

Jonasson, Erik

Working Paper

Earnings Differentials in the Rural Labour Market: Does Non-Agricultural Employment Pay Better?

Working Paper, No. 2008:7

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Jonasson, Erik (2009) : Earnings Differentials in the Rural Labour Market: Does Non-Agricultural Employment Pay Better?, Working Paper, No. 2008:7, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/259955>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Earnings Differentials in the Rural Labor Market: Does Non-agricultural Employment Pay Better? *

Erik Jonasson *

February 21, 2009

ABSTRACT

Rural non-agricultural employment (RNAE) is being increasingly emphasized as a potential pathway out of rural poverty for people who are unable to secure their income in agriculture. Although average earnings in the rural non-agricultural sector are higher than in agriculture, it is unclear whether income prospects are systematically better in non-agricultural activities than in agriculture. This paper tests for existence of earnings differentials between agricultural and rural non-agricultural employment, while controlling for worker and household characteristics. A theoretical farm household model is proposed that predicts that there will be no sectoral earnings differential for unskilled labor, whereas skilled labor will be better off in the non-agricultural sector. Based on Peruvian household data, the empirical findings do not support the notion that unskilled workers would earn a higher income by switching from agriculture to RNAE. Instead it tends to be the relatively well educated who might benefit from higher returns to education in RNAE than in agriculture, consistent with the predictions of the theoretical model.

Key words: non-agricultural employment, farm household model, wage differentials, Peru, Latin America

JEL classification: J24, J31, J43, O12, R23

1. INTRODUCTION

Peru continues to struggle with its persistently high rural poverty. Recent data suggest that 75 percent of the rural population fall below the ECLAC poverty line (defined as twice the cost of a daily food basket). In Latin America as a whole, rural poverty remains high, exceeding 58 percent (ECLAC, 2006). The corresponding rate in 1990 was 65 percent, indicating that the success of rural poverty alleviation efforts has been moderate over the past decade. Even though agricultural development traditionally has been the main ingredient in rural development strategies, scholars have for long emphasized the need for diversified approaches to fighting rural poverty in order to take the heterogeneity of the rural population into account. The message is that efforts to improve agricultural productivity should be concentrated to

* An earlier draft of this paper was presented under the title "Non-agricultural Employment as Pathway out of Rural Poverty? Evidence from Peru" at the 22nd Annual Congress of the European Economic Association, Budapest, August 28, 2007 and at the Annual International Conference of the German Economics Association, Research Committee Development Economics, Kiel, July 9, 2005. The paper has benefited from comments by Sonja Opper, Therese Nilsson, and Pernilla Johansson at the Department of Economics, Lund University. All remaining shortcomings of the paper are under the sole responsibility of the author.

* OECD Trade and Agriculture Directorate and Lund University. erik.jonasson@nek.lu.se, +33 637 832 388, +46 708 866 300.

viable farm households, and that alternative paths out of poverty should be stimulated for landless or non-viable farm households. These alternative paths could be migration for some and participation in the rural non-agricultural (RNA) sector for others. An attractive feature of rural non-agricultural employment (RNAE) is that it can provide a source of income for some of the rural landless and for those who cannot secure their income from agricultural wage labor. It also constitutes a source of complementary income for farm households. Diversifying into non-agricultural activities could be a response to insufficient farm income or a means to decrease the vulnerability associated with volatile agricultural incomes. Although migration to urban areas might be the most appropriate route out of poverty for some groups, RNAE could also have the potential to slow down rural-to-urban migration and the process of rural poor merely becoming urban poor (Lanjouw and Lanjouw, 2001). For the most vulnerable rural poor, poverty alleviation will require assistance through social transfers (de Janvry and Sadoulet 1993, 2000; Valdés and Wiens, 1996; Echeverría, 2000). The need for heterogeneous efforts to alleviate rural poverty in Latin America is well represented in recent strategy formulations (World Bank, 2003 and 2007; de Ferranti et al., 2005).

Even though average earnings in the RNA sector are higher than in agriculture, it is unclear whether income prospects are systematically better in non-agricultural activities than in agriculture (Lanjouw, 2008). In particular, is there a systematic earnings differential between RNAE and agricultural work, when controlling for other factors that we believe determine the earnings potential of an individual? Whether such an earnings differential exists is relevant, from a policy and strategic point of view, to determine whether rural non-agricultural employment should be included as a general element of rural development strategies or be promoted under certain conditions only. This paper adds to the RNAE literature by empirically testing for such a sectoral earnings differential between agriculture and non-agriculture. A household model with a dualistic rural labor market is introduced to guide the empirical analysis.

Few studies explore earnings differentials between agriculture and RNAE. The empirical literature on RNAE focuses mainly on the determinants of participation in the RNA sector, either considered to be an occupational choice of the individual (Ferreira and Lanjouw, 2001; Lanjouw, 2001) or part of a household income diversification strategy (Barrett et al., 2001; Reardon et al., 2000). Studies are also concerned with the determinants of the income of those who participate in the RNA sector (de Janvry and Sadoulet, 2001; Isgut, 2004). A reason that few studies examine income differentials, or wage gaps, is the empirical challenge of isolating the sector effect from unobserved factors that affect income and sector choice simultaneously. Dabalen et al. (2004) estimate returns to participation in the RNA sector relative to the agricultural sector in Rwanda. They use the method of propensity score matching to test whether people with similar attributes, but in different sectors, earn different incomes, and find that the self-employed in the non-agricultural sector earn significantly more than farm workers. McCulloch et al. (2007), in a study on pathways out of rural poverty in Indonesia, use panel data to trace the income changes of people switching from agriculture to non-agricultural activities. They find that increased engagement of rural farmers in non-agricultural businesses has been the most promising path out of rural poverty. It is unclear, however, to what extent these findings can be generalized to the Latin-American context, which differs from South-East Asia and Sub-Saharan Africa by its higher level of per-capita income, its much lower population density, and its high degree of wealth and income inequality.

This paper used Mincer-type income regressions, in which sector of employment is treated as exogenous. This OLS approach is complemented with an instrumental-variable approach to adjust for the potential effects of endogenous sector choice. The empirical analysis was undertaken using the 1994 Peruvian *Encuesta Nacional de Hogares Sobre Medición de Vida*, which is the survey source for the World Bank Living Standard Measurement Study (LSMS) for Peru that year. There is little support in the results for the notion that an unskilled worker

would earn a higher income in RNAE than in agriculture. The results do, however, suggest that returns to education are higher in RNAE and hence that skilled people tend to do better in RNAE than in agriculture. This finding is robust across most of the regression specification, including the instrumental-variable approach.

The next section provides an overview of rural poverty and employment in rural Peru. Section 3 introduces the theoretical model, followed by the empirical findings in Section 4. Section 5 concludes.

2. A PROFILE OF RURAL POVERTY AND EMPLOYMENT IN PERU

Peru, with its 27 million citizens, is the fifth largest country in Latin America in terms of population. According to the demographic census of 2005, 26 percent of the population lives in rural areas, which is close to average for the region. About 50 percent of the population lives in the coastal region (*Costa*), including Lima; about 37 percent live in the highlands (*Sierra*); and the remaining 10-15 percent of the population live in the lowland jungle of the Amazon basin (*Selva*).

One of the biggest economic and social challenges for Peru is the large share of the rural population that lives in poverty. The World Bank (2005) estimates that 72.5 percent of the rural population fall below the national poverty line (defined as twice the cost of a daily food basket; approximately 2 US dollar per day). With a poverty rate similar to that of the 1980s, the evolution of rural poverty is discouraging. Although economic growth led to a poverty decline in the 1990s, poverty increased again with the economic recession at the turn of the century. Programs specifically targeting rural poverty amount to 450 million US dollar per year but have not shown positive long-term results. Escobal (2004) notes that, even though this is a large commitment of resources, the majority of the programs consist of safety nets and temporary relief, and that little is spent to overcome the structural causes of poverty. Table 1 shows that the poverty rate is highest in the less developed Sierra and Selva regions, where people fall lowest beneath the poverty line and to which most of the anti-poverty resources are directed.

[Table 1]

It is evident that the path out of poverty for most rural households will have to be accompanied by continued political efforts to invest in rural infrastructure and to promote institutional change to the advantage of the poor (World Bank, 2003). The question is: What opportunities do the rural households have on their own to improve their income prospects? In particular, is it likely that the household will increase its income through diversifying its sources of income by engaging in RNAE? Agriculture is still the main sector of employment in rural Peru. According to the 1994 survey *Encuesta Nacional de Hogares Sobre Medición de Vida* (henceforth referred to as the Peruvian LSMS 1994), more than 70 percent of rural household labor was engaged in agriculture and the remaining 30 percent in non-agricultural activities. More recent statistical sources suggest that the general structure of the rural labour market was stable during the 1990s and the early 2000s.¹ The degree of diversification of income sources in rural households is generally high. If employment specialization is defined as spending 90 percent or more in one sector, then about 36 percent of rural households are specialized in agriculture, whereas only

¹ Escobal (2001), using the LSMS of 1994 and 1997, estimates the shares of the rural labor force engaged in non-agricultural employment at 31.6 and 30.5 percent for these two years respectively. The World Bank (2005) reports that 72.8 percent of household labor hours are spent on agricultural work and 27.2 percent on non-agricultural work. For Latin America as a whole, about 40 percent of the rural labor force is involved in non-agricultural activities, which is an increase by 5-10 percentage points since the early 1990s (Dirven, 2004).

5 percent are specialized in non-agricultural activities. The remaining 59 percent are pluriactive households. Table 2 shows that RNAE is most prevalent in the coastal region, which is the more developed region in terms of average income, infrastructure, and labor market participation. In this region, more than 37 percent of the labor force is engaged in the non-agricultural sector. The survey data, however, do not tell whether a rural resident also works in a rural area. Some non-agricultural workers are likely to be rural residents who commute to urban areas.² This causes some over-estimation of the size of the 'truly' rural non-agricultural labor force, particularly in more densely populated coastal region. Self-employment, as opposed to wage labor, is the dominant form of employment in both agriculture and non-agriculture, with the exception of the non-agricultural sector in the coastal region.

[Table 2]

The World Bank (2005) estimates that the poverty rate is 80 percent among people employed in the agricultural sector in Peru, and 60 percent among those employed in the RNA sector. Wage workers in agriculture are those who are most likely to be poor, followed by farmers (Lopéz and della Maggiora, 2000). This is confirmed by the distribution of individual earnings among wage labor and self employed in agriculture and non-agriculture. When not controlling for other factors, there is a statistically significant difference in average earnings between the agricultural and non-agricultural sectors, with earnings being higher in the latter. Table 3, which reports monthly income from principal employment, shows that this difference is driven mainly by different average wages in the wage labor market. Average earnings for the self-employed are not different between the two sectors.³ Neither is there any difference in average earnings between the self-employed and wage laborers within the non-agricultural sector.

[Table 3]

In this paper, RNAE includes all activities other than agricultural work at the own farm or another farm.⁴ A considerable share of the RNA economy consists of activities closely related to agriculture, such as food processing, transportation, and marketing of agricultural goods. It also includes activities such as mining, construction, domestic services, and tourist-related services, with little or no connection to agriculture. Table 4 shows the composition of employment in the RNA sector. Although some 28 percent of household labor hours are spent on RNAE, the share of the rural labor force with RNAE as the principal form of employment is only 20 percent. Thus, many households have RNAE as a secondary source of earned income. Commerce and manufacturing stand out as the biggest sectors in terms of employment, absorbing more than 55 percent of RNAE. The most common manufacturing activities are food processing and textile work. Among self-employed in non-agriculture, more than 80 percent are engaged in commerce and manufacturing. Wage labor is less concentrated in a certain sector, but more than 30 percent is found in education and other forms of public-sector employment. For non-agricultural wage labor, manufacturing, domestic services, and construction work are the biggest private-sector employers.

² Urban areas are defined as all towns and cities with 2,000 or more inhabitants.

³ Income estimates for the self-employed are subject to a higher degree of measurement error than estimates for wage labor. Two sources of possible measurement error are the volatility in income flows for the self-employed and the fact that earned income is not adjusted for expenditure related to these business activities.

⁴ There is no consensus in the literature on whether to include auxiliary farm activities, such as fishing and hunting in RNAE. Saith (1992), for example, considers these activities as non-agricultural since they do not fall under the constraint of land use. Non-agricultural work should not be confused with *off-farm* work, which is a broader concept used to denote all work (agricultural or non-agricultural) performed outside the own farm.

[Table 4]

As Peruvian rural households get wealthier they tend to spend less labor time on agriculture and more time on non-agricultural activities. Table 5 shows that households in the lowest quintile spend around 20 percent of their labor on RNAE, whereas households in the highest income quintile spend around 35 percent on RNAE. One can think of reasons for positive as well as negative correlations between household wealth and engagement in RNAE. Reardon et al. (2000) suggest that wealthier households are likely to possess the assets that make non-agricultural employment profitable (the necessary capital for business start up or education for well-paid employment), giving them 'pull' incentives to increase their share of RNAE as their wealth increases. To the extent that poor people are 'pushed' into RNAE, as an income source of last resort and as a backup for low agricultural incomes, one would predict a negative relationship between wealth and RNAE. According to Ellis (2000), the negative relationship between wealth and degree of multi-activity is observed in regions where landownership distinguishes the well-off from the poor. The positive relationship tends to be observed in regions where livestock and human capital are the main assets distinguishing the better off from the poor. Reardon et al. (2000) also suggest a U-shaped relationship between household wealth and degree of multi-activity. They explain this relationship in terms of push factors forcing the landless poor to undertake a high degree of non-agricultural employment, whereas the middle-income households are well off enough to survive on farm production or agricultural wage labor alone. High-income households are able to engage in well-paid non-agricultural activities due to high skills or asset holdings. In Peru pull incentives appear to dominate push incentives for participation in RNAE, as suggested by the increasing share of RNAE from low-income to high-income household quintiles.

[Table 5]

3. A HOUSEHOLD MODEL WITH DUALISTIC LABOR MARKETS

In this section, a basic farm household model is proposed in order to put the livelihood strategy of rural households into a theoretical framework. The model is chosen deliberately to capture some of the characteristics of rural households described in the previous section: The dominant source of employment and income for rural households in Peru is agriculture, and agricultural work most often takes the form of self-employment (peasant farming) rather than wage labor. The model is a slight extension of the static farm household model developed by Bardhan and Udry (1999). It is extended to take a dualistic feature of the rural labor market into account. In particular, it is assumed that there exists an agricultural labor market, in which labor is treated as homogenous, and a non-agricultural labor market, in which workers are compensated according to their skills.

Households are assumed to engage in agricultural production, but can supply all or parts of their labor in the competitive agricultural and non-agricultural labor markets. Households differ by their level of skill, which determines their labor productivity in the non-agricultural labor market, but not in the agricultural labor market. Let δ denote a household-specific skill parameter, which is normalized to 1 for unskilled households. Skilled households – provided that they supply their labor in the non-agricultural labor market – are able to supply a multiple δ of "unskilled-labor equivalents". The price of one unit of unskilled labor is equal in the two markets.

In the absence of input and output market failures, the farm household's consumption and production decisions are separable and can be made independently of each other (Taylor and Adelman, 2003). The constrained optimization problem that the household faces consists of

maximizing household composite utility with respect to consumption and leisure, given its total income from farm production and off-farm labor:

$$\max_{c,l} U(c, l) \quad (1)$$

subject to the following constraints:

$$p_c c + wL_A^h + rK^h \leq p_F F(K, L_A) + w(L_A^m + \delta L_N^m) + rK^m \quad (2a)$$

$$L_A = L_A^f + L_A^h \quad (2b)$$

$$K = K^f + K^h \quad (2c)$$

$$E^L = L_A^f + L_A^m + L_N^m + l \quad (2d)$$

$$E^K = K^f + K^m \quad (2e)$$

$$c, l, L_A^f, L_A^m, L_N^m, K^f, K^m \geq 0; \delta \geq 1 \quad (2f)$$

In the utility function (1), c and l denote the household's composite consumption and leisure. In the budget constraint (2a) prices of consumption goods, farm production, labor, and capital are denoted p_c , p_F , w , and r , respectively. $F(K, L_A)$ is the household farm production function with capital (K) and agricultural labor (L_A) as inputs. For simplicity, assume that capital includes productive land as well as physical equipment. The budget constraint states that expenditures on consumption, hired labor, and capital cannot exceed the revenues from production and from marketed labor (L^m) and capital (K^m). The subscripts A and N on marketed labor distinguish labor supplied on the agricultural labor market from labor supplied on the rural non-agricultural labor market. Identities (2b) and (2c) state that labor and capital in farm production are provided from the own household (f) or hired in the factor markets (h). Household labor and capital endowments are given by E^L and E^K and are allocated according to identities (2d) and (2e). The household can devote its time to four different activities: work on the own farm, marketed off-farm wage labor in the agricultural or non-agricultural labor market, and leisure. Non-negativity constraints are listed in (2f).

The separability assumption allows for profits from farm production be maximized independently of household preferences:

$$\pi^*(w, r) = \max_{L_A, K} [F(K, L_A) - wL_A - rK] \quad (3)$$

The household labor allocation decision depends on preferences for leisure and on labor productivity (δ). For the unskilled household, for which δ equals one, the opportunity cost of leisure is the wage rate w . Since the marginal revenue product of labor in own farm production equals the going wage rate, the unskilled household is indifferent to supplying its labor to farm production or on the agricultural or non-agricultural labor markets. The skilled household ($\delta > 1$) faces an opportunity cost of leisure equal to δw , provided that it can supply all its labor on the non-agricultural labor market. A skilled household will therefore neither allocate its labor to its own farm nor participate in the agricultural labor market, in which returns to labor are no higher than w , even for skilled labor. Thus, for skilled households the labor allocation identity (2d) reduces to $E^L = L_N^m + l$.

Re-arranging (2a) and substituting in (3), the full-income constraint for the unskilled and skilled households become, respectively:

$$p_c c + wl \leq \pi^*(w, r) + wE^L + rE^K \quad (4a)$$

and

$$pc + w\delta l \leq \pi^*(w, r) + w\delta E^L + rE^K \quad (4b)$$

Two implications of particular interest follow from the model. First, in the absence of market failures, no household has an incentive to diversify its income sources by participating in both labor markets. Second, an unskilled agricultural worker will not earn a higher income by switching to non-agricultural work; he or she will be paid the same competitive wage rate, w . Thus while earnings in the non-agricultural sector are higher due to higher average labor productivity, there is no “unconditional” earnings differential between the sectors once worker skills are taken into account – unconditional in the sense of not being conditional on having a certain level of education. This hypothesis is subject to empirical evaluation in the next section. Before turning to the empirical analysis, the limitations of this “naive” model are briefly elaborated on.

Under missing or imperfect markets the situation will be different, and the household problem cannot be represented by (1) – (3). Various forms of market failures are commonly observed in product and factor markets in rural areas of developing countries (Stiglitz, 1998). The roots of limited market access are usually high transactions costs (de Janvry et al., 1991; Key, et al., 2000; Sadoulet et al., 1998). Incomplete information usually causes inefficiency in factor markets (Stiglitz, 1988), which are further impeded in their functioning by poorly defined property rights (de Soto, 2001). If there are barriers to enter the land and capital markets, the production possibility of the household will be largely determined by its factor endowments. In the extreme case of complete factor market inaccessibility, the factor input identities (2b) and (2c) reduce to $L_A = L_A^f = (E^L - l)$ and $K = K^f$, respectively.⁵ If a household is landless and cannot easily gain access to productive land through land rental or other arrangements, the agricultural production function is no longer part of the budget. The household is then constrained to whatever income it can earn in the labor market. Similarly, when access to financial or productive capital other than land is limited, chances to engage in non-agricultural business are limited. If the household lacks endowment of, and access to, productive capital, then the budget constraint (2a) reduces to $p_c c \leq w(L_A^m + \delta L_N^m)$. Yet worse, if unemployment emerges in the agricultural labor market in slack seasons, and if access to the non-agricultural labor market is obstructed by entry barriers, not even labor income will be a secure means of income for households lacking productive assets. Thus, as soon as we allow for imperfections in markets, it is no longer obvious how households will allocate their labor time.

4. EARNINGS DIFFERENTIALS IN THE RURAL LABOR MARKET: AN EMPIRICAL ASSESSMENT

Given that average earnings in the rural non-agricultural sector are higher, and that wealthier households devote more of their labor time to non-agricultural activities, is rural non-agricultural sector a potential pathway out of rural poverty? Little can be said with descriptive statistics. If skilled and highly educated labor is systematically drawn to the RNA sector in search of the highest returns to labor, then the earnings differential is just a skill-compensating wage differential. As suggested in the theoretical model, such a case would give unskilled wage labor in the agricultural sector small chances to increase income by switching to non-agricultural employment. This section empirically assesses the extent to which the sectoral income differential between agricultural and non-agricultural employment is observed when a range of other factors are controlled for that might determine the individual’s earnings potential.

⁵ In this case, household-specific shadow prices of inputs and outputs jointly determine consumption and production, i.e. separability is no longer maintained.

4.1. Data and empirical method

The empirical analysis is based on data from the Peruvian LSMS of 1994. A comparison with more recent data (ENAHO 1997 and 2002) suggests that the general structure of the rural labor market has remained largely unchanged since the survey year (see footnote 1). The survey sample is nationally representative and consists of 3,623 households, of which 1,336 reside in rural areas. For the 1994 LSMS there was also a community survey carried out, which provides information about 204 population centers where the household survey was carried out. Most of these population centers were small villages of 100 households or less. This information was used in the analysis to control for local characteristics that are likely to influence the employment outcome of rural households, yet not directly affecting their income. The unit of analysis was the rural worker, and after the exclusion of unpaid family members and individuals below the age of 12, there were 1,681 individuals in the sample. The sample gives a fair representation of the rural labor force, with the exception that the exclusion of unpaid employees increases the share of people in RNAE from 20 to 35 percent and the share of males in the labor force from 55 to over 70 percent. Table 6 provides descriptive statistics for the variables that were included in the analysis, and covers the sample that was used in the regression analysis.

[Table 6]

A Mincerian income equation served as basis for testing the presence of earnings differentials between the agricultural and rural non-agricultural employment. In this approach, sector of employment was treated as an exogenous choice. Income regressions were estimated according to the following setup:

$$y_{ij} = \beta_0 + \beta_1 RNAE_{ij} + x_{ij} \beta_x + h_j \beta_h + \varepsilon_{ij} \quad (5)$$

where y_{ij} is the logarithm of earned income from the principal employment of individual i in household j . The variable of our main interest is $RNAE$, the binary variable that distinguishes non-agricultural employment from agricultural employment. Its corresponding coefficient β_1 is the estimate of the conditional sectoral earnings differential. Individual and household characteristics were controlled for by the sets of variables x_{ij} and h_j , described below. The residual ε has expected zero mean and constant variance σ^2 . Model (5) was estimated by OLS and was used as a benchmark for comparison with alternative model specifications. Various interaction effects were tested for and the exogeneity assumption of $RNAE$ was relaxed by using a two-stage least square approach.

Among the individual variables (x), age, age-squared, and years of education were included as proxies for individual work experience and human capital acquirement. Gender was included due to the large earnings differential between men and women. This earnings gap is likely due to a combination of self-selection into low-paid employment, lower average work hours, and gender discrimination in the labor market (Abramo and Valenzuela, 2005). A dummy variable for non-Spanish mother tongue controlled for ethnicity. The Quechua and Aymara languages are common among the indigenous people, who are most likely to be poor. Economic marginalization of indigenous groups is most likely a result of several factors, such as language barriers, ethnical discrimination, and adverse geographical location (Escobal, 2004). To capture mobility and the ability of an individual to respond to economic opportunities, the analysis included a variable that controls for whether the individual has migrated for work. Employment form (self-employed versus wage labor) controlled for the different conditions these two groups might face. In the labor market, earnings depend on the going wage rate, whereas for the self-employed a wider spectrum of factors will determine earnings. In combination with other productive assets, self-employment could provide higher returns to labor than wage income. Without such assets, however, self-employment could be a result of unemployment or

difficulties in entering the labor market, and provide earnings lower than wage incomes. To control for regional differences in economic opportunities, regional dummy variables for the highlands and the jungle regions were included (with the coastal region being the reference region). The household variables (h) consisted of three measures for productive assets: land ownership, livestock, and other productive equipment. Productive assets should matter little for wage labor, but for the self-employed these will constitute factors of production.

According to the theoretical model, RNAE and skills (education and work experience) are complements in generating income. This implies that the unskilled worker does not gain by switching from agriculture to RNAE, and that there are positive returns to skills in RNAE but not in agriculture. To test for interaction effects between RNAE and education, model (5) was modified as follows:

$$y_{ij} = \beta_0 + \beta_1 RNAE_{ij} + \beta_2 edu_{ij} + \beta_3 (RNAE_{ij} \times edu_{ij}) + x_{ij}^* \beta_x^* + h_j \beta_4 + \varepsilon_{ij} \quad (6)$$

Education (edu) was lifted out from the set of individual characteristics x_{ij} (hence the asterisks on x and β_x) to make the interpretation easier. The $RNAE$ indicator was interacted with the worker's years of education (edu). We can think of four scenarios, in which the corresponding interaction coefficient β_3 is positive. First, the theoretical model hypothesizes that β_1 and β_2 are both zero and that β_3 is positive: Education only has positive returns in RNAE, and RNAE is only beneficial given some level of education. Call this scenario S_1 . Second, it could be the case that there are positive returns to education in both sectors, that returns are higher in RNAE, but that there is no remaining (unconditional) benefit of RNAE. In this case β_1 is zero and β_2 and β_3 are positive (S_2). Third, returns to education could be positive in both sectors, higher in RNAE, and there could be an unconditional earnings premium in RNAE, in which case all three coefficients are positive (S_3). Last, there could be an unconditional earnings premium in RNAE while returns to education are only positive in RNAE. In that case, β_1 and β_3 are positive while β_2 is zero (S_4).

The test for interaction effects was then extended to allow differences in returns across sectors in *all* individual and household characteristics:

$$y_{ij} = (1 + x_{ij} + h_j) \beta^{AG} + [RNAE_{ij} \times (1 + x_{ij} + h_j)] \beta^{RNAE} + \varepsilon_{ij} \quad (7)$$

where the β^{RNAE} coefficients constitute structural differences between agriculture and RNAE.

The analysis that was carried out according to models (5) – (7) relied on the assumption that sector of employment is exogenously determined. There are reasons to assume that sector of employment is a choice that at least partly is determined by observed and unobserved personal and household characteristics. This is shown repeatedly in empirical studies on RNAE, such as Lanjouw (1998) for Ecuador, de Janvry and Sadoulet (2001) for Mexico, and Ferreira and Lanjouw (2001) and Jonasson and Helfand (2009) on Brazil. If sector of employment, or any of the other right-hand side variables, is endogenously determined in model (5), the OLS estimators are not consistent. A two-stage least square approach was applied in order to adjust for this potential endogeneity of $RNAE$. Instrumental variables were needed, which are conditionally correlated with $RNAE$, yet uncorrelated with the error term ε , and not a direct determinant of income y . Two sets of instruments were used: two household variables and three community variables.

First, if hired labor is not a perfect substitute for household farm labor (for example, due to monitoring costs), or if there are transactions costs involved in the labor market, then the labor allocation decision could be influenced by the size of the household. Since larger households have a larger labor endowment they might have better opportunities to let one or more household members work off the farm. The number of household members, as a proxy for labor

endowment, served as the first instrument. Second, if farming involves some degree of management and monitoring that is usually under the responsibility of the household head, he or she might be less likely than other household members to take an off-farm job. An indicator for household head was used as the second instrument. Neither household size nor household position is likely to have any direct relation with unobserved characteristics that affect earnings.

Jonasson and Helfand (2009) provide some empirical evidence that the economic geography to a large extent affects the RNAE opportunities for a rural worker. They find that RNAE opportunities are higher where distance to population centers are shorter, where rural infrastructure is better, and where the local market size is larger. Their findings on the relation between geographical factors and non-agricultural income are less conclusive. Only to some lesser extent does geography seem to directly affect earnings prospects once a worker is engaged in non-agricultural work. Based on these findings, three local characteristics were used as instruments for RNAE. The first is blunt indicator for infrastructure – a binary variable indicating the existence of a paved road in the community. The second community variable is population size of the nearest population center, as a proxy for local market size. The third variable is distance measure to the nearest permanent market place (which might be outside the local community).

4.2. Estimation results

The estimation results of model (5) and its extensions (6) and (7) are reported in Table 7 through 9 and are discussed in turn below. The primary focus is on the main variable RNAE, but coefficients of some of the control variables are also discussed.

Table 7 contains seven specifications, of which three were estimated on the full sample (columns 1–3), two on the wage labor sample (columns 4–5), and two on the self-employed sample (columns 6–7). The results suggest several structural differences between laborers and self-employed, so that treating the entire rural labor force as one homogenous group could give an incomplete understanding of the relationship between income and personal characteristics.

The specification in column 1 only contains individual-specific variables. Column 2 adds household productive assets, and column 3 adds interaction effects (as suggested in model 6). The reason that results are reported both with and without household assets is that these assets could be endogenous (just like employment outcome), and determined by the activity that the household is engaged in. If factor markets functioned seamlessly this would be a major concern. Under imperfect factor markets, it is probably less endogenous to employment and income but could still be influenced by unobserved household characteristics. Adding household productive assets does not majorly alter the coefficient estimates of other the variables, hence endogeneity of these might not be a major concern in this case.

The estimated RNAE coefficient (β_1) in column 1 is statistically significant and ranges between 0.41 and 0.45. This means that the estimated earnings premium of changing from the agricultural to the non-agricultural sector is approximately between 42 and 45 percent, keeping other factors constant.

Column 3 shows the results of a regression with interaction effects, specified according to model (6). The coefficients suggest that there are positive returns to education in both sectors, higher returns to education in RNAE than in agriculture, and an unconditional earnings premium (not conditional on education) in RNAE. This is consistent with scenario S_3 discussed above. The RNAE coefficient is smaller (0.29) but still significant, and the interaction coefficient is positive. It suggests that the returns to an additional year of schooling are more than twice as high as in agriculture.

Some interesting deviations emerged when the subsample of laborers were used for the same regressions (the sample was reduced to only contain 552 observations). Without interaction effects, the *RNAE* and education coefficients in column 4 are fairly similar to those in column 1 (with the full sample). As soon as *RNAE* was interacted with education [again, according to model (6)], both the *RNAE* and the educational coefficients became insignificant and only the interaction coefficient is significant. This is consistent with scenario S_1 , that is, that there are no returns to education in agriculture, there is no unconditional *RNAE* premium, and the only way to gain positive returns to education is to work outside agriculture. This is what the “naive” theoretical model predicts as well. The interaction coefficient of about 0.06 suggests that the average-educated laborer would increase her income by 30 percent by shifting out of agricultural wage labor into non-agricultural wage labor. For the uneducated laborer, however, there would be no effect on income.

The results from the regressions on the subsample of self-employed tell yet a different story. The *RNAE* coefficient in column 6 (without interaction effects) is of about same magnitude as for the full sample. The educational coefficient is positive and significant but about half the magnitude compared to laborers. In column 7, which includes the interaction effect, the *RNAE* coefficient is significant but both the educational and the interaction coefficients are statistically insignificant. This is not consistent with any of the four scenarios discussed in the previous subsection, but instead suggests that, for self-employed, there is an earnings premium in the non-agricultural sector that is not conditional on education. Moreover, the results suggest that there is no returns to education in either sector, for self-employed. The insignificant educational coefficients, along with positive and significant coefficients for land and livestock ownership, suggest that productive assets other than human capital play the most important role for the income of the self-employed. But this conclusion should be drawn with caution. Laszlo (2005) suggests that although education may have little effect on individual earnings for the self-employed in rural Peru, there is a positive *allocative* effect of education on the income of the household. The allocative effect means that educated households are better than uneducated households at making optimal choices of income-generating activities. Once the optimal labor allocation is determined, education has little effect on labor productivity, Laszlo argues. Yang and An (2002) find positive returns to human capital in both agricultural and non-agricultural activities for households in rural China, and they too emphasize the role of human capital in the process of allocating factors of production among activities.

[Table 7]

Some other results can be noted in Table 7. There is a strong gender effect in the results of the various regression specifications. For the labor force as a whole, the effect on earnings of being male is around 65 percent. The effect is more evident among the self-employed than among wage labor. Somewhat to the contrary, López and della Maggiora (2000), in their household income analysis for rural Peru, find that female-headed households are at a disadvantage among non-agricultural households but not among farming households. The results further suggest that there is a strong negative effect of having a non-Spanish mother tongue. This ethnical effect, however, differs between wage labor and the self-employed. Among the self-employed the estimated negative effect is more than 50 percent, but in the wage labor sample the ethnicity coefficient is statistically insignificant. This finding, combined with the fact that indigenous groups are under-represented in the labor market, suggests that ethnicity is more of an entry barrier to the labor market than a source of wage discrimination once in the labor market. To what extent the economic marginalization of indigenous people is due to language barriers and ethnical discrimination in the labor market (Griffin et al. 2002), or merely adverse geographic location (López and della Maggiora, 2000), is an issue for further research. Despite structural differences in income determination between the self-employed and wage labor, no earnings differential between them is captured in the full-sample regressions when keeping other things

constant. The coefficient of the self-employment indicator is insignificant in all the full-sample specifications (see also footnote 3). As for regional differences in earnings, there is a negative effect, of about 30 percent, of living in the mountain or jungle regions, compared to the coastal region. This effect is stronger among the self-employed but is not observed for wage workers in the lowly populated jungle region. Productive assets increase income for the self-employed but not for wage labor, in line with what we would expect. The coefficient for household equipment has a negative sign. This is unexpected, but is most likely an effect of the high correlation between the household asset variables (ranging between 0.47 and 0.66). When these variables were included one at the time, the equipment coefficient was not negative.

Table 8 reports estimation results of model (7), which includes RNAE interaction effects with all other variables. The lower half of the table shows the interaction coefficients. A statistically significant interaction coefficient suggests that there is a structural difference in the parameter of the respective variable between the agricultural and non-agricultural sectors. Again, the regressions were first run on the whole sample, then on laborers, and then on self-employed. Throughout all six specifications in Table 8, the *RNAE* coefficient is statistically insignificant, which partly deviates from the results reported in Table 7. The differences between laborers and self-employed need emphasis. Let us therefore focus in particular on columns 4 and 6, which contain the results of the “full” model for laborers and self-employed, respectively. Column 4 suggests that for laborers there is no unconditional earnings premium in RNAE and there are no returns to education in agriculture. Work experience (represented by *age*) matters in agricultural wage labor, but still has twice higher returns in RNAE ($0.048+0.049$). There is no significant difference in magnitude of the gender premium for wage laborers between the two sectors. Column 6 contradicts parts of the findings about the self-employed in Table 7. Once all interaction effects are taken into account, there is no unconditional earnings premium for self-employed in the non-agricultural sector. Self-employed with a non-Spanish mother tongue appears to be at a much lower disadvantage in RNAE than in agriculture. There is also a strong regional effect in the results, indicating that farm income is much lower in the jungle lowlands than in the coastal region. At the same time, non-agricultural self-employment in this region appears to be considerably more lucrative than in agriculture in either region. The results in columns 1 and 2 of Table 8, containing the full sample, are the outcome of these differences between laborers and self-employed.

[Table 8]

A major potential weakness of applying the Mincerian wage equation approach to test for earnings differentials between sectors is that sector of employment is likely to be endogenous. The results from the instrumental-variable analysis (two-stage least squares), which was undertaken in an attempt to adjust for this endogeneity, are reported in Table 9. There are a total of nine columns in the table. The first five are for the whole sample, columns 6–7 on laborers, and columns 8–9 on self-employed. Columns 1–3 show the results of the most basic second-stage specification under three different sets of instrumental variables (the coefficients of x_{ij} and h_j in the first-stage regression are not shown). First, only the two household variables *household head* and *household size* are included. Only the second is statistically significant. Second, the set of community instruments are included: paved road distance to market, population size, of which two are significant. Third, all five instrumental variables are included, of which three are statistically significant. A linear combination of all the five variables was assumed to be the best instrument at hand, and was used for the other specifications in Table 9.

This attempt to control for endogeneity of *RNAE* creates inconclusive results regarding the *RNAE* coefficient. In column 3 it is negative and statistically significant. Adding household assets (column 4) and the interaction term (column 5) does not alter this result. The interaction variable (*RNAE X education*) in column 5, which was not instrumented for, has a positive coefficient. This suggesting that, while RNAE and education separately do not have any positive

effect on income, education has some positive returns in RNAE. The instrumental variables show less statistical significance for the subsample of laborers. Columns 6 and 7 show an insignificant *RNAE* coefficient, and positive or insignificant returns to education. The results for the subsample of self-employed are in line with the results for the full sample. In particular, column 9 shows negative *RNAE* and educational coefficients and a positive interaction coefficient.

[Table 9]

The empirical results in Tables 7–9 can be summarized as follows: The OLS results suggest that laborers only have positive returns to education in RNAE, and to benefit from RNAE they need some level of education. The results for the self-employed are slightly different. The results suggest that they have low or no returns to education in either sector, but that a switch of sector, out of agriculture into RNAE, is associated with increased income, given their level of productive assets. This “unconditional” premium, however, is not evident once a whole series of interaction effects were accounted for. The 2SLS results challenge the OLS results and suggest that, after adjusting for endogeneity of sector choice (RNAE), there is a negative income premium for RNAE for the uneducated but possibly a positive premium for the educated. Thus, in sum the empirical results give little support for any unconditional earnings premium in RNAE (unconditional in the sense of existing for educated as well as uneducated). Instead, and consistent with the naive theoretical model, only the educated seem able to gain a potential RNAE income premium. Returns to education seem overall to be higher in RNAE than in agriculture, particularly for laborers.

5. CONCLUSION

The underlying question that motivated this study was to what extent the rural non-agricultural sector can serve as a potential pathway out of poverty for some rural households. Over 70 percent of the rural population in Peru was estimated to live below the national poverty line in 2005. An equally large share of the rural population was engaged primarily in agriculture. This does not mean that agriculture equals poverty; yet the typical household in rural Peru is a poor farm household. Based on these characteristics, a simple farm household model was proposed to predict the earnings potential in the agricultural and non-agricultural sectors for the typical rural household. The key assumption was that there are positive returns to education in RNAE but not in agriculture. Therefore, with well-functioning markets, only skilled people will gain by working in the RNA sector and unskilled workers will have little to gain. These implications, however, are not necessarily valid as soon as market failures are allowed for.

The empirical results are somewhat mixed and might reveal some of the limitations of using the Mincerian wage-equation approach on cross-sectional data to assess earnings differentials between sectors. The ideal empirical study would use panel data to trace workers over time and assess income changes for workers changed sector compared to workers that remained in the same sector. The data source is from the mid-1990s, but comparison with descriptions of more recent data on Peru reveals that there has not been any major structural change in the rural labor force since then. Thus the results of using the same empirical method on more recent data are unlikely to differ in any large extent from the results presented here. While the limitations of the empirical approach should be kept in mind when interpreting them, the empirical results does not reject the basic predictions of the model. There is little support in the results for the notion that an unskilled worker would earn a higher income in RNAE than in agriculture. The results do, however, suggest that returns to education are higher in RNAE and hence that skilled people tend to do better in RNAE than in agriculture. This finding is robust across most of the regression specification, including the instrumental-variable approach.

Strategies aiming at strengthening the non-agricultural sources of income for the rural people need to go beyond educating the rural population and to eradicating potential entry barriers to the non-agricultural labour market. In particular, before we can establish with confidence for whom RNAE is the appropriate path out of poverty and a viable livelihood strategy, we need deeper understanding of several factors that are closely tied to RNAE. These factors include 1) the role of location for the viability of the RNA sector, 2) the importance of access to marketing channels, and 3) backward- and forward-linkages between agriculture and the RNA sector. If the local demand for non-agricultural goods and services is low, education and facilitation of access to RNAE are unlikely to suffice as a rural poverty alleviation strategy. With good marketing channels (including well-developed rural infrastructure), however, the RNA sector is less dependent on local demand, but can benefit from comparative advantages in the form of cheaper inputs compared to the urban economy.

REFERENCES

- Abramo, L. & M. E. Valenzuela (2005), "Women's Labor Force Participation Rates in Latin America", *International Labor Review*, 144(4), 369–399.
- Bardhan, P. & C. Udry (1999), *Development Microeconomics*. Oxford University Press, New York.
- Barrett, C. B., T. Reardon, & P. Webb (2001), "Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics, and Policy Implications", *Food Policy*, 26(4), 315–331.
- Dabalen, A., S. Paternostro, & G. Pierre (2004), "The Returns to Participation in the Nonfarm Sector in Rural Rwanda", World Bank Policy Research Working Paper 3462.
- de Ferranti, D., G. E. Perry, W. Foster, D. Lederman, & A. Valdés (2005), *Beyond the City – The Rural Contribution to Development*. World Bank, Washington D.C.
- de Janvry, A., M. Fafchamps, & E. Sadoulet (1991), "Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained", *The Economic Journal*, 101(409), 1400–1417.
- de Janvry, A. & E. Sadoulet (1993), "Rural Development in Latin America: Relinking Poverty Reduction to Growth", in *Including the Poor*, M. Lipton & J. van der Gaag (eds). World Bank, Washington D.C.
- de Janvry, A. & E. Sadoulet (2000), "Rural Poverty in Latin America: Determinants and Exit Paths", *Food policy*, 25(4), 389–409.
- de Janvry, A. & E. Sadoulet (2001), "Income Strategies Among Rural Households in Mexico: The Role of Off-farm Activities", *World Development* 29(3), 467–480.
- de Soto, H. (2001), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Black Swan, London.
- Dirven, M. (2004), "Rural Non-farm Employment and Rural Diversity in Latin America", *CEPAL Review*, 83, August.
- ECLAC (2006), *Statistical Yearbook for Latin America and the Caribbean, 2005*. United Nations Publication, Santiago.
- Echeverría, R. G. (2000), "Options for Rural Poverty Reduction in Latin America and the Caribbean", *CEPAL Review*, 70, April.
- Ellis, F. (2000), *Rural Livelihoods and Diversity in Developing Countries*. Oxford University Press, New York.
- Escobal, J. (2001), "The Determinants of Nonfarm Income Diversification in Rural Peru", *World Development* 29(3), 497–508.
- Escobal, J. (2004), "Rural Poverty in Peru: The Role of Agriculture and Non-agriculture Sectors", Background paper to World Bank report "Peru – Opportunities for All. Peru Poverty Assessment". Grade, Lima.
- Ferreira, F. H. G., & P. Lanjouw (2001), "Rural Nonfarm Activities and Poverty in the Brazilian Northeast", *World Development*, 29(3), 509–528.
- Griffin, K., A. R. Khan, & A. Ickowitz (2002), "Poverty and the Distribution of Land", *Journal of Agrarian Change*, 2(3), 279–330.

- Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1), 153-161.
- Isgut, A. E. (2004), "Non-farm Income and Employment in Rural Honduras: Assessing the Role of Locational Factors", *Journal of Development Studies*, 40(3), 59-86.
- Jonasson, E. & S. Helfand (2009), "How Important Is Economic Geography for Rural Non-agricultural Employment? Lessons from Brazil", working paper 2008:3, Department of Economics, Lund University.
- Key, N., E. Sadoulet, & A. de Janvry (2000), "Transactions Costs and Agricultural Household Supply Response", *American Journal of Agricultural Economics*, 82, 245-259.
- Lanjouw, J. O., & P. Lanjouw (2001), "The Rural Non-farm Sector: Issues and Evidence from Developing Countries", *Agricultural Economics*, 26, 1-23.
- Lanjouw, P. (1998), "Rural Nonagricultural Employment and Poverty in Ecuador", *Economic Development and Cultural Change*, 48(1), 91-122.
- Lanjouw, P. (2001), "Nonfarm Employment and Poverty in Rural El Salvador", *World Development*, 29(3), 529-547.
- Lanjouw, P. (2008), "Does the Rural Nonfarm Economy Contribute to Poverty Reduction?". In *Transforming the Rural Non-farm Economy: Opportunities and Threats in the Developing World*, S. Haggblade, P. Hazell, & T. Reardon (eds). John Hopkins University Press. Baltimore.
- Laszlo, S. (2005), "Self-employment Earnings and Returns to Education in Rural Peru", *Journal of Development Studies*, 41(7), 1247-1287.
- López, R., & C. della Maggiora (2000), "Rural Poverty in Peru: Stylized Facts and Analytics for Policy". In *Rural Poverty in Latin America*, R. López & A. Valdés (eds). Saint Martin's Press, New York.
- McCulloch, N., J. Weisbrod, & P. Timmer (2007), "Pathways out of Poverty during an Economic Crisis: An Empirical Assessment of Rural Indonesia", World Bank Policy Research Working Paper 4173.
- Reardon, T., J. E. Taylor; K. Stamoulis; P. Lanjouw; & A. Balisacan (2000), "Effects of Non-Farm Employment on Rural Income Inequality in Developing Countries: An Investment Perspective", *Journal of Agricultural Economics*, 51(2), 266-288.
- Sadoulet, E., A. de Janvry, & C Benjamin (1998), "Household Behaviour with Imperfect Labor Markets", *Industrial Relations*, 37(1), 85-108.
- Saith, A. (1992), *The Rural Non-farm Economy: Processes and Policies*. International Labor Office, Geneva.
- Stiglitz, J. E. (1988), "Economic Organization, Information, and Development". In *Handbook of Development Economics, Vol 1*. H. Chenery & T. N. Srinivasan (eds).
- Stiglitz, J. E. (1998), "Markets, Market Failures, and Development". In *International Agricultural Development, Third Edition*, C. K. Eicher & J. M. Staatz (eds). John Hopkins University Press, Baltimore.
- Taylor, J. E. & I. Adelman (2003), "Agricultural Household Models: Genesis, Evolution, and Extensions", *Review of Economics of the Household*, 1(1-2), 33-58.
- Valdés, A. & T. Wiens (1996), "Rural Poverty in Latin America and the Caribbean". In *Annual World Bank Conference on Development in Latin America and the Caribbean 1996*, S. J. Burki, S-R. Aiyer, & R. Hommes (eds). World Bank, Washington D.C.
- World Bank (2003), *Reaching the Rural Poor – A Renewed Strategy for Rural Development*. World Bank, Washington D.C.
- World Bank (2005), "Peru – Opportunities for All. Peru Poverty Assessment". Report No. 29825. World Bank, Washington D.C.
- World Bank (2007), *World Development Report 2008 – Agriculture for Development*. World Bank, Washington, D.C.
- Yang, D. T. & M. Y. An (2002), "Human Capital, Entrepreneurship, and Farm Household Earnings", *Journal of Development Studies*, 68(1), 65-88.

TABLE 1. *Poverty indicators, Peru 2004*

| | Poverty (head count ratio) | Poverty gap | Extreme poverty |
|-------------------|----------------------------|-------------|-----------------|
| National | 51.6 | 18.0 | 19.2 |
| Urban | 40.3 | 12.4 | 7.9 |
| Rural | 72.5 | 28.3 | 40.3 |
| Urban Costa | 37.1 | 10.6 | 6.2 |
| Metropolitan Lima | 36.6 | 10.4 | 3.4 |
| Rural Costa | 53.5 | 16.4 | 14.6 |
| Sierra | 67.7 | 27.2 | 36.5 |
| Selva | 59.5 | 19.7 | 26.4 |

Note: The extreme poverty line is around 110 Nuevo Sol (PEN), but varies by region and is estimated as the cost of a daily minimum food basket. Exchange rate of June 1994: 1 PEN=0.46 USD; PPP conversion rate 1994, 1 PEN=0.84 USD. The non-extreme poverty line is approximately twice the extreme poverty line, around 200 PEN per month, which is close to the US\$2/day standard. *Source:* World Bank (2005).

TABLE 2. *Rural household labor allocation by region, percent of weekly labor hours*

| | Costa | Sierra | Selva | Rural, total |
|------------------------|-------|--------|-------|--------------|
| <i>Agriculture</i> | 62.7 | 74.3 | 70.2 | 71.8 |
| Wage labor | 16.5 | 6.6 | 8.9 | 8.6 |
| Self-employment | 46.2 | 67.7 | 61.3 | 63.2 |
| <i>Non-agriculture</i> | 37.3 | 25.6 | 29.8 | 28.2 |
| Wage labor | 23.4 | 11.8 | 11.2 | 13.5 |
| Self-employment | 13.9 | 13.8 | 18.6 | 14.7 |

Source: Author's calculation based on the Peruvian LSMS, 1994.

TABLE 3. *Earned income in the rural workforce, 1994, local currency*

| | Wage laborer | Self-employed | Total |
|-----------------|--------------|---------------|-----------|
| Agriculture | 125 (4.9) | 202 (9.2) | 183 (7.1) |
| Non-agriculture | 199 (7.9) | 200 (11.9) | 200 (7.6) |
| Total | 161 (4.9) | 201 (7.4) | 188 (5.4) |

Note: Peruvian Nuevo Sol. Standard error is within parentheses. Unpaid family members are excluded. *Source:* Author's calculations based on the Peruvian LSMS, 1994.

TABLE 4. *Rural non-agricultural employment by sector*

| <i>Sector of employment</i> | <i>Percentage of non-agr. labor force</i> | <i>Percentage of self-employed</i> | <i>Percentage of wage labor</i> |
|--|---|------------------------------------|---------------------------------|
| Commerce | 34.4 | 51.1 | 5.9 |
| Manufacturing | 22.7 | 28.2 | 15.4 |
| Construction | 7.0 | 5.0 | 12.1 |
| Education | 6.5 | <1 | 17.6 |
| Domestic services | 5.9 | 1.8 | 13.5 |
| Transportation | 5.6 | 4.4 | 8.4 |
| Public administration | 4.4 | <1 | 11.5 |
| Hotels and restaurants | 4.0 | 3.1 | 4.5 |
| Fishing | 3.9 | 3.2 | 1.5 |
| Other public and private services | 1.7 | 1.8 | 2.3 |
| Social services | 1.6 | <1 | 3.6 |
| Real estate and business services | 1.1 | <1 | 2.7 |
| Other | 1.0 | <1 | 1.0 |
| <i>Non-agricultural sectors, total</i> | <i>100</i> | <i>100</i> | <i>100</i> |
| <i>Employment form:</i> | | | |
| Self-employed | 48.7 | | |
| Wage labor | 37.1 | | |
| Unpaid family member | 14.1 | | |

Note: Principal occupations only. Twenty percent of the paid rural labor force has a non-agricultural principal occupation. *Source:* Author's calculations based on the Peruvian LSMS, 1994.

TABLE 5. *Rural household labor allocation by income quintile (percent of weekly labor hours)*

| | Lowest | Second | Third | Fourth | Highest |
|------------------------|-------------|-------------|-------------|-------------|-------------|
| <i>Agriculture</i> | <i>78.4</i> | <i>75.6</i> | <i>71.6</i> | <i>68.1</i> | <i>64.6</i> |
| Wage labor | 8.8 | 11.6 | 8.6 | 7.1 | 6.8 |
| Self-employment | 69.6 | 64.0 | 63.0 | 61.0 | 57.8 |
| <i>Non-agriculture</i> | <i>21.6</i> | <i>24.4</i> | <i>28.5</i> | <i>31.9</i> | <i>35.5</i> |
| Wage labor | 12.6 | 13.8 | 13.4 | 15.0 | 12.8 |
| Self-employment | 9.0 | 10.6 | 15.1 | 16.9 | 22.7 |

Source: Author's calculation based on the Peruvian LSMS, 1994.

TABLE 6. Summary statistics of regression sample

| Variable | Sample mean | Standard deviation | Definition |
|-------------------------------|-------------|--------------------|---|
| Dependent variable | | | |
| Earned income | 217 | 272 | individual's earned income from principal employment (log), dep. variable |
| Independent variables | | | |
| RNAE | 0.35 | 0.48 | has principal occupation in the non-agricultural sector (d) |
| Age | 39.3 | 15.7 | individual's years of age |
| Male | 0.71 | 0.45 | gender, 1 for male, 0 for female (d) |
| Education | 5.0 | 3.67 | Individual's years of education |
| Non-Spanish | 0.31 | 0.46 | mother tongue other than Spanish (d) |
| Work migrant | 0.13 | 0.34 | individual has migrated for work (d) |
| Self-employed | 0.67 | 0.47 | self-employed in principal occupation (d) |
| Land ownership | 11.5 | 64.3 | hectares of owned land (log) |
| Livestock | 3.22 | 5.29 | owned livestock, cow equivalents, (log) |
| Equipment | 418 | 3463 | value of owned farm equipment, local currency (log) |
| Costa | 0.26 | 0.44 | individual lives in the coastal region (d) |
| Sierra | 0.51 | 0.50 | individual lives in the highlands (d) |
| Selva | 0.23 | 0.42 | individual lives in the jungle region (d) |
| Instrumental variables | | | |
| Household head | 0.61 | 0.49 | Individual is head of the household (d) |
| Household size | 5.97 | 2.75 | number of members in the household |
| Population size | 117 | 144 | number of households in nearest population center (log) |
| Paved road | 0.46 | 0.50 | paved or improved road in local population centre (d) |
| Distance to market | 10.7 | 19.3 | Distance (km) to the nearest market from local population center (log) |

Note: All variables were generated from the Peruvian LSMS, 1994. The number of observations is 1,680. Dummy variables are indicated by (d), taking value 1 if true, 0 otherwise. (log) indicates that the variable enters in log form in regressions.

Table 7. Estimation results – earned income (OLS)

| | All | | | Laborers | | Self-employed | |
|--------------------|-----------|-----------|-----------|-----------|-----------|---------------|-----------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| RNAE | 0.411*** | 0.448*** | 0.288*** | 0.428*** | 0.055 | 0.387*** | 0.466*** |
| years of education | 0.042*** | 0.039*** | 0.024** | 0.052*** | 0.013 | 0.027** | 0.025 |
| RNAE X education | | | 0.030** | | 0.063*** | | -0.004 |
| age | 0.037*** | 0.037*** | 0.037*** | 0.064*** | 0.063*** | 0.031** | 0.029** |
| age squared | -0.000*** | -0.000*** | -0.000*** | -0.001*** | -0.001*** | 0 | 0 |
| male | 0.660*** | 0.669*** | 0.677*** | 0.499*** | 0.546*** | 0.769*** | 0.786*** |
| non-Spanish | -0.425*** | -0.420*** | -0.426*** | -0.116 | -0.113 | -0.558*** | -0.552*** |
| work migrant | 0.026 | 0.032 | 0.021 | 0.189** | 0.137 | -0.075 | -0.057 |
| self-employed | 0.064 | 0.051 | 0.073 | | | | |
| Sierra | -0.289*** | -0.344*** | -0.345*** | -0.202*** | -0.217*** | -0.423*** | -0.505*** |
| Selva | -0.299*** | -0.307*** | -0.309*** | -0.015 | -0.002 | -0.499*** | -0.525*** |
| land ownership | | 0.016** | 0.015** | | 0.012 | | 0.019* |
| livestock | | 0.022** | 0.022** | | 0.008 | | 0.031*** |
| equipment | | -0.019** | -0.019** | | -0.013 | | -0.023* |
| Constant | 3.342*** | 3.506*** | 3.571*** | 2.837*** | 3.104*** | 3.716*** | 3.942*** |
| Observations | 1680 | 1680 | 1680 | 552 | 552 | 1128 | 1128 |
| R-squared | 0.164 | 0.17 | 0.172 | 0.309 | 0.331 | 0.155 | 0.163 |
| F statistic | 35.37 | 28.08 | 28 | 26.34 | 25.47 | 24.96 | 18.26 |

Note: Dependent variable is the log of earned income. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity.

Table 8. Estimation results – earned income (OLS with interaction terms)

| | <i>All</i> | | <i>Laborers</i> | | <i>Self-employed</i> | |
|-----------------------|------------|-----------|-----------------|-----------|----------------------|-----------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| RNAE | -0.378 | -0.63 | -0.618 | -0.59 | -0.089 | -0.369 |
| education | 0.026** | 0.023* | 0.023 | 0.021 | 0.027* | 0.021 |
| age | 0.032*** | 0.031** | 0.049*** | 0.048*** | 0.029* | 0.017 |
| age squared | -0.000* | 0 | -0.000*** | -0.000*** | 0 | 0 |
| male | 0.641*** | 0.628*** | 0.556*** | 0.565*** | 0.734*** | 0.688*** |
| non-Spanish | -0.526*** | -0.506*** | -0.074 | -0.067 | -0.661*** | -0.654*** |
| work migrant | 0.021 | 0.031 | 0.169 | 0.168 | -0.037 | -0.045 |
| self-employed | 0.143* | 0.073 | | | | |
| Sierra | -0.339*** | -0.447*** | -0.393*** | -0.375*** | -0.389*** | -0.524*** |
| Selva | -0.443*** | -0.457*** | 0.007 | 0.024 | -0.671*** | -0.642*** |
| land ownership | | 0.013 | | -0.007 | | 0.029** |
| livestock | | 0.047*** | | 0.015 | | 0.070*** |
| equipment | | -0.019 | | -0.009 | | 0.048 |
| RNAE X education | 0.029* | 0.030* | 0.046** | 0.042** | -0.003 | 0.005 |
| RNAE X age | 0.03 | 0.033* | 0.044* | 0.049** | 0.02 | 0.032 |
| RNAE X age squared | -0.000* | -0.000** | -0.001** | -0.001** | 0 | 0 |
| RNAE X male | 0.031 | 0.033 | -0.009 | -0.022 | 0.102 | 0.131 |
| RNAE X non-Spanish | 0.290** | 0.254** | -0.107 | -0.108 | 0.349** | 0.286* |
| RNAE X work migrant | -0.009 | -0.046 | -0.122 | -0.142 | -0.035 | -0.059 |
| RNAE X self-employed | -0.107 | -0.063 | 0 | 0 | 0 | 0 |
| RNAE X Sierra | 0.121 | 0.255** | 0.424*** | 0.395** | -0.092 | 0.107 |
| RNAE X Selva | 0.446*** | 0.512*** | -0.039 | 0.009 | 0.633*** | 0.668*** |
| RNAE X land ownership | | 0.011 | | 0.038** | | -0.016 |
| RNAE X livestock | | -0.056*** | | -0.015 | | -0.089*** |
| RNAE X equipment | | 0 | | -0.009 | | -0.066* |
| Constant | 3.560*** | 3.847*** | 3.210*** | 3.247*** | 3.848*** | 4.083*** |
| Observations | 1680 | 1680 | 552 | 552 | 1128 | 1128 |
| R-squared | 0.175 | 0.188 | 0.35 | 0.36 | 0.166 | 0.201 |
| F statistic | 23.57 | 18.77 | 23.31 | 18.07 | 15.11 | 13.11 |

Note: Dependent variable is the log of earned income. “RNAE X...” denotes interaction of the RNAE binary variable with the other respective independent variable. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity. A significant interaction coefficient indicates that coefficient estimates differ between agricultural and non-agricultural workers.

Table 9. *Empirical results - two-stage least squares*

| | <i>All</i> | | | | | <i>Laborers</i> | | <i>Self-employed</i> | |
|--------------------|------------|-----------|-----------|-----------|-----------|-----------------|-----------|----------------------|-----------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| RNAE | -1.467* | -0.866 | -1.100** | -1.154** | -2.228*** | 0.36 | -2.366 | -0.810* | -1.916** |
| education | 0.090*** | 0.075*** | 0.081*** | 0.082*** | -0.094** | 0.054** | -0.112 | 0.034** | -0.071** |
| RNAE X education | | | | | 0.319*** | | 0.33 | | 0.281*** |
| age | 0.048*** | 0.045*** | 0.046*** | 0.047*** | 0.041*** | 0.065*** | 0.063*** | 0.025* | 0.037*** |
| age squared | -0.000*** | -0.000*** | -0.000*** | -0.000*** | -0.000*** | -0.001*** | -0.001*** | 0 | -0.000** |
| male | -0.078 | 0.158 | 0.066 | 0.058 | 0.374*** | 0.484*** | 0.586*** | 0.284 | 0.377** |
| non-Spanish | -0.592*** | -0.539*** | -0.560*** | -0.576*** | -0.575*** | -0.108 | -0.134 | -0.718*** | -0.732*** |
| work migrant | 0.077 | 0.06 | 0.067 | 0.064 | -0.06 | 0.173** | 0.016 | -0.062 | -0.065 |
| self-employed | -0.169 | -0.095 | -0.124 | -0.103 | 0.189** | 0 | 0 | 0 | 0 |
| Sierra | -0.019 | -0.106 | -0.072 | -0.029 | -0.158 | -0.198 | -0.045 | -0.364*** | -0.321** |
| Selva | -0.273*** | -0.281*** | -0.278*** | -0.261*** | -0.301*** | 0.012 | -0.008 | -0.546*** | -0.527*** |
| land ownership | | | | 0.012 | 0.01 | 0.016 | -0.015 | 0.020* | 0.025** |
| livestock | | | | -0.017 | -0.008 | 0.005 | 0 | -0.002 | -0.006 |
| equipment | | | | -0.01 | -0.008 | -0.015 | 0.018 | -0.041*** | -0.033** |
| Constant | 4.233*** | 3.948*** | 4.059*** | 4.036*** | 4.519*** | 2.903*** | 3.863*** | 4.864*** | 4.929*** |
| Instruments | | | | | | | | | |
| household size | 0.006 | | 0.006 | 0.008** | 0.005* | 0 | 0.005 | 0.016*** | 0.008** |
| household head | -0.114*** | | -0.111*** | -0.133*** | -0.091*** | 0.072 | -0.091 | -0.221*** | -0.123*** |
| paved road | | 0.048** | 0.045** | 0.006 | 0.008 | 0.013 | 0.008 | 0.007 | -0.005 |
| distance to market | | -0.006 | -0.006 | -0.007 | -0.003 | 0.019* | -0.003 | -0.014* | -0.003 |
| population size | | 0.04*** | 0.041*** | 0.031*** | 0.016*** | 0.046*** | 0.016 | 0.021** | 0.019*** |
| Observations | 1680 | 1680 | 1680 | 1680 | 1680 | 552 | 552 | 1128 | 1128 |

Note: Dependent variable is the log of earned income. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity.

APPENDIX: CORRELATION MATRIX

Table A1. *Correlations between variables used in the empirical analysis*

| | inc | rnae | edu | age | male | nonsp | workm | self | land | live | equip | costa | sierra | selva | hhhead | hhsz | pop | pave | dist |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|--------|-------|-------|-------|------|
| income | 1.00 | | | | | | | | | | | | | | | | | | |
| rnae | -0.01 | 1.00 | | | | | | | | | | | | | | | | | |
| education | 0.09 | 0.23 | 1.00 | | | | | | | | | | | | | | | | |
| age | 0.17 | -0.22 | -0.37 | 1.00 | | | | | | | | | | | | | | | |
| male | 0.16 | -0.36 | 0.11 | 0.12 | 1.00 | | | | | | | | | | | | | | |
| nonspan | -0.13 | -0.08 | -0.15 | 0.05 | 0.00 | 1.00 | | | | | | | | | | | | | |
| workmig | 0.03 | -0.09 | -0.09 | 0.10 | 0.13 | -0.05 | 1.00 | | | | | | | | | | | | |
| selfemp | 0.13 | -0.14 | -0.26 | 0.34 | -0.05 | 0.09 | 0.03 | 1.00 | | | | | | | | | | | |
| land | 0.09 | -0.12 | -0.02 | 0.16 | 0.09 | 0.10 | -0.04 | 0.22 | 1.00 | | | | | | | | | | |
| livestock | 0.06 | -0.20 | -0.03 | 0.15 | 0.07 | 0.06 | -0.05 | 0.22 | 0.47 | 1.00 | | | | | | | | | |
| equip | 0.08 | -0.15 | -0.06 | 0.16 | 0.11 | 0.05 | 0.00 | 0.27 | 0.63 | 0.66 | 1.00 | | | | | | | | |
| costa | 0.13 | -0.02 | 0.09 | -0.02 | -0.02 | -0.27 | 0.08 | -0.17 | -0.20 | -0.21 | -0.19 | 1.00 | | | | | | | |
| sierra | -0.09 | 0.09 | -0.01 | 0.07 | -0.03 | 0.36 | -0.26 | 0.11 | 0.19 | 0.24 | 0.13 | -0.60 | 1.00 | | | | | | |
| selva | -0.02 | -0.09 | -0.08 | -0.07 | 0.06 | -0.15 | 0.22 | 0.04 | -0.02 | -0.07 | 0.04 | -0.32 | -0.57 | 1.00 | | | | | |
| hhhead | 0.17 | -0.35 | -0.16 | 0.48 | 0.54 | 0.09 | 0.20 | 0.27 | 0.08 | 0.05 | 0.06 | -0.08 | 0.01 | 0.06 | 1.00 | | | | |
| hhsz | -0.08 | 0.07 | -0.02 | -0.17 | -0.01 | 0.00 | -0.07 | -0.14 | -0.02 | 0.07 | 0.06 | -0.01 | -0.01 | 0.03 | -0.25 | 1.00 | | | |
| population | -0.04 | 0.13 | 0.09 | 0.01 | -0.02 | -0.01 | -0.03 | -0.02 | -0.02 | -0.12 | -0.05 | 0.06 | 0.13 | -0.21 | 0.02 | -0.03 | 1.00 | | |
| paved road | -0.06 | 0.09 | 0.09 | -0.06 | -0.03 | 0.11 | 0.05 | -0.09 | -0.13 | -0.27 | -0.25 | 0.06 | -0.06 | 0.01 | -0.06 | -0.03 | 0.11 | 1.00 | |
| distance | -0.05 | 0.00 | -0.03 | 0.01 | 0.00 | -0.10 | 0.08 | 0.07 | 0.06 | 0.03 | 0.04 | -0.10 | -0.01 | 0.11 | 0.03 | 0.04 | -0.05 | -0.08 | 1.00 |

Note: The sample is the same as in Table 6 ($N=1,680$). See Table 6 for definitions.