pakistan (Hussain, 1989; Ahmad, Chaudhry, and Iqbal, 2002), which argued that most of the progressive farmers adopting superior technology have education higher than the primary or middle levels. We interpret our results as showing that the main contribution of education to farm value-added comes from a more efficient crop choice. In order to be sensitive to market returns, a jump from no education to formal, primary education may matter more than a marginal gain from schooling above the primary or middle levels. In other words, farmers who have primary or higher education can behave in a more market-oriented way than those who have never attended schools.

The results, therefore, shed new light on the controversy on the effects of education on farm productivity (Lockheed, Jamison, and Lau, 1980; Jamison and Lau, 1982; Yang, 1998). First, its effects are likely to be non-linear. Our results suggest a possibility that in farm production, a jump from no education to literacy matters the most. If this is the case, applying a model that includes only a linear term of schooling years may result in the insignificance of education. Second, its effects are likely to differ at different levels of aggregating farm output. Our results suggest that at a higher level of aggregation, the effects of education can be depicted more distinctly, possibly due to the superiority of educated farmers in allocating factors efficiently.

## 4.4 Job Stratification and Returns to Labor

The previous subsections have demonstrated the contrast between the response to higher education of farm returns and that of non-farm productivity — the farm returns are the most sensitive to the literacy whereas the non-farm labor markets remunerate higher education with a higher wage.[20] Because of this reason and the diminishing return to labor in self-employment on the farm, which is captured by a coefficient on the labor input significantly smaller than unity in Table 6, we expect that more educated households have more diversified labor force, spanning a number of non-farm activities. Then how much can these differences in labor returns alone explain the observed allocation of labor force? To examine this question, we simulate labor force allocation predicted by the results in Tables 5-6 but ignoring selection terms.

In the simulation, we would like to allocate individual $i$ in household $h$ to sector $j$ where his marginal labor return is the highest. Let $f_{hj}(L_{ij})$ be his net-return-to-labor function. For wage sectors, we assume that $\ln(\partial f_{hjt}/\partial L_{ijt}) = \ln W_{ijt}$. We thus calculate a fitted value or out-of-sample forecast value from estimation results of equation (3) for the simulation, namely,

$$\hat{\ln}(\partial f_{hjt}/\partial L_{ijt}) \equiv X_{it}\hat{\beta}_j, \quad j = 1, 2. \tag{5}$$

For self-employment, what we have estimated is $\ln q_{hjt}$, the value-added from household $h$'s activity $j$. Based on the approximation $\partial f_{hjt}/\partial L_{ijt} \approx \partial q_{hjt}/\partial L_{hjt} = b_{j1} q_{hjt}/L_{hjt}$, where $b_{j1}$ is a coefficient on the log of labor in equation (4), we calculate

$$\hat{\ln}(\partial f_{hjt}/\partial L_{ijt}) \equiv \ln \hat{b}_{j1} + \hat{b}_{j0} + (\hat{b}_{j1} - 1) \ln L_{hjt} + \hat{b}_{j2} \ln H_{hjt} + X_{hjt}\hat{c}_j, \quad j = 3, 4, \quad \forall i \in h. \tag{6}$$

This value is calculated only for those individuals belonging to a household, where $L_{hjt}$, $H_{hjt}$, and $X_{hjt}$ are available, i.e., a household with self-employment activities.

We thereby obtain $\hat{\ln}(\partial f_{jt}/\partial L_{ijt})$, for each individual $i$ in year $t$, where $j = 1$ (non-agricultural wage), 2 (agricultural wage), 3 (self-employment in non-agriculture), and 4 (self-employment in agriculture). Then each individual is assigned a "predicted" job whose $\hat{\ln}(\partial f_{jt}/\partial L_{ijt})$ is the highest among the four activities.

Predicted patterns of labor allocation are summarized in Table 7. Diagonal cells show the number of correctly predicted individuals. Among 1,612 males engaged in one of the four sectors, 896 or 55.6% are predicted correctly, which is a reasonably high percentage as a whole, considering that a substantial part of the information included in household attributes $X_{ht}$ used in estimating the multinomial logit model (1) is ignored (the multinomial logit results in Table 4 predict labor allocation correctly for 990 or 61.4% of the same individuals). The reasonably good performance of the simulation in Table 7 implies that the difference in individuals' productivity due to different education levels underlies the job stratification with a substantial income disparity. The stratification is likely to be re-produced over generations because credit and insurance markets in the study region are very incomplete (Kurosaki, forthcoming; Kurosaki and Khan, 2004).

Predictions for agricultural wage jobs are less precise. This could be attributable to a social stigma associated with agricultural wage employment as a primary job. In the study region, full time farm laborers are found only among those households belonging to the lowest social rank. Incorrect prediction for several individuals in Table 7 could also be attributable to household risk aversion. Agriculture is risky, especially rainfed farming in Village A (Kurosaki, forthcoming).

# 5 Conclusions

This paper investigated the effects of human capital on farm and non-farm productivity using micro panel data of rural households in NWFP, Pakistan, where a substantial job stratification is observed in terms of income and education. To clarify the mechanism underlying this stratification, the human capital effects are estimated both for wages (individual level) and for self-employed activities (household level) on the one hand and both for farm and non-farm sectors on the other hand.

Estimation results of returns-to-labor regression models can be summarized as follows. First, private returns to education are significantly positive in non-farm wages for males, which increase with education at an increasing rate. Second, the effects of human capital are weak on agricultural wages. Third, the effects of education on non-farm enterprise productivity are positive with acceleration in reward, as in the case for non-agricultural wages. Fourth, the effects of primary education on crop productivity are positive but the additional gain from higher education is small. Fifth, the effects of education on crop productivity are more significant at more aggregate levels in farm production, possibly reflecting the efficiency of educated farmers in factor allocation. The non-linearity and aggregation issues regarding the effects of education could be one of the reasons for the mixed results in the literature on the effects of education on farm productivity in developing countries.

Thus a clear contrast was shown between farm and non-farm sectors — wages and productivity in non-farm activities rise with education at an increasing rate, whereas those in agriculture respond only to the primary education. The contrast implies that more educated household members have comparative advantages in non-farming. This implication was confirmed by comparing observed labor force allocation with simulated labor force allocation predicted by the difference in labor returns. In other words, the difference in individuals' comparative advantages due to different education levels underlies the job stratification, which is likely to be re-produced over generations under imperfect credit markets in the study region.

The findings of this paper could justify a policy to give high priority to primary education in rural Pakistan, because the provision of quality primary education has efficiency enhancing effects on various rural activities. Since the private returns to higher education are sufficiently high for males in non-farm sectors, the priority of public intervention into these levels might be lower than the case for primary education.

# References

Ahmad, Munir, Ghulam Mustafa Chaudhry, and Mohammad Iqbal, "Wheat Productivity, Efficiency and Sustainability: A Stochastic Production Frontier Analysis," *Pakistan Development Review* 41 (2002): 643-63.

Ahmed, Akbal S., *Pukhtun Economy and Society: Traditional Structure and Economic Development in a Tribal Society*, London: Routledge and Kegan Paul, 1980.

Alderman, Harold, Jere R. Behrman, David R. Ross, and Richard Sabot, "The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market," *Oxford Bulletin of Economics and Statistics* 58 (1996): 29-55.

Alderman, Harold, Jere R. Behrman, Victor Lavy, and Rekha Menon, "Child Health and School Enrollment: A Longitudinal Analysis," *Journal of Human Resources* 36 (2001): 185-205.

Card, David, "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card (ed.), *Handbook of Labor Economics, Volume 3*, Amsterdam: North Holland, 1999.

Fafchamps, Marcel and Agnes R. Quisumbing, "Human Capital, Productivity, and Labor Allocation in Rural Pakistan," *Journal of Human Resources* 34 (1999): 369-406.

Glewwe, Paul and Hanan G. Jacoby, "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana," *Journal of Human Resources* 29 (1994): 843-64.

Hoodbhoy, Pervez (ed.), *Education and the State: Fifty Years of Pakistan*, Karachi: Oxford University Press, 1998.

Huffman, Wallace E., "Farm and Off-Farm Work Decisions: The Role of Human

Capital," *Review of Economics and Statistics* 62 (1980): 14-23.

Hussain, S., "Analysis of Economic Efficiency in Northern Pakistan: Estimation, Cases and Policy Implication," Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1989.

Jamison, Dean T. and Lawrence J. Lau, *Farmer Education and Farm Efficiency*, Baltimore: Johns Hopkins University Press, 1982.

Jolliffe, Dean, "Whose Education Matters in the Determination of Household Income? Evidence from a Developing Country," *Economic Development and Cultural Change* 50 (2002): 287-312.

Kurosaki, Takashi, "Consumption Vulnerability to Risk in Rural Pakistan," *Journal of Development Studies* (forthcoming).

—— and Marcel Fafchamps, "Insurance Market Efficiency and Crop Choices in Pakistan," *Journal of Development Economics* 67 (2002): 419-53.

—— and Humayun Khan, "Effects of Human Capital on Farm and Non-Farm Productivity and Occupational Stratification in Rural Pakistan," COE Discussion Paper No.46, Institute of Economic Research, Hitotsubashi University, Tokyo (http://hi-stat.ier.hit-u.ac.jp/english/research/discussion/index.html), November 2004.

Lanjouw, Jean O. and Peter Lanjouw, "The Rural Non-farm Sector: Issues and Evidence from Developing Countries," *Agricultural Economics* 26 (2001): 1-23.

Lanjouw, Peter, "Rural Non-agricultural Employment and Poverty in Ecuador," *Economic Development and Cultural Change* 48 (1999): 91-122.

Lee, Lung-Fei, "Generalized Econometric Models with Selectivity," *Econometrica* 51 (1983): 507-12.

Lockheed, Marlaine E., Dean T. Jamison, and Lawrence J. Lau, "Farmer Education and Farm Efficiency: A Survey," *Economic Development and Cultural Change* 29 (1980): 37-76.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka (ed.) *Frontiers in Econometrics*, New York: Academic Press, 1974.

Newman, John L. and Paul J. Gertler, "Family Productivity, Labor Supply, and Welfare in a Low Income Country," *Journal of Human Resources* 29 (1994): 989-1026.

Nielsen, Helena Skyt and Niels Westergård-Nielsen, "Returns to Schooling in Less Developed Countries: New Evidence from Zambia," *Economic Development and Cultural Change* 49 (2001): 365-94.

Sawada, Yasuyuki and Michael Lokshin, "Household Schooling Decisions in Rural Pakistan," Policy Research Working Paper, No.2541, World Bank, 2001.

Schultz, Theodore W., "Investment in Human Capital," *American Economic Review* 51 (1961): 1-17.

Yang, Dennis Tao, "Education and Off-Farm Work," *Economic Development and Cultural Change* 45 (1997): 613-32.

——, "Education in Production: Measuring Labor Quality and Management," *American Journal of Agricultural Economics* 79 (1998): 764-772.

—— and Mark Yuying An, "Human Capital, Entrepreneurship, and Farm Household Earnings," *Journal of Development Economics* 68 (2002): 65-88.

World Bank, *Pakistan Poverty Assessment – Poverty in Pakistan: Vulnerability, Social Gaps, and Rural Dynamics*, Report No. 24296-PAK, October 2002.

# Notes

1 See Glewwe and Jacoby (1994) for such an example applied to schooling decisions in developing countries.

2 In the recent literature, Jolliffe (2002) estimated the effects of several alternative measures of household education on household income, differentiated into farm and non-farm income. We adopt more detailed decomposition of household income sources than he did. Nielsen and Westergård-Nielsen (2001) estimated the effect of education on individual earnings, differentiated into wage and self-employment income sources. Unlike their work, we allow returns to labor to be non-linear and impute income from consumption of own farm products and returns to assets properly.

3 A household is defined as a unit of coresidence and shared consumption.

4 It was found that the attrition occurred more for households living in Village A and whose heads were more educated but the attrition bias was not significant in the household-level regression results (Kurosaki and Khan, 2004).

5 During the three years since the first survey, Pakistan's economy suffered from macro-economic stagnation with rising poverty (World Bank, 2002), which severely hurt the NWFP economy. Reflecting these macroeconomic shocks, the general living standard declined in the study villages.

6 Below the age of 15, no female children had a primary occupation, while 37 male children, aged 10-14, or 8.4% of that age group, had a primary occupation.

7 Achievement in female education was much lower than that for males. In 1996, average schooling was 0.5 years and the literacy rate was 7.6% for female counterparts.

8 Alternative approaches may include adopting a multinomial probit framework or modeling sequential decision making in which the household first allocates its mem-

ber to the wage or the self-employment sector and then allocates him to agriculture or non-agriculture. Relaxing the assumption that the choices are exclusive is also worth exploring, since several individuals have secondary jobs (see note *c*, Table 2). Robustness of our results under these approaches is left for a future investigation.

9 The correction term is defined as $\hat{\lambda}_{ijt} \equiv \frac{\phi[\Phi^{-1}[\hat{Prob}(z_{it}=j)]]}{\hat{Prob}(z_{it}=j)}$, where $\phi[.]$ and $\Phi[.]$ are density and distribution functions for a standard normal variable and $\hat{Prob}(z_{it}=j)$ is obtained from the first-stage multinomial logit model. If at least one variable in $X_{ht}$ in (1) does not affect wages directly but affects it indirectly through the activity choice, the second-stage wage regression is identified.

10 With household panel data, we can control for $\alpha_h$ by either fixed or random effect specification. Since the fixed effect specification may exaggerate measurement error problems, we adopt the random effect specification as long as Hausman test cannot reject at 1% level the null hypothesis that $X_{it}$ and $\alpha_h$ are uncorrelated.

11 Alternatively, we can estimate directly the system of Kuhn-Tucker equations that equate marginal returns to labor with marginal rates of substitution (Newman and Gertler, 1994). Since our interests are on the effects of education on returns to labor, a simpler approach of production functions is adopted, which allows an intuitive comparison among the four economic activities and with previous studies.

12 Under the assumption of normality of the error terms to the value-added function, the correction term becomes $\hat{\lambda}_{hjt} \equiv \frac{\phi[\Phi^{-1}[\hat{Prob}(z_{ht}=j)]]}{\hat{Prob}(z_{ht}=j)}$, where $\hat{Prob}(.)$ is given in (2).

13 Allocative efficiency may improve either because a farm manager with higher human capital is more able to allocate resources in a way closer to what maximizes the expected profit, than a manager with lower human capital, or, because a farm household with higher human capital would behave in a less risk averse way thanks to

its higher ability to cope with risk, even when both types of farms are equally able to adopt the expected profit maximizing plan.

14 See Kurosaki and Khan (2004) for the results using a specification with schooling years and their squared terms, and for the definition and statistics of the empirical variables. The non-linearity of education effects discussed below was robustly found with significant coefficients on the squared terms.

15 $X_{ht}$ in (1) serve as identifying variables for the selection term. We assume that household asset variables such as land holding do not directly affect wages paid by others but only indirectly through activity choices. Although it is possible that these variables may capture unobservable ability of individuals in implementing wage work so that they directly affect the wage level, our field observations suggest that this is unlikely. For example, the nutrition-based efficiency wage theory suggests that individuals from a landed family should be paid higher due to their superior nutrition conditions. This is unlikely among villagers in the study areas, since no difference was observed in calorie intake across land holding classes.

16 The returns to schooling reported in this paper could be an overestimate for rates of return expected from education investment on a random basis, if more able children are selected by the parents or by the community to receive higher education (Card, 1999; Alderman et al., 1996; 2001). The bias may not be large since we control for household-level unobservables by $\alpha_h$.

17 Initially, the Mincerian model in (3) was re-estimated separately for the two types. Since the difference of the coefficients was not statistically significant except for the intercept, we merged them with an employment type dummy. To control for the earning difference due to a difference in the intensity of employment in a month, daily

earnings multiplied by the standard number of monthly working days are employed as the dependent variable for the casually hired, not the observed monthly earnings.

18 Possible choices include the maximum of education among household members, the average (or median) of household members, the average (or median) of those household members who work in the household self-employment business, the education level of the household head, and so on (Jolliffe, 2002; Yang, 1998).

19 Our finding is similar to Jolliffe's (2002) that the average education among the household members is the best determinant of household productivity in Ghana. On the other hand, ours is in sharp contrast to Yang's (1997, 1998) finding that the household maximum education matters the most in determining farm productivity in China. Yang (1997) argued that more educated members of a Chinese farm household, even when they have non-farm jobs, can contribute to decision making on the farm. This effect is missing in our case of Pakistan because more educated members with non-agricultural jobs are usually indifferent to farm management, due to a strong preference for non-manual, non-agricultural work, a larger household size that enables them to be specialized in non-farm activities, and a relatively low share of agricultural income in the total household income.

20 To examine the robustness of these results, the empirical models were re-estimated using alternative selection correction formulas and different functional forms (Kurosaki and Khan, 2004). Results were qualitatively the same with those reported here.

**Table 1. Sample Villages and Sample Households (NWFP, Pakistan)**

| | Village A | Village B | Village C |
|---|---|---|---|
| Characteristics of the sample villages | | | |
| Irrigation | Rainfed | Rainfed/ Irrigated | Irrigated |
| Distance to main roads | 10km | 4km | 1km |
| Population (1998 Census) | 2,858 | 3,831 | 7,575 |
| Number of households (1998 Census) | 293 | 420 | 1,004 |
| Adult literacy rates (%, 1998 Census) | 25.8 | 19.9 | 37.5 |
| Number of the sample households (1996) | | | |
| Total | 119 | 116 | 120 |
| Non-farm households | 38 | 40 | 41 |
| Farm households, total | 81 | 76 | 79 |
| Owner farm households | 48 | 38 | 39 |
| Owner-cum-tenant farm households | 17 | 18 | 16 |
| Pure tenant farm households | 16 | 20 | 24 |
| Number of the sample households (1999) | | | |
| Total | 117 | 115 | 120 |
| Replacement sample households | 26 | 4 | 13 |
| Resurveyed households | 91 | 111 | 107 |
| Complete and comparable panel hhs. | 83 | 111 | 105 |
| Divided households | 8 | 0 | 0 |
| Households with incomplete information | 0 | 0 | 2 |
| Characteristics of the complete and comparable panel households | | | |
| Average household size in 1996 | 10.7 | 8.4 | 9.0 |
| Average household size in 1999 | 11.1 | 7.9 | 9.3 |
| Average landholding size in 1996 (ha)[a] | 2.23 | 0.52 | 0.58 |
| Average landholding size in 1999 (ha)[a] | 2.26 | 0.52 | 0.60 |
| Average per-capita income, 1996 ($)[b] | 194 | 231 | 337 |
| Average per-capita income, 1999 ($)[b] | 148 | 165 | 212 |
| Average per-capita consumption, 1996 ($)[b] | 134 | 157 | 201 |
| Average per-capita consumption, 1999 ($)[b] | 133 | 143 | 198 |

[a] "Average landholding size is the average over the total of complete panel including landless households.

[b] "Average per-capita income (consumption)" is the average over individuals included in the complete panel and its unit is US $ in nominal values.

**Table 2. Labor Force Allocation and Household Income**

A. Distribution of working household members by employment status and sector

| | Household | Employee | | Self-employed | |
|---|---|---|---|---|---|
| | work | Non-ag. | Agri. | Non-ag. | Agri. |
| 1996 Survey | | | | | |
| Males | 13 | 383 | 58 | 78 | 284 |
| Females | 762 | 6 | 0 | 3 | 4 |
| 1999 Survey | | | | | |
| Males | 10 | 459 | 25 | 125 | 200 |
| Females | 774 | 9 | 0 | 1 | 3 |

B. Level and composition of household income excluding transfers and remittances

| | Mean of | Composition shares (%) | | | |
|---|---|---|---|---|---|
| | household | Employee | | Self-employed | |
| | income[a] | Non-ag. | Agri. | Non-ag. | Agri. |
| 1996 Survey | | | | | |
| All sample households | 70,468 | 41.4 | 6.0 | 25.0 | 27.6 |
| Non-farm households | 58,839 | 49.3 | 9.4 | 33.8 | 7.5 |
| Owner farm households | 81,986 | 35.4 | 2.4 | 22.3 | 39.9 |
| Owner-cum-tenant farm hh. | 75,346 | 38.4 | 4.1 | 23.1 | 34.4 |
| Pure tenant farm hh. | 65,389 | 46.0 | 11.4 | 18.1 | 24.5 |
| 1999 Survey | | | | | |
| All sample households | 61,796 | 40.4 | 4.2 | 26.5 | 28.8 |
| Non-farm households | 50,120 | 59.5 | 6.1 | 26.5 | 7.9 |
| Owner farm households | 72,527 | 31.8 | 1.8 | 29.6 | 36.7 |
| Owner-cum-tenant farm hh. | 84,513 | 20.2 | 3.0 | 28.0 | 48.8 |
| Pure tenant farm hh. | 53,548 | 40.7 | 7.9 | 15.5 | 35.9 |

[a] Mean of the sum of the four sources of household income is shown. It is in nominal Pakistan Rupees (US$ 1.00 = Rs. 33.57 during the 1996 survey's reference period, i.e., the fiscal year of 1995/96, and 46.79 during the 1999 survey's reference period, i.e., the fiscal year of 1998/99). Income from other sources (net transfer receipt, net remittances receipt, and other unearned income) is equivalent to 11.9% (1996) and 21.6% (1999) of the total reported in this table. See Kurosaki and Khan (2004) for the definition of each income source.

[b] The sample for Panel A of this table is those household members who were working (including household work) and whose age was 15 years and above.

[c] Each worker is assigned to an employment status in Panel A based on his/her primary jobs. Approximately 12% of these workers reported their secondary jobs in 1996.

**Table 3. Human Capital Characteristics of Working Males**

| | Total | Employee | | Self-Employed | |
|---|---|---|---|---|---|
| | | Non-ag. | Agri. | Non-ag. | Agri. |
| 1996 Survey | | | | | |
| Mean age | 35.99 | 31.25 | 38.02 | 33.64 | 42.52 |
| Mean schooling years[b] | 3.68 | 4.94 | 1.90 | 4.28 | 2.19 |
| Literacy rates (%) | 42.9 | 53.8 | 25.9 | 51.3 | 29.2 |
| 1999 Survey | | | | | |
| Mean age | 34.55 | 31.04 | 31.76 | 34.78 | 43.05 |
| Mean schooling years[b] | 4.04 | 4.69 | 1.96 | 4.44 | 2.72 |
| Literacy rates (%) | 47.7 | 53.1 | 28.0 | 52.8 | 37.0 |

[a] The sample is the male subset shown in Table 2.
[b] Since repetition is common and skipping is also possible for bright students (Hoodbhoy, 1998; Sawada and Lokshin, 2001), years measured in Pakistan's standardized education system were used in converting completed grades into the schooling years.
[c] For the total working males, age is distributed between 15 and 80 with standard deviation (s.d.) 15.11 in 1996; between 15 and 83 with s.d. 15.12 in 1999. Schooling year is distributed between 0 and 16 with s.d. 4.63 in 1996; between 0 and 16 with s.d. 4.67 in 1999.

**Table 4. Estimation Results of the Multinomial Logit Model**

| | Marginal effects on the probability of choosing $j$: | | | | |
|---|---|---|---|---|---|
| | $j = 0$ (household work) | $j = 1$ (non-farm wage) | $j = 2$ (farm wage) | $j = 3$ (non-farm self-emp.) | $j = 4$ (farm self-emp.) |
| Intercept | -0.011 | 0.499 | -0.185 | -0.149 | -0.154 |
| Human capital variables | | | | | |
| Education stage dummies | | | | | |
| Primary | -0.018 | 0.014 | -0.006 | 0.041 | -0.031 |
| Middle | -0.021 | 0.077 | -0.017 | 0.061 | -0.101 |
| Higher than middle | -0.057 | 0.323 | -0.059 | 0.021 | -0.228 |
| Age | -0.002 | 0.002 | 0.003 | 0.000 | -0.003 |
| $\text{Age}^2/100$ | 0.002 | -0.011 | -0.003 | 0.000 | 0.012 |
| Household asset variables | | | | | |
| Adult males # | 0.011 | 0.028 | 0.026 | -0.023 | -0.042 |
| Household size # | -0.006 | -0.009 | -0.009 | 0.034 | -0.011 |
| Dummy for land own. | 0.001 | -0.176 | -0.039 | -0.008 | 0.221 |
| Land size # | 0.007 | 0.079 | -0.136 | 0.000 | 0.049 |
| Livestock value # | 0.005 | -0.047 | -0.018 | -0.025 | 0.085 |
| Other assets # | 0.003 | 0.012 | -0.005 | 0.000 | -0.010 |
| Village fixed effects | | | | | |
| Village B | 0.008 | -0.152 | 0.057 | 0.064 | 0.024 |
| Village C | 0.017 | -0.225 | 0.065 | 0.062 | 0.080 |
| Log-likelihood | | -1620.33 | | | |
| $LR$ statistics[c] | | 593.14 | | | |

[a] Variables with # are standardized as ($X$ - mean)/standard deviation. "Other assets" mainly include the value of vehicles and electric appliances.
[b] Only those explanatory variables whose $\gamma_j$ is statistically significant at least 10% for some $j$ are included. Full regression results including individual parameter estimates of $\gamma_j$ on $X_i$ and $X_h$ and its standard errors are available on request.
[c] "$LR$ statistics" shows the likelihood ratio test statistics for zero slope. It is statistically significant at 1% level.
[d] The sample is the male subset shown in Table 2 (the number of observations is 1,635).

**Table 5. Estimation Results of Wage Equations for Males**

| | Non-agriculture | | | | Agriculture | |
|---|---|---|---|---|---|---|
| | Without employment type | | With employment type | | | |
| Intercept | 6.002 | *** | 6.239 | *** | 6.700 | *** |
| | (32.51) | | (30.66) | | (16.23) | |
| Dummy for 1999 | -0.271 | *** | -0.310 | *** | -0.325 | *** |
| | (-7.137) | | (-5.911) | | (-3.394) | |
| Human capital variables | | | | | | |
| Education stage dummies | | | | | | |
| Primary | 0.154 | ** | 0.107 | * | 0.185 | *** |
| | (2.334) | | (1.686) | | (2.638) | |
| Middle | 0.268 | *** | 0.189 | *** | 0.017 | |
| | (4.018) | | (2.973) | | (0.200) | |
| Higher than middle | 0.494 | *** | 0.299 | *** | 0.090 | |
| | (6.629) | | (4.993) | | (0.526) | |
| Age | 0.068 | *** | 0.059 | *** | 0.026 | ** |
| | (8.364) | | (7.649) | | (2.707) | |
| $Age^2/100$ | -0.081 | *** | -0.063 | *** | -0.031 | *** |
| | (-7.276) | | (-6.162) | | (-2.782) | |
| Employment type dummy for the regularly hired | | | 0.474 | *** | | |
| | | | (8.570) | | | |
| Selection correction | 0.528 | *** | 0.211 | * | 0.152 | |
| | (3.919) | | (1.801) | | (1.074) | |
| $R^2$ | 0.236 | | 0.327 | | 0.297 | |
| $\bar{R}^2$ | 0.230 | | 0.320 | | 0.232 | |
| Hausman test statistics | 11.51 | | 13.55 | | 10.23 | |
| NOB | 841 | | 841 | | 83 | |

[a] Values of $t$-statistics are reported in the parentheses with *** significant at 1%, ** at 5%, * at 10% (two-sided test).
[b] Estimated by an unbalanced panel method with random household effects.
[c] The sample is the subset of the male household members described in Table 2, who are employed by others. One sample employed in non-agriculture is deleted since its wage information is incomplete.
[d] The dependent variable is natural log of monthly wage.
[e] Hausman test statistics are distributed as $\chi^2$ with degrees of freedom seven, eight, and seven, respectively.

**Table 6. Estimation Results of Value-Added Functions for Self-Employment**

| | Non-agriculture | | | | Agriculture | | | |
|---|---|---|---|---|---|---|---|---|
| | Without business type | | With business type | | Wheat only | | All crops | |
| Intercept | 3.150 | *** | 3.245 | *** | 5.462 | *** | 5.712 | *** |
| | (3.990) | | (4.168) | | (10.47) | | (12.75) | |
| Dummy for 1999 | -0.238 | | -0.119 | | 0.098 | | 0.014 | |
| | (-0.928) | | (-0.458) | | (0.736) | | (0.200) | |
| Basic production factors | | | | | | | | |
| log of labor | 0.708 | *** | 0.688 | *** | 0.261 | *** | 0.276 | *** |
| | (11.32) | | (11.03) | | (3.845) | | (6.417) | |
| log of capital/land | 0.127 | *** | 0.136 | *** | 0.598 | *** | 0.498 | *** |
| | (4.463) | | (4.777) | | (7.734) | | (9.955) | |
| Human capital variables | | | | | | | | |
| Share of family labor with the following education | | | | | | | | |
| Primary | 0.100 | | 0.044 | | 0.112 | | 0.201 | |
| | (0.812) | | (0.352) | | (0.909) | | (1.565) | |
| Middle | 0.328 | ** | 0.284 | ** | 0.080 | | 0.299 | * |
| | (2.465) | | (2.141) | | (0.515) | | (1.886) | |
| Higher than middle | 0.558 | *** | 0.452 | *** | 0.133 | | 0.289 | ** |
| | (4.132) | | (3.212) | | (0.973) | | (2.151) | |
| Household experience | | | | | | | | |
| Head's age | -0.029 | | -0.032 | * | 0.003 | | 0.002 | |
| | (-1.612) | | (-1.827) | | (0.273) | | (0.171) | |
| (Head's age)$^2$/100 | 0.031 | * | 0.034 | ** | -0.003 | | -0.006 | |
| | (1.798) | | (2.030) | | (-0.224) | | (-0.543) | |
| Control variables for production | | | | | | | | |
| Business space dummy | | | 0.211 | ** | | | | |
| | | | (2.275) | | | | | |
| Irrigation | | | | | 1.076 | *** | 1.809 | *** |
| | | | | | (9.161) | | (13.66) | |
| (Sharecropping ratio) times (Village A dummy) | | | | | 0.430 | *** | 0.522 | *** |
| | | | | | (3.304) | | (3.347) | |
| Selection correction | -0.006 | | -0.009 | * | -0.003 | | -0.128 | |
| | (-0.063) | | (-0.090) | | (-0.042) | | (-1.505) | |
| $R^2$ | 0.635 | | 0.648 | | 0.489 | | 0.601 | |
| $\bar{R}^2$ | 0.614 | | 0.626 | | 0.470 | | 0.590 | |
| Hausman test statistics | 12.17 | | 12.36 | | 7.86 | | 10.73 | |
| NOB | 170 | | 170 | | 323 | | 413 | |

[a] See Table 5.

[b] Estimated by a 2SLS unbalanced panel method with random household effects. Basic factors and sharecropping ratios are replaced by their fitted values using other right-hand-side variables, the acreage of agricultural land owned by the household, the number of adult males, the net value of household assets (transportation and durable consumption goods), and the value of livestock (both levels and logs) as instruments.

[c] The dependent variable is natural log of value-added from self-employment enterprises.

[d] Hausman test statistics are distributed as $\chi^2$ with d.o.f 9, 10, 11, and 11, respectively.

**Table 7. Observed and Simulated Labor Allocation**

| | Simulated labor force allocation | | | | |
|---|---|---|---|---|---|
| Observed labor force allocation | (1) | (2) | (3) | (4) | Total |
| 1996 Survey | | | | | |
| (1) Non-agricultural wage | **203** | 37 | 62 | 81 | 383 |
| (2) Agricultural wage | 35 | **6** | 8 | 9 | 58 |
| (3) Non-agricultural self-employment | 1 | 0 | **77** | 0 | 78 |
| (4) Agricultural self-employment | 74 | 16 | 44 | **150** | 284 |
| Total | 313 | 59 | 191 | 240 | 803 |
| 1999 Survey | | | | | |
| (1) Non-agricultural wage | **214** | 12 | 88 | 145 | 459 |
| (2) Agricultural wage | 13 | **2** | 4 | 6 | 25 |
| (3) Non-agricultural self-employment | 6 | 0 | **106** | 13 | 125 |
| (4) Agricultural self-employment | 14 | 2 | 46 | **138** | 200 |
| Total | 247 | 16 | 244 | 302 | 809 |

[a] Bold face figures show correct predictions.
[b] Based on parameter estimates in Table 5 without employment type for non-agriculture and those in Table 6 without business type for non-agriculture and all crops for agriculture.