View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Public Disclosure Authorized
Public Disclosure Authorized

Policy Research Working Paper 4728

Determinants of Choice of Migration Destination

Marcel Fafchamps
Forhad Shilpi

The World Bank
Development Research Group
Sustainable Rural and Urban Development Team
September 2008

Abstract

Internal migration plays an important role in moderating regional differences in well-being. This paper analyzes migrants' choice of destination, using Census and Living Standard Surveys data from Nepal. The paper examines how the choice of a migration destination is influenced by income differentials, distance, population density, social proximity, and amenities. The study finds population density and social proximity to have a strong significant effect: migrants move primarily to high population density areas where many people share their language and ethnic background. Better access to amenities is significant as well. Differentials in expected income and consumption expenditures across districts are found to be relatively less important in determining migration destination choice as their effects are smaller in magnitude than those of other determinants. The results of the study suggest that an improvement in amenities (such as the availability of paved roads) at the origin could slow down out-migration substantially.

This paper—a product of the Sustainable Rural and Urban Development Team, Development Research Group—is part of a larger effort in the department to understand the determinants of migration. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at fshilpi@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Determinants of Choice of Migration Destination*

Marcel Fafchamps†

Forhad Shilpi‡

* We thank Mans Soderbom and seminar participants at University of Gothenburg for their excellent comments. We are very grateful to Prem Sangraula and Central Bureau of Statistics of Nepal whose assistance with data was essential for the success of this endeavor. Financial support for this research was provided by the World Bank. The views expressed here are those of the authors and should not be attributed to the World Bank.

† Department of Economics, University of Oxford, Email: marcel.fafchamps@economics.ox.ac.uk.

‡ DECRG, The World Bank.

1 Introduction

There has been a long tradition of research on migration issues in the development literature (Greenwood 1975, Borjas 1994). Recent research has highlighted the methodological issues in estimating returns to migration, in assessing role of migration network in actual migration flows, and in evaluating effect of migration on economic well being. This literature has contributed significantly to the understanding of migration process and its impacts. But, with the exception of some on-going studies, there is little evidence on how migrants choose their destination, particularly in the context of developing countries.[1] This paper seeks to fill this gap in the literature. By focusing on the choice of destination, this research seeks to shed light on the respective role of various locational attributes in the choice of migration destination.

The literature on migrations maintains that differences in income and infrastructure – suitably corrected for price differentials – play a dominant role in the choice of a place to live. To investigate this issue, we develop an original empirical strategy focusing on the choice of destination conditional on the migration decision. This approach offers the advantage of eliminating possible biases resulting from unobserved individual heterogeneity. To allow for network effects, we also correct for correlation in the destination choice of migrants originating from the same location.

The econometric analysis seeks to identify the main factors influencing the choice of migration destination. We limit our analysis to adult males who have migrated outside their birth district for work reasons. We begin by constructing a measure of expected income differentials between the place of origin and all the possible migration destinations. These differentials are allowed to vary depending on observable migrant characteristics believed to affect labor market outcomes,

[1] For instance, Lall and Timmins (2008) are examining the factors that influence individuals' migration decisions in a number of developing countries. This study, among other things, focuses on hetergeneity in migration costs among different socio-economic groups and the role played by different amenities in the migration dicisions of different groups.

such as education and caste. We also construct measures of social proximity between a migrant's place of birth and each possible destination, using detailed available data on ethnicity, caste, language, and religion.

We also investigate a number of factors that may influence the choice of migration destination but have not received much attention in the existing literature. Fafchamps and Shilpi (2009) have shown that the subjective welfare cost of geographical isolation is high. To investigate this issue, we include regressors controlling for population density and for the average distance to various amenities. Fafchamps and Shilpi (2008) have further shown that migrants are concerned with their welfare relative to that of their birth district as well as to that in their destination location. We examine whether relative welfare considerations influence the choice of migration destination. Additional controls include distance and prices.

The empirical analysis is conducted using LSMS survey data as well as the 2001 population Census data from Nepal. The diverse terrain of Nepal along with geographical variation in amenities makes it ideal for our study. The mountainous nature of Nepal means that the country faces daunting challenges in the provision of transport and energy infrastructure. These challenges are unique to Nepal, however. Similar constraints are faced by many developing countries – or regions within such countries. There are also many non-mountainous countries that nevertheless suffer from serious geographical isolation because of the lack of roads. This applies, for instance, to much of sub-Saharan Africa. Many of the same factors are likely to affect migration patterns in these countries as well.

It has long been observed that migrants often are better educated than non-migrants.[2] Migrants may differ from non-migrants in terms of unobservables as well. A number of recent studies have sought to estimate returns to migration that are immune to selection on unob-

[2] A related strand of work points out that migration prospects raise investment in education (de Brauw and Giles, 2006; Batista and Vicente, 2008).

servables (Gabriel and Schmitz,1995; Akee, 2006; and Mckenzie, Gibson and Stillman, 2006). Their results suggest that simply comparing the earnings of migrants and non-migrants overestimates the return to migration. For instance, Mckenzie, Gibson and Stillman (2006) use an experimental design to show that ignoring selection bias leads to an overestimation of the gains from migration by 9 to 82 percent. Similar evidence is reported by researchers investigating the relationship between education and migration (Dahl, 2002).[3] Our empirical strategy sidesteps individual selection issues by controlling for individual fixed effects and by focusing on the choice of destination conditional on migrating, rather than on the decision to migrate itself.

The role of networks in the migration process has also attracted significant recent attention among economists. Carrington et al. (1996) argue that the presence of a large migrant population in the place of destination reduces migration costs and generates path dependence. They use this to explain the Great Black Migration of 1915-1960 in the US. In the same vein, Munshi (2003) investigates the role of interpersonal networks in helping Mexican migrant workers in the US. A similar conclusion is reached by Winters, de Janvry and Sadoulet (2001), also using Mexican migrants to the US, and by Uhlig (2006) for Germany.[4] Network effects also matter at the place of origin. Munshi and Rosenzweig (2005), for instance, show that strong mutual assistance networks in the place of origin discourages migration. Mora and Taylor (2006) reach similar conclusions.

We do not have data on social networks and therefore cannot control for network effects directly. We therefore seek to control for network effects indirectly. Network effects at the

[3] The view that it is the better educated and more able who migrate has not gone unchallenged, however (Borjas, 1994). According to Borjas' negative selection hypothesis, the less skilled are those most likely to migrate from countries/locations with a high skill premia and earnings inequality to countries/locations with a low skill premia and earnings inequality. Chiquiar and Hanson (2005) test and reject this hypothesis for Mexican immigrants in the US and conclude instead for intermediate selection.

[4] Using data on refugees resettled in various parts of the US, Beaman (2006) proposes a more complex story in which an influx of refugees initially overwhelms the network as it struggles to provide job relevant information, but has longer term positive effect as new migrants find their way into employment.

place of destination tend to favor migrants who are better connected with local residents – and therefore may have easier access to jobs, credit, information, etc. To capture such effects, we construct variables that measure social proximity between the migrant and the population mix at the destination. These variables proxy for network effects but also for possible discrimination. Network effects also generate correlation in migration decisions among individuals originating from the same place. This induces correlation in residuals for migrants having the same districts of origin, and can seriously affect inference. To correct for these effects, we cluster residuals by district of origin.

Results show that population density, social proximity, and access to amenities exert a strong influence on migrants' choice of destination. These results confirm earlier work on the factors affecting the subjective welfare cost of isolation (Fafchamps and Shilpi, 2008). Differentials in income and consumption expenditures play a less important role than anticipated.

The paper is organized as follows. The conceptual framework and testing strategy are presented in Section 2. The data is discussed in Section 3, together with the main characteristics of the studied population. Econometric results are presented in Section 4. Conclusions follow.

2 Conceptual framework

Geographical differences in welfare are expected to induce people to relocate. Migrations patterns thus provide valuable evidence regarding income differences – or more generally welfare differences – across space.

Where do these welfare differences come from? A frequent explanation of the migration flow in response to income differences is derived from the Roy's (1951) model of job selection where workers move to the location which provides the highest return to their skill and talent ("unobserved ability") (Gabrial and Schmitz, 1995; Dahl, 2002). According to the recent economic geography literature (Henderson, 1988; Fujita, Krugman and Venables, 1999), agglomeration economies resulting from learning externality and increasing returns cause certain activities to concentrate in a few urban locations which in turn attract workers to those locations. Lucas (2004) recently revisited the issue in the context of low income economies during the post-war period, focusing on the historical issue of rural-urban migration patterns in relation with urbanization. In his analysis, Lucas emphasizes the role of cities as places in which new immigrants can accumulate and earn returns on the skills required by modern production technologies. In this approach, differences in welfare across space are driven by differences in technology – and differences in technology result from agglomeration effects leading certain industries to locate in cities and to take the form of large-scale, modern firms (Fafchamps and Shilpi, 2003 and 2005). The predominance of large firms and the emphasis on modern technology would explain why returns to education are higher in cities and why migrants hoping to move there seek to acquire more education (e.g., de Brauw and Giles, 2006).

These observations are the starting point for our work. We are interested in the factors that incite people to move to a specific location. Standard migration models predict that some of these factors have to do with the gain from moving, others have to do with the cost – or risk – of moving. More formally, let us assume that individuals derive a different utility from residing in different locations. Let utility of individual h in location i be denoted U_i^h. The probability of migrating from i to s is expected to increase in the difference between $U_s^h - U_i^h$ and to fall with the cost C_{is}^h of moving from i to s. Our empirical strategy is to construct estimates of U_s^h and C_{is}^h for *all* locations to which a migrant h might have relocated within the study country, and to test whether migrants' choice of destination follows $U_s^h - U_i^h$ and C_{is}^h.

Following the literature, let us assume that utility U_i^h is a function of the income y_i^h (or consumption) that the individual can achieve in location i, of the prices p_i he or she faces, and a vector of location-specific amenities A_i (Bayoh, Irwin and Haab, 2006):

$$\begin{aligned} U_i^h &= U^h(y_i^h, p_i, A_i) \\ &\approx y_i^h - \alpha p_i + \beta A_i \end{aligned}$$

The above linear approximation forms the basis of our empirical estimation. Income y_i^h in turn depends on observable z^h and unobservable μ^h characteristics of individual h:

$$y_i^h = \delta_i + \eta_i z^h + \gamma_i \mu^h + \varepsilon_i^h \tag{1}$$

where ε_i^h is a disturbance independent of z^h and μ^h. Note that parameters η_i and γ_i vary across locations. This captures the idea that returns to talent differs with the mix of activities undertaken in that location (Fafchamps and Shilpi, 2005).

Individuals choose the location that gives them the highest expected utility. Let M_{is}^h describe h's choice of destinations: $M_{is}^h = 1$ if individual h migrates from location i to location s, and 0 otherwise. By construction, each individual only migrates to a single location. We have to control for the cost of migrating. If people are credit constrained, or if they are risk averse and there is friction in the circulation of information, they would not want to travel too far. There is also the issue of social interaction with neighbors and friends in the place of destination (for entertainment, mutual support, marriage market, etc.). As recent papers by Munshi (2003) and Beaman (2006) have shown, social networks also play a role in finding employment. Social distance may thus discourage movement.

We therefore assume that the cost of moving from i to s depends on the physical and social distance between i and s (e.g., including differences in religion, language, or caste). Let d_{is}^h denote a vector of physical and social distances, where we recognize that social distance depends on characteristics of individual h. We have:

$$\begin{aligned}\Pr(M_{is}^h = 1) &= \lambda\left(E(U_s^h - U_i^h | z^h, \mu^h) - \omega d_{is}^h\right) \\ &= \lambda(\delta_s - \delta_i + (\eta_s - \eta_i)\, z^h + (\gamma_s - \gamma_i)\, \mu^h \\ &\quad -\alpha(p_s - p_i) + \beta(A_s - A_i) - \omega d_{is}^h) \end{aligned} \tag{2}$$

where $\lambda(.)$ is the logit function. Since we condition on migrating, the dependent variable takes value 1 for one and only one destination. This means that we can only identify the effect of differences between destinations, not the likelihood of migrating itself. This is standard in multiple discrete choice estimation (Train, 2003).

In practice, we do not observe individual h in two locations at the same time. How can we estimate (2)? We proceed as follows. We begin by estimating equation (1), separately for each location. This yields an estimate of:

$$E[\widehat{y_s^h - y_i^h} | z^h] = \widehat{\delta}_s - \widehat{\delta}_i + (\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$$

for each possible destination. We then use $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ to estimate equation (2) for migrants only. If income differences drive migration, the coefficients of $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ should be positive and significant, and they should be equal.

How adequately does this approach take care of unobserved heterogeneity? We begin by noting that, in general $E[z^h \mu^h] \neq 0$: observable and unobservable talents are correlated. For those who wish to estimate the return to a specific individual characteristic z^h, this correlation is problematic. For our purpose, this correlation is good news. To see this, consider the extreme case in which μ^h is a deterministic function of z^h:

$$\mu^h = \lambda z^h$$

Inserting in (1), we get:

$$y_i^h = \delta_i + (\eta_i + \gamma_i \lambda) z^h + \varepsilon_i^h$$

In this case the estimated coefficient of z^h also captures the effect of unobserved heterogeneity on income:

$$E[\widehat{\eta}_i] = \eta_i + \gamma_i \lambda$$

and $\left(\widehat{\eta}_s - \widehat{\eta}_i\right) z^h$ in equation (2) controls for *both* observed and unobserved heterogeneity.

What happens if z^h and μ^h are only imperfectly correlated? Say we have:

$$\mu^h = \lambda z^h + v^h$$

with $E[v^h] = 0$ and $E[z^h v^h] = 0$. Inserting in (1), we get:

$$y_i^h = \delta_i + (\eta_i + \gamma_i \lambda) z^h + \gamma_i v^h + \varepsilon_i^h$$

It follows that:

$$p\lim[\widehat{\delta}_i] = \delta_i + \gamma_i p\lim[v^h] = \delta_i$$

For the above to hold, we need to estimate (1) on all individuals, migrants and non-migrants. This is not possible, of course, since migrants are not observed in their place of origin. Fortunately, in the studied country, the overwhelming majority of household heads still reside in their birth village, probably because the economic and psychological costs of migrating are high.

This means that the distribution of unobserved talent μ^h among district residents corresponds roughly to the distribution of talent in the population at large. This implies that the bias in estimating δ_i is probably small when we estimate (1) using data on district residents.

What of equation (2)? It can be rewritten:

$$\begin{aligned}
\Pr(M_{is}^h &= 1) = f_+[\delta_s - \delta_i + (\eta_s - \eta_i + \lambda(\gamma_s - \gamma_i))\, z^h \\
&\quad -\alpha(p_s - p_i) + \beta(A_s - A_i) - \omega d_{is}^h + u_{is}^h] \qquad (3) \\
u_{is}^h &\equiv (\gamma_s - \gamma_i)\, v^h
\end{aligned}$$

which shows that since v^h is uncorrelated with z^h by construction, $(\widehat{\eta}_s - \widehat{\eta}_i) z^h$ is uncorrelated with the disturbances. The above can thus be used to consistently test whether income differences drive the choice of migration destination.

We have discussed unobserved heterogeneity in income generation. There can also be unobserved heterogeneity in migration costs. We are particularly concerned about the large proportion of surveyed households who still live in their birth district. This population includes households who chose not to migrate, but also many households for whom the cost – or the risk – of migrating were probably too high. Munshi and Rosenzweig (2005), for instance, have shown that mutual insurance within castes in India provides a strong disincentive to migrate. The same probably applies to our study country, which is neighboring India. It follows that the decision not to migrate at all – $M_{ii}^h = 1$ – is distinct from the choice of a destination, conditional on migrating. To minimize the bias that self-selection into migration may generate, we drop M_{ii}^h and estimate (3) with migrants only. Since we have no data on individuals who have left the country, our analysis is only pertinent to internal migrants.

Estimation of model (3) is achieved as follows. We begin by generating, for each migrant,

$N-1$ observations on M_{is}^h and the regressors, where N is the number of possible locations.[5] We then estimate (3) by logit.[6] Since the same individual appears $N-1$ times, we have to correct for correlation between the different choices for the same individual h. We do so first by adding individual fixed effects. This takes care of much of the correlation. We also correct standard errors for clustering by district of origin. This takes care of possible peer effects, as would arise if individuals from a given location all tend to migrate to the same destination. Robust standard errors that cluster by district of origin also correct for negative correlation in errors across choices for the same individual, a possibility that fixed effects do not control for. Negative correlation is a serious issue here, a point that is discussed in more detail in the next section.

We worry about possible circularity resulting from general equilibrium effects (Dahl, 2002; Hojvat-Gallin, 2004; Borjas, 2006; Bayer, Khan and Timmins, 2008). If many people migrate to a specific location, such as the capital city, this is likely to affect wages, incomes, and access to amenities in that location.[7] This would generate a potential endogeneity bias due to the fact that incomes and amenities in that location result in part from the decision of many migrants to locate there.

To eliminate this bias, we use past data to estimate the income regression. More precisely, let T be the period for which we have income information and $T+t$ the period at which we

[5] The dropped observation corresponds to the location of origin M_{ii}^h which, as explained earlier, we do not include in the analysis since including M_{ii}^h would mean de facto including the decision of whether to migrate or not.

[6] McFadden (1974) has shown that, in multiple choice problems of the kind studied here, the application of logit estimation is justified if (1) the errors in each latent choice equation follow the extreme value distribution and (2) errors are independent across choices. See Train (2003), Chapter 3 for a detailed discussion. The estimation of models with correlated errors across choices requires either multiple integration or the use of Bayesian estimation techniques relying on Gibbs sampling. With a choice of over 70 possible destinations, multiple integration is out of the question. Gibbs sampling remains a possibility but would require extensive programming. We choose instead to keep the logit approach but to correct the standard errors for possible correlation in errors across choices. In our case the possible efficiency gain achieved by Bayesian methods does not appear to justify the programming cost.

[7] The effect could be negative – e.g., congestion – or positive – e.g., agglomeration externalities.

observe migrants. The income regression is estimated using data for period T. Migrants are defined as those who migrated between T and $T+t$. This implies that migration decision are assumed to be taken based on income differentials at time T, that is, prior to the time at which migrants choose their destination.[8] This appears to be a reasonable assumption given that most migrants in our dataset come from rural areas of Nepal and are unlikely to be particularly good at forecasting differential income trends in multiple locations.

We also examine whether migrants consider relative incomes – rather than absolute incomes – when deciding where to migrate. This point was already touched upon by Stark and Taylor (1991) who showed that households' relative deprivation in their village reference group is significant in explaining migration to destinations where a reference group substitution is unlikely and the returns to migration are high. More recent work in economics and psychology has shown that subjective well-being depends on relative achievement, of which one dimension is income (see Fafchamps and Shilpi, 2008 and 2009 for brief surveys of the literature). This raises the question of whether people choose the migration destination that, on the basis of their individual characteristics, promises them a high income relative to that of others in that location. To this effect, we replace y_i^h with $y_i^h/\overline{y}_i$ in equation (1) and proceed as outlined above. If migration decisions are based on relative rather than absolute income, then the coefficients of $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ should be positive and significant only when they are computed using $y_i^h/\overline{y}_i$.

In addition to relative and absolute income differences, the analysis also examines the respective roles of various location characteristics such as housing and food prices, availability of public services, and density of human settlement.

[8] An alternative strategy for the estimation of pre-migration income distribution in cross-section data is suggested by Bayer, Khan and Timmins (2008).

3 The data

Having described the conceptual framework and estimation strategy, we now present the data. The data used in this paper come from two sources: living standard household surveys, and the population census.

The living standard data come from two rounds of Nepal Living Standards Survey (NLSS). The first round was conducted in 1995/96 while the second took place in 2002/3. The NLSS surveys collected detailed information on households and individuals using nationally representative samples. The 1995/96 NLSS survey is used as source of detailed information about locally available amenities. It is also used to estimate the income regression (1).

Survey data are complemented with information from the 2001 population census. The short population census questionnaire was administered to the whole population. It contains information about ethnicity and caste. For a randomly selected 11% of the census population, additional information was collected using a second, longer questionnaire. This questionnaire collected information on district of current residence, district of residence 5 years prior to the census, and district of origin. Detailed information is also available on gender, age, education, unemployment, occupation, and motive for migration, if any. The Nepalese Central Bureau of Statistics was kind enough to merge the short and long questionnaire datasets for the 11% of the population covered by the long questionnaire. This provides a very large data set on which we estimate the migration regression (3).

Nepal is divided into 75 districts and further subdivided into 3,915 VDCs and 35,235 wards. The 11% population census covers approximately 2.5 million individuals in 520,624 households. 345,349 of these individuals are living in a district other than their district of residence and 119,475 have moved in the five years preceding the census, that is, in the period between the 1995/96 NLSS and the 2001 census. Most of these individuals have moved for reasons other than work. Marriage is the dominant reason for moving among women; study is the dominant reason for moving among children and youths. In contrast, of the adult males who migrated during last 5 years, 69% moved for work reasons.

Because our focus is on work migration, we restrict our attention to adult males. Among those, 16,850 are recorded as having moved in the five years preceding the census specifically for work reasons. These individuals are the focus of our analysis. We note that, by construction, this approach excludes those who have migrated outside Nepal. Our focus is thus on internal migrants. We do not have data on India but since there is no big Indian city within 200 km of Nepalese border, commuting to India for work while residing in a Nepalese district is rare, making it unlikely that economic opportunities in neighboring India affected the choice of migration destination within Nepal.

Figures 1 and 2 show the geographical distribution of work migrants in terms of district of residence and origin. Districts with a high concentration of work migrants relative to non-migrant adult males appear in red, those with a low concentration appear in blue. We see that a small number of destination districts have a high proportion of work migrants. In contrast, districts of origin are distributed widely across the country. This reflects the fact that much work migration is from remote rural areas to towns and cities.

The main characteristics of work migrants are reported in Table 1, together with those of non-migrant adult males. We see that work migrants are on average younger and better educated. The census contains detailed information about ethnicity, language, and religion. In the Nepal census, the term 'ethnicity' is used to capture a hodgepodge of caste and tribal distinctions. The census distinguishes up to 103 ethnic categories. Most of these categories only account for a tiny proportion of the total population. In terms of the total adult population, the most common ethnic categories are Chhetri, Brahmin, and Newar who, together, account for 35% of adult males in the 11% census. All three categories are regarded as upper castes. As we see from Table 1, migrants are much more likely to be upper caste than non-migrants.

The census distinguishes 84 different languages. The main ones are Nepali and Maithili, spoken by 58% of the population. In Table 1 we see that work migrants are much more likely to speak Nepali, the main language in the country. While the Nepalese population is heterogeneous in terms of ethnicity and language, it is relatively homogeneous in terms of religion: 81% of adult males are Hindu and 11% are Buddhist. We see in Table 1 that work migrants are predominantly Hindu.

The dependent variable M_{is}^h in our main regression of interest, regression (3), is constructed as follows. We begin by creating, for each of the 16850 work migrants h identified in the 11% census, 75 M_{is}^h observations corresponding to each of the possible 75 district destinations s. We set $M_{is}^h = 1$ if migrant h moved from district i to district s in the 5 years preceding the census, and 0 otherwise. We then drop M_{ii}^h since we focus on migrants. By construction a migrant reside in one district. For each migrant, variable M_{is}^h thus takes value 1 once and value 0 73 times.

Since the migrant can only move to a single destination, the 74 M_{is}^h observations are not independent and residuals in (3) are correlated. Dependence across M_{is}^h observations combines negative and positive correlation. To illustrate this point, imagine for a moment that all destinations are equivalently attractive to the migrant. The probability $\Pr(M_{is}^h = 1)$ of selecting one of them is thus 1/74. Further assume that one of them is selected at random; for this observation, we have $u_{is}^h = 1 - \Pr(M_{is}^h = 1) = 73/74$. For all other observations, the residual $u_{is}^h = -1/74$. We see that, for individual h, the observation in which $M_{is}^h = 1$ is negative correlated with observations in which $M_{is}^h = 0$. We also see that observations in which $M_{is}^h = 0$ are positively correlated with each other. This combination of positive and negative correlation means that a standard fixed or random effect approach is not sufficient to ensure correct inference; clustering standard errors by individual is necessary. This is what we do.

Having described how the dependent variable is constructed, we turn to regressors. We begin by describing how we construct an estimate of $\widehat{E[y_s^h|z^h]}$, the level of income (or consumption) y_s^h that a migrant with characteristics z^h can expect to earn in district s. To construct such estimate, we use the 1995/96 NLSS data. The reason for using the 1995/96 data instead of the 2002/3 NLSS survey is to avoid reverse causation, i.e., migration causing a change in income patterns. Migrants are unlikely to be able to accurately predict the evolution of incomes in each district over time. Income and consumption levels observable before migration are thus a reasonable starting point.

Using the NLSS data we begin by estimating a regression of the form:

$$y_s^k = \delta_s + \alpha(a_s^k - \overline{a}) + \beta_s(E_s^k - \overline{E}_s) + \chi_s(H_s^k - \overline{H}_s) + v_s^k \tag{4}$$

where y_s^k is the log of income (or consumption) of household k residing in district s, coefficients δ_s, β_s and χ_s vary by district, a_s^k stands for the age and age squared of the household head, E_s^k is the education level of the head measured in years of completed education, and $H_s^k = 1$ if the head belongs to what we have earlier classified as a high caste (i.e., Brahmin, Chhetri or Newar). Since income or consumption are expressed in logs, β_s and χ_s can be thought of as education and high caste premia, respectively. Female headed households are excluded from the regression since the focus is on migrant males. Vector $\overline{a}$ denotes the average age and age squared of observations across the sample. Variables $\overline{E}$ and $\overline{H}_s$ denote the district-specific averages of E_s^k and H_s^k. By demeaning regressors, we ensure that $\widehat{\delta}_s$ measures the unconditional, district-specific average of y_s^k. Marital status, household size, and other household characteristics are not included because they are possibly affected by migration.[9] In contrast, age, education, and caste status can be regarded as exogenous to the migration decisions of adult males. Equation (4) is estimated using correct sampling weights.[10]

Regression estimates for equation (4) are summarized in Table 2 where we show α as well as the average and standard error of δ_s, β_s and χ_s. The coefficients $\widehat{\delta}_i$ and $\widehat{\eta}_i$ are large and jointly significant. There is considerable variation across districts not only in average log income and consumption but also in the income or consumption premia associated with education and high caste. These results are used to construct, for each of the 16,000 or so work migrants in the census, a measure of the income or consumption they can expect to achieve in each of the possible destination districts. Formally, this measure is calculated as:

$$E[\widehat{y_s^h|z^h}] = \widehat{\delta}_s + \widehat{\beta}_s(E_s^h - \overline{E}_s) + \widehat{\chi}_s(H_s^h - \overline{H}_s) \tag{5}$$

where E_s^h and H_s^h are the education and high caste dummy for migrant h. Age is ignored from the calculation since work migrants typically migrate around the same age, i.e., in early adulthood.

Formula (5) can be decomposed into two parts: $\widehat{\delta}_s$, which measures the average income level in district s, and $\widehat{\eta}_s z^h \equiv \widehat{\beta}_s(E_s^h - \overline{E}_s) + \widehat{\chi}_s(H_s^h - \overline{H}_s)$ which captures individual-specific variation in income. Migration models predict that, other things being equal, the choice of migration destination should depend on $E[\widehat{y_s^h|z^h}]$. This means that if we regress the choice of destination separately on $\widehat{\delta}_s$ and $\widehat{\eta}_s z^h$, they should have the same coefficient.

The same methodology is used to construct other variables that may affect the choice of

[9] The literature has often emphasized that migrations often serve an important role in household formation. For migrants, the prospect of forming a large, successful household is likely to be one of the purposes of migration.

[10] The 1995/96 NLSS survey adopted the following sampling strategy. Within each district a small number of wards were selected at random. Within each ward, 12 randomly selected households were interviewed. Because the wards differ widely in terms of population, applying sampling weights is essential in order to obtain consistent estimates of δ_s.

destination. Building on a growing literature documenting the relationship between subjective welfare and relative income, Fafchamps and Shilpi (2008) show that Nepalese households care about their consumption level relative to that of others in the same location. If this is the case, it is conceivable that migrants choose their destination not so much for the absolute gain in income it may provide but for the gain in relative status that would ensue. For instance, if returns to education and ability are higher in an urban setting, an educated individual may improve his relative position in society by moving from a rural to an urban setting. To investigate this possibility, we estimate equation (4) using the log of relative income (or relative consumption) as dependent variable and construct a predicted relative income measure using the same formula (5). These are shown in the second panel of Table 1.

Theories of work migration predict that individuals move to increase their utility or welfare. The 1995/96 NLSS asked respondents a number of questions regarding their subjective satisfaction level with various dimensions of consumption – namely, food, clothing, housing, health care, and child schooling. They were also asked their subjective satisfaction with their level of total income. We apply the same methodology to these data – i.e., we estimate a regression of the same form as (4) and apply formula (5) to construct an expected subjective satisfaction index. If migrants correctly anticipate the subjective satisfaction they will enjoy from moving to different destinations, these subjective satisfaction measures may offer a better way of controlling for expected welfare differences across destinations.

To control for migration costs, we construct variables proxying for geographical and social distance. For geographical distance between districts, we use the arc distance between the district of origin and each possible district of destination, computed from the longitude and latitude of each districts' administrative center. We expect the cost and risk of migration to increase with physical distance.

Social distance is proxied by the proportion of individuals in the district who share the same language, religion, and ethnic group. This is implemented as follows. From the census we have information on ethnic, religious, and language diversity in all districts of the country. From these we construct an index of similarity between individual h and the population of each district. Let m denote a specific trait – e.g., ethnicity, religion or language – and let p_s^m be the proportion of the population of district s that has trait m. Consider the trait m_h of individual h. We expect h's chances of finding a job, etc, to increase in the proportion of individuals in the district of destination who share the same trait. We therefore construct, for each destination and each migrant, a variable $p_s^{m_h}$ equal to the proportion of members of h's with trait m_h. For this migrant, the social distance between two locations i and s is $p_s^{m_h} - p_i^{m_h}$. The idea behind this measure is that individual h 'fits' better in district s if the proportion of like individuals is higher than in his district of origin. We construct similar indices for language and religion. Note the similarity between $p_s^{m_h}$ and the commonly used index of ethno-linguistic fractionalization (ELF). The ELF index measures the probability that two individuals taken at random belong to the same ethnic or linguistic group. Variable $p_s^{m_h}$ measures the probability that an individual taken at random belongs to the same ethnic or linguistic group as the migrant and is thus the individual-equivalent of the ELF index for groups.

We seek to control for price differences across locations. This is difficult because we do not have detailed price data. We are mostly concerned about housing costs and prices of common household goods.

We use the price of rice as a proxy for the price of common household goods. This is not entirely satisfactory but in the absence of a district-level consumer price index this is the best we can do. Given the mountainous nature of the country, rice cannot be grown in many parts of the country. The price of rice thus tends to rise with altitude and geographical isolation, as we expect the prices of many manufactures to do as well. The 1995/96 NLSS collected information on the quantity and price paid for rice by individual households. From this we compute a unit price per Kg. The log of the district median is used as our price index proxy.

To construct an index of housing costs, we take advantage of a section of the 1995/96 NLSS survey focusing on housing. The survey collected information on hypothetical and actual house rental values of each household together with house characteristics such as square footage, number and type of rooms, quality of materials, and the availability of various utilities. We use these data to construct an hedonistic index of housing costs for each district. Let r_s^k be the house rental price paid (or estimated) by household h in district s and let x_s^h denote a vector of house characteristics. We estimate a regression of the form:

$$\log r_s^k = a_s + bx_s^h + e_s^k$$

to obtain estimates of $\widehat{a}_s$, the housing cost premium in each district s. Regression results are shown in Table A1 in appendix. Many house characteristics are significant with the expected sign, e.g., larger, better built houses with better in-house amenities are worth more. District price differentials are large and jointly significant. Since the dependent variable is in log form, $\widehat{a}_s$ measures the housing cost premium in each district.

To the extent that people are mobile, housing price differentials capture, in a reduced form, the effect of location attributes such as proximity to jobs and access to public amenities. It is therefore possible for migrants to be attracted by districts which command a high housing price premium. To further control for access to amenities, we include travel time to the nearest road (a measure of market access) and to the nearest bank (a measure of financial and commercial development).

We include a number of regressors to control for geographical isolation. Fafchamps and Shilpi

(2009) have shown that, in Nepal, subjective welfare is negatively associated with geographical isolation. Census data on total population and population density in each district are used as proxies for urbanization and geographical proximity: the denser the population, the less geographically isolated individuals are likely to be. We also include data on the average elevation in each district. Nepal being a mountainous country, the higher the average elevation of a district, the more costly it is to build roads, raising transport and delivery costs to the district. Ceteris paribus, we expect migrants to seek out districts with a higher population density and a lower elevation.

4 Econometric results

4.1 Univariate analysis

We now investigate the choice of migration destination. We begin with simple univariate analysis. Variables are of the form $\Delta_{is}^h = x_s^h - x_i^h$ where i is the district of origin of migrant h and s is each of 74 possible districts of destination. We examine the average value of Δ_{is}^h for the destination district and compare it to the value of Δ_{is}^h for alternative destinations. For instance, let x_s^h be population density in district s. The average value of Δ_{is}^h for the actual destination of the migrant tells us whether the destination district is more densely populated than the district of origin. The comparison between Δ_{is}^h for actual and hypothetical destinations tells us whether the actual district of destination is more densely populated than alternative destinations.

Results are presented in Table 3 for all variables used in the analysis. We begin with district log income $\widehat{\delta}_s$. We have two estimates of $\widehat{\delta}_s$, one obtained using reported income data, and the other based on reported consumption data. Given that most respondents to the NLSS survey are self-employed, measurement error is typically larger for income than for consumption. We see that our estimates of log income and consumption $\widehat{\delta}_s$ are on average 20% and 8% higher in the district of destination than in the district of origin, respectively. Migrating to one of the 73 alternative destinations would, on average, have reduced income and consumption relative to the district of origin. The difference in anticipated income and consumption between actual and hypothetical destinations is strongly significant. Migrants thus tend to move to districts where consumption and income are higher.

Next we examine whether there are significant differences in returns to individual characteristics $\widehat{\eta}_s z^h$. Surprisingly, results for income show that $\widehat{\eta}_s z^h$ is on average lower in the district of destination than in the district of origin. The difference is large enough to be statistically significant. This implies that better educated, high caste migrants are expected to gain relatively less from migrating to actual destination districts than less educated, lower caste migrants. In contrast, $\widehat{\eta}_s z^h$ estimates based on consumption data show an increase relative to the district of origin. This suggests that better educated, high class migrants would gain more from migrating. We also observe a slightly stronger increase for the actual destination than in the alternatives. The difference is not statistically significant, however.

Differences in relative log income and consumption are displayed next. Predicted relative log income and consumption are generated using the same formula $\widehat{\delta}_s + \widehat{\beta}_s(E_s^h - \overline{E}_s) + \widehat{\chi}_s(H_s^h - \overline{H}_s)$ used for log income, except that, by construction, $\widehat{\delta}_s = 0$ always. We see that relative income falls between the district of origin and the district of destination while it would have risen in alternative destinations. The difference is statistically significant. In contrast, relative consumption is higher in the destination district than in the district of origin or in alternative destinations but the difference between actual and hypothetical destinations is not significant.

We then turn to differences in subjective welfare. The equivalent of $\widehat{\delta}_s$ is used as for log income. We begin with subjective perceptions regarding the adequacy of total income. Relative to their district of origin, the average subjective satisfaction with total income is found to rise between the district of origin and the district of destination. Whether this is fully anticipated by migrants is unclear. Fafchamps and Shilpi (2008) show that in assessing their subjective satisfaction migrants still compare themselves to those in their district of origin.

Results regarding subjective satisfaction from the consumption of food, clothing, housing, health care, and schooling are shown next. We see that in all cases the district of destination has a much larger level of subjective satisfaction, both relative to the district of origin and relative to other possible destinations. We also compute the equivalent of $\widehat{\eta}_s z^h$ and find it to be negative in five out of six cases. This is consistent with the fall in returns to education and high caste that was found for income between the districts of origin and destination.

We then turn to prices and amenities. We observe on average an 9% fall in the median price of rice between the districts of origin and destination. Migrating to alternative destinations would have raised the price of rice instead of reducing it. This is consistent with our interpretation that the price of rice in part captures differences in delivery costs driven by isolation. In contrast, we find a 38% average increase in the rental cost of housing between the districts of origin and destination. Moving to an alternative destination would also have raised average housing costs but by less than that in the actual destination district. Travel time to various facilities and infrastructures falls uniformly between the district of origin and that of destination. Since these differences are strongly correlated with each other, we only report two: travel time to the nearest road, and travel time to the nearest bank. Both fall massively between district of origin and destination, and both would have risen had the migrant moved to an alternative destination.

We observe a strong negative difference in elevation between the district of origin and district of destination. Moving to an alternative destination would, on average, have resulted in a higher elevation than the district of origin. This implies that migrants on average move down from the mountains. They also tend to go to districts with a larger and more dense population than the district of origin and alternative destinations. Migration is thus primarily from rural to urban areas.

In terms of social proximity, we see that migrants on average face a population that is more different from them in terms of both language and caste/ethnicity than it would be in their district of origin. This is true for the actual destination district but also for alternative districts. We do not observe the same pattern for religion; if anything, migrants are more likely to face someone of their religion in their district of destination. The difference is small, however. Finally, the geographical distance between the district of origin and the actual destination is on average smaller than that between the district of origin and alternative destinations: if anything, migrants tend to go to a district that is closer. The difference is statistically significant but it is not large, however.

To summarize, simple bivariate analysis shows that migrants tend to move to a district with: a larger population and population density; a lower elevation and closer proximity to the district of origin; a higher average income and consumption; higher subjective consumption adequacy; lower rice prices and higher housing costs; better access to public amenities. In contrast, migrants move to districts where they have a lower relative income compared to their district of origin. They also tend to move to districts where fewer people speak their language or share their religion.

4.2 Multivariate analysis

We have seen that there are strong differences between actual and alternative migration destinations. Many of these characteristics are correlated with each other, however. To disentangle them we turn to multivariate analysis and estimate the migration regression (3). As explained in the previous section, regressors include: prices as described above; geographical and social distance; and access to amenities. We also include the log of total population, population density, and average elevation as additional controls.

We begin by estimating (3) with $\widehat{\delta}_s - \widehat{\delta}_i$ computed from the log income data. Results are shown in the first column of Table 4. As discussed earlier, reported results include individual fixed effects and standard errors clustered by district of origin. The univariate analysis showed that income was significant on its own. Once we control for distance, population, prices and amenities, however, the difference in expected income is no longer significant. Most of other variables are, though. Distance has the expected negative sign – on average the migration destination is closer to the district of origin than alternative destinations. The destination district also has a significantly larger population and population density, a lower elevation, and a lower rice price. Housing costs in contrast are higher in the destination district than in alternative destinations, possibly because they control for the availability of amenities and other public goods. We also see that the destination district has a significantly shorter average travel time to the nearest road. Once we control for road distance, travel time to the nearest bank is no longer significant.

The univariate analysis showed that migrants on average move to destinations where they are on average less likely to find people like them. The results presented in Table 4 present a different picture. Conditional on the other regressors, the ethnicity and language proximity indices are significant with the anticipated sign: social proximity between the migrant and the population of the destination district is higher than in alternative destinations. The religion proximity index is not significant. Taken together, these results suggest that, conditional on material benefits from migration, migrants prefer to move to a destination where they integrate more easily – and possibly enjoy network benefits in terms of access to jobs and housing (Munshi 2003, Beaman 2006).

It is surprising that income differences are not significant once we control for geography, population, prices and amenities. This may be because we have not included individual-specific income differentials across districts. We therefore reestimate (3) with $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ as well as $\widehat{\delta}_s - \widehat{\delta}_i$. Results are shown in column 2 of Table 4. They remain non-significant. In column 3 we replace absolute differences in log income with relative differences. The constructed regressor, which by construction depends only on $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$, remains non-significant. Finally in column 4 we compute $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ using answers to the question regarding the subjective adequacy of total income. Estimate coefficients are significant, but with opposite signs: only the $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ part as the anticipated positive sign.

It is conceivable that these surprising and disappointing results are driven by measurement error in income. It is indeed well known that income is notoriously difficult to measure in poor, primarily self-employed populations. To investigate this possibility, we reestimate (3) using NLSS consumption data to construct $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$.

Results, shown in Table 5, are more in line with expectations. Although average log consumption in the district is not significant, the coefficient of the $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ is strongly significant, and so is the coefficient of the combined $\widehat{\delta}_s - \widehat{\delta}_i + (\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ variable. We also find a significant positive coefficient when the combined $\widehat{\delta}_s - \widehat{\delta}_i + (\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ variable is constructed using relative rather than absolute log consumption. If we include $\widehat{\delta}_s - \widehat{\delta}_i + (\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$ computed both from absolute and relative income, neither of them is significant, probably because they are too strongly correlated. We cannot therefore discern whether it is absolute or relative standards of living that affect the choice of migration destination.

We also estimate similar regressions using subjective consumption adequacy questions to construct $\widehat{\delta}_s - \widehat{\delta}_i$ and $(\widehat{\eta}_s - \widehat{\eta}_i)\, z^h$. Results, not shown here to save space, are generally non-significant. The only exception is food consumption but, as we found in column 4 of Table 4, estimated coefficients have opposite signs so the results are difficult to interpret.

4.3 Robustness checks

We conduct numerous robustness checks. We first try to understand the contradiction between the univariate and multivariate results. To this effect, we estimate a series of simple regressions that include $\widehat{E[y_s^h|z^h]}$ (measured in terms of income or consumption) together with one of the additional regressors appearing in Tables 4 and 5. We find that $\widehat{E[y_s^h|z^h]}$ remains highly significant with all regressors with a single exception: as soon as the average travel time to the nearest road is included in the regression, $\widehat{E[y_s^h|z^h]}$ loses all significance. We already know from Fafchamps and Shilpi (2008) that income is strongly negatively correlated with geographical isolation. What this suggests is that once we control for geographical isolation, income differentials no longer matter. Similar findings are reported for Brazil and Mexico by Timmins (2008), using a different methodology.

Next, we investigate in different ways whether our failure to find a significant income effect in Tables 4 and 5 is due to income mismeasurement. The income regression (4) does not control for household size and composition. The rationale for doing so is that (1) household size and composition may be endogenous to the migration decision – e.g., individuals who migrate to the city may opt to have a smaller household – and that (2) migrants may derive satisfaction from the total income jointly earned by the household they head. However, not correcting for household size and composition a higher predicted income $\widehat{E[y_s^h|z^h]}$ in districts where households are larger and there are more work opportunities for household dependents – typically rural districts. To investigate whether this is responsible for the low income coefficients, we include the log of household size and the share of adult males and females in the income regression (4) and we replicate the analysis using the revised $\widehat{E[y_s^h|z^h]}$. The results, which are not shown here to save space, are virtually undistinguishable of those reported in Tables 4 and 5.

Central to our estimation are estimates of income and consumption levels enjoyed by households in various districts. To check the robustness of our results, we reestimate all income and consumption regressions (4) using non-migrants only. The reason for doing so is that non-migrants represent the bulk of the population and thus $E[v^h|\text{do not migrate}] \approx E[v^h]$. Regression results, not shown here to save space, are disappointing: if anything, income and consumption variables are even less significant.

This strategy does not control for possible self-selection: if more talented individuals migrate, remaining households may be less productive. As a result, they may earn less than migrants in the same location. To correct for the self-selection of non-migrants we need variables that affect the decision to migrate but are unlikely to affect income. Family background variables such as the education and occupation of the father may serve this purpose because they affect the ability of the migrant's father to help finance the cost of migration. Given that most migrants migrate early in their adult life, it is reasonable to expect that parental influences play a role in the decision to migrate – and in the financing of migration costs. We use the education and occupation of the father to construct two selection correction terms for the income regressions – one selection term for migrants, and one for non-migrants (Wooldridge, p. 631):

$$\begin{aligned} y_s^k &= \delta_s + \alpha(a_s^k - \overline{a}) + \beta_s(E_s^k - \overline{E}_s) + \chi_s(H_s^k - \overline{H}_s) \\ &\quad + \rho_1 m \frac{\phi(z\widehat{\theta})}{\Phi(z\widehat{\theta})} + \rho_2(1-m)\frac{\phi(z\widehat{\theta})}{1-\Phi(z\widehat{\theta})} + v_s^k \end{aligned} \tag{6}$$

where $\phi(z\widehat{\theta})$ and $\Phi(z\widehat{\theta})$ are the normal density function and cumulative distribution from the selection regression of migrant status m on determinants z.

The selection regression is shown in Table A2 in Appendix. Other variables are the same as those appearing in the income and consumption regressions (4). We see that family background variables are significant. Using this selection regression we construct the two Mills ratio shown in equation (6), one for migrants and one for non-migrants, and we reestimate the income and consumption regressions with these additional regressors, obtain corrected $\widehat{\delta}_s$ and $\widehat{\eta}_s$ estimates, and reestimate the destination choice regressions. Results are nearly indistinguishable from those reported in Tables 4 and 5. They are omitted here to save space.

When constructing $E[\widehat{y_s^h|z^h}]$ we implicitly assume that migrants are well informed about incomes in all potential destinations. But it is possible that they are better informed about certain destinations, for instance, destinations chosen by migrants from their district in the past. Failing to control for this possibility may lead to an attenuation bias in the income coefficient. To investigate this possibility, we interacted the income variable with a proxy for the availability of income information. If migrants only respond to income differences for those districts on which they have more accurate information, the coefficient of the interacted term should be significant even if the uninteracted term is not. As proxy for the availability of information, we use the number of adult males who migrated more than 5 years ago (that is *before* the migrants themselves) from the district of origin to each of the districts of destination. The coefficient of the interacted term is minuscule in magnitude and uniformly non-significant. The same finding obtains whether we use all migrants or only work migrants.

As a final robustness check, we reestimate the model using migrant data from the NLSS 2002/03. The number of migrants is significantly smaller, so results may be less precise. The advantage of this approach is that it serves as cross-validation. Results are presented in Tables 6 and 7. Table 6 should be compared with Table 4, and Table 7 with Table 5.

Comparing Tables 6 and 4, we again find that anticipated income, whether absolute or relative, is either non-significant or negative. Most of our other results obtain. Exceptions include the rice price – which appears with the wrong sign but is only marginally significant – and elevation and population density – which are no longer significant. Comparing Tables 7 and 5, we find that in the smaller NLSS 2002/3 dataset none of the anticipated consumption variables is statistically significant. Other results are as before.

4.4 Magnitude

To assess the relative magnitude of our results, we multiply coefficients estimated in Tables 4 and 5 by the standard deviation of their respective regressors. We then average over the various regressions reported in Tables 4 and 5. Calculations are summarized in Table 8. The larger the value, the more influence the regressor has on the choice of a destination district.

We see that the most important regressors in terms of magnitude are travel time to the nearest road, elevation, language similarity, and the price of rice. Consumption variables have an effect on migration destination that is smaller in magnitude: a one standard deviation increase in anticipated relative consumption, for instance, has an effect on destination that corresponds to a third of the effect of a one standard deviation in elevation – and one-sixth of a one standard deviation in distance from the nearest road. Income variables have a negligible effect on migration decisions. These calculations confirm our earlier assessment.

5 Conclusion

Combining data from a household survey and an 11% census of the population, we have estimated destination choice regressions for Nepalese internal migrants. Results show that population density, social proximity, and access to amenities exert a strong influence on migrants' choice of destination. These results confirm earlier work on the factors affecting the subjective welfare cost of isolation (Fafchamps and Shilpi, 2008).

Differentials in income and consumption expenditures across districts are significant in univariate comparisons but are found to be less important than expected once we control for covariates. Income variables, whether measured in absolute or relative terms, are either not statistically significant or have the wrong sign. Consumption expenditure variables are significant with a positive sign in some regressions, but the data do not enable us to distinguish whether migrants respond to gains in absolute or relative consumption. Results are robust to different specifications and datasets.

The analysis reported here is based on one critical maintained assumption, namely, that income and consumption levels obtained by district residents in the recent past can be used as proxy for the anticipations of subsequent migrants. Undoubtedly it would be better to have direct measurements of what migrants actually anticipate to earn and consume in different districts upon migration. Unfortunately such data is not available – and would be difficult to collect.

Taken together, our results suggest that an urban environment and access to amenities are key considerations when internal migrants choose a migration destination. Anticipated income and consumption expenditures, whether absolute or relative, appear secondary. This does not imply that income differentials do not affect the decision to migrate, an issue that we have sidestepped by focusing on the choice of destination conditional on migrating.

It is difficult to draw causal inference from observational data. This study is no exception. The results presented here are nevertheless sufficiently suggestive to cast doubt on the theory that the choice of migration destination is driven primarily by income differentials. Other factors seem to play a strong – and probably more important – role.

References

1. Adams, Richard, *Remittances, Investment, and Rural Asset Accumulation in Pakistan*,

International Food Policy Research Institute, Washington, D.C., 1997

2. Akee, Randall K.Q., "Deciphering Immigrant Self-Selection: New Evidence from a Developing Country," Kennedy School of Government, Harvard University, Cambridge Mass., 2006 (mimeograph)

3. Bayer, Patrick J., Shakeeb Khan, and Christopher Timmins, "Nonparametric Identification and Estimation in a Generalized Roy Model," NBER Working Paper No. W13949. March 2008

4. Bayoh Isaac, Elena G. Irwin and Timothy Haab, "Determinants of Residential Location Choice: How Important Are Local Public Goods in Attracting Homeowners to Central City Locations?," *Journal of Regional Science*, 46(1), 97-120, February 2006.

5. Beaman, Lori A.,"Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the US," Department of Economics, Yale University, New Haven, 2006 (mimeograph)

6. Borjas, George J., "The Economics of Immigration," *Journal of Economic Literature*, 32(4), 1667-1717, December 1994

7. Borjas, George J., Native Internal Migration and the Labor Market Impact of Immigration, *Journal of Human Resources*, 41(2), 221-58, Spring 2006

8. Borjas, George J., Stephen G. Bronars and Stephen J. Trejo, "Assimilation and the Earnings of Young Internal Migrants," *Review of Economics and Statistics*, 74(1), 170-75, February 1992

9. Carrington William J., Enrica Detragiache and Tara Vishwanath, "Migration with Endogenous Moving Costs," *American Economic Review*, 86(4), 909-30, September 1996

10. Chiquiar, Daniel and Gordon H. Hanson, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2), 239-81, April 2005

11. Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367-2420, November 2002

12. de Brauw, Alan and John Giles, "Migrant Opportunity and the Educational Attainment of Youth in Rural China," IFPRI and Department of Economics, Michigan State University, Washington DC, September 2006 (mimeograph)

13. de la Briere, Benedicte, "Household Behavior toward Soil Conservation and Remittances in the Dominican Republic," University of California, Berkeley, 1996 (unpublished PhD thesis)

14. Fafchamps, Marcel and Forhad Shilpi, "The Spatial Division of Labor in Nepal," *Journal of Development Studies*, 39(6), 23-66, 2003.

15. Fafchamps, Marcel and Forhad Shilpi, "Cities and Specialization: Evidence from South Asia," *Economic Journal*, 115(503), 477-504, April 2005.

16. Fafchamps, Marcel and Forhad Shilpi, "Subjective Welfare, Isolation, and Relative Consumption," *Journal of Development Economics*, 2008 (forthcoming)

17. Fafchamps, Marcel, and Forhad Shilpi, "Isolation and Subjective Welfare: Evidence from South Asia," *Economic Development and Cultural Change*, 2009 (forthcoming)

18. Fujita, Masahisa, Paul Krugman and Anthony J. Venables, *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge and London, 1999

19. Gabriel, Paul E. and Suzanne Schmitz, "Favorable Self-Selection and the Internal Migration of Young White Males in the United States," *Journal of Human Resources*, 30(3), 460-71, Summer 1995

20. Greenwood, Michael J., "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 13(2), 397-433, June 1975

21. Harris, John and M. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *Amer. Econ. Rev.*, 60, 126-142, 1970

22. Henderson, J. Vernon, *Urban Development: Theory*, Fact, and Illusion, Oxford University Press, New York, 1988

23. Hojvat-Gallin, Joshua, "Net Migration and State Labor Market Dynamics," *Journal of Labor Economics*, 22(1), 1-21, January 2004

24. Jackman, Richard and Savvas Savouri, "Regional Migration in Britain: An Analysis of Gross Flows Using NHS Central Register Data," *Economic Journal*, 102(415), 1433-50, November 1992.

25. Lall, Somik and Christopher Timmins, "Rural-Urban Migration: Successful Integration or just "Bright Lights"? Evidence from Brazil and Mexico," Department of Economics, Duke University, 2008 (mimeograph)

26. Lokshin, Misha, Mikhail Bontch-Osmolovski and Elena Glinskaya, "Work Migration and Poverty in Nepal," World Bank, mimeograph, 2007.

27. Lucas, Robert E. Jr, "Life Earnings and Rural-Urban Migration," *Journal of Political Economy*, 112(1), S29-59, Feb. 2004

28. Lucas, Robert E.B. and Oded Stark, "Motivations to Remit: Evidence from Botswana," *J. Polit. Econ.*, 93 (5), 901-918, October 1985

29. Mansuri, Ghazala, "Migration, School Attainment and Child Labor: Evidence from Rural Pakistan," DECRG, The World Bank, Washington DC, April 2006 (mimeograph)

30. McCall, B.P and J.J. McCall, A Sequential Study of Migration and Job Search, *Journal of Labor Economics*, Oct. 1987, 5(4), 452-76

31. McCormick, Barry and Jacqueline Wahba, "An Econometric Model of Temporary International Migration and Entrepreneurship," Department of Economics, University of Southampton, Southampton, 2001 (mimeograph)

32. McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp.105-42, 1974

33. McKenzie, David, John Gibson and Steven Stillman, "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration," *IZA Discussion Papers 2087*, Institute for the Study of Labor, 2006

34. Mesnard, Alice and Martin Ravallion, "Is Inequality Bad for Business? A Nonlinear Empirical Model of Entrepreneurship," ARQADQ, University of Toulouse and DECRG, The World Bank, Toulouse and Washington D.C., 2001 (mimeograph)

35. Munshi, Kaivan and Mark Rosenzweig, "Why Is Mobility in India so Low? Social Insurance, Inequality and Growth," Economic Growth Center, Yale University, New Haven, July 2005 (mimeograph)

36. Munshi, Kaivan, "Networks in the Modern Economy: Mexican Migrants in the US Labor Market," *Quarterly Journal of Economics*, 118(2), 549-99, May 2003

37. Stark,Oded and Robert E. Lucas, "Migration, Remittances, and the Family," *Economic Development and Cultural Change*, 36, no.3, 465-481, Apr. 1988

38. Stark, Oded and J. Edward Taylor, "Migration Incentives, Migration Types: The Role of Relative Deprivation," *Economic Journal*, 101(408), 1163-78, September 1991

39. Roy, Andrew D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3: 135-46, 1951

40. Timmins, Christopher, "Estimating Spatial Differences in the Brazilian Cost of Living with Household Location Choices," *Journal of Development Economics*, 80(1): 59-83, June 2006

41. Train, Kenneth E., *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, 2003

42. Uhlig, Harald, "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification," *American Economic Review*, 96(2), 383-87, May 2006.

43. Winters, Paul, Alain de Janvry and Elisabeth Sadoulet, "Family and Community Networks in Mexico-U.S. Migration," *Journal of Human Resources*, 36(1), 159-84, Winter 2001

44. Woodruff, Christopher, "Remittances and Microenterprises in Mexico," Graduate School of International Relations and Pacific Studies, UCSD, La Jolla, August 2001 (mimeograph)

45. Wooldridge, Jeffrey M., *Econometric Analysis of Cross-Section and Panel Data*, MIT Press, Boston, 2002

Table 1: Summary Statistics

	Work Migrant	Adult Male
Age		
Mean	35.3	43.9
Standard Deviation	10.6	13.9
Education (years)		
Mean	8.0	3.0
Standard Deviation	5.0	4.3
Ethnicity (Percentage)		
Brahmin	34.5	11.7
Chhetri	21.5	15.6
Newar	7.4	7.9
Tharu	3.1	6.7
Magar	6.1	6.0
Tamang	4.2	5.9
Other	23.2	46.2
Language (Percentage)		
Nepali	73.9	45.3
Maithili	6.2	13.2
Bhojpuri	1.3	7.3
Newar	4.4	6.1
Tharu	2.0	5.8
Tamang	3.7	5.5
Other	8.5	16.8
Religion (Percentage)		
Hindu	89.6	81.0
Buddheism	7.2	11.7
Muslim	0.9	3.7
Kirat	1.5	2.9
Christian	0.6	0.3
Others	0.2	0.4

Table 2. Income and Consumption regressions using NLSS 95/96

	Age		Age Squared/10000		District Fixed Effect		District Level Premium for Education		District Level Premium for High caste	
Absolute:	coef	t-stat	coef	t-stat	Mean	SD	Mean	SD	Mean	SD
Log income	0.042	6.456	-3.055	-4.479	10.289	0.340	0.218	0.200	0.145	0.405
Log Consumption	0.038	7.916	-2.974	-5.873	10.325	0.340	0.196	0.138	0.184	0.304
Relative:										
Relative log income	0.004	6.388	-0.292	-4.422	n.a.		0.021	0.020	0.014	0.039
Relative log consumption	0.004	7.860	-0.285	-5.826	n.a.		0.019	0.014	0.018	0.030
Consumption Adequacy Index:										
Food	0.008	1.922	-0.345	-0.801	1.496	0.213	0.130	0.121	0.120	0.249
Clothing	0.006	1.593	-0.293	-0.752	1.357	0.196	0.052	0.101	0.059	0.217
Housing	0.007	1.685	-0.220	-0.549	1.404	0.184	0.105	0.103	0.096	0.264
Healthcare	0.004	0.990	-0.063	-0.164	1.412	0.198	0.077	0.112	0.063	0.237
Children's Schooling	-0.006	-1.426	0.900	2.069	1.444	0.201	0.051	0.120	0.043	0.302
Total Income	0.006	1.944	-0.307	-0.920	1.251	0.156	0.069	0.098	0.067	0.195

Each line corresponds to a different regression. The estimator is weighted least squares, using sampling population weights.

Table 3. Comparing the actual destination to alternative destinations

All figures are relative to the district of origin	Actual Destination	Mean in Alt. Destin.	Diff. in mean t-stat
Income and consumption			
Average income (log)	0.195	-0.037	-61.840
Differential in log income due to education and high caste	-0.035	0.007	9.031
Average consumption expenditures (log)	0.075	-0.046	-33.561
Differential in log consumption due to education and high caste	0.020	0.018	-0.832
Relative log income	-0.003	0.001	8.915
Relative log consumption	0.002	0.002	-1.001
Subjective consumption adequacy			
Average consumption adequacy index: total income	0.054	-0.008	-35.523
Differential due to education and caste: total income	0.002	-0.016	-8.508
Average consumption adequacy index: food	0.094	-0.010	-44.127
Differential due to education and caste: food	-0.008	0.014	7.844
Average consumption adequacy index: clothing	0.076	-0.019	-42.983
Differential due to education and caste: clothing	-0.002	-0.019	-7.700
Average consumption adequacy index: housing	0.070	-0.028	-47.457
Differential due to education and caste: housing	0.002	-0.004	-2.195
Average consumption adequacy index: health care	0.081	-0.022	-46.605
Differential due to education and caste: health care	-0.010	-0.009	0.255
Average consumption adequacy index: children schooling	0.093	-0.011	-45.711
Differential due to education and caste: children schooling	-0.003	-0.022	-6.701
Prices and amenities			
Log of rice price	-0.089	0.021	47.802
Housing price premium (log)	0.377	0.210	-12.221
Time travel to nearest paved road	-0.746	0.103	79.767
Time travel to nearest bank	-0.373	0.091	71.345
Population and distance			
Population density	0.281	-0.033	-86.131
Log(population)	0.330	-0.207	-74.129
Elevation in meters	-0.317	0.166	57.156
Ethnic/caste similarity index	-0.042	-0.060	-13.664
Language similarity index	-0.123	-0.101	7.427
Religion similarity index	0.008	-0.017	-13.816
Distance in '000 Km	0.261	0.281	13.822

Table 4. Income and the choice of migration destination

District difference in:	coef	t-stat	coef	t-stat	coef	t-stat	coef	t-stat
Income								
Average log income	-0.185	-0.946	-0.188	-0.965				
Differential in log income due to education and high caste			-0.035	-0.359				
Relative log income controlling for education and high caste					-0.089	-0.087		
Average consumption adequacy index: total income							-0.958**	-2.078
Differential due to education and caste: total income							0.479***	3.041
Prices and amenities								
Log of rice price	-1.909**	-2.001	-1.883**	-1.977	-1.849**	-1.973	-1.921**	-2.163
Housing price premium (log)	0.188***	3.005	0.188***	2.993	0.182***	2.916	0.182***	2.924
Time travel to nearest paved road	-0.951***	-9.579	-0.955***	-9.451	-0.920***	-8.893	-0.950***	-9.268
Time travel to nearest bank	0.107	0.430	0.118	0.473	0.146	0.639	-0.033	-0.119
Elevation in meters	-0.575**	-2.359	-0.579**	-2.386	-0.630***	-2.855	-0.457*	-1.857
Population								
Population density	0.828***	5.967	0.823***	5.744	0.791***	5.579	0.797***	5.837
Log(population)	0.372**	2.046	0.376**	2.029	0.348*	1.912	0.400**	2.130
Ethnicity similarity index	1.685***	7.170	1.686***	7.169	1.701***	7.039	1.668***	7.163
Language similarity index	1.519***	10.544	1.515***	10.390	1.496***	10.307	1.483***	10.498
Religion similarity index	-0.576	-1.376	-0.588	-1.427	-0.604	-1.468	-0.462	-1.037
Distance								
Distance above 100 Km	-0.842***	-2.726	-0.845***	-2.733	-0.829***	-2.667	-0.779**	-2.414
Log-Likelihood	-57,089.55		-56,898.28		-56,910.66		-56,786.67	
Number of observations	1,076,556		1,072,804		1,072,804		1,072,804	
Pseudo R2	0.155		0.155		0.155		0.157	

The estimator is Fixed Effect Conditional Logit. Standard errors are corrected for clustering across district of origin. *** p<0.01, ** p<0.05, * p<0.1

Table 5. Consumption and the choice of migration destination

Consumption	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**
Average consumption expenditures (log)	0.140	0.635	0.232	1.007						
Log consumption differential due to education and high caste			0.566***	6.937						
Combined average and differential					0.457***	4.264			0.227	0.997
Relative log consumption controlling for education and ethnicity							5.609***	7.143	3.548	1.584
Prices and amenities										
Log of rice price	-1.929**	-2.100	-2.016**	-2.340	-2.083**	-2.378	-1.920**	-2.154	-2.015**	-2.340
Housing price premium (log)	0.186***	2.887	0.188***	2.900	0.195***	3.196	0.180***	2.871	0.187***	2.865
Time travel to nearest paved road	-0.881***	-9.904	-0.850***	-9.393	-0.790***	-8.492	-0.920***	-8.932	-0.854***	-9.412
Time travel to nearest bank	0.139	0.635	0.092	0.422	0.096	0.452	0.101	0.442	0.093	0.425
Elevation in meters	-0.674***	-2.752	-0.693***	-2.957	-0.777***	-3.524	-0.607***	-2.949	-0.687***	-2.921
Population										
Population density	0.783***	5.504	0.878***	6.290	0.842***	6.239	0.887***	6.489	0.878***	6.292
Log(population)	0.334*	1.866	0.265	1.529	0.252	1.412	0.295*	1.719	0.266	1.536
Ethnicity similarity index	1.719***	7.155	1.719***	7.230	1.742***	7.138	1.703***	7.177	1.723***	7.244
Language similarity index	1.485***	10.484	1.589***	11.198	1.548***	10.771	1.608***	11.181	1.592***	11.210
Religion similarity index	-0.578	-1.373	-0.529	-1.267	-0.508	-1.249	-0.552	-1.311	-0.526	-1.261
Distance										
Distance above 100 Km	-0.811***	-2.601	-0.826***	-2.627	-0.796**	-2.506	-0.851***	-2.733	-0.827***	-2.632
Log-Likelihood	-57,096.58		-56,776.16		-56,794.57		-56,785.89		-56,773.81	
Number of observations	1,076,556		1,072,804		1,072,804		1,072,804		1,072,804	
Pseudo R2	0.155		0.157		0.156		0.157		0.157	

The estimator is Fixed Effect Conditional Logit. Standard errors are corrected for clustering across district of origin. *** p<0.01, ** p<0.05, * p<0.1

Table 6. Income and the choice of migration destination -- using migrants from the NLSS 2002/3

District difference in:	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**
Average log income	-0.888**	-2.011	-0.888**	-2.140				
Differential in log income due to education and high caste			-0.002	-0.007				
Relative log income controlling for education and ethnicity					0.860	0.283		
Average consumption adequacy index: total income							-2.177***	-2.613
Differential due to education and caste: total income							-0.384	-0.636
Prices and amenities								
Log of rice price	2.160*	1.923	2.160*	1.907	1.873*	1.706	1.711	1.550
Housing price premium (log)	0.353***	3.056	0.353***	3.068	0.329***	3.048	0.383***	3.483
Time travel to nearest paved road	-1.658***	-3.327	-1.658***	-3.245	-1.386**	-2.541	-1.464***	-2.973
Time travel to nearest bank	0.939	1.356	0.940	1.284	0.979	1.385	0.744	1.123
Elevation in meters	0.105	0.273	0.105	0.278	-0.065	-0.156	0.231	0.603
Population								
Population density	-0.384	-0.859	-0.384	-0.854	-0.531	-1.109	-0.571	-1.254
Log(population)	2.776***	4.738	2.776***	4.804	2.744***	4.325	2.898***	4.918
Ethnicity similarity index	0.915*	1.731	0.915*	1.698	0.925*	1.691	1.017*	1.794
Language similarity index	1.832***	2.812	1.833***	2.795	1.855***	2.711	1.380*	1.859
Religion similarity index	-0.743	-1.120	-0.743	-1.117	-1.066	-1.562	-0.392	-0.492
Distance								
Distance above 100 Km	-11.355***	-6.698	-11.355***	-6.724	-11.482***	-6.805	-11.697***	-6.963
Log-Likelihood	-620.47		-620.47		-623.10		-617.00	
Number of observations	16,214		16,214		16,214		16,214	
Pseudo R2	0.390		0.390		0.388		0.394	

The estimator is Fixed Effect Conditional Logit. Standard errors are corrected for clustering across district of origin. *** p<0.01, ** p<0.05, * p<0.1

Table 7. Consumption and the choice of migration destination -- using migrants from the NLSS 2002/3

Consumption	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**	**coef**	**t-stat**
Average consumption expenditures (log)	-0.163	-0.238	-0.161	-0.236						
Log consumption differential due to education and high caste			0.069	0.236						
Combined average and differential					-0.007	-0.026			-0.108	-0.160
Relative log consumption controlling for education and ethnicity							0.484	0.155	1.566	0.188
Prices and amenities										
Log of rice price	1.991	1.445	1.999	1.438	1.852*	1.685	1.854*	1.656	1.949	1.411
Housing price premium (log)	0.328***	3.000	0.327***	2.990	0.329***	3.081	0.328***	3.064	0.327***	3.000
Time travel to nearest paved road	-1.469**	-2.402	-1.469**	-2.400	-1.394**	-2.560	-1.391**	-2.573	-1.444**	-2.379
Time travel to nearest bank	1.049	1.514	1.040	1.513	1.008	1.523	0.999	1.511	1.029	1.504
Elevation in meters	-0.048	-0.116	-0.047	-0.115	-0.065	-0.158	-0.065	-0.158	-0.053	-0.126
Population										
Population density	-0.504	-1.095	-0.496	-1.100	-0.532	-1.105	-0.527	-1.129	-0.509	-1.126
Log(population)	2.733***	4.492	2.726***	4.539	2.748***	4.288	2.743***	4.344	2.734***	4.508
Ethnicity similarity index	0.897*	1.699	0.907*	1.722	0.912*	1.728	0.921*	1.733	0.910*	1.728
Language similarity index	1.917***	2.981	1.926***	3.044	1.860***	2.727	1.865***	2.811	1.906***	3.021
Religion similarity index	-1.036	-1.418	-1.036	-1.421	-1.065	-1.530	-1.066	-1.555	-1.045	-1.432
Distance										
Distance above 100 Km	-11.458***	-6.827	-11.454***	-6.830	-11.473***	-6.783	-11.470***	-6.781	-11.460***	-6.818
Log-Likelihood	-623.08		-623.06		-623.14		-623.13		-623.10	
Number of observations	16,214		16,214		16,214		16,214		16,214	
Pseudo R2	0.388		0.388		0.388		0.388		0.388	

The estimator is Fixed Effect Conditional Logit. Standard errors are corrected for clustering across district of origin. *** p<0.01, ** p<0.05, * p<0.1

Table 8. Relative magnitude of effect of regressors on choice of migration destinatior

Income and consumption	Standard deviation	Relative effect
Combined income effect	0.76	0.02
Relative log income controlling for education and ethnicity	0.06	-0.01
Combined consumption effect	0.56	0.19
Relative log consumption controlling for education and ethnicity	0.03	0.17
Prices and amenities		
Log of rice price	0.29	-0.56
Housing price premium (log)	1.74	0.32
Time travel to nearest paved road	1.00	-0.92
Time travel to nearest bank	0.83	0.09
Elevation in meters	1.08	-0.67
Population		
Population density	0.47	0.38
Log(population)	0.92	0.32
Ethnicity similarity index	0.17	0.28
Language similarity index	0.38	0.58
Religion similarity index	0.23	-0.12
Distance		
Distance above 100 Km	0.18	-0.15

Relative effect of a one standard deviation calculated as coefficient x standard deviation, averaged over the different regressions reported in Tables 4 and 5.

Table A1. Hedonistic regression of house rental value

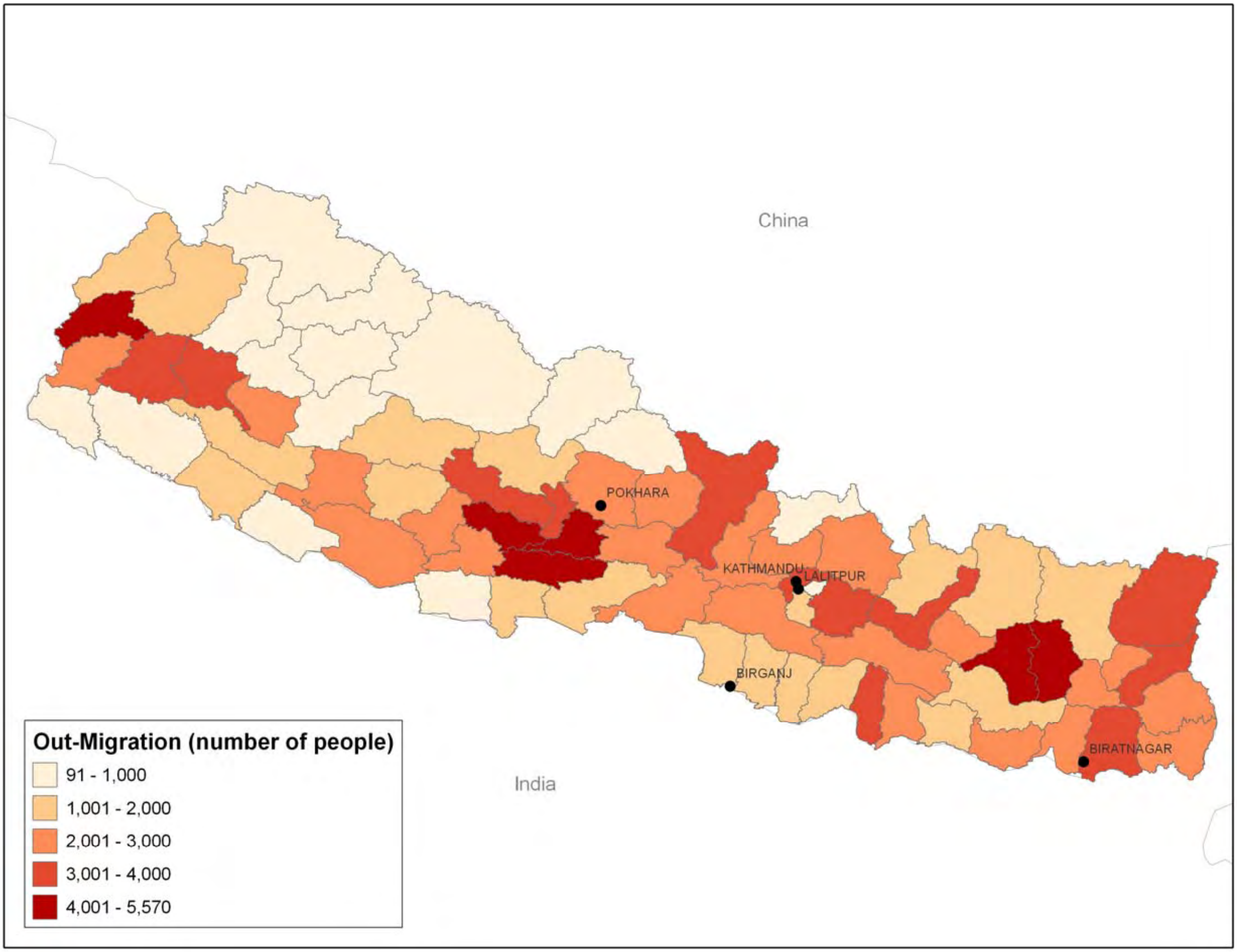
	Coef.	t-stat
Area of dwelling		
Log(sq.ft of the dwelling)	0.179	(3.08)**
Log(sq.ft of the plot)	-0.093	(1.91)
Kitchen garden (yes=1)	-0.202	(2.72)**
Number of rooms and room composition		
Log(number of rooms)	0.553	(6.37)**
Share of Kitchen	-1.467	(0.69)
Share of toilet/bathroom	-2.619	(1.21)
Share of bedrooms	-2.113	(1.00)
Share of living/dinning room]	-1.517	(0.72)
Share of office	-1.185	(0.55)
Share of mixed use room	-2.256	(1.07)
Share of other rooms	-2.358	(1.11)
Construction material of outside wall		
Mud Bricks/stone (yes=1)	-0.197	(1.66)
Wood/branches (yes=1)	-0.369	(2.36)*
Other (yes=1)	-1.455	(7.90)**
Floor material		
Wood, Stone,Cement/tile or other (yes=1)	0.461	(3.66)**
Roof material		
Galvanized Iron (yes=1)	0.823	(6.75)**
Concrete, Cemnet(yes=1)	0.882	(4.90)**
Tiles/slate(yes=1)	0.44	(4.79)**
Characteristics of windows		
Shutters (yes=1)	0.379	(4.43)**
Screen/glass(yes=1)	0.496	(2.64)**
Other (yes=1)	-0.602	(2.32)*
Drinking water source		
Covered Well/Hand Pump	-0.25	(1.99)*
Open Well	-0.309	(1.80)
Other (yes=1)	-0.474	(3.27)**
Amenities		
Sanitary System (yes=1)	0.115	(0.88)
Garbage Disposal (yes=1)	0.121	(0.78)
Non-Flush/Communal Toilet (yes=1)	-0.48	(2.90)**
No toilet (yes=1)	-0.596	(3.47)**
Electric Light (yes=1)	-0.003	(0.08)
District dummies	Yes	

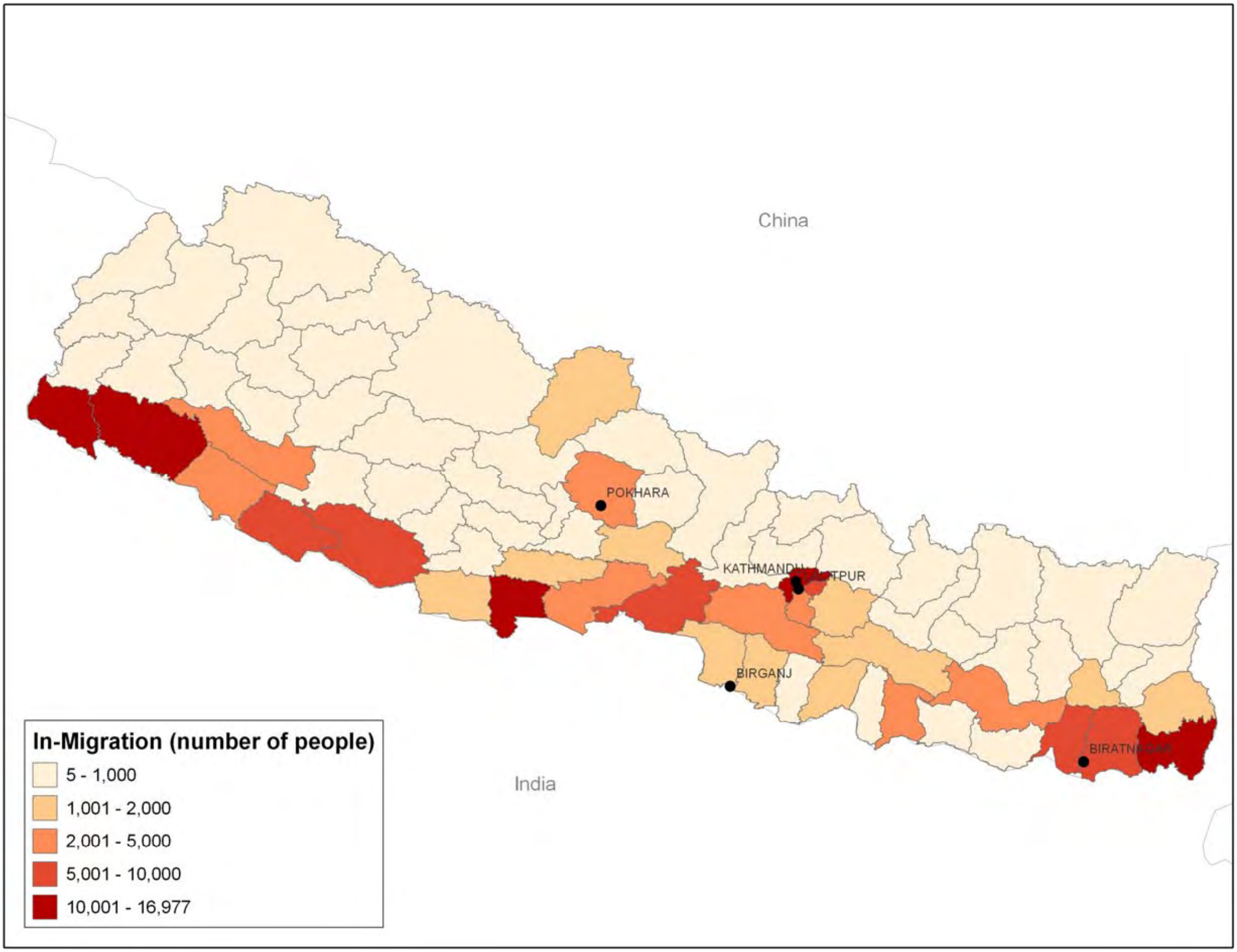
The dependent variable is the log of the rental value of the dwelling.
Rental value is either actual or estimated in case of owner occupation.
Based on NLSS 1995/96.

Table A2. Migration Selection Equation

	Coef/z-stat
Age	0.011
	(0.78)
Age squared	-0.000
	(0.41)
Father's education level	0.036
	(2.60)**
Father's employment in non-farm sector	0.344
	(3.61)**
High caste dummy	0.253
	(3.78)**
Education	0.033
	(0.87)
Constant	-1.532
	(4.58)**
Observations	2762

The dependent variable is 1 if head was born outside district of residence
Robust z statistics in parentheses
* significant at 5%; ** significant at 1%


China
POKHARA
KATHMANDU
LALITPUR
BIRGANJ
BIRATNAGAR
India
Out-Migration (number of people)
91 - 1,000
1,001 - 2,000
2,001 - 3,000
3,001 - 4,000
4,001 - 5,570


China
POKHARA
KATHMANDU
LALITPUR
BIRGANJ
BIRATNAGAR
India
In-Migration (number of people)
5 - 1,000
1,001 - 2,000
2,001 - 5,000
5,001 - 10,000
10,001 - 16,977