Intra-household Selection into Migration : Evidence from a Matched Sample of Migrants and Origin Households in Senegal

Isabelle Chort, Jean-Noël Senne

Intrahousehold Selection into Migration: Evidence from a Matched Sample of Migrants and Origin Households in Senegal

Isabelle CHORT[*] Jean-Noël SENNE[†]

Abstract

Migrant's selection issues are addressed by a great number of articles since the founder paper by Borjas (1987), which applies to international migration the Roy model of self-selection. However, most migration models usually regard location choices as an individual income-maximizing strategy and do not consider the collective dimension of the decision to migrate. In this paper, we therefore try to fill the gap in the literature between individual selection models and household-based migration decisions. We thus extend the Roy theoretical framework in order to account for household-based migration decisions and derive its implications on migrant selection. Assuming that the household maximizes its earnings including further remittances when choosing the one among its members who is to migrate, migrant selection in this case may differ from what is predicted by an individual decision model. We specifically tackle the so far under-explored issue of *intra-household* selection into migration and aim at determining which component of the household utility - earnings, remittances or non-monetary factors - mostly drives location choices. We provide empirical evidence from a unique matched sample of 900 Senegalese migrants in three destination countries - France, Italy and Mauritania - and their origin households in Senegal.

KEYWORDS : migration, remittances, intrahousehold allocation, selection models

JEL - CODES : F22, F24, D13, C51

[*]PSL, Université Paris-Dauphine, LEDa, UMR DIAL, 75016 Paris

[†]Paris School of Economics (PSE), CREST(ENSAE) and UMR DIAL. Address: 4 rue d'Enghien, 75010 Paris, Phone: +33 153241469, E-mail: jean-noel.senne@ensae.fr

1 Introduction

The estimation of migrant self-selection is a key issue in the migration literature. Self-selection into migration merely reflects the fact that migrants differ from non migrants with regard to both observed (education) and unobserved (motivation, psychic costs) characteristics. As a consequence, any study intending to evaluate the impact of migration on a range of outcomes, either in destination countries (labor market insertion) or in origin countries (income, health, consumption), needs to go beyond mere comparisons between migrant and non-migrant individuals or households and take selection bias into account[1].

Unsurprisingly, selection issues are tackled by a great number of articles since the founder paper by Borjas (1987) applying to international migration the Roy model of self-selection (Roy, 1951). In this theoretical framework, location choices depend on individuals' comparative advantage and observable characteristics, but also on their unobserved characteristics. All these papers derived from the Roy model of self-selection into different sectors of activity thus explicitly share an individualistic approach (see among others Chiswick (1999), Orrenius and Zavodny (2005), McKenzie and Rapoport (2010)), in line with the first models of migration developed by economists (Harris and Todaro (1970); Sjaastad (1962)). Indeed, in these models, individuals choose where to live and work according to their actual or expected gains at each location, once accounted for migration costs.

However no paper has yet addressed the issue of migrant selection *within* the origin household. The first migration models developed by Harris and Todaro (1970) and Sjaastad (1962) are individual ones. Subsequently though, following Stark and Bloom (1985) and based on socio-anthropological observations, the decision to migrate has been modelled at the household level. In this paper we derive the implications of a household model for migration on migrant selection: indeed, if migration is decided on collectively

[1] Indeed, in a paper on the impact of migration on inequality using household survey data collected in Nicaragua, Barham and Boucher (1998) illustrate the crucial importance of selection biases. Whereas a naive estimation considering that remittances are exogenous suggests a negative impact of migration on inequalities, they find a positive impact if remittances are considered to be substituted to the income migrants would have earned had they remained in their origin household.

within the household, so should be selected the migrant member. The selection of one or more migrants among household members may not be equivalent to individual self-selection into migration. We thus choose to extend the Roy model of selection to account for a household model of migration. We *de facto* specifically address the issue of intra-household selection into migration, that is to say, the question of who, among household members, is to migrate and live abroad.

This question has received very few interest to date[2] , one reason being probably the lack of demanding empirical data to address this issue and the econometric challenges it raises. The issue of migrant selection when migration choices are taken at the household level is indeed made much more complex by enlarging first the set of potential destinations, and more importantly by considering that households may split, and that the double choice of who migrates where and who stays is not random. Note already that the number of available alternatives for a given household depends on the size of the pool of potential migrant members, which varies across households.

In this paper, we extend the individual Roy model to account for the household allocation decision[3] of its members in different countries, including the home country, and solve the econometric issue raised by a varying number of alternatives across observations in a mixed logit model. We finally aim at determining which components of the household utility - earnings, remittances or individual characteristics affecting non-monetary utility - mostly drive location choices.

In the empirical section we build on Dahl (2002) who enriched the theoretical framework inherited from Roy (1951) in particular by implementing a semi-parametric method to correct for selection when allowing for multiple alternatives (see DeVreyer, Gubert, and Roubaud (2010) for an application to location choices in Africa).

[2]Note that the implications of a household approach to migration decision on migrant selection have been first addressed by Borjas and Bronars (1991). However they do not depart from the individual selection approach: the nature of migrant selection in their model mechanically results from the comparison of individual income distributions in both countries.

[3]We do not intend to settle the question of the collective *versus* unitary nature of the household migration decision (Fortin and Lacroix, 1997) and for the sake of simplicity choose a very simple household utility function of the unitary type.

Our model is estimated in the empirical part of the paper using an original matched sample of Senegalese migrants and their origin households. These data, collected in 2009-2010 as part of the MIDDAS project[4], provide information on migrants' characteristics in three of the top destination countries of Senegalese migrants (France, Italy, and Mauritania) as well as detailed information on all the members of their origin household in Senegal.

2 Theoretical model and empirical specification

In this section, we investigate selection issues and location choices within an extended Roy model to account for a household-based decision of migration. We consider that location choices of household members result from the maximization of the household utility whose monetary component includes earnings of each member in its relevant location and remittances from members abroad.

2.1 A household selection model for migration

The household can be regarded as a "portfolio" of members whose geographical allocation is decided on collectively. Each household has the choice to send members in any country or to have them stay in Senegal. Note that for the sake of simplicity, we first investigate the rationality of migrant selection and location, conditional on the fact that households are yet selected into migration. Indeed, we want to focus the analysis on intrahousehold selection, that is to say on the choice of the specific member who is to migrate and live abroad, once the decision to have a migrant has been taken[5]. Second, we also ignore the fact that one household can have several migrants and further focus the analysis on the rationality of one specific - observed - migrant allocation.

[4]MIDDAS is three-year project standing for "Migration and development in Senegal: an empirical analysis using matched data on Senegalese migrants and their origin households (MIDDAS)" funded by the French Agence Nationale de la Recherche and the Agence Française de Développement.

[5]Moreover, this will be consistent with the following empirical application and the structure of the matched data we use, which is exclusively composed of migrant households (see section 4)

We assume that households maximize a stochastic additively separable utility function which depends on the household's total home earnings and on the migrant's earnings at the chosen location. We additionally assume that the migrant's earnings can be decomposed into consumption at destination [6] and remittances to his origin household. Therefore, if we consider I potential migrant members within the household and J possible destination country, the random utility of household h having a migrant i in country j writes:

$$\begin{aligned} U_{hij} &= \alpha y_h + \beta(y_{ij} - r_{ij}) + \gamma r_{ij} + z_i'\delta + \epsilon_{hij} \qquad (\forall i = 1, ..., I \; ; \; \forall j = 1, ..., J) \\ &= \alpha y_h + \beta y_{ij} + (\gamma - \beta) r_{ij} + z_i'\delta + \epsilon_{hij} \end{aligned} \tag{1}$$

where y_h is the total log home earnings of remaining household members, y_{ij} the log earnings of migrant i in country j, r_{ij} the log amount of remittances to household h from migrant i in country j and z_i a vector of migrant i characteristics. $z_i'\gamma_j + \epsilon_{hij}$ corresponds to what Dahl (2002) defines as a taste vector, which represents the non-wage determinants entering the utility function, including in particular the fixed costs of moving and any other non-monetary or psychic costs and benefits for household h of having a member i in country j. It comprises a mean component which depends on migrant i characteristics z_i, and an unobserved individual-specific error component ϵ_{hij} for deviations from mean tastes.

The set of parameters $(\alpha, \beta, \gamma, \delta)$ are assumed to be identical across households. Moreover, the β and γ parameters are further assumed to be homogenous across destination countries. In other words, an increase in labor market earnings abroad or remittances provides identical gains in terms of utility, independently of the migrant's country of residence[7]. Also note that even if we assume that households are altruistic in the sense that

[6] Individual savings can enter this component without loss of generality.

[7] As noted by DeVreyer, Gubert, and Roubaud (2010), this might be too strong an assumption if large differences exist between countries in the set of available goods and their prices (for instance public services), so that the living standards of migrants with equal incomes but residing in different country would be indirectly impacted. However, we can credibly assume that households are not in a position to take this dimension into account in their utility. Moreover, earnings will be converted into Purchasing

they take into account earnings of members abroad, the weights on home earnings, α), on non-remitted earnings abroad (migrant's consumption and savings), β, and on remittances, γ, are allowed to differ in the utility function[8]. This very last point is crucial in the following analysis since this paper aims at identifying which component of the above defined household utility mostly drives location choices.

Households choose among $I \times J$ alternatives the member's geographical allocation that maximizes their collective utility, so that household h decides to locate the member i in country j according to:

$$M_{hij} = \begin{cases} 1 & \text{if} \quad U_{hij} = max(U_{h11}, ..., U_{h1J}, ..., U_{hI1}, ..., U_{hIJ}) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where M_{hij} is a indicator variable which is equal to one if household h locates the member i in destination country j and all remaining members in the home country[9]. The selection equations can alternatively be written as:

$$M_{hij} = \begin{cases} 1 & \text{if} \quad U_{hij} > U_{hkl} \qquad \forall (k,l) \neq (i,j) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The selection rule is such that each households' home earnings and migrants' earnings and remittances are observed for the household utility-maximizing allocation choice. In other words, a household can only locate each member in one specific destination, so that earnings and remittances are not observed for each member in every location but only if all $I \times J$ selection equations in (3) are satisfied simultaneously. Equations (1) to (3) define

Power Parity (PPP) units in the following empirical application, to allow relevant comparisons.

[8]We could conversely assume that households are not altruistic and only take remittances into account in their utility in order to determine the optimal location of their members. This would nevertheless put too much power in the hand of the household and cancel out the migrant's individual rationality. We consequently give a very general form to the utility without any restrictions on parameters that are allowed to vary across each monetary component considered.

[9]Formally, each household h faces a $I \times J$ number of alternatives, so that $I \times J$ binary variables M_{hkl} can actually be defined, corresponding to $I \times J$ selection equations. M_{hij} equals one if alternative $\{ij\}$ is chosen and observed; consequently all the remaining M_{hkl} equal 0 since only one allocation can be chosen by assumption. In other words, exactly one of the set of binary variables $M_{h11}, ..., M_{h1J}, ..., M_{hI1}, ..., M_{hIJ}$ is non-zero for each household h.

an extended Roy model of earnings, remittances and mobility, such as in Dahl (2002), but in which location choices are the result of household utility-maximizing strategy.

2.2 Estimation strategy

Since the central question is whether differences in potential earnings, remittances and non-monetary components of the utility determine location choices, we are particularly interested in estimating the set of structural parameters $(\alpha, \beta, \gamma, \delta)$ in equation 1. However, following the extended Roy model of selection that we described above, identification issues stem from the fact that earnings and remittances are only observed at one location for each household member. A selectivity bias results from the fact that households choosing a specific utility-maximizing geographical allocation are not a random sub-sample of the population. As a consequence, earnings and remittances in other locations must be imputed, taking into account the fact that location choices are not random but partially driven by observed and unobserved characteristics explaining earnings and remittance gaps.

Estimation is conducted in several steps. First, we assume that each migrant i living in destination country j faces a Mincer-type earnings equation:

$$y_{ij} = x_{1i}^{'}\theta_j + \eta_{ij} \qquad (\forall j = 1, ..., J) \tag{4}$$

where x_{1i} is a set of migrant i characteristics explaining log (destination) earnings y_{ij} and η_{ij} an individual-specific error component.

Second, we further assume that remaining household members in Senegal face that same type of Mincer-type earnings equation and consider that household h total home earnings can basically be computed as the sum of individual log earnings of remaining working members s[10], so that:

$$y_h = \sum_{s \neq i} y_s = \sum_{s \neq i} x'_{1s}\theta_0 + \mu_h \tag{5}$$

where x_{1s} is a set of household member s characteristics explaining log (home) earnings y_s and μ_h an household-specific error component[11].

Third, we similarly define a remittance equation for each migrant i living in destination country j:

$$r_{ij} = x'_{2i}\rho_j + \nu_{ij} \qquad (\forall j = 1, ..., J) \tag{6}$$

where x_{2i} is a set of migrant i characteristics explaining the log amount of remittances r_{ij} sent to origin household h and ν_{ij} an individual-specific error component.

Finally, we can substitute the above expressions of y_{ij}, y_h and r_{ij} in equation (1) to get the household utility in a reduced form:

$$U_{hij} = \alpha(\sum_{s \neq i} x'_{1s}\theta_0) + \beta(x'_{1i}\theta_j) + (\gamma - \beta)(x'_{2i}\rho_j) + z'_i\delta + \xi_{hij} \qquad (\forall i, j) \tag{7}$$

where $\xi_{hij} = \alpha\mu_h + \beta\eta_{ij} + (\beta - \gamma)\nu_{ij} + \epsilon_{hij}$.

Equation (7), together with the selection rule in equation (3), depicts the general framework of an additive random utility model (ARUM). Under the statistical assumption that error components ξ_{hij} have a type-1 extreme value distribution, it can be shown that the probability P_{hij} that household h locates a member i in country j:

$$P_{hij} = P(M_{hij} = 1) = P(U_{hij} > U_{hkl}) \qquad \forall (k, l) \neq (i, j) \tag{8}$$

[10]This aggregation rule is disputable, especially in agricultural rural areas where households might derive their income from a collective production function. However, for the sake of simplicity in such cases, we will either consider that earnings are individually separable or basically rely on per capita earnings.

[11]Formally, $\mu_h = \sum_{s \neq i} \mu_s$ where μ_s is an individual-specific error component. We define μ_h as the aggregated household-specific error component for ease of notation.

can be written[12]:

$$P_{hij} = \frac{exp[-\alpha(x'_{1i}\theta_0) + \beta(x'_{1i}\theta_j) + (\gamma - \beta)(x'_{2i}\rho_j) + z'_{hi}\delta_j]}{\sum_{k=1}^{I}\sum_{l=1}^{J} exp[-\alpha(x'_{1k}\theta_0) + \beta(x'_{1k}\theta_l) + (\gamma - \beta)(x'_{2k}\rho_l) + z'_{hk}\delta_l]} \tag{9}$$

These are conditional probabilities derived from a standard conditional logit model with a $I \times J$ fixed number of alternatives which corresponds to each possible intrahousehold members' allocation[13]. In this specific setting, the reduced-form probability that a member be selected as a migrant appears to depend both on his own individual characteristics and on the characteristics of all other potential migrant members within the household.

Yet, one important but non-standard issue for estimation is that households in the sample are not necessarily of equal size. As a consequence, the number of potential migrant members I should vary across households so that each household h actually faces a varying $I_h \times J$ number of alternatives. If we further assume that the set of parameters to estimate is identical across households, a solution consists in writing conditional probabilities and the log-likelihood function for a household single observation conditional on the specific number of alternatives available to the household, as follows:

$$\mathcal{L}_h = \ln(L_h) = \sum_{i=1}^{I_h}\sum_{j=1}^{J} M_{hij} \ln P_{hij} \tag{10}$$

where I_h is the number of potential migrant members in household h, J the fixed number of possible destination countries for migrant member i, M_{hij} a dummy equal to one if household h has a member i in country j and P_{hij} the associated conditional probability from equation (9) but whose denominator now depends on a household-specific $I_h \times J$ number of allocation choices. Equation (10) is an immediate generalization of the log-

[12]Note that $U_{hij} > U_{hkl}$ writes: $\alpha(\sum_{s \neq i} x'_{1s}\theta_0) + ... + \xi_{hij} > \alpha(\sum_{s \neq k} x'_{1s}\theta_0) + ... + \xi_{hkl}$ where sums on both sides of the inequality reduce: $-\alpha(x'_{1i}\theta_0) + ... + \xi_{hij} > -\alpha(x'_{1k}\theta_0) + ... + \xi_{hkl}$ and yields the simplified expression in equation (9).

[13]The conditional logit model is an immediate generalization of the usual multinomial logit model where the set of explanatory variables is alternative-specific but the whole set of parameters to estimate is alternative-invariant.

likelihood function from a standard conditional logit model where the set of alternatives is allowed to vary across observations. The log-likelihood function for a sample of N households then writes as usual:

$$\mathcal{L}_N = \sum_{h=1}^{N} \mathcal{L}_h = \sum_{h=1}^{N} \sum_{i=1}^{I_h} \sum_{j=1}^{J} M_{hij} \ln P_{hij} \tag{11}$$

Standard maximization routines, adapted to a varying number of alternatives, can then be applied to get consistent estimates of the reduced form set of parameters.

Results from this first step maximum-likelihood estimation are used in a second step to correct for endogenous selection in earnings and remittances equations, following the method suggested by Dahl (2002). The idea is to use the results of the above defined polychotomous choice model to compute, for each household, a set of predicted choice probabilities, then to correct parameters for selectivity bias by adding a polynomial control function of these probabilities as additional explanatory variables in equations (4), (5) and (6)[14]. As noted by Dahl (2002), a potential drawback to the conditional logit model is its independence of irrelevant alternatives (IIA) property and its reliance on a parametric framework, so that a non-parametric estimation of choice probabilities should be preferred. However, this is only feasible with a large number of observations. Moreover, based on Monte-Carlo simulations, Bourguignon, Fournier, and Gurgand (2007) later pointed out that, even when the IIA assumption is severely at odds, selection bias correction based on the conditional/multinomial logit can be considered a reasonable alternative when the focus is to estimate an outcome over selected populations. Therefore, results should not be affected by the choice of the conditional logit model at this stage.

In order to finally recover the set of parameters $(\alpha, \beta, \gamma, \delta_j)$ in the structural model, a last step is needed. Using unbiased estimates $\hat{\theta}_0$, $\hat{\theta}_j$ and $\hat{\rho}_j$ from selectivity-corrected equations (4), (5) and (6), we can compute unconditional average earnings and remittance predictions for each individuals in each possible location, and then estimate the following

[14] In practice, Dahl suggests to add the probability of the first-best location choice, the highest predicted probability and/or the retention probability.

structural conditional logit model with $I_h \times J$ alternatives:

$$P_{hij} = \frac{exp[-\alpha(x'_{1i}\hat{\theta}_0) + \beta(x'_{1i}\hat{\theta}_j) + (\gamma - \beta)(x'_{2i}\hat{\rho}_j) + z'_i\delta]}{\sum_{k=1}^{I_h}\sum_{l=1}^{J} exp[-\alpha(x'_{1k}\hat{\theta}_0) + \beta(x'_{1k}\hat{\theta}_l) + (\gamma - \beta)(x'_{2k}\hat{\rho}_l) + z'_k\delta]} \quad (12)$$

$$= \frac{exp[-\alpha\hat{y}_{hs} + \beta\hat{y}_{ij} + (\gamma - \beta)\hat{r}_{ij} + z'_i\delta]}{\sum_{k=1}^{I_h}\sum_{l=1}^{J} exp[-\alpha\hat{y}_{hk} + \beta\hat{y}_{kl} + (\gamma - \beta)\hat{r}_{kl} + z'_k\delta]}$$

2.3 Empirical specification and identification

In the next section, we provide an empirical application using a unique matched sample of Senegalese migrants in three different destination countries - namely France, Italy and Mauritania - and their origin households in Senegal. The originality of this dataset is that it records information on both migrants at destination and non-migrant members from their origin household in the home country. A detailed description of the survey design and the resulting matched dataset is provided in the next section. Following the above theoretical framework, empirical estimation proceeds in three main steps which rely on additional assumptions for identification that need to be properly discussed.

Step 1: Intrahousehold allocation choice - Reduced form

We first use the sample of migrants and non-migrant members from these migrants' origin household to estimate an intrahousehold selection equation using the reduced form specification of the conditional logit model of allocation choices from equation (9). To keep estimation tractable and further focus the analysis on the selection of the migrant member, we consider that household allocation choices are driven by average expected earnings abroad and therefore pool the three destination countries. One potential alternative thus refers to one (migrant) member abroad and all remaining members in Senegal. The number of alternatives varies according to the number of potential migrants within the household[15]. We define as potential migrants all working-age household members

[15]We could indeed extend the analysis by considering the destination choice as an additional parameter in the household decision. This would basically enlarge the set of alternatives available to the household. This would also require large samples to allow robust inference, so that this simplifying implementation

(i.e aged 18-59). The dependent variable is then a dummy equal to one for the chosen (observed) household member allocation among all possible (non-observed) allocations. Independent variables are by definition alternative-specific and include the gender, age and educational attainment (measured by the last grade completed) of the selected migrant member.

Estimation results from this stage are used to compute appropriate choice probabilities that are added to the second step earnings and remittance equation to correct for endogenous selection. Identification at this second stage consequently relies on the inclusion in the first step regression of at least one variable that explains household allocation choices but do not affect earnings and remittances. Following Munshi (2003) and Pugatch and Yang (2011), we exploit rainfall data as an exogenous source of variation in emigration from Senegal[16]. Indeed, precipitations in origin regions may affect emigration flows, although their net effect is ambiguous: precipitations lower than average may damage local economic conditions and generate or increase incentives to emigrate; but a negative shock on household income may also negatively affect emigration flows if household are credit-constrained. Although both reference papers find in the case of Mexico a negative effect of rainfall on migration (lower than average rainfall resulting in a larger stock of Mexican immigrants in the U.S.), the Senegalese context being much different, first-step results are of special interest.

In practice, we use variations in levels of rainfall by defining normalized yearly precipitation variables (z-scores) as observed precipitations minus a long term average (1970-2009), divided by the long-term standard deviation, and include as an additional relevant explanatory variable in the first step the average z-score over the 5 years previous to the observed migrant's year of departure. We can reasonably argue that rainfall in the home country has no impact on earnings at destination. We further argue that it has no effect on *current* earnings in Senegal nor on remittances from abroad since estimation relies on

choice was always guided by practical matters.

[16]Rainfall data are derived from gridded datasets of monthly precipitations published by the Climate Research Unit from the University of East Anglia: `http://www.cru.uea.ac.uk/fr`.

past levels of precipitations at the time of reported migration[17].

Step 2: Earnings and remittances equations

We estimate in a second step earnings equations (in Senegal and pooled destination countries), as well as a remittance equation (from pooled destination countries) using Mincer-type specifications from equations (4), (5) and (6). We run separate OLS regressions where the dependent variable is respectively the log of yearly earnings in Senegal, the log of yearly earnings in destination countries and the log of yearly remittances. For comparison purposes, all amounts are expressed in purchasing power parity (PPP). This conversion is also needed in the third step of the estimation where expected earnings and remittances are allowed to influence allocation choice probabilities[18]. Basic explanatory variables include gender, age (as a proxy for potential experience) and education level (as defined above). The home earnings equation is estimated on the sample of non-migrant household members with positive earnings and the destination earnings and remittance equation are estimated on the sample of migrants with positive earnings and remittances[19].

As above mentioned, we add to the set of explanatory variables a polynomial of choice probabilities obtained from step 1 in order to correct estimations from endogenous selection. In practice, we use the first best choice probability, that is the probability that the observed allocation was chosen and the highest predicted probability in the home earnings equation (for the “stayers”), and a polynomial of the first best choice probability in earnings abroad and remittance equations (for the “movers”)[20].

[17] As we do not observed the household composition at the year of the migrants’ departure, note that we also have to implicitly assume that actual non-migrant household members were relevant alternatives available to the household at the migrants’ year of departure (i.e were potential migrants at that time).

[18] We use the conversion factors published by the World Bank in its World Development Indicators. PPP factors for private consumption in 2009 (country currency units buying the same amount of consumption goods as 1 USD in the U.S.) were 0.85 for Italy, 0.92 for France, 143.03 for Mauritania and 332.56 for Côte d’Ivoire (http://data.worldbank.org/indicator/PA.NUS.PRVT.PP).

[19] An issue raised by both sample restrictions is that we do not properly take into account additional selection on the labour market and into remittances. This would add to much complexity to the matter at hand. We leave it for further research

[20] Theoretically, all choice probabilities could enter the control function. However, this would raise

Step 3: Intrahousehold allocation choice - Structural form

Parameters estimates from the second step are finally used in a third step to identify the effect of expected earnings and remittance differentials on allocation choices. Practically, using unbiased - i.e corrected for selection - estimates from earnings and remittance equations, we compute counterfactual earnings in Senegal of migrants, had they not migrated, and counterfactual earnings and remittances of non-migrants from their household, had they migrated abroad. Imputed earnings and remittances are then included as alternative-specific explanatory variables in the structural form of the conditional logit from equation (12). Additional explanatory variables include gender and age.

Identification at this stage relies on the exclusion from the structural allocation choice equation (12) of at least one variable that enters the earnings and remittance equations (4), (5) and (6). We argue here that the education level strongly affect earnings and remittances but not allocation choice, once earnings and remittances are accounted for. This might not be the case if households benefit directly from having an educated member at home, through externalities on other members for instance. However, it not clear whether non-monetary utility gains from the residence of an educated member in the household should be large. We could reasonably argue that this direct effect of education on allocation choices is negligible in comparison with its indirect (monetary) effect through earnings and remittances, so that the specified structural model should consistently identify the set of structural parameters (α, β, γ).

3 Data

This paper uses data from surveys conducted within the framework of the MIDDAS project[21]. Using Senegal as a case-study, this research project aims at increasing the

multicollinearity issues so that we have to restrict the set of included probabilities. Results are nevertheless robust to alternative specifications. We present here the one that best fit the data, following the method suggested by Dahl (2002)

[21] This three-year project (2008-2011) entitled "International Migration and Development: an Analysis using Matched data on Migrants and Origin Households in Senegal (MIDDAS)" was funded by

common knowledge on the relationships between migration, remittances and development, and holds part of its originality in the unique dataset that was collected, on which the subsequent analysis is based.

First, most of the existing studies on migration issues are based on data that is generally truncated: it is indeed collected either among migrants in host countries, thus providing only indirect and partial information on origin households, either among households in home countries, giving in this case very few and imprecise insights on the characteristics of migrants. The main objective of the MIDDAS project was therefore to build an original dataset matching representative samples of Senegalese migrants in host countries with their origin households in Senegal, in order to collect accurate information on both "sides" of migration[22].

Second, migrants were surveyed across four countries in two distinct receiving areas, namely France and Italy for the analysis of South-North migration to Europe, and Mauritania and Côte d'Ivoire for the analysis of South-South migration to Africa. These countries were selected as being the top-two destination countries in their respective area, and the top-four destination countries in the world for Senegalese migrants[23].

The resulting matched and multi-sited dataset provides the unique opportunity to investigate original issues that are not or poorly tackled by the existing literature due to a lack of appropriate data, such as the role of origin families in migrants' behavior and (intra-household) selection into migration for the case under study. It also helps bringing new insights on these topics through comparative analysis of migrations in various contexts.

---

the French Agence Nationale de la Recherche (ANR) and the Agence Française de Développement (AFD). Overall coordination was carried out by a research team from the Institut de Recherche pour le Développement (IRD-DIAL), with fieldwork support from researchers from the Forum Internazionale ed Europeo di Ricerche sull' Imigrazione (FIERI) in Italy and the Agence Nationale de la Statistique et de la Démographie (ANSD) in Senegal. For futher details on the global objectives and the institutional setting of the MIDDAS project, see `http://www.dial.prd.fr/dial_enquetes/dial_enquetes_middas.htm`.

[22]To our knowledge, Osili (2007) is the only other empirical study based on a matched sample of migrants in Chicago and origin households in Nigeria.

[23]Excluding the Gambia, due to its peculiar landlocked position within the Senegalese territory, migration flows to these four countries accounted for 65% of total emigration flows between 1995 and 2002, according to the last 2002 Senegalese census. Spain and the United States are the two other countries that attract a non-negligible share of the remaining flows (17% of total).

3.1 Survey design

The MIDDAS project was phased in two successive stages. First, surveys were conducted among representative samples of Senegalese migrants in the four above-mentioned destination countries. Second, their origin households were tracked and interviewed in Senegal, thanks to contacts provided by the migrants themselves. All origin households have been tracked, except those of migrants residing in Côte d'Ivoire, due to budgetary constraints[24]. We thus focus the following analyses on the French, Italian and Mauritanian migrant-household matched samples.

Migrant surveys

Stage 1 migrant surveys took place in three waves between 2009 and 2010, successively in France, Italy and Mauritania, and using common sampling design, fieldwork procedures and questionnaire. Any attempt to carry out a representative survey focused on international migrants faces the issue that they usually represent a very small proportion of the population in a given country and that no survey frame is available[25]. To mitigate these two problems, a survey methodology similar to the one applied by Lydié, Guilbert, and Sliman (2008) in their survey on Sub-Saharan Africans in Greater Paris was adopted. First, the most recent population censuses were used in each country to identify regions, and within them cities and districts, hosting significant populations of Senegalese migrants. When the number of potential districts to survey within a region or city was important, three strata were constructed according to the density of the Senegalese population in each district. Districts to be surveyed were then randomly drawn within each stratum with probabilities proportional to the number of Senegalese in those districts. The number of migrants to be interviewed was finally determined using the relative weight of the relevant district in the total Senegalese population. This sampling

[24]Moreover, due to rising political instability in the country preceding the first round of the 2010 presidential election, stage 1 migrant survey only took place in Abidjan, so that the migrant sample from Côte d'Ivoire is only representative at the capital city level.

[25]For a detailed discussion on issues raised by migrant surveys and a comparison of the performance of alternative survey methods, see McKenzie and Mistiaen (2009)

method aimed at achieving representativeness firstly at the geographic level[26].

To get as various a sample as possible, surveyors were then sent in the selected cities/districts and tasked with getting randomly in contact with migrants in the public space (streets, markets or shopping centers, public transport stations, etc.). To be eligible, interviewees had to meet three compulsory criteria: (i) being aged 18 and over; (ii) residing in the relevant city/district; and (iii) either being a Senegalese national or a former Senegalese national. To further ensure representativeness, surveyors were also asked to keep diversity according to gender, age, education and migration history. Contacting groups and snowballing were prohibited. For comparative purposes, the same full questionnaire was administered in each country, with nonetheless marginal adaptations depending on the context. It was designed to cover a wide range of quantitative and qualitative aspects of the migrant experience: socio-demographic characteristics, housing and living conditions, employment and earnings, migration history, links to the origin household, remittance behavior, return and investment projects in Senegal, insertion and social networks in the host country.

Migrant sample size and composition by country are given on top of table 1. 300 migrants were interviewed in France (24.3% of which are women), 302 in Italy (22.9% of which women) and 326 in Mauritania (36.5% of which are women), mostly in the capital and/or the main cities of the country. Note that surveyors were also asked to record as much information as possible on those overall 41.9% of eligible migrants who refused to be interviewed. If one excludes gender (women being more reluctant than men to answer the questionnaire), no significant differences can be found in terms of age distribution and date of arrival between those who accepted and those who refused to be part of the survey.

[26] In practice, this stratified sampling process was implemented for cities within the Paris region for France and within the northern regions of Italy, which host in both cases the vast majority of the Senegalese population. In Mauritania, migrants were much more concentrated in a small number of districts within the capital city, Nouakchott, and the two other main cities of the country, so that all could be included in the sample. For further details on sampling and fieldwork procedures in each countries, see Gubert and Senne (2010)

Origin household surveys

Stage 2 origin household tracking surveys took place in Senegal in 2009 for the French migrant sample and in 2010 for the Italian and Mauritanian ones, using contacts provided by the migrants at stage 1. Indeed, a whole section of the migrant questionnaire was dedicated to the household defined by the migrant as being his household of origin, and primarily designed to record self-reported basic information such as household demographics, socio-economic characteristics and wealth measures. However, at the end of this section, the migrant was also asked to give the address and telephone number of a referent member in the origin household, so that the latter could be tracked and more accurately surveyed to build a matched sample.

The questionnaire that was administered to successfully tracked origin households was taken from the PSF survey, which is a nationally representative household survey conducted in Senegal in 2007[27]. Reasons leading to this choice were threefold. First, this rich questionnaire covers a full range of demographic and socio-economic household characteristics. Second, this innovative survey was designed to grasp the complex structure of Senegalese households. It therefore records precise information not only on each resident household members, but also on those who left and on the relationships they maintain with the remaining members. Migrants' position and role within this structure can thus be accurately comprehended. Third, the 2007 survey round provides useful information on an additional representative sample of Senegalese households, including migrant and non-migrant ones.

Origin household sample size and composition by country are given on bottom of 1. 92 households were successfully tracked from migrants surveyed in France, 62 from those surveyed in Italy and 174 from those surveyed in Mauritania, that is an overall matching

[27]The "Pauvreté and Structure Familiale" survey was designed and carried out by a research team from the INRA-LEA and Paris-Dauphine University, with fieldwork support from the Senegalese Agence Nationale de la Statistique et de la Démographie (ANSD) and funding from the International Development Research Centre (IRDC). It aims at studying the complex structure of the Senegalese household and how it is related to household and individual well-being. 1,800 households were surveyed in 2006-2007, while a follow-up round was implemented in 2012, in order to build a panel dataset and further investigate poverty dynamics. For further details on the PSF survey, see DeVreyer et al. (2008).

rate of 35.3%. This rate strongly varies depending on the destination country. This can be explained by two factors. First, some migrants squarely refused to provide any contact in their origin household at stage 1. Second, provided contacts sometimes proved insufficient or wrong - voluntarily or not. Mistrust was mainly observed for surveys in Europe, especially in Italy where the matching rate only amounts to 20.5% due to a highly tense climate surrounding migration issues at the time of the survey. On the contrary, migrants in Mauritania were much less reluctant when it came to origin household questions, so that 53.4% of households could be tracked and surveyed[28]. Nevertheless note that, if the overall matching rate seems low prima facie, matching among available contacts amounts to ...% which is sensibly higher than any other (rare) survey of this type. Finally, 328 migrant-household pairs compose the matched sample, whose representativeness is assessed in the next section.

3.2 Sample representativeness

The sampling design of the migrant surveys at destination was aimed at constituting a representative sample of the population of Senegalese migrants living in these countries. The availability of census data for France and Italy allow us to assess the representativeness of our resulting migrant samples [29]. Only basic characteristics such as the place or residence, gender and education are available in censuses. The comparison of these characteristics for the MIDDAS sample and the Senegalese population according to census data (table 1 and 2 in Appendix) suggest in particular that men and young migrants are over-represented in our sample, especially in France. We are not able to provide the same analysis for Mauritania for lack of data.

Moreover, we need to adress selection issues resulting from imperfect matching. First, between 19% (in Italy) and 62% (in Mauritania) of all interviewed migrants refused or

[28]The geographic proximity between Mauritania and Senegal and the frequency of "physical" encounters between the migrant and the remaining household members also surely explain this highest matching rate.

[29]These data are made available by the OECD [ad ref.]

were not able to provide the address or contact of their origin household in Senegal. Second, among those who provided a contact, only 54 to 65% could actually be matched. Much of the variation in the rate of matching across destinations is due to the political context prevailing at the time of the surveys. [30] Note that although these matching rates may seem low on average, they are well above those achieved by other surveys conducted among migrants and using the same methodology (see Beauchemin and González-Ferrer (2011))

We explore the potential selection of the matched sample with regard to observable characteristics of migrants and their household (based on information collected in the migrant questionnaire) by running probit regression where the dependent variable is the probability to have been matched. Results are shown in table in Appendix. No systematic differences appear between the matched and unmatched samples, but we observe a few country-specific differences between the two subsamples. Matched migrants are older and less likely to belong to the Wolof ethnic group in France, and they are more likely to be in the highest income quartile for the Mauritanian sample.

Finally, we need to assess the representativeness of our matched sample of migrant households. Using data from the PSF survey, which were designed to be nationally representative at the household level, we show that our sample of migrant households resulting from the matching procedure is well enough representative of Senegalese migrant households. Tables 5 and 6 provide comparison of the characteristics of our matched sample of migrant households and those of migrant households included in the PSF sample. Although some differences are statistically significant, MIDDAS and PSF migrant households appear to be much similar according to basic characteristics such as size, demographic composition or education.

[30] The especially low matching rate in Italy is in part attributable to the climate of mistrust and insecurity generated by anti-immigration speaches and measures.

3.3 Descriptive statistics

The characteristics of migrant households depending on their migrant's location are shown in table 7. Households with a migrant in Italy (and to a lesser extent in France) are more urban, whereas households with a migrant in Mauritania are unsurprisingly more likely to be settled in the region of the Senegal river valley bordering Mauritania. Note that a relatively high share of households with a migrant in France are also located in the North-Eastern part of Senegal. This feature is explained by the historical migration flows from this region mainly inhabited by Haalpulaar'en. The Mouride brotherhood is dominant among households having a migrant in Italy. Migrants in Italy come more frequently from households whose head has at least some secondary education. The monthly earnings of the household head are twice lower when the household has a migrant in Mauritania rather than in Europe.

Individuals' characteristics depending on their migration status are given in table 8. Migrants are a few years older on average than non-migrants. 71.9% of migrants are men, whereas men represent only 42.6% of household members staying in Senegal. Migrants are on average more educated, more likely to be married and to be working at destination. Their PPP earnings are around 4.5 times higher than those of non migrants.

However, this global picture eclipse the observed differences between migrants' characteristics depending on their destination shown in table 9. Migrants are predominantly male, even though we are closer to gender balance in Mauritania. Migrants in Mauritania are more similar to non migrants than migrants in Europe as regards many of their characteristics, and especially education. The relatively low proportion of migrants currently working (70%) in Italy reflects the adverse labor-market conditions at the time of the survey. PPP earnings and remittances are almost the same in France and Italy. Both are much lower in Mauritania but remittances amount to a slightly larger share of migrants' income (33%, against 28% in France).

4 Results

Results from the reduced form intra-household selection equation are provided in table 10. The dependent variable in the model is a dummy which is equal to one for the observed allocation, the surveyed migrant being abroad and all other members of his origin household in Senegal. The positive coefficient on the male dummy means that households tend to choose on average the allocations in which men migrate. However we do not know whether this is because men earn more and remit more on average or because households have a preference for women to stay in the household (or for men to leave). We find similar results for education and age: allocations in which the migrant is older and more educated are more likely to be observed. Another way of interpreting these coefficient is the following: for a Senegalese aged 18-64, having at least a high school degree increases the probability that his household prefers the allocation in which he migrates rather than another member of the household.

The second column shows the results from the same model, with rainfall variables added to the set of regressors. All coefficients on interaction terms between precipitations and gender, age or education are positive: rainfall higher than the long-term average seem to accentuate the above discussed features. Men, older and more educated household members are even more likely to migrate when the household has been faced with a positive rainfall shock (which we may interpret as a positive income shock).

Results from estimations of earnings and remittance equations, both uncorrected and corrected for selection, are given in table 11. They exhibit quite usual patterns: men and educated migrants tend to earn more on average than respectively women and migrants that never went to school, whereas the effect of potential experience, as measured by age, is positive but with decreasing marginal returns. This is true for both home earnings and destination earnings. We find the same patterns for remittance equations. Yet, education level is only significant for migrants having at least completed high school. We also find a significant positive effect of koranic schooling, suggesting that those migrants seem to commit to solidarity norms conveyed by religion.

An interesting feature of those results is that selectivity-corretion terms are jointly significant in each relevant equation, which put forward the fact that not taking migrant's selection into account would induce a non-negligible bias in the estimations. Indeed, the comparison of uncorrected and corrected results reveals that coefficients on independent variables are in general slightly under-estimated in the uncorrected home earnings equation and largely under-estimated in the uncorrected destination earnings equation, whereas they are over-estimated in the uncorrected remittance equation.

Results from the structural form intra-household selection equation are provided in table 12. Estimation is conducted using here as independant variables imputed home earnings of migrants, had they not migrated, and imputed destination earnings and remittances from non-migrants of their origin household, had they migrated, from corrected earnings and remittance equations. Coefficients on these variables allow us to identify the weights households allocate to each of these components in their underlying structural utility. However note that, from equation (12), the coefficients on home earnings, destination earnings and remittances respectively identify $-\alpha$, β and $\gamma-\beta$. Simple computation then gives the desired paramaters that are reported at the bottom of table 12.

We finally find negative α and β estimated parameters on home and destination earnings and a positive γ parameter on remittances. This original result suggests that households tend to choose on average allocations in which they send abroad the member who induces the minimal loss in home earnings but not necessarily the one who has the highest expected earnings abroad; rather the one who will send the highest amount of remittances from his destination country. Another way of interpreting these results is the following: for a Senegalese aged 18-64, having a high earnings potential decreases the probability that his household prefers the allocation in which he migrates rather than another member of the household, whereas this choice probability is higher for those having high remittance potential. Consequently, remittances rather than income abroad seem to be a relevant and crucial parameter that matters in the intrahousehold selection of the migrant member.

5 Conclusion

Although tackled by a great number of papers, migrants' selection has mostly been modeled as the result of an individual income-maximizing strategy. However, since the New Economics of Migration, migration has been increasingly viewed as the result of a household decision, especially in developing countries. In line with this strand of literature we model migrants' selection as a household utility-maximizing strategy. Therefore, this paper aims at shedding a new light on the selection process of migrants by investigating the so far under-explored issue of intrahousehold selection into migration.

We first extend the seminal Roy model of self-selection to account for a household model for migration. In this framework, households base their migration decisions on the maximization of a collective utility whose components include earnings of non-migrants members in the home country but also earnings and remittances from migrant members abroad. Using observed allocation choices of household members, we develop a three-step estimation procedure to estimate the weight on each component in the structural intrahousehold selection decision. We provide an empirical application using a unique matched sample of Senegalese migrants in France, Italy and Mauritania and their origin households in Senegal.

Results show that the probability to be selected as a migrant member within the household depends negatively on earnings in the home country. More surprisingly, it also depends negatively on expected earnings at destination. Finally, it depends positively on remittances from abroad. Overall, these results seem to suggest that households select in migration those who have the highest remittance potential but not necessarily the highest potential earnings abroad. This stands in striking contrast with results derived from usual individual selection models which suggest that individuals tend to locate where they expect higher earnings.

References

Barham, B. and S. Boucher. 1998. "Migration, remittances, and inequality: estimating the net effects of migration on income distribution." *Journal of Development Economics* 55 (2):307–331.

Beauchemin, Cris and Amparo González-Ferrer. 2011. "Sampling international migrants with origin-based snowballing method: New evidence on biases and limitations." *Demographic research* 25.

Borjas, G.J. 1987. "Self-Selection and the Earnings of Immigrants." *The American Economic Review* 77 (4):531–553.

Borjas, G.J. and S.G. Bronars. 1991. "Immigration and the Family." *Journal of Labor Economics* 9 (2):123–148.

Bourguignon, F., M. Fournier, and M. Gurgand. 2007. "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons." *Journal of Economic Surveys* 21 (1):174–205.

Chiswick, B.R. 1999. "Are Immigrants Favorably Self-Selected?" *The American Economic Review* 89 (2):181–185.

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

DeVreyer, P., F. Gubert, and F. Roubaud. 2010. "Migration, Self-selection and Returns to Education in the WAEMU." *Journal of African Economies* 19 (1):52–87.

DeVreyer, P., S. Lambert, A. Safir, and M.B. Sylla. 2008. "Pauvreté et Structure Familiale, pourquoi une nouvelle enquête?" *Stateco* 102:261–275.

Fortin, B. and G. Lacroix. 1997. "A Test of the Unitary and Collective Models of Household Labour Supply." *The Economic Journal* 107 (443):933–9–55.

Gubert, F. and J.N. Senne. 2010. "Projet MIDDAS: Premiers résultats d'enquêtes." *Mimeo DIAL* .

Harris, J.R. and M.P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *The American Economic Review* 60 (1):126–142.

Lydié, N., P. Guilbert, and G. Sliman. 2008. "Exemple de méthodologie pour une enquête auprès des populations d'Afrique subsaharienne." In *Méthodes de sondage*. Hachette, p224.

McKenzie, D. and H. Rapoport. 2010. "Self-selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks." *The Review of Economics and Statistics* 92 (4):811–821.

McKenzie, D.J. and J. Mistiaen. 2009. "Surveying migrant households: a comparison of census-based, snowball and intercept point surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (2):339–360.

Munshi, K. 2003. "Networks in the modern economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics* 118 (2):549–599.

Orrenius, P.M. and M. Zavodny. 2005. "Self-selection among undocumented immigrants from Mexico." *Journal of Development Economics* 78:215–240.

Osili, U.O. 2007. "Remittances and savings from international migration: Theory and evidence using a matched sample." *Journal of Development Economics* 83 (2):446–465.

Pugatch, T. and D. Yang. 2011. "The Impact of Mexican Immigration on US Labor Markets: Evidence from Migrant Flows Driven by Rainfall Shocks." Tech. rep., Tech. rep.

Roy, A.D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2):135–146.

Sjaastad, L.A. 1962. "The Costs and Returns of Human Migration." *Journal of Political Economy* 70 (4):80–93.

Stark, O. and D.E. Bloom. 1985. "The New Economics of Labor Migration." *American Economic Review* 75 (2):173–78.

Appendix

Table 1: Sample size and composition by country

	France	**Italy**	**Mauritania**	**Pooled**
Stage 1: Migrant samples				
Nb of eligible migrants	579	616	402	1,597
Refusal rate (%)	48.2	51.0	18.9	41.9
Nb of surveyed migrants	**300**	**302**	**326**	**928**
...% of women	*24.3*	*22.9*	*36.5*	*28.1*
...% in capital/main cities	*72.3*	*48.0*	*73.0*	*64.0*
Stage 2: Origin household samples				
Nb of provided contact	158	114	266	538
Matching rate (%)				
... overall	*30.7*	*20.5*	*53.4*	*35.3*
... among provided contacts	*58.2*	*54.4*	*65.4*	*61.0*
Nb of tracked households	**92**	**62**	**174**	**328**
...% in Dakar	*46.7*	*54.8*	*21.3*	*34.8*

Source: MIDDAS Survey, 2009 - 2010

Table 2: Sample representativeness by country

		France		Italy	
		Census	**MIDDAS**	**Census**	**MIDDAS**
Gender (%)	Men	54.7	75.5	88.1	77.3
	Women	45.3	25.5	11.9	22.7
Age (%)	20-29 year	20.1	27.6	15.4	23.4
	30-39 years	22.3	35.0	49.4	40.5
	40-49 years	25.0	21.6	29.2	30.4
	50-60 years	20.4	12.6	4.4	5.7
	60+ years	12.2	3.2	1.6	0.0
Duration of stay (%)	up to 5 years	17.5	14.8	29.2	18.9
	5 to 10 years	12.2	33.8	26.7	35.1
	10+ years	70.3	51.4	44.2	46.0
Citizenship (%)	National	58.6	25.5	1.6	2.3
	Other country	41.4	74.5	98.4	97.7
Education (%)	ISCED 0/1/2	45.1	54.6	83.8	48.5
	ISCED 3/4	26.9	20.3	12.3	20.1
	ISCED 5/6	28.0	25.2	3.9	26.4
Labor force status (%)	Employed	54.8	74.8	79.5	70.5
	Unemployed	12.6	14.1	9.1	21.2
	Inactive	32.6	11.1	11.4	8.3
Observations		**93,076**	**286**	**28,030**	**299**

Source: DIOC 2005/06, OECD and MIDDAS Survey, 2009 - 2010
Note: OECD census data records information on all individuals born in Senegal, aged 20 and above and living in an OECD country. MIDDAS sample is restricted to this sub-population for comparison purpose. ISCED 0/1/2 corresponds to no formal education, primary and lower secondary education; ISCED 3/4 to upper secondary, vocational and technical education; ISCED 5/6 to tertiary education.

Table 3: Sample representativeness by country and gender

		France				Italy			
		Men		**Women**		**Men**		**Women**	
		Census	**MIDDAS**	**Census**	**MIDDAS**	**Census**	**MIDDAS**	**Census**	**MIDDAS**
Age (%)	20-29 year	18.1	26.4	22.5	31.4	12.8	22.9	34.8	25.0
	30-39 years	20.5	36.1	24.4	31.4	49.7	39.0	46.7	45.6
	40-49 years	22.1	21.3	28.5	22.9	31.4	31.6	13.2	26.5
	50-60 years	24.0	12.0	16.1	14.3	4.6	6.5	3.2	2.9
	60+ years	15.3	4.2	8.5	0.0	1.5	0.0	2.1	0.0
Duration of stay (%)	up to 5 years	16.7	16.4	18.4	9.9	26.6	19.3	48.0	17.4
	5 to 10 years	11.2	34.7	13.3	30.9	26.3	32.2	29.0	44.9
	10+ years	72.1	48.9	68.3	59.2	47.1	48.5	23.0	37.7
Education (%)	ISCED 0/1/2	44.4	51.4	45.9	64.3	84.8	47.6	76.4	51.5
	ISCED 3/4		26.6	29.2	27.3	11.6	22.9	17.8	32.3
	ISCED 5/6	29.0	28.2	26.8	15.7	3.6	29.5	5.8	16.2
Labor force status (%)	Employed	58.8	76.3	50.0	70.4	84.6	73.4	42.4	69.9
	Unemployed	10.4	12.3	15.1	19.7	8.3	19.7	13.9	26.1
	Inactive	30.8	11.4	34.9	9.9	7.0	6.9	43.7	13.0
Observations		**50,941**	**216**	**42,135**	**70**	**24,707**	**231**	**3,323**	**68**

Source: DIOC 2005/06, OECD and MIDDAS Survey, 2009 - 2010

Note: OECD census data records information on all individuals born in Senegal, aged 20 and above, and living in an OECD country. MIDDAS sample is restricted to this subpopulation for comparison purpose. ISCED 0/1/2 corresponds to no formal education, primary and lower secondary education; ISCED 3/4 to upper secondary, vocational and technical education; ISCED 5/6 to tertiary education.

Table 4: Probit Analysis of Matching Success

	France (1)	Italy (2)	Mauritania (3)	Pooled (4)
Migrant characteristics				
Male (d)	0.016	0.043	-0.022	0.000
	(0.078)	(0.049)	(0.076)	(0.044)
Age (c)	0.013***	0.004	0.008*	0.009***
	(0.005)	(0.003)	(0.004)	(0.002)
Ethnic group				
Wolof	-0.151**	-0.105	0.198**	-0.016
	(0.069)	(0.074)	(0.080)	(0.045)
Peul	-0.082	-0.020	0.063	-0.043
	(0.074)	(0.075)	(0.102)	(0.053)
Other	(ref)	(ref)	(ref)	(ref)
Religion/Brotherhood				
Murid	0.090	-0.064	0.109	0.014
	(0.154)	(0.143)	(0.127)	(0.080)
Tidjan	-0.015	-0.054	0.141	0.030
	(0.139)	(0.107)	(0.125)	(0.080)
Other muslim	0.015	0.206	0.166	0.062
	(0.133)	(0.264)	(0.127)	(0.081)
Other	(ref)	(ref)	(ref)	(ref)
Koranic schooling (d)	0.151*	-0.025	0.063	0.085*
	(0.080)	(0.067)	(0.089)	(0.047)
Formal schooling				
Elementary/Secondary	0.014	0.016	0.087	0.027
	(0.087)	(0.081)	(0.071)	(0.047)
Highschool/University	0.012	0.055	-0.064	0.027
	(0.093)	(0.084)	(0.120)	(0.056)
Vocational	-0.016	0.142	0.229*	0.056
	(0.138)	(0.172)	(0.134)	(0.085)
No formal schooling	(ref)	(ref)	(ref)	(ref)
Marital Status				
Married	-0.069	0.000	-0.062	-0.046
	(0.077)	(0.061)	(0.085)	(0.048)
Divorced/Widowed	-0.101	-0.099**	-0.068	-0.108*
	(0.095)	(0.041)	(0.132)	(0.062)
Single	(ref)	(ref)	(ref)	(ref)
Time since arrival (c)	-0.010**	0.003	-0.006	-0.006**
	(0.004)	(0.004)	(0.005)	(0.003)
Capital/Main cities (d)	-0.052	-0.077*	0.068	-0.014
	(0.072)	(0.044)	(0.076)	(0.038)
Labor status				
Working	0.105	0.748***	0.067	0.127
	(0.137)	(0.099)	(0.171)	(0.091)
Unemployed	0.221	0.998***	0.266*	0.240**
	(0.176)	(0.002)	(0.138)	(0.108)
Non-working	(ref)	(ref)	(ref)	(ref)
Income				
2nd quartile	0.155	0.302**	0.149	0.203***
	(0.102)	(0.134)	(0.093)	(0.059)
3rd quartile	-0.052	0.125	0.110	0.086
	(0.110)	(0.127)	(0.101)	(0.065)
4th quartile	-0.010	0.091	0.266***	0.171**
	(0.110)	(0.132)	(0.095)	(0.070)
Missing	0.080	0.169	-0.336	0.025
	(0.201)	(0.207)	(0.298)	(0.119)
1st quartile	(ref)	(ref)	(ref)	(ref)

(continued)

Table 4 (continued)

	France (1)	Italy (2)	Mauritania (3)	Pooled (4)
Origin household characteristics				
Rural (d)	0.089	-0.052	0.023	0.060
	(0.082)	(0.050)	(0.067)	(0.044)
Size (c)	0.004	0.004	0.006	0.005**
	(0.003)	(0.003)	(0.005)	(0.002)
Missing size (d)	0.084	-0.060	-0.228	-0.046
	(0.224)	(0.068)	(0.294)	(0.115)
Resident spouse/child (d)	-0.098	-0.035	0.127*	0.016
	(0.075)	(0.048)	(0.072)	(0.042)
Wealth score (c)	0.008	0.038***	0.010	0.029***
	(0.017)	(0.010)	(0.026)	(0.010)
Remittances in cash/kind (d)	-0.104	-0.036	0.033	0.006
	(0.111)	(0.065)	(0.082)	(0.048)
Remittances amounts (in euros) (c)	0.002	0.001	0.001	0.001
	(0.002)	(0.001)	(0.005)	(0.001)
Missing amounts	-0.014	-0.115***	-0.305***	-0.203***
	(0.118)	(0.037)	(0.107)	(0.047)
Country dummies				
Italy				-0.117**
				(0.051)
Mauritania				0.255***
				(0.056)
France	(ref)	(ref)	(ref)	(ref)
Observations	**300**	**302**	**326**	**928**

Source: MIDDAS survey, 2009-2010
Note: Marginal effects at the mean for continuous variables, at 0 for dummy variables. Robust standard errors in parenthesis. (d) stands for dummy variables, (c) for continuous variables
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Sample representativeness

	PSF with migrants (1)	MIDDAS (2)	χ^2/Fisher/t (1)-(2)	PSF without migrants
Household characteristics				
Size	10.6	11.9	1.92*	7.8
Dependency ratio	44.3	41.6	-1.66*	41.1
Proportion of (%)				
...*Male*	36.2	42.0	2.97***	47.5
...*Female*	63.8	58.0		52.5
Land tenure (%)				
...*Yes*	29.0	23.4	2.05	35.4
...*No*	71.0	76.6		64.6
Environment(%)				
...*Urban*	60.8	69.2	3.61*	55.2
...*Rural*	39.2	30.8		44.8
Region(%)				
...*Dakar*	41.9	37.2	21.98***	35.2
...*North and East*	21.5	32.3		14.4
...*South*	14.5	4.3		9.8
...*Center*	22.0	26.2		40.5
Household head characteristics				
Age (in years)	52.3	58.2	4.37***	49.9
Gender (%)				
...*Male*	52.7	62.2	4.49**	80.7
...*Female*	47.3	37.8		19.3
Religion (%)				
...*Murid*	23.7	27.4	0.81	34.6
...*Other*	76.3	72.6		65.4
Education (%)				
...*None*	60.2	54.8	2.32	62.8
...*Primary*	15.6	20.9		20.3
...*Secondary and higher*	24.2	24.3		16.9
Observations	**189**	**328**		**1524**

Source: MIDDAS Survey, 2009 - 2010

Notes: χ^2 / Fisher's exact tests for the equality of distributions for categorical and dummy variables, t-test for the equality of meansfor continuous variables, between columns (1)-(2)samples.

*** 1% significance level, ** 5% significance level, * 10% significance level.

Table 6: Sample representativeness

	PSF with migrants		MIDDAS		χ^2/Fisher/t		PSF
	Out of Africa	In Africa	France/Italy	Mauritania			
	(1)	(2)	(3)	(4)	(1)-(3)	(2)-(4)	without migrants
Household characteristics							
Size	10.9	10.3	12.8	11.1	1.75*	1.03	7.8
Dependency ratio	42.7	47.1	39.1	43.8	-1.55	-1.27	41.1
Proportion of (%)							
...*Male*	32.5	40.8	42.7	41.4	3.84***	0.22	47.5
...*Female*	67.5	59.2	57.3	58.6			52.5
Land tenure (%)							
...*Yes*	21.3	39.5	25.3	21.6	0.57	8.98***	35.4
...*No*	78.7	60.5	74.7	78.4			64.6
Environment (%)							
...*Urban*	67.6	51.9	74.7	64.3	1.57	3.36	55.2
...*Rural*	32.4	48.1	25.3	35.7			44.8
Region (%)							
...*Dakar*	49.1	30.9	53.9	22.2	0.91	33.82***	35.2
...*North and East*	17.6	27.2	18.2	45.0			14.4
...*South*	4.6	27.2	3.9	4.7			9.8
...*Center*	28.7	14.8	24.0	28.1			40.5
Household head characteristics							
Age (in years)	52.8	51.8	58.2	58.1	2.81***	3.41***	49.9
Gender (%)							
...*Male*	46.3	60.5	62.3	62.0	6.62***	0.07	80.7
...*Female*	53.7	39.5	37.7	38.0			19.3
Religion (%)							
...*Murid*	31.5	14.8	28.6	26.3	0.26	4.06*	34.6
...*Other*	68.5	85.2	71.4	73.7			65.4
Education (%)							
...*None*	58.3	63.0	44.8	63.7			62.8
...*Primary*	12.0	19.8	20.1	21.6	5.35*	0.36	20.3
...*Secondary and higher*	29.6	17.3	35.1	14.6			16.9
Observations	**108**	**81**	**154**	**174**			**1524**

Source: MIDDAS Survey, 2009 - 2010

Notes: χ^2 / Fisher's exact tests for the equality of distributions for categorical and dummy variables, t-test for the equality of means for continuous variables, between columns (1)-(3)/columns (2)-(4) samples.

*** 1% significance level, ** 5% significance level, * 10% significance level.

Table 7: Household characteristics by migrant's location

	France (1)	**Italy** (2)	**Mauritania** (3)
Household characteristics			
Zone (%)			
...*Dakar*	47.8	62.9	22.1
...*North/East*	29.3	1.6	45.3
...*South*	4.3	3.2	4.7
...*Center*	18.5	32.3	27.9
Environment (%)			
...*Urban*	70.7	80.6	64.0
...*Rural*	29.3	19.4	36.0
Land tenure (%)			
...*No*	63.0	91.9	77.3
...*Yes*	37.0	8.1	22.7
Composition (%)			
...*Children (15-)*	32.9	29.1	37.9
...*Adults (15-65)*	62.0	62.3	57.2
...*Elderly (65+)*	5.1	8.6	4.9
Size	14.9	9.7	11.2
Dependency ratio	38.0	37.7	42.7
Household head characteristics			
Age	57.6	59.4	58.4
Gender (%)			
...*Male*	66.3	58.1	64.0
...*Female*	33.7	41.9	36.0
Ethnic group (%)			
...*Wolof*	19.6	71.0	61.0
...*Serere*	9.8	11.3	9.3
...*Peul*	23.9	6.5	15.1
...*Diola*	5.4	3.2	9.3
...*Other*	41.3	8.1	5.2
Religion (%)			
...*Murid*	15.2	48.4	26.9
...*Tidjan*	47.8	32.3	61.1
...*Other*	37.0	19.4	12.0
Schooling (%)			
...*No schooling*	46.2	43.5	64.0
...*Primary*	23.1	12.9	21.5
...*Middle School*	18.7	17.7	9.3
...*High School and more*	12.1	25.8	5.2
Labour status (%)			
...*Unemployed/Non-working*	46.7	58.1	41.3
...*Working*	53.3	41.9	58.7
Monthly earnings (XOF)	149 283	165 558	82 546
Observations	**92**	**62**	**174**

Source: MIDDAS Survey, 2009 - 2010

Table 8: Individual characteristics by migration status

	Non-Migrants (1)	**Migrants (2)**
Age	33.9	36.2
Gender (%)		
…*Male*	42.6	71.9
…*Female*	57.4	28.1
Schooling level (%)		
…*No schooling*	40.4	23.6
…*Primary*	28.4	20.1
…*Middle School*	14.2	19.4
…*High School and more*	17.0	37.0
Marital status (%)		
…*Single*	39.6	29.0
…*Married*	52.8	59.5
…*Divorced*	4.1	9.4
…*Widowed*	3.6	2.1
Labour status (%)		
…*Unemployed/Non-working*	48.1	23.9
…*Working*	51.9	76.1
Monthly earnings (PPP)	216.7	983.9
Observations	**1,929**	**926**

Source: MIDDAS Survey, 2009 - 2010
Notes: Sample restricted to individuals aged 18-65. Non-migrants are non-migrant members from migrant households

Table 9: Individual characteristics by location

	Senegal (1)	**France (2)**	**Italy (3)**	**Mauritania (4)**
Age	33.9	36.9	35.9	35.8
Gender (%)				
…*Male*	42.6	75.7	77.2	63.6
…*Female*	57.4	24.3	22.8	36.4
Schooling level (%)				
…*No schooling*	40.4	18.0	10.9	40.6
…*Primary*	28.4	18.0	12.3	29.4
…*Middle School*	14.2	16.3	25.2	16.7
…*High School and more*	17.0	47.7	51.7	13.3
Marital status (%)				
…*Single*	39.6	37.3	22.5	27.2
…*Married*	52.7	47.7	70.2	60.6
…*Divorced*	4.1	13.7	6.3	8.4
…*Widowed*	3.6	1.3	1.0	3.8
Labour status (%)				
…*Unemployed/Non-working*	48.1	25.7	29.5	17.0
…*Working*	51.9	74.3	70.5	83.0
Monthly earnings (XOF/euros/MRO)	66 552.8	1 281.4	1 163.2	52 531.8
Monthly earnings (PPP)	216.7	1 408.1	1 368.5	367.4
Monthly remittances (XOF)	(/)	125 209.9	130 637.0	36 928
Monthly remittances (PPP)	(/)	407.7	425.3	120.2
Observations	**1,929**	**300**	**302**	**324**

Source: MIDDAS Survey, 2009 - 2010
Notes: Sample restricted to individuals aged 18-65. Sample from Senegal is composed of non-migrant members from migrant households

Table 10: Household allocation choice - Reduced form conditional logit estimates

	clogit (1)	clogit (2)
Male (d)	1.068***	1.266***
	(0.141)	(0.173)
Age	0.021***	0.034***
	(0.005)	(0.006)
Elementary school (d)	0.372*	0.416**
	(0.205)	(0.209)
Middle school (d)	0.953***	0.968***
	(0.225)	(0.231)
High school + (d)	1.509***	1.889***
	(0.235)	(0.273)
Rainfall z-score x Male		0.457**
		(0.205)
Rainfall z-score x Age		0.029***
		(0.007)
Rainfall z-score x High school +		0.770***
		(0.276)
Observations	2225	2159
χ^2-test for joint significance of rainfall variables		19.16*** (0.000)

Source: MIDDAS Survey, 2009-2010
Notes: Coefficient reported, robust standard errors in brackets.
(d) stands for dummy variables.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Earnings and remittance equations - Uncorrected and corrected OLS estimates

	Home earnings		Destination earnings		Remittances	
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
Male (d)	0.605***	0.617***	0.534***	0.688***	0.475***	0.495**
	(0.070)	(0.072)	(0.116)	(0.112)	(0.180)	(0.231)
Age	0.102***	0.102***	0.104***	0.108***	0.096**	0.102**
	(0.019)	(0.019)	(0.034)	(0.041)	(0.045)	(0.044)
Age squared	-0.001***	-0.001***	-0.001***	-0.001**	-0.001*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Elementary school (d)	0.102	0.096	0.368**	0.413***	0.224	0.190
	(0.084)	(0.086)	(0.145)	(0.159)	(0.244)	(0.235)
Middle school (d)	0.507***	0.513***	0.474***	0.557***	0.251	0.223
	(0.120)	(0.112)	(0.143)	(0.154)	(0.233)	(0.240)
High school+ (d)	1.056***	1.067***	0.988***	1.088***	0.837***	0.801***
	(0.128)	(0.125)	(0.142)	(0.138)	(0.210)	(0.208)
Rural (d)	-0.357***	-0.345***				
	(0.086)	(0.083)				
Capital/Main cities (d)			0.300***	0.298***		
			(0.105)	(0.086)		
Koranic schooling (d)					0.503**	0.475**
					(0.253)	(0.232)
1st best probability		0.814***		-3.982**		-1.491
		(0.266)		(1.637)		(1.166)
1st best probability^2				8.196*		2.231*
				(4.197)		(1.168)
1st best probability^3				-5.150*		
				(2.978)		
Highest probability		-0.640**				
		(0.299)				
Constant	2.034***	2.028***	3.585***	3.768***	1.873**	1.968**
	(0.361)	(0.363)	(0.629)	(0.682)	(0.909)	(0.771)
Observations	818	818	255	255	252	252
R^2	0.28	0.29	0.36	0.38	0.14	0.15
F-test for joint significance of selectivity variables		9.33***		9.37**		7.07**
		(0.002)		(0.024)		(0.029)

Source: MIDDAS Survey, 2009-2010
Notes: Coefficient reported, robust standard errors in brackets for uncorrected estimations, bootstrap standard errors in brackets for corrected estimations. (d) stands for dummy variables.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Household allocation choice - Structural conditionnal logit estimates

	clogit (1)	clogit (2)	clogit (3)
Log home earnings ($-\alpha$)	1.226***	1.225***	2.153***
	(0.347)	(0.416)	(0.511)
Log destination earnings (β)	-3.935***	-3.994***	-4.513***
	(0.538)	(0.519)	(0.545)
Log remittances ($\gamma - \beta$)	5.473***	5.463***	5.209***
	(0.539)	(0.545)	(0.496)
Male (d)		0.102	0.019
		(0.284)	(0.215)
Age			-0.025***
			(0.008)
Observations	2216	2216	2216
Estimated α	-1.226	-1.225	-2.153
Estimated β	-3.935	-3.994	-4.153
Estimated γ	1.538	1.469	0.696

Source: MIDDAS Survey, 2009-2010
Notes: Coefficient reported, bootstrap standard errors in brackets.
(d) stands for dummy variables.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$