INTERNAL MIGRATION AND BORROWING CONSTRAINTS: EVIDENCE FROM PERU[‡]

Sonia Laszlo
McGill University[§]

Eric Santor
International Department, Bank of Canada

JEL Classification: O15
Keywords: Economic Development, Migration, Credit Constraints.

ABSTRACT

The decision to migrate has received substantial attention from both theoretical and empirical perspectives. Underlying most analyses is the desire to understand why individuals relocate within their own country, or more drastically, migrate to another country. While there are numerous reasons to migrate, economists have focused their research on the notion that there are gains to be made from migration: ceteris paribus, migrants are expected to earn more than non-migrants (Todaro, 1989). This paper utilizes a rich data set from Peru to assess the determinants of migration. We find that, when controlling for self-selection, migrants do not earn more than "stayers."

[‡] We thank John Hoddinott, Sendhil Mullainathan, and participants at the North East Universities Development Conference 2001 in Boston, Econometric Society Meetings 2002 in Atlanta and at the Bank of Canada for their useful comments.
[§] Corresponding author: Sonia Laszlo, Department of Economics, McGill University 855 Sherbrooke Street West Montreal, Quebec, H3A 2T7 , Tel: 514-398-1924, Fax: 514-398-4938, e-mail:sonia.laszlo@mcgill.ca

1 INTRODUCTION

The decision to migrate has received substantial attention from both theoretical and empirical perspectives. Underlying most analyses is the desire to understand why individuals relocate within their own country, or more drastically, migrate to another country. While there are numerous reasons to migrate, economists have focused their research on the notion that there are gains to be made from migration: ceteris paribus, migrants are expected to earn more than non-migrants (Todaro, 1989).[1] Consequently, the determination of the earnings differential between migrants and non-migrants has occupied the bulk of the research agenda.

Recent empirical work on explaining the earnings differential has emphasized the need to control for the standard problems of self-selection (Borjas, 1990; Pessino, 1991; Lanzona, 1998; Tunali, 2000 and Ham, Li and Reagan (2003)). That is, since migrant households may have different levels of human capital, or unobservable characteristics that are correlated to higher earnings, failure to control for such individual level heterogeneity may lead to erroneous conclusions regarding the returns to migration. The empirical techniques utilized in such studies follow methodologies laid out, in particular, by Heckman (1974, 1976, 1978) and Lee (1978, 1979). The empirical strategy must therefore account for the fact that individuals sort themselves into migrant, non-migrant status based upon their expected earnings given their move status. The probability that an individual migrates is then estimated by reduced form methods and the resulting "propensity" score utilized to adjust a subsequent earnings equation. In this way, the propensity to migrate estimate can be used to control for the unobserved heterogeneity of migrants. This is the strategy utilized by Ham, Li and Reagan (2003) in their analysis of the impact of migration on wages. In our study, we compare standard instrumental variable and propensity score matching techniques in our analysis of the returns to migration.

[1] There are numerous other reasons why individuals may migrate. Risk diversification, familial obligations, the desire to leave tradition-bound localities or relocation owing to civil strife or natural disasters figure prominently in recent theoretical studies. See Carillo et al (1999) for a recent survey.

We follow the literature in addressing these issues in order to test the following hypothesis: do migrants earn more than non-migrants conditional on migration. This paper utilizes a rich data set from Peru to assess the determinants of migration and which groups benefit most from migration relative to 'stayers'. In particular, the effects of market failure on migration are explored: households' migration behaviour may be affected by borrowing constraints. Few papers have drawn a link between migration and credit constraints. Since much of the labour force in developing countries is engaged in some form of self-employment labour, credit constraints should not be considered independently of the migration decision. Rapoport (2002) and Mesnard (2004) argue that credit constraints are a determining factor in occupational choice and, since the migration decision is linked to occupational choice, their results imply a direct relationship between credit constraints and migration. We contribute to this small literature by investigating both directions of causality: the effect of credit constraint on migration as well as the effect of migration on access to credit.

The paper proceeds as follows. Section 2 outlines the empirical model and some econometric issues. Section 3 describes the data and Section 4 presents regression results. Section 5 discusses and provides some preliminary results of matching methods. Finally, section 6 concludes.

2 An Empirical Model of the Returns to Migration

The primary question facing economists regarding migration behaviour is whether there are returns to migration. Conceptually, it is possible to estimate the return to migration by comparing the present discounted earnings of the individual before migration to the present discounted earnings of the individual after migration, adjusting for the costs of migration. Individuals will choose to migrate if the net expected earnings from migration are positive. An alternative, yet related, conceptualization of this relationship can be represented by equation (1):

$$Y_i = \mathbf{X_i}\boldsymbol{\beta} + \gamma M_i + u_i \qquad (1)$$

where Y_i are the individual *i*'s earnings, *i* indexing over all individuals (migrant and non-migrant), $\mathbf{X_i}$ represent a vector of individual controls (such as individual characteristics, demographics, schooling and other controls correlated to the individual's earnings), M_i is a dummy variable indicating whether individual *i* is a migrant, and u_i is an unobservable error component of the individual's earnings. The disturbance term, u_i, is for the moment assumed to be normally distributed with zero mean. The returns to migration are captured in the parameter γ.

There are, however, a number of issues that may cause the estimation of equation (1) to yield biased estimates of the returns to migration. The issue of primary concern in this paper is the likely endogeneity of migration. If there is a pay-off to migration, then equation (1) yields a positive estimate of γ: the act of migration has in itself caused the individual to experience higher earnings. However, individuals with higher earnings are more likely to migrate since they are more able to overcome the fixed costs of migration. Consequently, it is not clear in which direction the causation works.

In addition, it may be difficult to empirically consider migration independent of earnings. In answering whether the individual is made better off by migrating (as proxied by the return to migration, γ), it is possible the individual migrated because the migrant expected his wages to be higher. This paper attempts to correct for this source of endogeneity by utilizing a two stage least squares approach with instrumental variables.

First, consider the individual's migration decision. As discussed above, the decision to migrate is not an exogenous event for the individual. The individual's decision to migrate will be correlated with individual characteristics (such as age, gender, household size, occupation and education) and a number of other correlates which may not be in and of themselves directly

correlated with that individual's labour market earnings.[2] Equation (2) represents the individual's migration process:

$$M_i = \mathbf{Z_i}\boldsymbol{\alpha} + v_i \qquad (2)$$

The vector $\mathbf{Z_i}$ includes the variables in vector $\mathbf{X_i}$ as well as a vector of instrumental variables $\mathbf{IV_i}$ which are correlated to the decision to migrate but which are *a priori* uncorrelated with the error term in the individual earnings regression in (1).

A candidate instrument for migration is the expected relative pay-off to moving relative to staying. An individual may decide to migrate if he or she sees that the earnings of her neighbours who have migrated are higher than the earnings of her neighbours who have stayed. Let us define $\overline{Y}_{mbi}$ to be the mean earnings of all individuals (other than individual *i*) who have migrated away from birth place *b*. Then let $\overline{Y}_{sbi}$ be the mean earnings of all individuals (other than individual *i*) born in birth place *b*, and who have not migrated. Furthermore, let us define $\overline{Y}_{bi}$ as the ratio $\overline{Y}_{mbi}/\overline{Y}_{sbi}$. From the perspective of the potential mover, $\overline{Y}_{bi}$ may proxy as the anticipated earnings premium of migration, or the anticipated relative payoff of migration. Theoretically, there is no *ex ante* reason to expect that $\overline{Y}_{bi}$, which represent averages over all other individuals, is correlated to the individual's actual earnings. Intuitively, the individual migrates if and only if this individual anticipates higher earnings, once we control for all other possible correlates, including the costs of migration.

The estimation procedure is a two stage least squares with instrumental variables procedure as follows. Using the instrument discussed above ($\overline{Y}_{bi}$), we estimate the first stage equation (2).

[2] Examples of such correlates are the costs of migration and the individual's *anticipated* – not *actual* – return.

From the estimation of equation (2), we obtain predicted migration $\hat{M}_i$. We then substitute M_i with $\hat{M}_i$ in equation (1), as in (3):

$$Y_i = \mathbf{X_i}\boldsymbol{\beta} + \gamma\hat{M}_i + u_i \quad (3)$$

The first stage, using $\overline{Y}_{bi}$ as an instrument for migration, in and of itself provides us with useful insightful information regarding what is actually driving individuals to move. We, of course, anticipate that the effect of $\overline{Y}_{bi}$ to be positively correlated with M_i: the larger the difference between the earnings of the neighbours which have migrated and the earnings of the neighbours that have stayed, the more likely the individual may think that it will payoff for her to move as well. Additionally, this instrument provides us with a consistency check: should we observe $\overline{Y}_{bi} < 1$, then any prediction for migration other than M_i=0 may suggest that individuals are migrating for reasons other than earnings.

Technically, this strategy would require that we observe $\overline{Y}_{bi}$ at the precise moment that the individual chooses to migrate. Unfortunately, the data employed in this paper does not provide such detailed information. The contention arises since average earnings are changing over time; however, this only causes a problem in the event that there are regional differences in the growth of average earnings (that is, if $\overline{Y}_{mbi}$ grows at a rate different than $\overline{Y}_{sbi}$). Given the limitations of cross-sectional household surveys such as the one utilized here, careful treatment of these regional effects will be difficult to conduct other than by simply including regional fixed effects. In other words, the empirical analysis that follows will also test model (3) with the following error structure, as a sensitivity analysis:

$$v_i = v_{ib} = \eta_i + \varepsilon_b \quad (4)$$

In equation (4), η_i represents the individual specific error term and ε_b represents the 'place of birth' fixed effect.

Other Concerns

It is likely that there is a selection effect that may yield an upward bias to the returns to migration: individuals who migrate may simply have a higher propensity to migrate than stayers, all else constant, and a positive estimate for γ may simply be picking up the possibility that this innate higher propensity to migrate is correlated with the individual's ability to earn higher earnings. However, this bias should only be problematic in the case that we are not able to control for this innate ability with the observable variables. This paper presumes that the propensity to migrate is picked up by the observable individual characteristics in $\mathbf{X_i}$ and $\mathbf{Z_i}$. Many conventional unobservable characteristics are in fact observable given the rich nature of the data set.

Migration and Credit

There is considerable difficulty in assessing *a priori* the likelihood that an individual who migrates has differential access to credit. The interaction between migration and credit can be considered from two perspectives: the impact of credit on the ability to migrate and the ability of an individual, after migrating, to access credit. Each will be considered in turn. If the process of migration entails a significant fixed cost, then it is reasonable to assume that migrants, in overcoming these costs, are less credit constrained than stayers. That is, migrants were able to save the necessary funds to migrate (and therefore are more likely to have fixed assets for collateral purposes) or, if liquidity constrained, were able to access credit to overcome that constraint. In either case, migrants should be less credit constrained than stayers (see also Mesnard (2004) and Rapoport (2002)).

The impact of migration on credit, however, is less clear. If credit markets in less developed countries rely heavily on informal methods to assess credit history (in order to overcome the problems of asymmetric information in the absence of collateral), then stayers should have a

relative advantage over migrants. This is simply due to the fact that stayers have already established reputations, networks and other forms of "social capital" that can be utilized to access credit. Conversely, migrants should be at a considerable disadvantage with respect to these characteristics and thus are less likely to be able to access both formal and informal forms of credit. However, this relative disadvantage in terms of credit access may be mitigated by the nature of migrants themselves. First, since migrants were able to incur the fixed cost associated with migration, this would imply that they, as a group, possess the ability to accumulate the necessary funds to migrate or gain access to such funds from credit markets. Likewise, if migrants are a self-selected group, they would have those characteristics, both observable and unobservable, that would enable them to generate contacts, social networks and other forms of social capital that would allow them to have greater access to credit markets.[3] In sum, it is most likely the case that migrants have greater access to credit than stayers, given the former's proven ability to overcome the costs associated with migration. Furthermore, as a self-selected group of motivated individuals, migrants are more able to generate the contacts and social ties that play an important role in the process of accessing credit.

3 DATA

The primary source of data utilized in our analysis is a nationally representative household survey from Peru conducted by the Instituto Cuánto with technical and financial assistance from the World Bank. The survey is the 1991 round of the Peruvian Living Standards Survey (PLSS) and covers a multitude of topics from education, migration and health to labour market information for up to three jobs, non-farm self-employment business operations and agricultural production. The survey sampled 2,308 households from various regions, including 585 rural

[3] One would also expect that the initial reputational disadvantage associated with migration would dissipate with time.

households.[4] The average household consists of 5 to 6 members currently residing in the same dwelling, such that more than 11,000 individuals are represented in the 1991 PLSS.[5] More complete information regarding this data set is found in World Bank (1991a, b), Grosh and Glewwe (1995).

Migration

Respondents were asked to report their migratory history for all household members aged 15 and above. Migration is a common aspect of the Peruvian economy. Almost half of all surveyed individuals reported having migrated at least once. At a more aggregated level, almost three quarters of all surveyed households contain at least one migrant among their members. Though a complete migration history is not available, the 1991 PLSS does provide information on the place of birth, on the reasons for leaving the place of birth and on the details of the last migratory event. In addition to standard migration analyses, this information permits an investigation of both the geographic direction of migrants as well as motives for migration.

Table 1 presents means for the direction of migration and reasons for first leaving the place of birth. Of all migrants, 21 percent returned to their place of birth.[6] Of the remaining 79 percent, most migrants currently live in an urban area with a large proportion in Metropolitan Lima. Nonetheless, over 10 percent of migrants never returning to their place of birth has migrated to rural areas. Table 1 provides some indication of heterogeneity in terms of the direction of migration and indicates that migration from rural areas to the metropolitan centre (in this case Lima), the traditional study group for internal migration issues, only accounts for less than one quarter of all migration.

[4] Some departments in the Sierra and the Selva (Amazon jungle) were excluded from the study because of their remoteness or terrorist activities.
[5] 'Current' refers to the time of the survey.
[6] Unfortunately, the data are not informative regarding the migratory history of this group so that where these people have moved to is unknown to the researcher.

The reasons for first leaving the place of birth vary depending on whether the individual has returned to the place of birth. Migrants returning to their home province are more likely to have left for employment reasons than those who have never returned. The largest single reason for migration remains employment in either case. The striking result in this table is the large number of respondents reporting 'other' as their primary reason for migration. Given the political tension in the early 90s, it is possible that respondents were reluctant to specify the exact reason for having migrated.[7] Other major reasons reported for migration are 'education', 'more income' and 'marriage'.

Individual and Household Characteristics

Table 2 provides the means of the variables used in the empirical analysis to follow. The data is split into different samples according to their migratory behaviour. The first column provides the means for all individuals. Stayers and migrants are presented in the second and third columns. The last two columns split the sample of all migrants into those who have returned to their birthplace ('return migrants') and those who have not ('settled migrants'). For reasons that are explained when we discuss our instrument, we keep only observations for which the individual's place of birth (defined as the province for the rural sample and the department for the urban sample) is included in the PLSS.

Comparing the first two columns, we find interesting patterns between migrants and non-migrants ('stayers'). Since a majority of migrants leave their birthplace for employment reasons, it is not surprising to find that migrants have a higher participation rate in 'work for pay' than stayers. Individuals are considered to work for pay if they either work for a wage or are self-

[7] In fact, it is possible that the wording of the category 'terrorism' is too restrictive: the era of the Shining Path and counter-revolutionary activity was more widely known as *la violencia*, and not *terrorismo*, especially in the Andes.

employed. Excluded from this definition are individuals who do not receive any compensation for their labour and do not work for the family business.

Mean individual earnings of those working for pay are comparable among migrants and stayers. However, migrants are more likely to currently be involved in non-farm self-employment than stayers.[8] These differences become much more pronounced when we split the sample of migrants into return and settled migrants. Return migrants are far more likely to work for pay than stayers. However, those actually working receive the same earnings on average in both groups. Earnings of settled migrants, on the other hand, are higher than for any other group. Return migrants are more likely to be engaged in farming, reflecting the high likelihood that they currently reside in rural areas.

Educational patterns are similar across samples. However, return migrants have the highest mean completed years of formal schooling, suggesting that some of the human capital, perhaps acquired in the cities, returns to the area of origin.[9] In contrast, settled migrants have higher average household years of schooling than any of the other samples. Since the majority of settled households currently reside in urban areas, the observed educational gap between categories of migrants may simply reflect the urban composition of the four samples.

The independent variables of primary interest are income and access to credit. The survey specifically asks the respondents whether they currently have access to credit. Though the numbers are small relative to the entire sample, migrants have more access to credit than stayers (irrespective of whether this credit is considered formal or not). Unfortunately, the 1991 PLSS does not provide information on whether individuals had access to credit prior to the migration event. Since access is thus only observed at the time of the survey, the fact that migrants have

[8] Non-farm self-employment includes all activities that do not produce raw agricultural or livestock goods and services, and includes all individuals who work for their own account or in the family business. For additional details on non-farm self-employment in the 1991 PLSS see Laszlo (forthcoming).
[9] In fact, almost 17% of migrants from rural areas cited education as the reason for migration in the 1991 PLSS.

more access to credit than stayers may indicate an endogenous relationship between borrowing constraints and migration. Indeed the causal direction is unclear. On the one hand, the individual may have certain characteristics correlated with the decision to migrate that are also correlated with having access to credit. In this case, if these characteristics are time invariant, then it possible that these individuals would have qualified for credit prior to the migration event, and so may have been less credit constrained in their migration decision. On the other hand, individuals may have chosen to migrate to areas where credit is more easily acquired. To circumvent the lack of information about credit constraints at the time of migration, we utilize the proportion of households in the place of residence (province in rural areas or department in urban areas) with access to credit. If the relative geographic distribution in the changes in access to markets is time-invariant, this solution should work reasonably well.

Table 2 shows that most credit sources are informal: individuals are most likely to borrow from another individual or a business than through more formal mechanisms such as banks or *cooperativas*.[10] These patterns are common to all types of migrants and stayers: the most important source of credit for all individuals is from another individual, followed by a business, and only then from the formal banking sector. Migrants tend to have more access to credit (formal or informal) than stayers: they may have stronger social ties to their communities or may simply have more collateral.

Earnings of migrants of their households are higher than those of stayers. In particular, settled migrants have higher earnings than return migrants. Several hypotheses present themselves to explain this pattern. First, since the majority of settled migrants currently reside in urban areas, their higher earnings may simply be picking up a rural versus urban fixed effect: urban wages are larger than rural wages. Second, return migrants may have returned to their

[10] Informal banks (*banca paralela*) accounts for less than one percent of informal borrowing.

birthplace because they have not succeeded in their migration experience.[11] Third, though related to the second point, settled migrants may be successful in their migration for the same reasons that they are successful in the labour market. In other words, individuals who have a high propensity of success in the formal labour market self-selected themselves into the group of settled migrants. These reasons may be either observed (such as education) or unobserved (such as ability).[12] The empirical section that follows shall disentangle many of these effects.

In summary, the preliminary evidence provided by the descriptive statistics show that there are substantial differences in outcomes and characteristics of different types of migrants. The next section applies the model outlined in section 2 in order to determine whether much of these observed differences can be explained by differences in borrowing constraints and endogenous migration.

Constructing the instrument – Expected Returns

The variable that we propose to instrument for migration is the expected earnings gain for the migrant. Recalling the discussion in section 2 above, we construct the expected earnings as the ratio of the average earnings of those individuals who have migrated from the birthplace to the average earnings of the individuals who have remained in the place of birth. Intuitively, the individual will look at the average earnings of the villagers who have moved away (assuming, of course, that he is able to observe these – an assumption that is plausible if the migrant maintains

[11] This effect is reported in Pessino (1991). More intuitively, rural individuals may have migrated to Lima in search of employment or higher wages but were unsuccessful and so had no recourse but to return to their rural homes.

[12] 'Return' migration is only related to this success versus failure hypothesis if the ex-ante migration decision was a permanent one. It remains possible that return migration results from an ex-ante decision to temporarily migrate. For example, children may be sent to school in the city with the understanding that they return to their rural homes. Additionally, living in the city for a few years may be considered a rite of passage, and migrants return to their birthplace with a higher social standing as a result of their 'urban' experience.

social and family ties in the village) and compare that to the average village earnings.[13] If this individual observes a significant difference between the two, then he may consider migration if he views this difference as being the expected return of migration. Consequently, we expect that the individual is more likely to migrate if he observes a ratio larger than one than if the ratio is less than or equal to one.

Table 3 provides some evidence that the observed pattern between migration behaviour and the magnitude of the instrument is consistent with the prediction mentioned above. Individuals whose expected return (as measured by the ratio of average earnings of migrants to the average earnings of stayers) is less than or equal to one have a rate of migration of 16.7%. Those who instead observe a ratio greater than one have a rate of migration of 24.6%. In addition, for 75% of observations, the ratio itself is larger than one. The interpretation is as predicted: those who observe a higher expected return are more likely to migrate than those whose expected returns are negative. The fact that 16.7% migrate if the ratio is less than one may represent individuals who migrate for reasons other than earning or, more broadly, for reasons not related to labour markets.[14]

Several caveats must be highlighted. First, while it is possible that the 16.7% of migrants whose ratio is less than or equal to one have migrated for reasons not related to the labour market, it is also possible that they were not able observe the actual ratio as precisely as we (the researchers) are. The surveyors have access to earnings information that is confidential, thus it is possible, and likely in many cases, that this is information unknown to the individual. Nonetheless, we are confident that this information is a reasonable proxy for the ratio actually

[13] In fact, it is often the case, at least in some parts of the central Andes, that a migrant who has 'made it big' in the city temporarily returns to the village to host a '*fiesta*' to show off his wealth by redistributing some of it during the *fiesta*.
[14] Recall from Table 1 that only 6.86% of all migrants have migrated for more income, and only 22.7% have migrated for employment reasons.

observed by individuals.[15] Second, the construction of this variable relies on the place of birth. If a particular place of birth is located in a province or department that has not been surveyed in the PLSS, then the construction for the ratio will be impossible for this birthplace. For this reason, we have 'dropped' from the analysis all observations in which migrants have were born in a region that was not surveyed in the 1991 round of the PLSS. Recall from table 2 that the resulting sample of migrants is similar in means to the settled migrant group that includes all observations. Thirdly, and more worrisome, is the jurisdictional identification of the birthplace. For rural individuals (either born in or currently residing in a rural area), it is possible to identify the province of birth.[16] However, the urban identification is more complicated. If an individual currently resides in an urban area, the survey identifies the department of residence, but not the actual city (with the exception of metropolitan Lima-Callao). In other words, the instrument is at the provincial level for rural dwellers but at the department level for urban residents. Intuitively, this is not likely to cause a problem: there are few large cities in each of Peru's 25 departments and urban dwellers are more likely to look at the 'city' wage than a more disaggregated wage.

Finally, table 4 provides the first moments of the variables that are included in the empirical analysis presented next in section 4. The most striking aspect in table 4 pertains to the instrument. The mean is 1.8, meaning that, on average, the mean earnings of migrants born in region (province or department) *b* is 1.8 times the mean earnings of those who have stayed in region (province or department) *b*. This ratio is bounded by 0.299 to 6.598 in the sample.

4 OLS AND IV RESULTS

The estimated model for earnings (equation (1) in section 2), not corrected for endogenous migration, is presented in table 5. The OLS regression includes weights by household size in

[15] Ideally, we would want to have the difference of observed earnings at the time that the migration decision is made. However, the data required to do this in our case is not available.
[16] A province in Peru is a jurisdictional level below that of a department which itself is below that of the country as a whole. A province then includes a number of districts and populated centres and villages.

order to control for household specific heteroscedasticity. The standard errors reported in this procedures are also Huber-White corrected to account for any potential remaining sources of heteroscedasticity. Migrants have higher earnings, on average, than non-migrants, as suggested by the positive and significant coefficient on the migration dummy variable. Age is positively related to earnings, but at a decreasing rate, consistent with the standard life-cycle interpretation of earnings. Men make on average more than women, a trend common to most countries, but is particularly acute in Peru. Education is positively related to the individual's earnings.

Whether the household is engaged in farm self-employment positively affects the individual's earnings, while if the household is engaged in farm-wage employment, then the estimated effect is negative.[17] Intuitively, farm self-employment ventures may be more lucrative occupations and farm wage-employment occupations are not. Since this sample includes a large urban component, the farm self-employment ventures are most likely larger operations than the subsistence farm and the farm wage-employment is often some form of piece-rate work. These relative occupations at the household level may have direct effects on the individual earnings (for instance, in this case, whether the individual is actually working in the household's farm self-employment venture), or through spillover effects.[18] The analysis that follows will attempt to determine whether these migration estimates (which suggest a positive return to migration) are resulting from an endogeneity bias. The credit constraint effect, captured by the coefficient on the percentage of households in the place of birth who have access to credit, is statistically insignificant. Household size and the average years of household schooling (for adults) are included to capture the household human capital that may spillover to the individual's earnings profile.[19]

[17] Note that these dummy variables are not mutually exclusive – a household may operate a non-farm self-employment venture concurrently with running a farm.
[18] See Laszlo (forthcoming) for an in-depth discussion of these effects and their interpretations.
[19] See Basu et al. (2001) and Laszlo (forthcoming) for a discussion of this spillover effect.

To account for the potentially important regional effects, fixed effects models are also estimated. First, current place (province or department) of residence fixed effects are included. Comparing to the OLS estimates, this fixed effects model produces strikingly different results since now migrant status has no effect. Controlling for the place of residence, migrants appear similar to their 'non-migrant' counterparts. Second, place (province or department) of residence at birth effects are included instead. Now, migrant status has a significant positive effect of a larger magnitude than that estimated by OLS. Controlling for birthplace, migrants fare significantly better than their 'non-migrant' counterparts. Based on an estimated coefficient of 0.2300, migrants earn on average 1.259 time what 'non-migrants' earn.

Finally, one must acknowledge the potential self-selection bias of including only individuals currently employed (for a positive wage or salary). This may lead the returns to migration to be biased and is easily corrected for using standard Heckman techniques. This is done in table 6. The probability of working for pay (either by self- or wage-employment) is identified using the presence of young children in the household (ages 0 to 4 and ages 5 to 15), since young children are more likely to affect labour market participation than earnings. Table 6 provides the results of the first stage in the first column. The individual level variables are consistent with intuition: migrants, males, and individuals with more education are more likely working, an effect increasing at a decreasing rate with age. The presence of young children increases participation as it increases the number of dependents for whom to provide. The second stage results are presented in the second column. Comparing these results to the uncorrected OLS results in table 5, we find that controlling for selectivity in the labour market does not change our results. While the estimated lambda (Inverse Mills Ratio) indicates positive selection, it is insignificant.

The migration decision

Table 7 provides the results of estimating the probability of migration. Since migration is a binary choice variable, the models estimated and presented in the table is a variant of equation (2):

$$\Pr ob(M_i = 1) = \mathbf{Z_i}\boldsymbol{\alpha} + v_i \qquad (2')$$

where $\Pr ob(M_i = 1)$ is the probability that the individual migrates. We estimate this model using binary choice models such as the probit and logit.

The first two columns provide the probit results for the migration decision. The first includes the instrument. Of the individual characteristics, the only variables that predict migration is whether the individual currently resides in a rural area. The household characteristics fare better than the individual characteristics in the migration decision in terms of statistical significance.

Of the household occupational dummy variables, farm wage-employment activities are negatively related to the decision to migrate while non-farm wage-employment is positively so. This is to be anticipated if we consider that households engaged in wage farming may not be sufficiently wealthy to 'afford' the migration costs. However, if this interpretation is true, then we would expect that credit availability at the birth place would alleviate the migration costs and would increase the proportion of individuals moving. Yet, what we observe is the opposite. Another interpretation of the negative coefficient on farm wage-employment is that people (and entire households) are migrating away from rural areas (recall table 1), thus reducing the proportion of individuals (and households) engaged in farming. This is corroborated by the positive coefficient on whether the household engages in non-farm wage-employment: rural individuals migrate to the cities precisely because they are looking for (and obtaining) a more lucrative non-farm wage job replacing their previous, subsistence, farming job.

One of the most interesting findings of this exercise is the statistical significance of our variable of access to credit at the place of birth. We find that this variable has a statistically

significant and negative effect on the probability of migration.[20] If we believe that this variable is picking up access to credit at the time of migration, then our results suggest that borrowing constraints, rather than acting as a deterrent, are an incentive to migrate: credit constraints may be interpreted as indicating the lack of opportunities. These results suggest that credit access is not related to migration via the costs of migration. Rather, access to credit becomes a motivating factor. We will turn our attention in what follows on the determinants of access to credit as a function of migration to validate or refute this interpretation.

Finally, the instrument (the expected relative pay-off to migration) is a statistically significant and positive predictor of migration: the higher the average earnings of individuals who have migrated from my birthplace relative to the average earnings in my birthplace, the more likely I can interpret this as a signal that I may experience a similar return if I move, thus increasing my probability of migration. The second column repeats the analysis by excluding the instrument from the probit estimation. The Hausman test for the excluded variable is statistically insignificant, failing to reject that the model which includes the instrument is inconsistent.[21]

To check whether the results obtained in column (1) are sensitive to the error structure imposed by the probit model, equation (2') is also estimated with the logit model as well as the linear probability model. The results (presented in table 7, columns (3) and (4) respectively) are generally consistent with the probit estimates. This suggests that the migration decision results are robust to the empirical model chosen.

Access to Credit

We now turn our attention to the reverse causal direction between migration and credit. There are two alternative scenarios explaining how migration may affect credit. First, migrants

[20] In fact, we also ran the same regressions including also current access to credit for the individual (results not shown here), and found that current access to credit was insignificant.
[21] Note also the magnitude of the standard error relative to the estimated coefficient.

might have better access to credit because they are a positively self-selected group and what makes them successful migrant also makes them successful in obtaining credit. We would then expect a positive effect of migrant status on access to credit. However, our previous analysis of the determinants of migration (table 7) showed that none of the typical individual variables (such as education) was a predictive of migration behaviour. Second, and conversely, we might think that migration worsens the probability of obtaining credit: moving often means that social ties are cut and lenders in the migrants current area of residence have less information about the potential borrower's credit history and relative risk. Worswick (1999) finds evidence that recent immigrants to Canada are more likely to be credit constrained than their native born Canadian counterparts. Extrapolating to our case, we would expect internal migrant status to have a negative effect on access to credit.

The results from a probit model (presented in table 8) of access to credit support the second scenario: migrant have less access to credit than non-migrants. In addition, table 8 presents the results from multinomial estimation of access to credit by source, to see whether migration status has a differential effect on where individuals find sources of credit. The last three columns of the table provide the determinants of obtaining credit from a bank, a cooperative or 'financiera', or from a business, relative to obtaining credit from an individual. The multinomial logit results show that migrants are more likely to obtain credit from a cooperative or a business than from other individuals than non-migrants. The likely interpretation of this effect is that, in support of the second scenario above, migrants have cut social ties with their original communities and thus are less likely to find 'informal' sources of credit because members of their new communities have less information about their credit histories and their risk of default than about their non-migrant counter-parts.

Endogenous migration

Table 9 provides the results of the estimation of equation (3). Column (1) repeats the ordinary least squares (OLS) estimates of equation (1) originally presented in table 5. Column (2) presents the Two Stage Least Squares Instrumental Variable (2SLS-IV) results. The two models estimate very similar coefficients on all control variables other than migration. The migration dummy (instrumented with the expected relative returns) becomes statistically insignificant. The interpretation of this surprising result is that the returns to migration are soaked up by the observable characteristics causing the individual to migrate. This could be related to a selective migration story. Recall that not all individuals whose relative expected return is greater than one ($\overline{Y}_{bi} \geq 1$) migrate (see table 3 and the discussion of table 3 in section 3). It is possible that individuals who are able to perceive the potential returns because of some innate characteristic (say 'ability') will choose to migrate. This innate ability could also be related to the individual's ability to earn a higher wage. In other words, there is an unobserved characteristic (innate ability) that is positively correlated with both the decision (and ability) to migrate and their labour market performance. While the 2SLS estimates an insignificant migration effect, the Hausman tests fails to reject that OLS is the more consistent estimator for the regression of equation (3).

5 MATCHING METHODS

The measure of the returns to migration generated from the use of standard instrumental variable and self-selection techniques depends heavily on the notion that the migrant (treatment) and stayer (control) groups share common supports for the distribution of individual characteristics. That is, individuals in the respective groups are comparable across a range of characteristics, such as age, gender and education. However, if these characteristics differ significantly, in that the density function of the distribution of characteristics do not share common supports, then standard non-experimental self-selection corrections techniques can lead

to inaccurate assessments of the effects of migration (Heckman *et al,* 1996). This is due to the fact that standard non-experimental techniques rely on the strong assumption that there is a common effect across all individuals in the treatment group. But it could be the case that the treatment group may respond differently to the treatment, given its differing characteristics. For example, the treatment group may consist of young, well-educated individuals with high access to credit, while the control group may be older, poorly educated individuals operating their own businesses in rural areas. Consequently, the effect of migration may differ substantially between migrants and stayers, thus biasing standard selection model results.

More formally, a treatment effects model estimates the difference, γ , between migrants and stayers as:

$$\gamma = E(Y_1 \mid M = 1) - E(Y_0 \mid M = 0) \tag{4}$$

where Y_1 is earnings in the migration state, Y_0 is earnings in the stayer state, and $M = 1$ indicates the individual migrated and $M = 0$ indicates the individual stayed. But one needs to calculate the effect of the treatment (migration) on the treated (those who migrated):

$$\gamma_T = E(Y_1 \mid M = 1) - E(Y_0 \mid M = 1) \tag{5}$$

That is, the valid comparison is the outcome of those individuals that migrated to a group of individuals that are otherwise identical, except for the fact that they did not migrate (but could migrate). Unfortunately, most data sets do not contain the second term of the right hand side of (5), since it is not observed.[22]

A potential solution to the problem of assessing the effects of migration, given significant differences between migrants and stayers, is to create the counterfactual $E(Y_0 \mid M = 1)$ by

[22] A solution is for the researcher to create the right hand side of (5) through the implementation of a randomized experiment: individuals would be randomly relocated and compared to stayers. This would create a true control group sample analogue that could be used to determine the difference between the outcomes of migrants and the outcomes of stayers. While the implementation of randomized experiments has been successfully executed in certain settings, evaluation techniques of this sort would be simply inappropriate A large literature has evolved around the use of randomized experiments to evaluate job training programs. See Heckman et al (1999) for a complete survey.

matching migrants and stayers along observable characteristics. For every individual who migrates, one needs to find an individual that is identical in every respect except for the act of migration. For instance, if the migrant group consisted of young, urban-based and highly educated individuals, one would like to find similarly educated individuals from the same cohort and location. Rosenbaum and Rubin (1984) propose a solution to this problem, known as "matching methods." Instead of matching along individual characteristics X, one can match along $P(X)$, the probability that the individual migrated. In this way, one can generate consistent and unbiased estimates of the effect of migration on migrants.[23]

The limitation of utilizing the propensity score as a measure of "comparability" is determined by the availability of sufficient conditioning variables. If the decision to migrate is poorly estimated, the treatment and control groups will be poorly matched. In this way, the bias caused by selection on unobservables can be accentuated by the matching process (Smith and Todd, 2001). Given these considerations, we propose to report the results from the application of matching methods in addition to the standard descriptive statistics and regression results.

Matching methods are a useful tool to assess the validity of the results generated by the non-experimental techniques utilized in section 2 and reported in section 4. The first step is to determine whether the sample of migrants and stayers exhibit covariate "imbalance". That is, are the two groups comparable in terms of the observable X's. Table 10 presents t-statistics and standardized differences for the unmatched migrant and stayer groups. As can be readily seen, there are considerable differences between the two groups. For instance, the standardized differences typically exceed twenty percent and simple t-statistic tests reveal that along most dimensions, the migrant group is statistically different than the control group.

[23] See Smith and Todd (2001) and Ham, Li and Reagan (2003) for a thorough discussion of matching method techniques.

Table 11 presents results from implementing nearest neighbour matching techniques with replacement.[24] The propensity score is estimated with a probit regression using the same set of regressors as reported in Table 7, column (1). The results are striking. Comparing the earnings between migrants and stayers for the un-matched sample suggests that migrants had substantially higher earnings than stayers. However, implementation of simple nearest neighbour matching indicates that the difference between migrant and stayer earnings is insignificantly different than zero, as shown by the 95% confidence interval of the difference in earnings. This is true when the number of nearest neighbours is varied from 1 to 10, common supports are imposed, and the density is trimmed. Similarly, Table 12 checks the robustness of these results by implementing local linear regression matching. Again, there appears to be no significant difference in the earnings of migrants versus stayers.

These results differ from those using OLS and 2SLS. In section 4, we estimated a positive and significant migration effect. Correcting for endogeneity, the 2SLS results found no migration effect, though we could not rule out that the OLS results were more consistent than the 2SLS. In fact, the 2SLS standard error on the migration coefficient becomes very large, leading to a very imprecisely estimated coefficient. While our instrument appears to satisfy the requirements for an instrumental variable, it is still possible that it is a weak one. Similarly, the results from propensity score matching suggest that there is no statistically significant difference in the earnings between migrants and stayers.

Our analysis suggests that the returns to migration results are sensitive to the chosen estimation method. Naturally, this raises the question as to which model provides the most

[24] First, run a probit regression to generate $P(X)$, the propensity to migrate. Then, sort the data according to the estimate of $P(X)$ from highest to lowest. For each individual in the migrant group match it to an individual in the stayer group, in descending order and repeat until each migrant is matched with a stayer from the control group. Individuals from the treatment and control group that do not have a "nearest neighbour" are excluded from the sample, while some control individuals may act as the control for more than one treatment individual. Then, the earnings outcome is compared between the treatment group individuals, and the newly selected control group individuals. The exercise is repeated for the average of the closest three, five and 10 nearest neighbours.

reasonable estimate of the treatment effect of migration. The OLS estimates cannot be immediately ruled out, since the Hausman test failed to reject that OLS is the more consistent estimator with respect to 2SLS. Furthermore, while the matching method and 2SLS results suggest that the effects of migration are statistically zero, their respective confidence intervals encompass the point estimate from OLS estimation, even despite the improved precision of the matching estimator. Nevertheless, the statistical and intuitive properties of our 2SLS instrument suggest that endogeneity is indeed a problem, and therefore 2SLS estimates are appropriate in this setting. Similarly, the large degree of covariate imbalance augurs for a matching methods approach. Confidence in the implementation of this technique is confirmed by the strong probit results regarding estimation of the propensity to migrate. In both cases, the results are consistent with the notion that migration in of itself does not contribute to higher earnings for migrants. The OLS estimates, therefore, can be viewed in the following way: the reported positive effects of the migration dummy do not necessarily represent the return to migration, but rather, reflect the fact that those individuals who choose to migrate are predisposed towards higher earnings in general.

6 CONCLUSIONS

The evidence presented here finds that the returns to migration are positive, until we control for individual and household characteristics and for endogeneity: estimation by 2SLS and matching methods suggests that migration itself does not contribute to higher earnings. Access to credit has an ambiguous predicted effect on migration, and this ambiguity is substantiated by the failure to find a statistically significant coefficient either on the migration decision to migrate or on labour market earnings. Meanwhile, migration seems to have a significant effect on access to credit. In particular, migrants are less likely to have access to credit, but those that do are more likely to borrow from cooperatives and businesses than from more informal sources – we attribute this effect to the loss of social ties on the part of the borrower and reduced credit history information on the part of the lenders.

REFERENCES

Borjas, G. (1990) "Immigration and self-selection," in R. Freeman and J, Abowd, eds. *Immigration and the Labour Market* (Chicago: University of Chicago Press), 29-76

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy* (Baltimore: Johns Hopkins University Press).

Dehajia, R. and S. Wahba (1998) "Propensity Score Matching Methods for Non-Experimental Causal Studies." NBER Working Paper No. 6829.

Grosh, M. E., and P. Glewwe (1995) "A Guide to Living Standards Measurement Study Surveys and Their Data Sets," LSMS Working Paper 120. Washington, D.C.: World Bank.

Ham, John C. , Xianghong Li and Patricia B. Reagan. "Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men." Department of Economics and CHRR working paper, Ohio State University, November 2003.

Harris, J. and M. Todaro (1970) "Migration, Unemployment and Development: a Two-Sector Approach," *American Economic Review* 60, 126-142.

Heckman, J. (1974) "Shadow Prices, Market wages, and Labour Supply," *Econometrica* 42, 679-94.

_____ (1978) "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica* 46, 931-59

Heckman, J., Ichimura, H., Smith, J. and P. Todd. "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method." *Proceedings of the National Academy of Science* 93(23), 13146-420.

Heckman, J. and V. Hotz (1989) "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association* 84, 862-74.

Heckman, J. and R. Robb (1985) "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds. *Longitudinal Analysis of Labour Market Data* (Cambridge: Cambridge University Press), 156-245.

Katz, E. and O. Stark (1986) "Labour Migration and Risk Aversion in Less Developed Countries," *Journal of Labour Economics* 4, 134-149.

Lanzona, L. (1998) "Migration, Self-Selection and Earnings in Philippine Rural Communities," *Journal of Development Economics* 56, 27-50.

Laszlo, S. (forthcoming) "Self-Employment Earnings: An Empirical Analysis of Non-Farm Activities in Rural Peru," *Journal of Development Studies*

Lee, L. (1978) "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19, 415-33.

_____ (1982) "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies* 49, 355-372.

Mesnard, Alice. "Temporary Migration and Capital Market Imperfections." *Oxford Economic Papers* Vol. 56 (2) (April 2004): 242-62.

Pessino, C. (1991) "Sequential Migration Theory and Evidence from Peru," *Journal of Development Economics* 36, 55-87.

Rapoport, Hillel. "Migration, Credit Constraints and Self-Employment: A Simple Model of Occupational Choice, Inequality and Growth." *Economics Bulletin* Vol. 15 (7), 1-5

Rosenbaum, P. and D. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrica* 70, 41-55.

Rosenzweig, M and K. Wolpin (1988) "Migration Selectivity and the Effects of Public Programs," *Journal of Public Economics* 37, 265-289.

Ruud, P. (1983) "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification in Multinomial Discrete Choice Models," *Econometrica* 51, 225-8.
Schultz, T. (1988) "Heterogenous preferences and migration: self-selection, regional Prices and programs, and the behaviour of migrants in Colombia," *Res. Pop. Econ*. 6, 163-181.
Smith, J. and P. Todd (2000) "Does Matching overcome Lalonde's Critique of Nonexperimental Estimators." Mimeo, University of Western Ontario.
Tunali, I. (2000) "Rationality of Migration," *International Economic Review* 41, 893-920.
World Bank (1991) *Basic Information. Peru: Living Standards Measurement Survey (PLSS) 1991*. Washington D.C.: World Bank Poverty and Human Resources Division. (a)
_____ (1991) *Surveyor's Manual*. Lima: Cuánto S.A. (b)
Worswick, Christopher. "Credit Constraints and the Labour Supply of Immigrant families in Canada." *Canadian Journal of Economics* Vol. 32(1), (February 1999): 152-170.

Table 1 - Means of direction of migration and sources of credit

	All Migrants	Return Migrants	Settled Migrants
Number of Observations	3295	708	2587
Direction of Migration (% of Migrants)			
Rural to Rural	5.40	44.07	6.88
Rural to Urban	39.70	0.00	50.56
Urban to Rural	3.07	0.00	3.90
Urban to Urban	30.35	55.93	38.65
Rural to Metropolitan Lima-Callao	22.21	0.00	28.30
Urban to Metropolitan Lima-Callao	17.78	0.00	21.49
Reason First Left Birth Place (%)			
More Income	6.86	6.07	7.07
Employment	22.70	32.63	19.98
Education	17.45	19.21	16.97
Marriage	6.46	5.79	6.65
Terrorism	0.15	0.14	0.15
Other	46.37	36.16	49.17

Source: Authors' own calculations from 1991 PLSS

Table 2 - Variable means of migrants versus stayers

	All Observations	Never Migrated (Stayers)	All Migrants	Return Migrants	Settled Migrants
Number of Observations	2436	1551	885	403	482
Individual Characteristics					
Has migrated at least once	0.36	0	1	1	1
Works for pay	0.55	0.49	0.65	0.67	0.63
Log Individual Earnings (If working only)	6.49	6.32	6.71	6.44	6.95
Age	37	35	40	38	42
Gender (Male=1)	0.49	0.47	0.53	0.59	0.47
Marital Status (Married=1)	0.45	0.42	0.50	0.46	0.53
Language Spoken by Household Head is Indigenous	0.12	0.14	0.09	0.11	0.07
Years of Education	7.20	6.82	7.86	8.04	7.71
Currently Resides in Rural Area	0.47	0.57	0.29	0.52	0.10
Currently Resides in Metropolitan Lima	0.19	0.12	0.32	0.15	0.46
Is Non-Farm Self-Employed	0.25	0.21	0.34	0.31	0.36
Is Farm Self-Employed	0.03	0.03	0.04	0.07	0.02
Is Farm Wage-Employed	0.15	0.17	0.11	0.19	0.04
Is Non-Farm Wage-Employed	0.18	0.13	0.26	0.24	0.28
Born in Rural Area	0.59	0.59	0.60	0.64	0.56
Migrated Rural to Rural	0.01	0	0.04	0	0.07
Migrated Rural to Urban	0.10	0	0.26	0	0.48
Migrated Urban to Rural	0.00	0	0.01	0	0.02
Migrated Urban to Urban	0.08	0	0.23	0	0.42
Years Since Migration (Migrants only)	17.94	0	17.94	11.79	23.09
Household Characteristics					
Log Per Capita Household Earnings	5.67	5.46	6.06	5.76	6.30
Household Size	5.81	5.83	5.76	5.59	5.89
Average Household Years of Schooling	7.27	6.80	8.08	7.57	8.51
% of Household Members who have Migrated at Least Once	0.36	0.15	0.72	0.72	0.72
Household Engages in Non-Farm Self-Employment	0.53	0.48	0.61	0.59	0.62
Household Engages in Farm Self-Employment	0.09	0.10	0.09	0.13	0.05
Household Engages in Farm Wage-Employment	0.36	0.44	0.21	0.29	0.09
Household Engages in Non-Farm Wage-Employment	0.46	0.39	0.59	0.35	0.67
Household has Access to Credit	0.138	0.123	0.163	0.191	0.139
% of Households in Birth Place with Access to Credit	0.138	0.128	0.155	0.171	0.142
Source(%):					
Individual	54.86	54.97	54.86	66.23	41.79
Bank	11.81	16.75	11.81	14.29	8.96
Cooperative, 'Financiera'	9.72	7.85	9.72	5.19	14.93
Business	23.61	20.42	23.61	14.29	34.33
Banca Paralela	0.00	0.00	0.00	0.00	0.00

Note: The sample size of migrants is smaller than in Table 1 because this table only includes observations with complete data for all variables. We include in this table only those whose birth place is included in the sample, with all observations (except for the income and years since migration) non-missing. Source: Authors' own calculations from 1991 PLSS

Table 3 - Instrument for migration - Anticipated (or expected) earnings

	$\overline{Y_{bi}} \leq 1$	$\overline{Y_{bi}} > 1$
Migrate	55 (16.7%)	261 (24.6%)
Stay	274 (83.3%)	743 (75.4%)
Number of Observations	329 (100%)	1004 (100%)

Table 4 - Descriptive statistics for variables used in regressions

	Means	Standard Error	Min Value	Max Value
Number of Observations		*1333*		
Individual Characteristics				
Log Individual Earnings (from all sources)	6.489	1.235	1.386	11.657
Has Migrated and Settled	0.227	0.419	0	1
Age	37.837	13.983	16	81
Gender (Male=1)	0.653	0.476	0	1
Marital Status (Married=1)	0.519	0.500	0	1
Years of Education	7.770	4.806	0	19
Currently Resides in Metropolitan Lima	0.196	0.397	0	1
Born in Rural Area	0.583	0.493	0	1
Household Characteristics				
Household has Access to Credit	0.140	0.347	0	1
% of Households in Birth Place with Access to Credit	0.145	0.106	0	0.364
Average Years of Education of Household members (aged 15+)	7.499	3.717	0	16.667
Household Size	5.653	2.441	1	16
Household Engages in Non-Farm Self-Employment	0.632	0.483	0	1
Household Engages in Farm Self-Employment	0.107	0.309	0	1
Household Engages in Farm Wage-Employment	0.350	0.477	0	1
Household Engages in Non-Farm Wage-Employment	0.508	0.500	0	1
Instrument				
Y_{bi}	1.828	1.304	0.299	6.598

Source: Authors' own calculations from 1991 PLSS

1333 observations with non-missing variables (except for the income variable).

Settled migrants whose birth place is not included in the PLSS are excluded.

Table 5 - Raw OLS estimates

Dependent Variable	Log of Individual Earnings		
	OLS	Fixed Effects for Current Place of residence	Fixed Effects for Place of residence at Birth
Individual Characteristics			
Has Migrated and Settled	0.1591**	-0.0268	0.2300***
	(0.081)	(0.133)	(0.094)
Age	0.0764***	0.0713***	0.0720***
	(0.013)	(0.012)	(0.012)
Age Squared	-0.0008***	-0.0007***	-0.0007***
	(0.000)	(0.000)	(0.000)
Gender (Male=1)	0.5380***	0.4911***	0.4772***
	(0.078)	(0.064)	(0.064)
Marital Status (Married=1)	0.0555	0.1145*	0.1042
	(0.074)	(0.066)	(0.066)
Years of Education	0.0652**	0.0504**	0.0566**
	(0.028)	(0.025)	(0.025)
Years of Education Squared	-0.0012	-0.0002	-0.0004
	(0.002)	(0.001)	(0.001)
Rural	-0.5512***	--	-0.3531**
	(0.093)	--	(0.141)
Household Characteristics			
% of Households in Birth Place with Access to Credit	0.2538	0.7478	--
	(0.344)	(0.676)	--
Average Years of Education of Household members (aged 15+)	0.0231	0.0244	0.0219
	(0.018)	(0.016)	(0.015)
Household Size	-0.0118	-0.0215*	-0.0206*
	(0.015)	(0.013)	(0.012)
Household Engages in Non-Farm Self-Employment	0.0903	0.1163	0.1187
	(0.081)	(0.074)	(0.073)
Household Engages in Farm Self-Employment	0.2963**	0.2481**	0.2627**
	(0.127)	(0.105)	(0.103)
Household Engages in Farm Wage-Employment	-0.3454***	-0.2485***	-0.2402***
	(0.095)	(0.088)	(0.087)
Household Engages in Non-Farm Wage-Employment	-0.0030	0.0443	0.0536
	(0.076)	(0.070)	(0.069)
Constant Term	4.1420***	4.0302***	4.1792***
	(0.0329)	(0.302)	(0.288)
R-Squared	0.2871		
F-Statistic	28.71***		
Within Group R-Squared		0.1584	0.2116
Between Group R-Squared		0.6402	0.7399
Overall R-Squared		0.2440	0.2858
F-Statistic		17.18***	24.75***

Source: Authors' own calculations from 1991 PLSS

Only individuals who are earning positive incomes. 1333 observations. OLS regressions are weighted by the household size and the standard error (in brackets) are Huber-White corrected. ***: significant at 1%, **: significant at 5%. *: significant at 1%

Table 6 – Testing for self-selection in the earnings equation

	First Stage	Self-Selection Corrected OLS
Dependent Variable	**Earnings>0**	**Log Individual Earnings**
Individual Characteristics		
Has Migrated and Settled	0.2048***	0.1646***
	(0.094)	(0.081)
Age	0.1615***	0.0826***
	(0.013)	(0.015)
Age Squared	-0.0018***	-0.0008***
	(0.000)	(0.000)
Gender (Male=1)	1.1253***	0.5752***
	(0.072)	(0.095)
Marital Status (Married=1)	0.1318*	0.0597
	(0.077)	(0.073)
Years of Education	0.0474*	0.0667**
	(0.028)	(0.028)
Years of Education Squared	0.0013	-0.0012
	(0.002)	(0.001)
Rural	-0.2775***	-0.5579***
	(0.105)	(0.092)
Household Characteristics		
% of Households in Birth Place with Access to Credit	0.7087**	0.2789
	(0.360)	(0.345)
Average Years of Education of Household members (aged 15+)	-0.0575***	0.0212
	(0.017)	(0.018)
Household Size	-0.1577***	-0.0141
	(0.020)	(0.014)
Household Engages in Non-Farm Self-Employment	1.0029***	0.1191
	(0.082)	(0.083)
Household Engages in Farm Self-Employment	0.0803	0.2982**
	(0.129)	(0.125)
Household Engages in Farm Wage-Employment	0.6549***	-0.3276***
	(0.098)	(0.099)
Household Engages in Non-Farm Wage-Employment	0.4588***	-0.0220
	(0.076)	(0.075)
Number of children in the household aged 0 to 4 years	0.1914***	
	(0.043)	
Number of children in the household aged 5 to 15 years	0.1692***	
	(0.031)	
Lambda		0.0701
		(0.098)
Constant Term	-3.967***	3.9429***
	(0.311)	(0.439)
Number of Observations	2436	
Censored Observations	1103	
Uncensored Observations	1333	
Wald $\chi^2(14)$	410.12***	
Wald $\chi^2(1)$ test of independence across equations	0.51	

Source: Authors' own calculations from 1991 PLSS

Only individuals who are earning positive incomes. Regressions are weighted by the household size and the standard error (in brackets) are Huber-White corrected. ***: significant at 1%, **: significant at 5%. *: significant at 1%

Table 7 - First stage results and sensitivity analysis for migration decision

Dependent Variable	Prob(Individual has migrated)			
	Probit		Logit	Linear Probabity
	(1)	(2)	(3)	(4)
Individual Characteristics				
Age	0.0238	0.0278	0.0389	0.0069*
	(0.018)	(0.018)	(0.033)	(0.004)
Age Squared	-0.0001	-0.0001	-0.0002	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Gender (Male=1)	-0.0500	-0.0394	-0.0799	-0.0053
	(0.097)	(0.095)	(0.168)	(0.023)
Marital Status (Married=1)	0.1595	0.1248	0.2997*	0.0413*
	(0.100)	(0.097)	(0.173)	(0.023)
Years of Education	-0.0262	-0.0372	-0.0656	-0.0024
	(0.038)	(0.037)	(0.067)	(0.009)
Years of Education Squared	0.0005	0.0014	0.0019	-0.0001
	(0.002)	(0.002)	(0.004)	(0.000)
Rural	-1.2671***	-1.2107***	-2.4667***	-0.2787***
	(0.148)	(0.143)	(0.299)	(0.031)
Household Characteristics				
% of Households in Birth Place with Access to Credit	-1.3740***	-2.4785***	-2.2581**	-0.3357***
	(0.514)	(0.482)	(0.912)	(0.112)
Average Years of Education of Household members (aged 15+)	0.0290	0.0145	0.0412	0.0082
	(0.023)	(0.022)	(0.041)	(0.005)
Household Size	0.0123	0.0175	0.0056	0.0008
	(0.018)	(0.018)	(0.032)	(0.004)
Household Engages in Non-Farm Self-Employment	0.0471	0.0645	0.0618	0.0044
	(0.110)	(0.108)	(0.192)	(0.025)
Household Engages in Farm Self-Employment	0.0930	0.0455	0.0790	0.0038
	(0.176)	(0.174)	(0.337)	(0.036)
Household Engages in Farm Wage-Employment	-0.4224***	-0.4097***	-0.9390***	-0.0819***
	(0.146)	(0.143)	(0.284)	(0.030)
Household Engages in Non-Farm Wage-Employment	0.4900***	0.4474***	0.8594***	0.1130***
	(0.107)	(0.103)	(0.192)	(0.024)
Instrument				
Y_{bi}	0.2526***		0.4810***	0.0624***
	(0.039)		(0.073)	(0.009)
Constant Term	-1.7940***	-1.1550***	-2.8617***	-0.0076
	(0.456)	(0.437)	(0.811)	(0.103)
Pseudo R-Squared [Adjusted R-Squared]	0.2519	0.2226	0.2603	[0.2253]
LR Chi-Squared [F-Statistic]	359.36***	317.49***	371.31***	[26.82***]
Hausman Test for the instrument	18.46			

Source: Authors' own calculations from 1991 PLSS. Only individuals who are earning positive incomes. 1333 observations. Standard errors (in brackets) are Huber-White corrected. ***: significant at 1%, **: significant at 5%. *: significant at 1%.

Table 8 – Access to Credit

Dependent Variable	Has Access to Credit	Type of Credit Access (Relative to Loans Made From Individuals)		
	Probit	Multinomial Logit		
		Bank	Cooperative, 'Financiera'	Business
Individual Characteristics				
Log Individual Income	0.0156	0.3669	0.4645	0.1407
	(0.042)	(0.290)	(0.302)	(0.177)
Has Migrated and Settled	-0.1927*	0.1599	1.4223*	1.6989***
	(0.116)	(0.999)	(0.756)	(0.532)
Age	0.0110	0.1842	0.0912	-0.0398
	(0.019)	(0.118)	(0.152)	(0.101)
Age Squared	-0.0001	-0.0023	-0.0015	0.0002
	(0.000)	(0.001)	(0.002)	(0.001)
Gender (Male=1)	-0.1925*	-0.5796	-0.5491	-0.4410
	(0.099)	(0.682)	(0.580)	(0.427)
Marital Status (Married=1)	-0.0464	-1.6062**	-0.9523	-0.4555
	(0.100)	(0.751)	(0.705)	(0.502)
Years of Education	0.0188	-0.1938	0.3961	0.3907
	(0.041)	(0.249)	(0.304)	(0.288)
Years of Education Squared	-0.0002	0.0124	-0.0218	-0.0284*
	(0.002)	(0.014)	(0.014)	(0.015)
Rural	-0.1615	-0.0008	0.0609	-0.3251
	(0.147)	(0.815)	(0.779)	(0.688)
Household Characteristics				
Average Years of Education of Household members (aged 15+)	0.0644	0.3820**	0.0624	0.4048***
	(0.024)	(0.158)	(0.158)	(0.137)
Household Size	0.0578***	0.5786***	0.3119***	0.2048**
	(0.018)	(0.120)	(0.113)	(0.086)
Household Engages in Non-Farm Self-Employment	0.1205	-1.1830	0.5150	0.9686*
	(0.112)	(1.041)	(0.752)	(0.549)
Household Engages in Farm Self-Employment	0.5611***	2.0357**	-32.4307***	0.9722
	(0.148)	(1.034)	(0.650)	(0.790)
Household Engages in Farm Wage-Employment	-0.0059	2.6034***	1.2812*	1.3742*
	(0.138)	(0.907)	(0.720)	(0.764)
Household Engages in Non-Farm Wage-Employment	-0.1294	-2.9609***	-0.1899	-0.1113
	(0.099)	(0.721)	(0.655)	(0.607)
Constant	-2.1962***	-12.3028***	-9.8977***	-7.2120**
	(0.492)	(3.389)	(3.564)	(3.334)
Pseudo R-Squared	0.0799		0.2737	
LR Chi-Squared	75.82***		11704.71***	
N	1333		187	

Source: Authors' own calculations from 1991 PLSS. Only individuals who are earning positive incomes. Standard errors (in brackets) are Huber-White corrected.

Table 9 - 2SLS-IV estimates versus OLS estimates

Dependent Variable	Log of Individual Earnings	
	OLS	**2SLS-IV**
Individual Characteristics		
Has Migrated and Settled	0.1591**	-0.2708
	(0.081)	(0.553)
Age	0.0764***	0.0794***
	(0.013)	(0.014)
Age Squared	-0.0008***	-0.0008***
	(0.000)	(0.000)
Gender (Male=1)	0.5380***	0.5406***
	(0.078)	(0.079)
Marital Status (Married=1)	0.0555	0.0706
	(0.074)	(0.077)
Years of Education	0.0652**	0.0647**
	(0.028)	(0.029)
Years of Education Squared	-0.0012	-0.0013
	(0.002)	(0.002)
Rural	-0.5512***	-0.6649***
	(0.093)	(0.181)
Household Characteristics		
% of Households in Birth Place with Access to Credit	0.2538	0.0441
	(0.344)	(0.403)
Average Years of Education of Household members (aged 15+)	0.0231	0.0242
	(0.018)	(0.018)
Household Size	-0.0118	-0.0100
	(0.015)	(0.015)
Household Engages in Non-Farm Self-Employment	0.0903	0.1007
	(0.081)	(0.082)
Household Engages in Farm Self-Employment	0.2963**	0.3072**
	(0.127)	(0.130)
Household Engages in Farm Wage-Employment	-0.3454***	-0.3720***
	(0.095)	(0.106)
Household Engages in Non-Farm Wage-Employment	-0.0030	0.0134
	(0.076)	(0.095)
Constant Term	4.1420***	4.1815***
	(0.0329)	(0.342)
R-Squared	0.2871	0.2697
F-Statistic	28.71***	27.29***
Hausman Test Chi-Squared	1.07	

Source: Authors' own calculations from 1991 PLSS. Only individuals who are earning positive incomes. 1333 observations. OLS regressions are weighted by the household size and the standard error (in brackets) are Huber-White corrected. ***: significant at 1%, **: significant at 5%. *: significant at 1%

Table 10 – Un-matched Migrant and Stayer Groups

Select Variables	Migrants	Stayers	Standard Difference[+]	Two Sided T-statistic
Individual Characteristics				
Age	39.5	36.6	20.39	3.67
Gender (Male=1)	0.68	0.63	9.68	1.75
Marital Status (Married=1)	0.55	0.50	10.14	1.83
Years of Education	8.5	7.2	27.34	4.97
Rural	0.28	0.54	54.58	9.80
Household Characteristics				
% Household has Access to Credit	16.0	13.3	25.01	4.51
Average Years of Education	8.3	6.9	36.21	6.54
Household Size	5.8	5.6	4.38	1.37
Household Engages in Non-Farm Self-Employment	0.70	0.58	24.43	4.40
Household Engages in Farm Self-Employment	0.10	0.11	2.93	0.90
Household Engages in Farm Wage-Employment	0.23	0.44	46.82	8.36
Household Engages in Non-Farm Wage-Employment	0.63	0.42	43.22	7.80

+: The standardized difference in percent is the mean difference as a percentage of the average standard deviation: 100(x1-x2)/[(s12-s22)]/2 where each variable x1 and x2 are the sample means in the treated group and the control group and s12 and s22 are the corresponding sample variances. Standard differences and Two Sided T-Statistics are presented in absolute values. 'Source: Authors' own calculations from 1991 PLSS. N=1333.

Table 11 - Nearest Neighbour Matching

	Log Earnings			
	Migrants	Stayers	Migrants-Stayers	95% Confidence Interval
Un-Matched Data	6.71	6.32	0.39	0.25, 0.52
Nearest Neighbor Matching				
Nearest Neighbor (n = 1)	6.71	6.77	-0.06	-0.27, 0.15
Nearest Neighbor (n = 3)	6.71	6.73	-0.02	-0.20, 0.16
Nearest Neighbor (n = 5)	6.71	6.75	-0.04	-0.20, 0.12
Nearest Neighbor (n = 10)	6.71	6.78	-0.07	-0.22, 0.08
Nearest Neighbor Matching, Common Supports				
Nearest Neighbor (n = 1)	6.67	6.73	-0.05	-0.26, 0.15
Nearest Neighbor (n = 3)	6.67	6.69	-0.02	-0.19,0.15
Nearest Neighbor (n = 5)	6.67	6.72	-0.05	-0.20, 0.11
Nearest Neighbor (n = 10)	6.67	6.74	-0.07	-0.21, 0.07
Nearest Neighbor Matching, Common Supports, Trimming (0.02)				
Nearest Neighbor (n = 1)	6.70	6.76	-0.06	-0.26, 0.14
Nearest Neighbor (n = 3)	6.70	6.72	-0.02	-0.19, 0.15
Nearest Neighbor (n = 5)	6.70	6.74	-0.04	-0.20, 0.11
Nearest Neighbor (n = 10)	6.70	6.77	-0.07	-0.22, 0.08

Source: Authors' own calculations from 1991 PLSS. Nearest Neighbour macthing methods implemented. n is the number of neighbors used to calculate the control group average outcome. Common support indicates thast treatment observations with propensity scores greater (less) than the maximum (minimum) propenisty score for controls are dropped. Trimming eliminates 2% of the obersvations of the treatment group at the lowest segment of teh density of control group observations.

Table 12 - Local Linear Regression Matching Results

	Log Earnings			
	Migrants	**Stayers**	**Migrants-Stayers**	**95% Confidence Interval**
Un-Matched Data	6.71	6.32	0.39	0.25, 0.52
LLR	6.71	6.75	-0.04	-0.19, 0.11
LLR, Common Supports	6.67	6.72	-0.05	-0.19, 0.09
LLR, Common Supports, Trimmimg (0.02)	6.70	6.75	-0.05	-0.18, 0.09
LLR, Common Supports, Trimmimg (0.02), Bandwidth (0.06)	6.70	6.75	-0.05	-0.18, 0.09
LLR, Common Supports, Trimmimg (0.02), Bandwidth (0.08)	6.70	6.75	-0.05	-0.19, 0.09
LLR, Common Supports, Trimmimg (0.02), Bandwidth (0.10)	6.70	6.75	-0.05	-0.19, 0.09

Source: Authors' own calculations from 1991 PLSS. Local Linear Regression (LLR) macthing methods implemented. Common support indicates thast treatment observations with propensity scores greater (less) than the maximum (minimum) propenisty score for controls are dropped. Trimming eliminates 2% of the obersvations of the treatment group at the lowest segment of the density of control group observations. Bandwidth indicates varaiation in the bandwidth of the LLR.