Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

UNIVERSITÉ LAVAL

Incentives to labor migration and agricultural productivity : The Bayesian perspective

Marie Albertine Djuikom [1]

Abstract

Understanding how internal labor migration affects the agricultural sector is important for all developing countries whose markets don't work well or are non-existent. In fact, even if the movement out of the agricultural sector can be viewed as a process to reach development for many African countries, this could lead to a negative effect on the rural economy. The availability of labor and the cost of hiring people to work on farms is an example of a problem that farmers may face in the presence of labor migration. This paper investigates the effect of internal labor migration on agricultural productivity of rural households in Uganda. Since households select themselves into migration this raise the endogeneity problem. In order to account for the endogeneity of the migration decision and the fact that the effect might be different from one household to another, I model the households decisions to participate in migration along with their investment in agricultural productivity using the Bayesian treatment analysis. This approach allows me to self-matching each household and to estimate a distribution for the counterfactual outcome. The results show that even if on average internal labor migration positively affects agricultural productivity, there are some households for which the effect is negative. Those households for which the effect is negative are mostly small farmers, and are therefore more likely to be poor. Moreover, the average effect of the labor migration tends to increase with the likelihood of participating in the internal labor migration. In parallel, I also examine to what extent previous migration rates, widely used in the literature as an instrument for the migration decision, are exogenous to the agricultural productivity. It turns out that previous households' decisions to participate in migration are intimately correlated with their current agricultural productivity.

Keywords : Labor migration, agriculture, Bayesian treatment analysis, instrumental variables, rural, Uganda.
JEL : O15, O13, O18, J61, Q12.

1. PhD Candidate in economics at université laval

1. Introduction

Understanding how internal labor migration affects the agricultural sector is important for all developing countries whose markets don't work well or are non-existent. In fact, even if the movement out of the agricultural sector can be viewed as a process to reach development for many African countries, this could lead to a negative effect on the rural economy. The availability of labor and the cost of hiring people to work on farms is an example of a problem that farmers may face in the presence of labor migration. Surprisingly, so far there are few empirical studies devoted to the issue(de Brauw (2010),Mwesigye & Matsumoto (2016),Mendola (2008),de Haan (1999)). From a theoretical point of view, labor migration can have positive or negative effect on agricultural production (de Brauw, 2010). In fact, the effect would be positive if the migrant-sending household can hire individuals on the local labor market to substitute for the migrant's labor force. If, on the contrary, the household cannot find a replacement labor force because of scarcity or the cost of the labor force [2], labor migration can negatively affect the agricultural production. For this reason, the effect of the labor migration is likely to be different across households depending on factors such as the migrant's productivity and how well the local markets are functioning.

Unlike the existing literature that assumes that the effect of internal labor migration is homogeneous across households, this study estimates the distribution of the effect of the internal labor migration on the agricultural productivity of households living in the rural areas of Uganda. I use the four survey rounds of the unique nationally representative panel surveys, the Uganda National Panel Survey (UNPS). To achieve my goal, I estimate simultaneously and over time the households' decisions to invest in agricultural production and to participate in labor migration using a Bayesian treatment approach. I am not the first using this approach to allow heterogeneity of the effect between individuals and to account for the endogeneity of their decision. For instance, by using the Bayesian approach, Carneiro *et al.* (2003) estimate the distribution of the return to school in terms of earnings for youths in the United States.

Furthermore, I consider that the labor migration decision is a common agreement among household members to send one or more member(s) outside the village (in another district) to find additional sources of income so as to increase household resilience to negative shocks that might occur. Indeed, for the households in the rural areas, the labor migration provides a kind of insurance in times of bad harvest and a source of financial support to smooth household consumption, increase household's business(es), launch a new business, or invest in education. This insurance is even more important for the poorest households because the access to credit is virtually non-existent without any collateral in the rural areas of most developing countries such as Uganda. Therefore, labor migration is not a unilateral decision taken by one person. Even if a household member can decide to migrate on his own initiative, it is less likely that it will be done without the consent of other household members. Indeed, the loss of available labor to the household directly affects the labor supply of the all members left behind in and out of the farm sector but particularly in farming since the agricultural sector is highly labor-intensive in developing countries. For instance, Mu & Van de Walle (2011) found that labor migration increases the time that the woman left behind spend doing domestic and farm work. In that way, this study is fully integrated into the *New Economics Labor Migration (NELM)* theory developed by Stark & Bloom (1985) for which the labor migration decision is a common agreement household members.

Using the Bayesian approach, I am able to account for the endogeneity of the migration decision and to estimate an average effect for each household. Moreover, I allow for the selection into migration to depend on households' time-invariant unobserved heterogeneity such as households' willingness to take risk since it is not clear that households participating in internal labor migration will have positive returns from it. These confounders affect both the labor migration decision and the investment in agricultural production. Most of the studies don't account for the self-selection of households into migration based on their unobserved variables, which can lead to mis-evaluate the impact of migration. As a result, this paper thus contribute to the literature by evaluating to what extent this type of selection is related to the effect I attempt to identify and investigates whether the internal labor migration decreases the agricultural productivity

2. It can be difficult to find a substitute if the migrant is highly productive, especially so if we consider villages with high migration rates.

of all of the households in the rural areas of Uganda. The Bayesian approach has recently been introduced in the treatment analysis and provides, from implementation point of view, an easier way to account for households' time-invariant fixed effects and to estimate a distribution of the counterfactual outcome for each household. For example, I am able to consistently estimate for households participating in the labor migration, the agricultural production in the case that they were not involved in migration.

To sum up, the contribution of this paper is twofold. First, to the best of my knowledge this is the first study that attempts to estimate the causal effect of internal labor migration on households' agricultural productivity in Uganda. Moreover, I go beyond estimating an average effect by estimating the distribution of the effect on households. Second, the methodology used here allows to verify the exclusion assumption on the instruments variables used to correct for the endogeneity of the participation in labor migration. In fact, it is generally admitted in the literature that the past history of households participation in internal labor migration affect the current and future migration participation but not the outcome of interest.

To meet the objective of the study, the remainder of this paper is organized as follows. The next section provides a brief review of the literature on internal labor migration in Africa and an overview on socio-economic environment in Uganda. In section 3, I describe the data and provide a preliminary analysis. Section 4 presents the empirical model and discusses the identification strategy. Section 5 presents the main results and section 6 is the conclusion.

2. Context and literature

In the literature, the incentives for internal labor migration and its (average) effect on the outcomes of the household members left behind as well as on the migrants have attracted less attention (Garip (2008), Mendola (2008), Mu & Van de Walle (2011), Garlick *et al.* (2016), de Haan (1999), Mwesigye & Matsumoto (2016), Muto (2009), Kuhn (2015), Stack & Taylor (1991), de Brauw (2010)). Moreover, it is hard to identify who migrates internally through the available data and further to evaluate how important the internal labor migration is in developing countries.

Existing studies reveal that the push factors of the internal labor migration range from the economic costs (household wealth), the household social capital network and relative deprivation, the variation in rainfall to the increase in the investment in human capital. Furthermore, labor migration enables households investing in the non-farm activity and in education, acquiring new land, improving health of household members, smoothing household consumption and more generally improving the well-being of the household left behind. However, the internal labor migration doesn't lead to a huge transformation in the agricultural sector but helps households to meet their basic needs, and lightly increases agricultural productivity (de Haan, 1999). Based on seven papers investigating the effect of migration on households agricultural production in the rural areas of six developing countries, Davis *et al.* (2010) finds out that migration facilitates a transition away from agriculture or leads to less labor-intensive agriculture. Empirical evidence then shows that the *NELM* failed in most cases since the labor migration does not necessary lead to an investment in the agricultural sector.

Given the expansion of internal labor migration in the developing countries, if migration negatively affects the largest producers, it might be urgent for the government to invest more in their agricultural sector in order to stabilize or even increase the domestic agricultural production. Besides, many recommendations in terms of development policies have long focused on investment in education and in non-farm sector as a natural way out of poverty. As a result, from the moment that households are able to have access to capital through migration, they are more likely to invest in non-farm businesses that lead to a decrease in total agricultural production.

The case of Uganda is interesting since it is an East African country emerging from decades of conflict and security challenges in the Northern part of the country. In 2006, a cessation of hostilities agreement was signed by all stakeholders. These conflicts have caused a massive displacement of people from the affected areas and delayed the development of these areas. However, many people have already returned to their home and to their day-to-day life. It is then likely that the average agricultural productivity will be lower in this part of the country. Nevertheless, northern Uganda is not

the region with the highest rate of migration (labor migration or other). In fact, similarly to central Uganda, only 50% of people who were born in the north continue to live in the north. Besides, this study covers the period 2009-2011, at least three years after the end of the conflicts.

Additionally, less is known about the incentives of internal labor migration and its effects on the well-being of households left behind in Uganda. In fact, most studies devoted to labor migration in Uganda are primarily descriptive. Rutaremwa (2011) and Bukuluki (2015) give the profile of migrants and describe the uses of remittances from both international and internal migration in Uganda. Their results show that the remittances are mainly used for school investment, savings and to invest in buildings. Besides, Rutaremwa (2011) pointed out that compared to international migrants, internal migrants come from poorer households and that remittances are lower in amount. This implies that the return to internal labor migration in Uganda may not balance the cost in terms of labor loss. Nevertheless, in the rural areas of Uganda, Muto (2009) finds that internal labor migration increases with the household's social network proxied by being a member of the larger ethnic group in the capital (Kampala). Jagger *et al.* (2012) find a positive effect of circular migration in the logging business on the households living in the community of origin of migrants. Their results reveal that migration reduces inequality in the community of origin, yet, the study focus only on households living in the southwestern part of Uganda. My paper then fills a gap in the literature by providing evidence of the effect of the internal labor migration using nationally representative data.

Moreover, Uganda is one of the poorest countries in Africa with 80% of the population living in rural areas and where the agricultural sector has contributed up to 28.3% of the GDP in 2011 and 25.5% in 2015 [3]. In addition, Uganda is the country with the highest proportion of its population being youth aged less than 30 ; they represent for around 78% of the entire population and about half of the population is under 16 years old. Meanwhile, school dropout is a big concern, Ssewamala *et al.* (2011) report that, in 2007, only one third of children enrolled in the first grade of primary school were likely to participate in the seventh year (last grade of primary school). In fact, as we can see in figure 4, even though the education level of people aged between 15 and 35 has increased over time [4], only 50% of these individuals have actually completed their primary education, which corresponds to seven years of education in the figure. Along with that, the internal labor migration rate at household level has increased over time. Indeed, the percentage of households involved in migration was 10% in 2005 versus 24% in 2011 and the share of household members migrating is increasing with the average education of members living in the same house. As a result, many more Ugandans are involved in labor migration now with the recent boom in education attainment. Moreover, I have information in my dataset on the duration of the labor migration and whether the migrants are still considered a member of the household are available. In my data, temporary labor migration accounts for around 95% of the internal labor migration, that is when a household member lives away from home for at least three months, is still considered a member of the household and is not a permanent mover. The average duration of migration within each household ranges from three to ten months. Consequently, my focus in this paper is on temporary migration.

Furthermore, there is evidence that migrants are more highly educated than non-migrants implying that households participating in labor migration have members with higher ability than non-migrants households. Figure 3 shows that the agricultural production increases with household head's years of education and the average years of education among the members of the household. Thus, the existence of unobserved confounders such as ability that affect both the agricultural production decision and incentives to involve in migration introduce here one source of endogeneity of household assignment into labor migration.

Another main source of endogeneity is the social capital network. In developing countries it is very common that people belonging to the same ethnic group, or belonging to the same family (called the *"strong ties"*) or living in the same village (called the *"weak ties"*), to help one another. In the case of migration, a household social network increases the propensity to participate in migration by reducing the uncertainty surrounding the expected gain from

3. Trading economics website :https://tradingeconomics.com/uganda/agriculture-value-added-percent-of-gdp-wb-data.html

4. This result might be due to the "*Universal Secondary Education (USE)*" implemented by the government in 2007 make tuition fees free for ordinary secondary schools. Asankha & Takashi (2011) find a positive effect of USE on girls in secondary school enrollment.

migration and by reducing the monetary costs. For instance, from its social network, household can get information about job opportunities in the potential destination, a place to live or even a ride to the desired place. The endogeneity comes from selection into social network and the fact that we do not observe the household's social network while it is likely that it affects both the migration decision and the household production.[5] Furthermore, when the household decides to participate in labor migration, it simultaneously decides the level of agricultural production given the land owned, the labor available (from household members and hired labor) and other inputs. Consequently, the selection into the migration is not exogenous of the agricultural production decision.

3. Data and Preliminary Analysis

This study uses the Uganda-National Panel Survey (UNPS) conducted by the Uganda Bureau of Statistics (UBOS) with the technical support of the Government of Netherlands and the World Bank Group. The units of interest are individuals, households, and community/facilities. The surveys cover multiple areas such as health, education, consumption, labor force, etc. with a special module on agricultural activities. This is the first national representative panel data that covers a large number of socio-economic and demographic indicators. Launched in 2009, the surveys are implemented over a 12-month period divided in two visits (over a period of six months each) in order to rule out or at least minimize measurement errors on agricultural inputs and outputs. The first visit is about, among others, inputs and outputs of the last cropping season which could be the short or long season cropping ; besides, in the same period, the consumption module is implemented on half of the sample. Four waves implemented in 2009-2010, 2010-2011, 2011-2012 and 2012-2013 are publicly available[6]. Henceforth, each panel wave will be identified by the first year of the corresponding survey.

Initially, 3220 households have been selected from 7426 households interviewed in the 2005-2006 Uganda National Household Survey (UNHS) and tracked and re-interviewed up to the third wave. Thereby, it is possible to connect data from the UNHS to the panel data. At the fourth round, some adjustments have been made to the initial sample. In fact, the households extracted from the Uganda census survey implemented in 2012 have replaced a part of the panel sample. Therefore, the sample has changed significantly and the sample weight has been corrected. The goal of the sample rotation was to correct for the attrition and random answers that might occur when the households are already used to being interviewed. The average attrition rate at household level is around 17% across waves. This attrition mainly comes from the fact that the new location of some households who have moved is unknown. Since in this study I focus on migration where a household sends a member outside but not the migration for which a household moves, it is less likely that this attrition raises issues of selection bias. Besides, the highest rate of attrition is observed among households living in the urban area and mainly in Kampala while the focus of this paper is on rural households.

The main goal of the UNPS is to provide reliable national representative data for the experimentation and assessment of the national policies and programs (UBOS (2013)). The sample is clustered at the community level and covers all of the four regions and 323 communities of Uganda. In this study, I focus on agricultural households living in rural areas where about 80% of households live. Agricultural households are defined as households with at least one member operating a holding (farming household) or for which the head is economically active in the agricultural sector (United Nations, 1984). In my sample, about 75% of households are actually involved in agricultural activities across waves of survey. Since there are movements in and out of agricultural sector at the household level and that the panel data is not balanced, the analysis will focus on the balanced sub-sample for which I have information on agricultural productivity for all periods ; I will call this sample the *sample A* through out this study. The sample of households for which the information on agricultural production is not missing at least for the first round of panel survey is noted *Sample B*.

5. It is well documented in the literature how social networks impact the way that people behave.
6. Data is available on the World Bank website :http ://microdata.worldbank.org/index.php/catalog/lsms

3.1 Labor migration prevalence and agricultural productivity

UNPS provides detailed information that helps define migrants and their profiles. In fact, it is possible to know why a household member was absent during a certain period in the past twelve months, the duration of his absence, the district of destination and so forth. In this paper, I am interested in migration decisions at the household level that is, I estimate the propensity of household to get involved in labor migration. Thereby, I identify a labor migrant-household (henceforth migrant-HHs) as a household with at least one member who has spent at least three months outside the household dwelling place in the past twelve months preceding the survey [7] and who is still considered a household member. [8] Moreover, it is not so common for a migrant not to be considered as a household member. This is consistent with the way that I treat the migration decision, that is, migration is a household decision rather than an individual one.

Labor migration prevalence

With the information available, it is possible to distinguish between labor migration and migration for other economic reasons. I have grouped these two types of migration together as they lead at the end to the same goal, which is to find a job upon arrival at the destination place. For simplicity, I will identify this group of households as labor migrant-HH or simply migrant-HHs. The data reveal that the migration participation rate among households increased over time from 11.2% in 2005 to 24.6% in 2011. However, it is not possible to know if it is rural-urban or rural-rural or urban-urban migration because only the district of destination is known and unfortunately many districts have both urban and rural areas in Uganda. Most often, migration is from one district to a different district and practically never in the same district.

Table 1 – *Migration Rate*

	Whole sample			Agricultural Households in rural areas					
Year	All	Rural	Urban	*Sample A*	*Sample B*	Central	East	Northern	Western
2005	11.2	10.2	14.5	10.0	10.5	10.0	9.2	7.3	15.3
2009	17.8	17.1	19.9	17.7	17.7	25.3	12.8	11.5	20.5
2010	23.9	23.3	27.2	25.8	24.7	27.6	20.7	19.9	31.0
2011	24.6	24.2	26.4	26.1	26.4	36.8	21.3	26.2	25.8
Total	22.3	21.8	24.1	23.3	23.0	29.0	18.5	19.3	26.2
Agricultural production per hectare									
2009	4,815.6	5,847.7	1,811.9	7,244.4	7245.2	5,966.7	4,336.9	6,289.8	11,291.8
2010	5,224.3	5,971.8	1,442.2	6,916.5	6277.9	5,146.9	7,226.6	4,511.7	7,370.6
2011	3,709.8	3,901.5	2,873.4	6,519.8	5899.6	6,477.7	3,707.0	5,559.0	7,676.1
Total	4,582.0	5,218.4	2,041.8	6,916.3	6477.6	5,792.6	5,409.5	5,325.9	8,689.3
Nb. of Obs	2,617	2,201	768	1,452	1,749	393	493	439	444

Notes :

1. Percentage is given in each cell.
2. The statistics given by region in the last four columns is computed on the *sample B*.
3. In the sub-table for agricultural production, I have reported the average production per household in kilograms per hectare.

In Uganda, migration rates vary across regions with highest rates in the center and in the north. Given the long history of instability in the north, the high migration rate can be driven by the less advantaged economic environment and the lack of job opportunities. Moreover, while the migration rate increases over time for other regions, there is a mitigated pattern for the western part of Uganda. In fact, the migration rate increased from 19% in 2009 to 31% in 2010 and decreased to 25% in 2011. In the meantime in Uganda, the poverty rate also increased between 2009 and 2010 and decreased between 2010 and 2011. On the contrary agricultural production has followed the opposite pattern in the western part by decreasing between 2009 and 2010 and increasing between 2010 and 2011. Therefore, it seems that

7. A period of three months is the most common use in the literature
8. That is to still keep the perspective of temporary migration.

agricultural production in the west of Uganda is more sensitive to the movement of people. Furthermore, the average duration that migrants spend outside their district is around two months and half and this duration increases with the years of education of household head and the average years of education of household members. It is also the case for the share of household members involved in migration. In other words, households where the members are more educated are also more likely to participate in labor migration for a longer period.

In our data, we also have households whose the head has migrated permanently for economic or education reason within the last ten years. This information is sometimes used to capture the migration rate among individuals. To avoid any selection bias due to the household head past migration, in Table 6 we have reported the household labor migration status, as defined in this study, regarding the head's past permanent migration. It emerges that there is no significant difference in the proportion of the head's past permanent migration by household migration status. Therefore, it is less likely that there is selection into the actual migration participation due to the past permanent migration decision of the head. Nevertheless, we introduce this variable as an independent co-variable in migration participation and agricultural production equations.

Agricultural productivity

The agricultural productivity reported in table 1 is the average production of the main crops planted a little everywhere in Uganda, that is, productions of maize, beans, coffee, peanuts, bananas and potatoes. As we can see, the agricultural production tends to decrease over time with on average higher production in the rural areas. Regarding the distribution by region, the higher level of agricultural production is found in western Uganda and except for the east, the production decreased between 2009 and 2010 and increased between 2010 and 2011. However, the level of production in 2011 is still lower than the production in 2009 meaning that the agricultural production has decreased over the period covered by the study. In the eastern Uganda, although we end up with the lowest productivity in 2011 (3707.0 kg/ha), there was an increase of the production between 2009 and 2010 (4336.9 vs 7226.6 kg/hectares). In addition, the average total area planted per household has also decreased with time with a large drop between 2010 and 2011, which can explain the lower production in the last year.

It sometimes emerges in the literature that the migration increases land-related conflicts because, due to cultural diversity induced by migration, it may be difficult to resolve conflict based on customary laws in the absence of formal legislative law. In my sample, about 12.4% of households have reported having a conflict about at least one land they possess ; however the share of households involved in land-related conflicts declined over time with only 10.5% of households reporting land-related conflicts in 2011. On whether the presence of land-related conflict can affect household investment in the agricultural sector, I find that households that faced a land-related conflict have lower agricultural productivity in the first two periods, yet in the last period there is no difference in agricultural production. Moreover, the average total area planted by households facing land-related conflicts is even higher.

I also test if the household's social network driven by ethnicity group can help increase the agricultural productivity by providing advice and tips. Figure 4 shows the agricultural productivity given the share of households in the same ethnic group in each district. As we can see, there is a non-linear relationship between production and ethnic concentration ; the production increases with the size of the ethnic group's social but after a threshold, agricultural productivity tends to decrease.

3.2 Household profile and migration status

Table 2 presents some households characteristics depending on their labor migration status. It appears that among households participating in labor migration, the migration propensity increases between the first and the second quartiles of household wealth [9] ; however, from the third quartile of wealth, the migration propensity starts to decrease. However, it is only for the households in the first quartile and on the last quartile of the wealth distribution that I have found

9. Household wealth is measured here by the total expenditures made by household for consumption, education, health,etc.

a significant difference in the share of households within each quartile given their migration status. In fact for the households belonging to the first quartile of wealth, the share of labor migrant Households is significantly lower than the share of non-migrant households and the share of labor migrant households is higher for the households belonging to the fourth quartile. It seems that the monetary cost of labor migration could be a disincentive to participate in migration for the poorest households and for households in the middle of the distribution, other important factors affect the decision to participate or not in migration. Moreover, the migration rate increases across survey rounds for households belonging to the first quartile while it decreases for households in other quartiles. This suggests that households readjust their migration decision over time by comparing the cost and the gain from migration and we expect that if the net gain is positive, they will be more likely to participate in migration. We can see that for the poorest household, the migration rate increased between survey rounds. Beyond the monetary costs, there are other factors that determine households' labor migration decisions. In fact, migration decisions are intimately related to households head's characteristics and the composition of the household.

Households headed by a widow or by a polygamous are more likely to participate in migration. It is also the case with a female or a highly educated household heads. In addition, the households heads of migrant-HHs are slightly older than the heads of non migrant-HHs. To investigate to what extent the missing values for a head's education can introduce selection bias, I have computed the percentage of households for which the head's education is missing. It turns out that few if not any migrant-HHs have heads with missing value for education, yet, there is a very small share (3%) of household heads for which the head has missing years of education among the non migrant-HHs. Generally, heads' characteristics for migrant-households and non migrant-households don't vary across survey rounds.

Table 2 – *Migrants Households Profile*

	2009		2010		2011	
	Migrant	Non-Migrant	Migrant	Non-Migrant	Migrant	Non-Migrant
Head characteristics						
Married monogamous	0.53	0.57	0.54	0.57	0.53	0.57
Married polygamous	0.20	0.16	0.22	0.18	0.18	0.19
Widow	0.17	0.15	0.16	0.14	0.18	0.13
Divorcee or sep	0.09	0.09	0.07	0.09	0.10	0.09
=1 if woman headed-HH	0.32	0.26	0.32	0.27	0.37	0.26
Head years of educ	5.59	4.57	5.74	5.08	5.44	4.72
head educ. missing	0.00	0.02	0.00	0.03	0.00	0.03
Head age	48.45	45.46	48.14	45.74	48.36	46.42
HH characteristics						
1st Quartile	0.16	0.32	0.25	0.34	0.26	0.32
2nd Quartile	0.29	0.29	0.27	0.30	0.24	0.32
3rd Quartile	0.28	0.24	0.26	0.23	0.23	0.24
4th Quartile	0.27	0.15	0.22	0.12	0.27	0.13
HH size	7.25	5.79	8.14	6.21	8.67	6.49
Share of memb aged $\leqslant 5$	0.17	0.20	0.16	0.19	0.15	0.18
Share of memb aged>65	0.04	0.03	0.03	0.03	0.05	0.04
Average age in HH	20.16	22.35	18.94	22.05	19.55	22.23
Average educ.of migrants	3.86	.	11.29	.	17.52	.
Average educ.of Non-migrants	2.39	2.60	6.91	6.91	11.13	11.13
Average age of migrants	22.37	.	20.58	.	23.59	.
Average age of Non-migrants	22.74	22.85	20.39	20.39	20.95	20.95
Average educ. in HH	5.29	5.04	4.84	4.41	5.43	4.91
=1 if conflict land-related	0.14	0.17	0.10	0.10	0.10	0.11
Ethnicity concentration in Uganda	0.09	0.07	0.06	0.06	0.05	0.06
Ethnicity concentration at district level	0.60	0.64	0.69	0.71	0.68	0.70
Agricultural production						
log Production(kg/ha)	8.20	8.00	8.08	7.95	8.15	7.94
Total area planted	4.71	3.97	4.98	4.18	2.90	2.50
Number of crops	15.49	13.38	11.63	10.60	11.18	10.21
Average days HH hired people to work on farm	7.69	5.47	5.04	6.27	12.07	7.22
Household Adult Equivalence Scale	4.77	4.10	4.10	3.91	4.04	3.89

Regarding the composition of the households, migrant-HHs have larger size households and lower infant-age de-

pendency ratios (share of children aged less than five years old). Besides, the gap of infant-age dependency between migrant-HHs and non migrant-HHs increases over time, meaning that the gain due to migration may not offset the labor cost when the household has many children aged less than five years old and then needs to readjust their migration decision over period. I have also reported in the table the average age and the education of individuals who migrate or not within each household. Results reveal that the years of education and age of persons left behind are almost identical within migrant and non-migrant households. However, individuals who migrate are in general more educated than those left behind and the average years of education has increased over period, from around four years in 2009 to seventeen years in 2011.

In order to see whether ethnic concentration or ethnic diversity constitutes a push or pull factor for migration, I have computed the average share of individuals in the same ethnic group of each household and it comes out that migrant-HHs have a slightly smaller ethnic networks than non migrant-HHs. This means that migrant households have less people from the same ethnic group in their network than non migrant households. We also look at agricultural productivity and find that migrant households have a higher average agricultural productivity than non migrant-HHs. This supports the *NELM* theory that the labor migration might actually help invest more in the agricultural sector. In addition, migrant-HHs planted in larger areas and many more crops than non migrant-HHs.

4. Setting and Empirical strategy

4.1 Preliminaries

In this study, I open the black box to understand how the internal labor migration decision takes place in a dynamic setting and how it affects the well-being of households left behind in the context of a developing country. In fact, I want to measure the effect of the migration decision on agricultural productivity. Since the labor migration reduces the labor force available to households, agricultural production can be negatively affected by migration mainly because in developing countries, the agricultural sector is highly labor-intensive. Therefore, the way that migration affects the agricultural production depends on whether or not the return to migration allows the household to invest enough in the agricultural sector to make up for the cost of lost of or unavailable labor. Besides, to compute the agricultural productivity, I have only accounted for the main crops cultivated in all regions of Uganda to avoid selection bias related to geographical advantages of one crop from another. Therefore, I have added up the productions per hectare of maize, beans, coffee, peanuts, cassavas, bananas and potatoes.

Let us take equation 1 as a baseline model where α is the coefficient of interest that measures the average effects of labor migration on Y_{it} which is the production in period t of the household i. LM_{it} takes value one if household i participate in labor migration in period t. X_{it} and Z_i are respectively time-variant and time-invariant characteristics of households.

$$Y_{it} = LM_{it}\alpha + X_{it}\gamma + Z_i\theta + \mu_i + \epsilon_{it} \tag{1}$$

The estimator of α using the Ordinary Least Squared (OLS) might be biased due to the endogeneity of the labor migration decision. Given the nature of the outcome per se, the main sources of endonegeity will be the existence of omitted variables and the simultaneity between migration decision and investment in the agricultural sector. In fact, when a household decides which member(s) to send outside, it simultaneously decides the level of agricultural production given the (anticipated) net gain of migration, the land owned and the available labor force (from remaining members and/or from labor they can hire in the village). Besides, I have found out that many households don't cultivate the entire area of the land that they possess. In equation 1, μ_i captures the time-invariant factors that might be correlated with LM_{it} and unobserved to the econometrician. For instance, households involved in labor migration might have more members with higher abilities, a larger social network and more willing to take risks since there is no guarantee that the investment in migration will produce a positive gain.

To solve the simultaneity issue, we can add to equation 1 a second equation that estimates the labor migration decision. Yet, this does not solve the issue of omitted and unobserved variables. Another source of endogeneity might be the error measurement on outcomes of interest. Sometimes, it is difficult to measure, with high precision, the households' agricultural productivity. Nevertheless, to reduce potential error measurement, the data on the agricultural production are collected in 6-month intervals, corresponding to the short and long cropping seasons. Even if measurement error is minimized in this way, we might still face a heteroscedasticity problem.

To correct for the potential endogeneity problem and the heteroscedasticity issue, I present in the next two sections an identification strategy that minimizes if not eliminates the bias.

4.2 Identification Strategy

Due to the selection into the labor migration, the OLS provides biased estimates of the effect. The selection problems arises from many sources discussed in section 4.1. In the literature, different approaches have been developed to correct for the endogeneity problem due to the selection bias introduced by observed and unobserved variables. Different methods such as propensity score matching and its variants enable correcting for the bias based on observed variables in the static and dynamic analyses [10]. However, it is more difficult to be convinced that we have corrected for the selection due to the unobserved heterogeneity. Nevertheless, the Instrumental variables (IV) approach attempts to correct for both types of selection by finding an exogenous shock that affects the independent variable, the source of endogeneity, and that is not directly correlated with the outcome of interest. However, finding such a variable is not an easy task. In the literature, some authors use the variations of rainfall as exogenous shock that motivates people in the rural area to move to find job elsewhere (Konseiga (2007), Lucas (1987)). In our data, we did not find significant variation in rainfall for each district and it seems not to affect migration incentives. Besides, the variation in rainfall may be correlated with the agricultural productivity.

The IVs approach has been largely implemented in the literature to estimate the return to education using an exogenous variation in the supply side of education , such as the reduction of the tuition fees, which is not correlated with the wages of individuals (see Card (2001) for a review). In the economics of migration, empirical studies reveal that the costs related to the migration are the main disincentive for households to involve in migration (Mendola (2008), Mckenzie & Rapoport (2007), Garip (2008)). Therefore, the instruments for labor migration are related to the social capital network of household, from which the household can have many resources(transportation to the destination place, a place to live and useful information about job opportunities). In this way, the monetary cost of migration and the uncertainty surrounding the potential returns from migration are significantly reduced. For this reason, the larger the household's social network is, the higher the propensity of migrating is. In this setting, authors distinguish the *"strong"* social network from the *"weak"* social network. The former is the network made up of household members and relatives who have experienced the labor migration and the latter is the network formed by people from the same village who have also experienced internal labor migration in the past. The weak social network is relevant in developing countries since it is common that people in the same village to help each other. Therefore, the social network is a strong push factor to labor migration when the household faces a monetary constraint. We then expect to have a higher and stronger effect from the *"strong"* social network.

Furthermore, other data sources may contain specific information about the people with whom the household members interact. Since in my data there are no specific questions about the household's social network, I follow the current literature by taking as a proxy for the household's *"strong"* social network the number of household members who have experienced internal labor migration in the past and for the *weak* social network, the labor migration rate at the district level. In this study, the household's *"strong"* social network is captured by the number of household members who experienced labor migration in 2005, that is at least four years before the beginning of the period covered by the analysis. I take the much earlier prevalence of migration as instruments to manage the potential correlation that might exist between the migration that occurred just before the period covered by the data and the agricultural productivity of the

10. See Lechner (2009) for the evidence of the sequential matching approach.

households. However, it is not excluded that there are still unobserved time-invariant and/or time-varying variables that affect the migration propensity and production function over time, that is $E\left[LM_{it} \times \mu_i\right] \neq 0$ and $E\left[LM_{it} \times \epsilon_{it}\right] \neq 0$. To provide an illustration for the unobserved time-varying variables, based on the theory of learning by doing, it is expected that the more a household participates in migration, the higher the probability of succeeding in the destination place is. That could then increase the return to labor migration. Thereby, the migration prevalence in 2005 may not verify the restriction assumption as instrument. This raises the issue of the trade-off between the power of the instrument and the restriction assumption. Conley *et al.* (2012) offers a way to achieve efficiency in the case that the restriction assumption fails. The idea is to introduce the instruments in the outcome equation (the second stage of the two-stage least squared estimations) in order to test if their parameters are significantly equal to zero and to have efficient confidence intervals for the causal effect. At this point, we don't know if the instruments for the household's social network (as presented in the literature) actually verify the exclusion restriction assumption, because the authors usually just assume that it is verified. I propose to test this assumption in this paper. Moreover, I will also test to see if the social network through ethnicity can be a push factor for the internal labor migration in Uganda.

To estimate the effect of migration over time on agricultural productivity, I specify the model as follows :

$$\begin{cases} MU^*_{it} & = Z_i\beta + W_{it}\alpha_m + \theta_i\gamma + \lambda_{it}^{-1/2}\varepsilon_i \\ Prod_{1it} & = X_{it}\alpha_1 + \theta_i\gamma_1 + \lambda_{it}^{-1/2}\epsilon_{1i} \\ Prod_{0it} & = X_{it}\alpha_0 + \theta_i\gamma_0 + \lambda_{it}^{-1/2}\epsilon_{0i} \end{cases} \tag{2}$$

The system of equations that tests for the exclusion restriction assumption for the set of instruments is given by :

$$\begin{cases} MU^*_{it} & = Z_i\beta + W_{it}\alpha_m + \theta_i\gamma + \lambda_{it}^{-1/2}\varepsilon_i \\ Prod_{1it} & = Z_i\beta^1 + X_{it}\alpha_1 + \theta_i\gamma_1 + \lambda_{it}^{-1/2}\epsilon_{1i} \\ Prod_{0it} & = Z_i\beta^0 + X_{it}\alpha_0 + \theta_i\gamma_0 + \lambda_{it}^{-1/2}\epsilon_{0i} \end{cases} \tag{3}$$

$\forall t \in \{1, 2, 3\}$ stands for the time period ; the first period corresponds to the survey implemented in 2009. MU^*_{it} is the latent variable representing the migration utility function which is related to the migration status, LM_{it}, by the preference relation :

$$LM_{it} = \begin{cases} 1 & if \;\; MU^*_{it} > 0 \\ 0 & if \;\; MU^*_{it} \leqslant 0 \end{cases}$$

For recall, LM_{it} is equal to one if household i has experienced the labor migration in period t. $Prod_{1it}$ stands for the agricultural production per hectare of households involved in migration in period t and $Prod_{0it}$ the counterpart production for the households who are not involved in labor migration in period t.

Z_i is the vector of instruments for labor migration. In addition to the prevalence of migration at household level and at district level, I have added as instrument the household's relative income deprivation computed by using the monthly household total expenditures and the reference group for each household in the district. [11] In the relative deprivation model of migration, Stack & Taylor (1991) argue that once we control for the absolute income gain from migration, relative income deprivation can be an incentive for households to involve in migration if migrant-HHs and its migrants feel less deprived. However, in the case where the migrant-HHs substitute its reference group with the group of households in the district of destination so that the income gain does not compensate the higher relative income deprivation given the new reference group, neither the household's absolute income, nor the relative income is going to be significant in the household's propensity to participate in temporary internal labor migration. Additionally, the equation 3 allows to test for the exclusion restriction assumption of the instruments. In this case, if the set of parameters β^1 and β^0 are significantly different from zero, it means that the exclusion restriction assumption failed.

11. I have reported in appendix the details about the way that I compute the Index of relative deprivation for each household.

Table 3 – *Set of variables*

Production input	- Hired labor(in terms of the number of days) - Adult equivalent - Total area planted - Percentage of households involved in agricultural sector within a radius of 5 km
Risk management	- Number of crops managed
Head attributes	- Marital status - Education - Age
Household characteristics	- Household wealth measured by the household total expenditures - Size - Share of members aged less than 5 years - Share of members aged more than 65 years - Share of members aged between 6 and 14 years - Share of female - Geographical deprivation
Instruments for migration decision	- Num of members involved in migration in 2005 - Migration rate at district level in 2005 - Wealth Deprivation

X_{it} is the vector of covariables that I have regrouped into four sub-groups as detailed in the table 3. These are households' head attributes, households' characteristics, inputs for production and risk management. In the absence of formal insurance on agricultural production, some households plant many crops to manage the risk related to negative shock, thereby, we expect to have a positive correlation between the number of crops and the agricultural productivity. Larson *et al.* (2015) report some irregularities in the information about the number of household members working on the farm, therefore I take as proxy for the household labor force the household adult equivalence scale. I also include some spatial variables such as the percentage of households involved in agricultural activities and the relative geographical deprivation [12]. The first variable can measure the extent to which households can learn new agricultural techniques from others around them. The set of variables W_{it} in the migration likelihood equation contains the same covariates as X_{it} except those related to the inputs of the agricultural production.

In the model specification, the distribution of λ_{it} allows to account for the heteroskedasticity induced by the measurement error in agricultural productivity or by the fact that the effect of some variables can be different among individuals. In fact, the household social capital can affect households' migration decisions differently.

The matrix format of the system 2 can be written as follows :

$$z_{it} = \left.\begin{pmatrix} MU^*_{it} \\ Prod^g_{1it} \\ Prod^r_{1it} \\ Prod^g_{0it} \\ Prod^r_{0it} \end{pmatrix}\right|_{\substack{\alpha_m, \alpha_0, \alpha_1, \gamma \\ \gamma_0, \gamma_1, \lambda_{it}, \theta_i}} \sim \mathcal{N}\left(\begin{bmatrix} Z_i\beta + W_{it}\alpha_m + \theta_i\gamma \\ X_{it}\alpha_1 + \theta_i\gamma_1 \\ X_{it}\alpha_0 + \theta_i\gamma_0 \end{bmatrix}, \lambda_{it}^{-1} \begin{pmatrix} \varsigma & 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_0^2 & 0 \end{pmatrix} \right) \tag{4}$$

With, $\epsilon_{si} \sim \mathcal{N}\left(0, \sigma_s^2\right)$, for $s \in \{0, 1\}$, and $\varepsilon_{it} \sim \mathcal{N}(0, \varsigma)$. Since ς does not have a known dimension, I set it to one. In addition, I assume that there are unobserved time-invariant variables θ_i, that differently affect the migration decision and the agricultural production.

One implication of the model is that in each period, conditionally to the observed and unobserved variables, the

12. I compute the geographical deprivation using information on the geo-spatial position of the household dwelling place to the main road, the main market, the border post, the administrative services, etc. in the district.

vector of loading factors $(\gamma, \gamma_1, \gamma_0)$ and the variance of the distribution of $(\theta_i)_{i=1}^N$ drive the correlation between the migration decision and the production. In fact, given the set of parameters,
$\forall\, t$,

$$cov(MU_t^*, Prod_{1t}) = \gamma \times \gamma_1 \times Var(\theta) \text{ and } cov(MU_t^*, Prod_{0t}) = \gamma \times \gamma_0 \times Var(\theta).$$

Besides, across periods,

$$cov(MU_t^*, MU_{t+1}^*) = \varsigma \times cov(\lambda_t, \lambda_{t+1}) \text{ and } cov(Prod_t, Prod_{t+1}) = \sigma_1 \times cov(\lambda_t, \lambda_{t+1}).$$

On the other hand, to identify all parameters, one loading factor for each outcome has to be set to one, I choose $\gamma_0 = 1$. The likelihood function is defined as follows :

$$\begin{aligned}
L(Prod_t, LM_t|B, \sigma\lambda, \theta) &= \textstyle\prod_{i=1}^N \prod_{t=1}^3 f(Prod_{1it}, LM_{it} = 1) \times f(Prod_{0it}, LM_{it} = 0) \\
&= \textstyle\prod_{i|LM_{it}=1} \prod_{t=1}^3 f(Prod_{1it}|X_{it}\alpha_1 + \theta_i\gamma_1, \lambda_{it}^{-1}\sigma_1)P(LM_{it} = 1|Z_i\beta + X_{it}\alpha_m + \theta_i\gamma, \lambda_{it}^{-1}) \\
&\times \textstyle\prod_{i|LM_{it}=0} \prod_{t=1}^3 f(Prod_{0it}|X_{it}\alpha_0 + \theta_i\gamma_0, \lambda_{it}^{-1}\sigma_0)P(LM_{it} = 0|Z_i\beta + X_{it}\alpha_m + \theta_i\gamma, \lambda_{it}^{-1}) \\
&= \textstyle\prod_{i|LM_{it}=1} \prod_{t=1}^3 f(Prod_{1it}|X_{it}\alpha_1 + \theta_i\gamma_1, \lambda_{it}^{-1}\sigma_1)\Phi(Z_i\beta + X_{it}\alpha_m + \theta_i\gamma, \lambda_{it}^{-1}) \\
&\times \textstyle\prod_{i|LM_{it}=0} \prod_{t=1}^3 f(Prod_{0it}|X_{it}\alpha_0 + \theta_i\gamma_0, \lambda_{it}^{-1}\sigma_0)\Phi(-Z_i\beta - X_{it}\alpha_m - \theta_i\gamma, \lambda_{it}^{-1})
\end{aligned}$$

$\Phi(.)$ is the standard normal cumulative function and $f(.)$ is a density function for a normal distribution. In the likelihood function $L(.|.)$, $\lambda = \left((\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N\right)$ and $B = (\beta, \alpha_m, \alpha_1, \alpha_0, \gamma, \gamma_1, \gamma_0)$, $\sigma = (\sigma_1, \sigma_0)$ are the set of parameters.

Considering the complexity of the likelihood function and because I want to estimate a mean effect of internal labor migration for each household, I use the Bayesian approach that assumes that each parameter of the model has a distribution with non-zero mean and variance. In the setting of treatment analysis, this approach provides a simple way, from the computational point of view, to account for the selection due to unobserved variables. Also, it enriches the analysis by enabling the effect of internal labor migration to be heterogeneous between all households, which is important in terms of public policies implications. In fact, public policies should be more efficient if the specific population who suffer from the labor migration is targeted.

Recently introduced in the treatment analysis, there are few empirical studies that have attempted to use this approach (See Heckman *et al.* (2012) and Chib & Hamilton (2002) for a review) and particularly in the literature of internal labor migration, there is not study so far. In the next section, the procedure used to implement the distribution for each component of $B = (\beta, \alpha_m, \alpha_1, \alpha_0, \gamma, \gamma_1)$, the distribution of σ, θ_i and $\lambda_i = (\lambda_{1i}, \lambda_{2i}, \lambda_{3i})$ is described.

4.3 Simulation procedure : Posterior distribution

First, I define a prior distribution for each parameter :

Parameters	**Prior distribution**
$\sigma = (\sigma_{1g}, \sigma_{1r}, \sigma_0)$	$\mathcal{N}_4(\ell_0, L_0)$
$B = (\beta, \alpha_m, \alpha_1, \alpha_0, \gamma, \gamma_1)$ B is $k \times 1$ vector	$\mathcal{N}_k(b_0, B_0)$
θ_i	$\mathcal{N}(\mu_0, \nu_0)$
λ_{it}, $t = 1, 2, 3$	$\mathcal{G}\left(\frac{\varphi_{0t}}{2}, \frac{\varrho_{0t}}{2}\right)$

Therefore the posterior distribution is given by :

$$\pi(B,\sigma,\theta,\lambda) = \pi(\sigma|\ell_0, L_0)\pi(B|b_0, B_0)\pi(\theta|\nu_0)\pi(\lambda|\frac{\varphi_{0t}}{2}, \frac{\varrho_{0t}}{2})L(Prod_t, LM_t|B,\sigma,\lambda,\theta). \tag{5}$$

To sample a distribution for each parameter, unobserved heterogeneity and time-varying scale, I follow the strategy proposed by Chib & Greenberg (1998), Chib & Hamilton (2002) and Lindley & Smith (1972) which can be resumed by the following steps [13]

1. Initialize $B, \sigma, \theta_i, \lambda_{i1}, \lambda_{i2}, \lambda_{i3}$

2. Sample σ from a Metropolis Hastings strategy. The posterior distribution is $h\left(\sigma|\ell_0, L_0, B, \theta, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N\right) = f(\sigma|\ell_0, L_0) \times L\left(Prod_t, LM_t|B,\sigma,\theta,\lambda\right)$. To sample σ from this distribution, the proposal density function in a multivariate-t student $q(\mu, V)$, where μ and V are respectively the mode and the inverse of the negative of the Hessian matrix of the function $h(.)$ evaluated at the mode.

3. Sample the unobserved component of the vector $z_{it}^* = \left(MU_{it}^*, Prod_{1it}^*, Prod_{0it}^*\right), \forall t = 1,2,3$.
   - if $LM_{it} = 1$ then sample first $MU_{it}^*|B,\sigma,\theta_i,\lambda_{1i},\lambda_{2i},\lambda_{3i}$ a normal distribution truncated to the interval $]0,+\infty[$. Instead, if $LM_{it} = 0$ then sample $MU_{it}^*|B,\sigma,\theta_i,\lambda_{1i},\lambda_{2i},\lambda_{3i}$ a normal distribution truncated to the interval $]-\infty, 0]$.
   - $\forall\, t \in \{1,2,3\}$, $i = 1,\cdots,n$, sample either $Prod_{1it}^*$ or $Prod_{0it}^*$, independently from i and t, from a normal distribution depending on whether LM_{it} is equal to 0 or 1.

4. Sample the set of parameters $B|z_{it}, b_0, B_0, \sigma, (\theta_i)_{i=1}^N, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N$, from a normal distribution.

5. Sample $\theta_i|z_{it}, B, \sigma, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N$, noted that the posterior mean and posterior variance for θ_i is different from one household to another.

6. Sample $\lambda_{it}|z_{it}, B, \sigma, (\theta_i)_{i=1}^N, \forall t = 1,2,3$. The posterior parameters are also intrinsic to each household.
7. Repeat steps 2 to 6 to get a full distribution of the posterior distribution.

With the posterior distribution in hand, it is possible to compute various estimators of the effect of labor migration on household's production.

4.4 Bayesian treatment effect of internal labor migration

At this stage, I assume that we have the posterior distribution for all parameters, the time-invariant unobserved variables and the time scale variation. Besides, from the third step of the simulation algorithm, we have the posterior distribution of the outcome and its counterfactual for each household, that is $\left(Prod_{1it}^*, Prod_{0it}^*\right)$ for which two of them are observed and the other two are simulated depending on the household's migration status. In my specification, the independence assumption is similar to the one posited in the standard matching analysis and can be expressed as follows :

Assumption 1. $Prod_{1it}^*, Prod_{0it}^* \perp\!\!\!\perp LM_{it} \mid X_{it}, W_{it}, \theta_i, \lambda_{it}\,;\, \forall\, t = 1,2,3$

This assumption states that in each period, conditional on the data and on the posterior distributions of θ_i and λ_{it}, the agricultural distribution is independent of the migration decision. Moreover, I assume here that there are no time-varying confounding factors that affect both the migration decision and the production decision. In fact, the time scale variation λ_{it} captures the deviation from the mean variance to correct for the measurement error and heterogeneous effect of some variables. Nevertheless, it goes beyond the assumption made in the potential outcome analysis because the independence assumption accounts for the presence of time-invariant confounders.

13. See appendix for the details of the sample algorithm

Under assumption 1, the effect of migration on household i in period t obtained from the posterior distribution is given by :

$$\rho_{it} = \begin{cases} Prod_{1it} - Prod^*_{0it} & if\ LM_{it} = 1 \\ Prod^*_{1it} - Prod_{0it} & if\ LM_{it} = 0 \end{cases}$$

, where $Prod^*_{0it}$ and $Prod^*_{1it}$ are the simulated components of the agricultural production and $Prod_{0it}$ and $Prod_{1it}$, the observed and actual agricultural productivity.

By assuming that I have achieved the convergence to the posterior distribution after Q iterations from the simulation process, $\bar{\rho}_{it} = \frac{1}{Q}\sum_{q=1}^{Q}\rho^q_{it}$ is the mean effect for the household i in period t. The connotation of the mean effect for $\bar{\rho}_{it}$ is because we averaged on the distribution of posterior distribution of parameters. We can also aggregate the effect across time periods : $\bar{\rho}_i = \frac{1}{3Q}\sum_{t=1}^{3}\sum_{q=1}^{Q}\rho^q_{it}$, is the mean effect of migration on the entire period covered by the analysis.

The Bayesian-average Mean Treatment Effect in period t ($BAMTE_t$) and over the sample period ($BAMTE$) can be expressed as follow :

$$BAMTE_t = \frac{1}{N}\sum_{i=1}^{N}\bar{\rho}_{it} \quad and \quad BAMTE = \frac{1}{N}\sum_{i=1}^{N}\bar{\rho}_i$$

At the same time, we can also compute the BAMTE on the treated ($BAMTET_t$) or on the non-treated ($BAMTENT_t$) in period t. For the first one,

$$\begin{aligned} BAMTET_t &= \tfrac{1}{N_1}\textstyle\sum_{i|LM_{it}=1}^{N}\bar{\rho}_{it} & BAMTET_t &= \tfrac{1}{3N_1}\textstyle\sum_{t=1}^{N}\sum_{i|LM_{it}=1}^{N}\bar{\rho}_{it} \\ BAMTENT_t &= \tfrac{1}{N_1}\textstyle\sum_{i|LM_{it}=0}^{N}\bar{\rho}_{it} & BAMTENT_t &= \tfrac{1}{3N_1}\textstyle\sum_{t=1}^{N}\sum_{i|LM_{it}=0}^{N}\bar{\rho}_{it} \end{aligned}$$

I also follow Chib & Hamilton (2002) by grouping households depending on the probability of experiencing labor migration in the period t conditional to covariate and unobserved heterogeneity, that is $P_{it} = \Phi(Z_i\beta + X_{it}\alpha_m + \theta_i\gamma, \lambda_{it}^{-1})$. At the qth iteration, $P^q_{it} = \Phi\left(Z_i\beta^q + X_{it}\alpha^q_m + \theta^q_i\gamma^q, \left(\lambda^q_{it}\right)^{-1}\right)$. Therefore, by discretizing the distribution of probability at each period and at each iteration per percentile, we can match households given that random probability inside each percentile. Let $D^q_{ht} = \left\{i|P^q_{it} \in \left(\frac{h-1}{10}, \frac{h}{10}\right)\right\}$, for $h = 0, 1, \cdots, 10$ and $t = 1, 2, 3$ be the different groups. As pointed out by Chib & Hamilton (2002), the matching of individuals based on P^q_{it} is well defined even at the bottom tail of the distribution because households are self-matched since we are able to compute counterfactual for each household. In the frequentist analysis, the individuals with extreme values of propensity score are generally dropped from the estimation. The average effect in each percentile group is given by :

$$\delta_{ht} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{M^q_h}\sum_{i\in D^q_{ht}}\rho^q_{it} \quad and \quad \delta_h = \frac{1}{3Q}\sum_{t}^{3}\sum_{q=1}^{Q}\frac{1}{M^q_h}\sum_{i\in D^q_{ht}}\rho^q_{it}$$

where $j \in \{g, r\}$ and $M^q_h = |D^q_{it}|$, $|.|$ is the cardinality function.

Moreover, we can also estimate the average effect for a group of households gathered given the households' characteristics and its head's characteristics. For example, one can be interested in the effect of migration on poorer households or female-headed households. Overall, the average effect for a group A is :

$$\delta_{At} = \frac{1}{Q\times A}\sum_{i\in A}\sum_{q=1}^{Q}\rho^q_{it} \quad and \quad \delta_A = \frac{1}{3Q\times A}\sum_{t}^{3}\sum_{i\in D_t}\sum_{q=1}^{Q}\rho^q_{it},$$

5. Results and Discussion

In this section, I discuss the issues of convergence of the posterior distribution. I also comment the parameters that intervene in the likelihood function and the distribution of the effect of the internal labor migration on household

agricultural production.

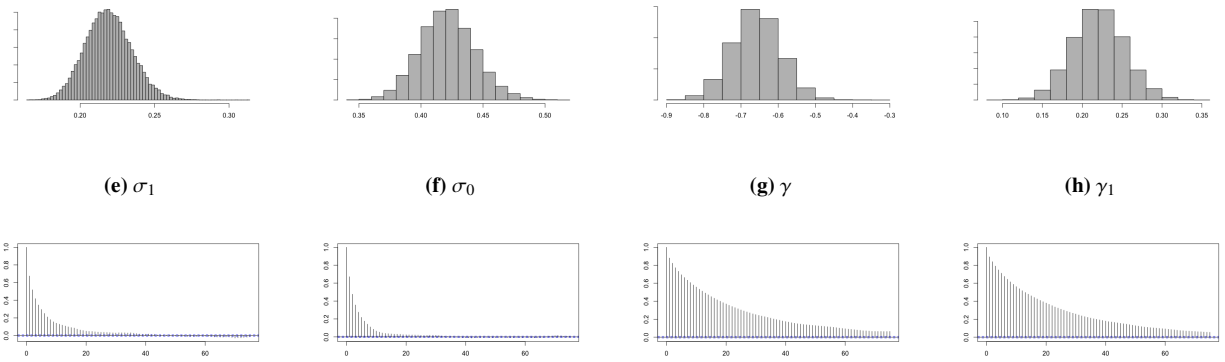
5.1 Convergence check and strength of instruments

Convergence check

I follow the algorithm reported in section 4.3 to reach the full set of posterior distribution for each parameter. To start with reliable parameters for each prior distribution and thereby achieve the convergence more rapidly, I first run 4000 iterations with 400 burn-ins from the simulation algorithm on a random training sample made up of a quarter of the initial sample (365 households). The algorithm is then run over 55,000 iterations with 5,000 burn-ins on the entire sample using the parameters obtained from the posterior distribution of the training sample. The figure 1 shows the distribution and the autocorrelation functions for each component of the variance-covariance matrix and the two loading factors.

As we can see, the autocorrelation functions reveal that our algorithm mixed well and the convergence is reached rapidly. Indeed, the autocorrelation over iterations is dropped very fast for all variances and a little less faster for the loading factors. Moreover, the posterior distribution of the loading factors in the migration decision equation is negative, meaning that households' time-invariant unobservables tend to decrease the probability of participating in migration. On the contrary, the sign (positive) of the loading factor in the production function suggests that households' unobservables tend to increase the agricultural production. Therefore, there is a negative correlation between the participation in migration and the investment in agricultural production due to time-invariant unobservables.

Figure 1 – *Distribution and auto-correlation of the Variance posterior distribution*



(e) σ_1 **(f)** σ_0 **(g)** γ **(h)** γ_1

To go further with the convergence check, I first run the stationary test proposed by Heidelberger & Welch (1983) and then next the Geweke's means convergence test on all parameters. The latter test states that a Markov chain Monte Carlo (MCMC) converges to the right posterior distribution if the mean of each parameter computing on a proportion of the sample on the top of the iteration is equal to the mean computing on the tail of the distribution. Geweke (1992) proposes using 10% on the top and 50% on the tail of the distributions. The convergence tests reveal that each variance posterior distribution reaches the convergence.

Strength of instruments : Testing the exclusion restriction

I have reported in tables 4 and 7 the mean and the standard deviation of the posterior distribution of all parameters for equations 2 and 3 respectively. For both specifications, the posterior distribution of the two components of the variance-covariance matrix are all significant at the 5% level, meaning that the 95% credibility interval doesn't include zero. Moreover, the variances are higher in the first specification (equation 2) than in the specification that allows the

instruments of migration to be correlated with the agricultural production. This might suggest that some instruments are correlated with the error term. Indeed, in table 2 we can see that the migration rate experienced in 2005 at household and district levels tends to significantly decrease the household production. Alternatively, when I don't allow the instruments to intervene in the production equation, the migration rate at the district level is strongly positively correlated with the migration decision (see table 3) ; however, when it appears in the production equation, it is no longer significant in the migration decision.

5.2 Posterior distribution

In this part, I comment on the way that different variables affect the households' likelihood to invest in labor migration and also how households' attributes impact the agricultural production. Upper and lower represent respectively the upper and lower bands of the interval of credibility at p-value of 5%. All results reported here are obtained from estimations on the balanced sample (*sample A*). To test if restraining our analysis to this sample could bias our results and in what way it does, I also run the model on the *sample B*. I find that the average effect of the labor migration on agricultural productivity is lower than the one obtained from *sample A*. This result is borne out by the fact that the average agricultural productivity in the *sample B* is lower than the agricultural productivity of its counterpart in *sample A*. It seems that households who are absence for some survey rounds tend to have lower agricultural productivity [14].

5.2.1 Migration likelihood

Unlike other studies that find a concave relationship between the probability to participate in migration and the level of households' wealth, I find that in the rural areas of Uganda this relationship is convex [15]. In fact, the propensity to invest in labor migration decreases with the household's wealth and, the squared of the log of the wealth positively affects the migration decision. This means that below a certain level of wealth, households are less likely to involve in migration, perhaps because their expected gain does not offset other costs of migration that they cannot absorb. An increase of 1% of wealth (measured in log) decreases the probability to participate in labor migration by 3.2%. Moreover, households headed by women are more likely to participate in migration ; yet, the marital status of the head does not seem to differently affect the migration participation, only the households headed by singles have lower probability compared with households headed by a monogamously married head. The composition of the household in terms of members and their education attribute seem to be the strong determinants of the migration decision. In fact, households with a higher proportion of children aged 5 or less have a significantly lower incentive to participate in labor migration. On the contrary, households with a higher proportion of adults aged more than 65 years have a higher incentive to get involved in migration. In the literature, some authors argue that once we control for the head's education, the average education among households' members might be a strong instruments for the labor migration. My results in this paper also confirm this finding as I find that the average education level in households significantly increases the participation to migration and has no effect on the agricultural productivity. Yet, the magnitude of the effect is too small, it only increases the probability by .006. In the same vein, the household's relative deprivation in terms of total hours worked in domestic tasks by adult members is a disincentive to get involved in migration and it has no significant effect on household's agricultural production at a 95% credibility interval. Again, this variable appears to be a strong instrument favoring migration.

14. Results are available on request
15. A two stage Heckman selection model leads to the same conclusion

Table 4 – *Posterior distribution of parameters : Sample A*

Variables	Migration likelihood			Agr. Production for MIG-hhs			Agr. Production for Non-MIG-hhs		
	mean	**lower**	**upper**	**mean**	**lower**	**upper**	**mean**	**lower**	**upper**
Log wealth	-0.75	-1.07	-0.44	0.05	-0.22	0.33	0.36	0.08	0.64
Log wealth sq. 10	0.06	0.03	0.09	0.01	-0.02	0.04	-0.02	-0.05	0.01
If head is									
Married polygamous	0.04	-0.11	0.19	-0.08	-0.19	0.04	0.07	-0.06	0.19
Divorced or sep	0.06	-0.18	0.30	-0.13	-0.30	0.04	0.01	-0.20	0.21
Widow	-0.02	-0.23	0.20	-0.08	-0.24	0.06	-0.01	-0.20	0.17
Single	-0.67	-1.18	-0.17	-1.30	-1.92	-0.73	0.02	-0.40	0.45
=1 if head is female	0.57	0.40	0.75	0.13	0.01	0.25	-0.27	-0.41	-0.12
log Head age	0.05	-0.15	0.24	0.44	0.29	0.60	0.12	-0.05	0.29
Head Educ	0.03	0.01	0.05	0.02	0.01	0.03	0.00	-0.02	0.02
Children less than 5	-1.15	-1.52	-0.78	-0.04	-0.35	0.27	-0.40	-0.70	-0.10
Individuals aged more than 65	1.10	0.62	1.57	-1.01	-1.46	-0.57	-0.43	-0.87	0.01
log HH size	1.44	1.28	1.60	0.33	0.18	0.48	0.26	0.12	0.41
Average educ in hh	0.04	0.01	0.08	0.00	-0.02	0.03	0.00	-0.02	0.03
Adult domes. lab DP.	-0.10	-0.17	-0.02	0.01	-0.05	0.08	-0.01	-0.06	0.03
Children domes.lab. DP.	0.05	-0.06	0.15	-0.04	-0.12	0.05	0.03	-0.04	0.09
GEO DP.	-0.07	-0.13	-0.02	-0.06	-0.10	-0.03	0.00	-0.03	0.03
Head-MIG	0.01	-0.25	0.27	-0.24	-0.48	-0.01	-0.05	-0.23	0.14
Ethnicity (at country)	-1.46	-2.04	-0.90	1.66	1.12	2.20	0.38	-0.19	0.96
Ethnicity (at district)	-0.05	-0.26	0.15	-0.24	-0.40	-0.09	0.04	-0.13	0.20
Center	0.02	-0.15	0.20	-0.17	-0.29	-0.04	-0.10	-0.25	0.06
East	-0.59	-0.76	-0.43	-0.22	-0.35	-0.10	-0.25	-0.39	-0.10
North	-0.41	-0.58	-0.24	-0.09	-0.24	0.06	-0.05	-0.21	0.11
Hired lab	-	-	-	-0.02	-0.06	0.01	0.03	0.00	0.06
Proxy of HH labor	-	-	-	-0.10	-0.22	0.03	-0.16	-0.30	-0.02
Area	-	-	-	-0.57	-0.64	-0.51	-0.49	-0.55	-0.44
Num crops	-	-	-	0.84	0.73	0.94	0.75	0.67	0.84
labor Agriculture in 1 km radius	-	-	-	0.00	-0.22	0.23	0.08	-0.19	0.34
Nb. of migrants in hh(2005)	0.30	0.20	0.41	-0.08	-0.14	-0.01	0.01	-0.10	0.11
Migration rate in the District	0.22	-0.39	0.83	-0.74	-1.34	-0.14	-2.15	-2.80	-1.51
Log Wealth DP	-0.00	-0.02	0.02	-0.01	-0.03	0.01	-0.00	-0.03	0.03
Intercept	-1.45	-2.03	-0.88	4.60	4.06	5.14	4.93	4.35	5.51
Loading fact.	-0.66	-0.79	-0.53	0.22	0.16	0.29	-	-	-
σ	-	-	-	0.22	0.19	0.25	0.42	0.38	0.47

Note :

- ⋆ Head-MIG equal 1 if the current place of residence of household is different from the place of household's head place of birth, that is when the household's head has permanently migrated in the past (less than 11 years and more than two years).
- ⋆ In the column of the likelihood migration, I have reported estimated parameters and they can't not be interpreted as a marginal effect. Only the sign is significant. However, in the text I will sometimes refer to the marginal effect computed for some variables.

The propensity of migrating also increases with the head's education, the size of the household and decreases with the head's age ; an increase in 1% of household size increases the migration probability by 0.98%. The geographical position of the household's dwelling place has a negative effect on migration participation as expected. I also test the hypothesis that the household's relative wealth deprivation can be a strong push factor for migration once controlling for the level of the wealth. It turns out that the relative wealth deprivation has no effect on the migration propensity. Moreover, all other instruments appear to be intimately related with household migration decisions. Having one more member who has experienced labor migration inside the household at least five years earlier increases the likelihood of migrating by .045 of percentage points, and an increase of the migration rate at the district level by 1% increases household migration by .033 of percentage points.

Moreover, the distribution of the unobservables shows that for some households, θ_i is positive and for others it is negative. I plot the inverse demand of migration as a function of unobservables. It emerges that the inverse demand of migration decreases with the time-invariant unobservables meaning that the distribution of θ_i captures the unobserved

distribution of the cost of migration that doesn't vary over time (see figure 7).

5.2.2 Agricultural Production

As mentioned above, the total production aggregates the production of maize, beans, coffee, peanuts, bananas and potatoes per hectare. Furthermore, these crops constitute the most cultivated crops within the country and the agricultural production is expressed in kilogram per hectare. We can see from the Table 2 that the investment process in the agricultural production differs according to the household migration status. Nevertheless, regardless of migration status, the agricultural productivity does not change given the marital status and household's relative deprivation in terms of wealth, the domestic hours worked by adult household members and the geo-spatial position of household dwelling place. On the contrary, both the migrant-HHs and non migrant-HH productions are positively affected by household size and the number of crops planted while negatively affected by the total area plotted. The number of crops managed is usually used as a proxy of how households manage the risk that might occur. The results suggest the importance of this factor for households involved in migration. In fact, cultivating one additional crops increases the agricultural productivity by 84% for migrant-HHs and 76% for non migrant-HHs. Moreover, an increase of one hectare of planted area decreases the agricultural productivity by 53% and 56% respectively for migrant-HHs and non migrant-HHs.

The results suggest that that agricultural productivity of female-headed households is higher in the migrant-HHs group than in the non-migrant HHs group. Since female-headed households are more likely to participate in migration, it seems that returns to migration appear to allow them to invest more in the agricultural sector than for households headed by males. Moreover, while belonging to the larger ethnic group at the district and at the country levels is respectively negatively and positively correlated with the migrant-HHs production, they have no effect on non migrant-HHs. Mwesigye & Matsumoto (2016) find in the case of Uganda that ethnicity diversity tends to lower agricultural production ; instead, my results suggest the contrary since the increase of the share of individuals belonging to the same ethnic group lowers the investment in agricultural production.

Regarding the labor hired to work on the farms, I find that the number of days that the households hired people to work on their farms does not affect sending-migrant households' production ; yet, it increases the non sending-migrant households' production. Moreover, while the agricultural productivity increases by 3% with each additional year of the education of the migrant-HHs head, head's education has no effect within households in the non-migrant households group. In parallel, the share of household members aged less than 5 years old tends to lower the agricultural production among the non migrant-HHs group while it has no effect on the sending-migrant households' productivity. This means that non sending-migrant households' agricultural productivity strongly depends on its household composition as compared to the sending-migrant households, which is in line with their choice to not participate in migration.

The distribution of the individual's variance scale, $(\lambda_{i1}, \lambda_{i2}, \lambda_{i3})$ reveals that the conditional distributions of the agricultural productivity of the households belonging to the non sending-migrants households are more heteroscedastic than the counterpart for the sending-migrant households. Regarding the source of heteroscedasticity, this result means that the measurement error of agricultural production and the omission of some time-variant unobservables in the production function are more likely to occur for the households that are not involved in labor migration.

5.3 Distribution of the effect of labor migration on agricultural production

In this section, we compare each household production to the counterfactual production obtained by simulation. Before getting there, I first evaluate to what extent my model accurately predicts the actual agricultural production distributions. Figure 6 shows that even if the model does not perfectly predict the distribution of the household's production, at least, the predicted distributions are very close to the actual distributions. Moreover, it is normal that the predicted distribution has a smaller variance than the observed one. Besides, the average gap between the actual and the predicted value does not attain a production of two kilograms per hectare.

Table 5 – *Average effect of internal labor migration on agricultural productivity*

		Migrant-HHs(*BAMET*)					Non Migrant-HHs (*BAMENT*)				
Average effect	All	All	Women headed	Male headed	Production>med	Production≤med	All	Women headed	Male headed	Production>med	Production≤med
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Period 1	0.26	0.27	0.39	0.21	0.89	-0.44	0.26	0.41	0.21	-0.50	1.00
	(0.03)	(0.07)	(0.14)	(0.08)	(0.07)	(0.10)	(0.03)	(0.07)	(0.04)	(0.03)	(0.04)
Period 2	0.17	0.54	0.69	0.47	0.99	0.01	0.02	0.23	-0.05	-0.55	0.56
	(0.03)	(0.05)	(0.09)	(0.05)	(0.04)	(0.07)	(0.03)	(0.06)	(0.04)	(0.04)	(0.04)
Period 3	0.16	0.30	0.37	0.26	0.78	-0.24	0.10	0.25	0.05	-0.51	0.68
	(0.02)	(0.04)	(0.08)	(0.05)	(0.04)	(0.06)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)
Total	0.20	0.37	0.48	0.31	0.89	-0.22	0.13	0.30	0.07	-0.52	0.75
	(0.03)	(0.05)	(0.10)	(0.06)	(0.05)	(0.08)	(0.03)	(0.06)	(0.04)	(0.03)	(0.04)

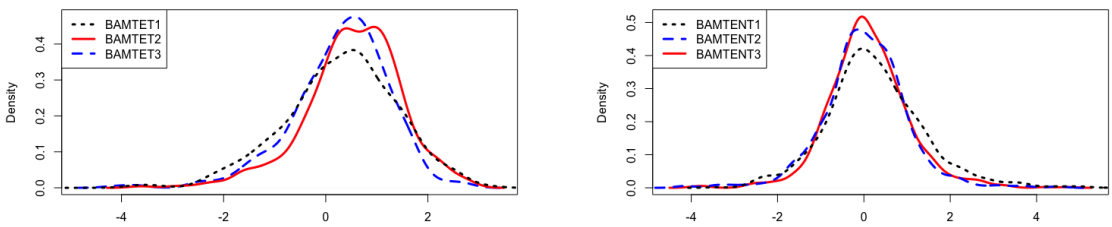
Note :

⋆ **med stands for the median of the distribution.**

Although the Bayesian average mean effect (*BAME*), the Bayesian average effect on the treated (*BAMET*) and the Bayesian average effect on the non treated (*BAMENT*) of the internal labor migration are all positives [16] over the entire sample (see columns (1), (2) and (3) of the Table 5), the distribution of the effect represented in the Figure 2 and the average effect on the specific subgroup tell us a different story. Indeed there are households for whom the effect is negative while for other households the effect is positive. This result then suggests that the internal labor migration affects the households' production differently, thus, aggregating the effect over the entire population could hide other facets of the actual impact of migration.

For all periods, the *BAMET* is higher than the average mean effect on the non treated (*BAMENT*). Furthermore, while the *BAMET* increases between the first and the second periods and decreases between the second and the third periods, the *BAMENT* decreases over time. The households headed by women always have a higher return to migration than households headed by men. Thereby, internal labor migration might be a way for female-headed households to increase their agricultural production, which is an interesting result since the literature on poverty usually depicts worse livelihoods condition of this group of households. Nevertheless, since female headship is highly correlated with being a migrant household, one might think that this is mostly due to the fact that the male (husband) has migrated. Thus, it is more likely that transfers of husband to household left behind to be higher. However, only about 23% of female-headed households participating in migration have a husband who has migrated.

Figure 2 – *Distribution of the effect of internal migration on agricultural production*



Percentage of HHs with BAMET>0 : 63.0%,76.8%,67.7%

Percentage of HHs with BAMTENT>0 : 56.6%,51.3%,52.7%

Among sending-migrant households, there are respectively 63%, 76.8% and 67.7% of the households for which the effect is positive from the first to the third period. Results in columns (5) and (6) reveal that, among households who

16. The average effect does not exceed the gain of two kilograms per hectare

decide to participate in migration, the larger farmers are those who benefit the most from the internal labor migration compared to the smaller farmers. This implies that the larger farmers are more likely to invest the return to migration in the agricultural sector. Among households belonging to the non-migrants group, we obtain the parallel result that larger farmers would have had positive return by investing in the labor migration. Instead, for the smaller farmers, the results suggest that the internal labor migration tends to decrease the agricultural production. Everything happens as if smaller farmers bear all of the labor cost induced by the migration and that the major part of the return to migration is devoted to non-agricultural production.

Regarding the distribution of the effect by region, we have a higher share of migrant households who benefit from the labor migration (about 75%) in the western part of Uganda and it is in this region that the agricultural production is also the highest. This result and the fact that the larger farmers are positively affected by the labor migration lead to think that there is no reason to believe that the labor migration will be a great threat to food security and the stability of the food prices in Uganda. Another thing we could to look at is if risk diversification (by planting many crops) allows households to increase the benefit from migration. Our results lead to a mitigated effects since the households who have planted many crops have a higher returns from migration for a period and sometimes they have the smallest return in the another period.

To go further, one might be interested in knowing how the effect changes when the likelihood of participating in migration increases. Figure 7.(c) shows that the effect of migration on agricultural productivity increases with the likelihood of participating in migration. Moreover, the dispersion of the effect within each decile increases when the households are more likely to participate in migration. Furthermore, the average effect of migration is higher for those with negative time-invariant unobservables compared to those with positive unobservables. Therefore, the selection into the participation in migration due to unobservables is also correlated with the average effect.

6. Conclusion and discussion

So far, there are few studies investigating the effect of internal labor migration on agricultural productivity. This might be due to the difficulty of identifying the causal impact since households select themselves into migration. This paper fill a gap in the literature by investigating the distribution of the effects of the temporary internal labor migration on households living in the rural areas of Uganda. The outcome of interest is the agricultural productivity in kilograms per hectare of six crops (maize, beans, coffee, peanuts, bananas and potatoes) planted in all regions of Uganda. I find that the average effect of the internal labor migration on the agricultural productivity is positive ; the average effect on the households participating in migration is around 0.37 in terms of logs of agricultural productivity, corresponding to a 44% increase in agricultural productivity. This is in line with the *NELM* theory that argues that migration enables households to invest in the agricultural sector. However, as I allow the effect to be heterogeneous between households, it emerges that even if the average effect is positive, there are some households for which the labor migration decreases their agricultural productivity. These households are mostly small farmers, and are therefore more likely to be poor. Moreover, about a half of the households that do not participate in the labor migration across rounds would have had higher levels of production if they had participated in the labor migration. This study then brings new insight into how the internal labor migration affect households' agricultural productivity. This kind of analysis is possible through the introduction of the Bayesian approach in the treatment analysis by allowing to self-match each household.

Moreover, the Bayesian framework enables to test the exclusion restriction assumption in the Instrumental Variables (IV) approach that is often used. Indeed, in the migration literature, some authors use the past observations on the migration prevalence at the household and community levels to instrument the current migration participation. The problem is that there is no evidence that these variables are exogenous, meaning that they are not correlated with the households' livelihoods. In this paper I test this hypothesis and it emerges that the migration decision taken five years earlier is highly correlated with the current agricultural productivity. Moreover, when I introduce the instruments in the

agricultural production, the effect of the time-invariant unobserved factors on the likelihood of participating in migration increases significantly while its effect on the agricultural productivity does not change.

I also estimate the average effect within each percentile of the probability of participating in the labor migration. It turns out that the effect of migration on agricultural productivity increases with the likelihood of participating in migration which is a good new in terms of achieving optimality.

Although I attempt to limit the bias on my estimation results, my results might suffer from many weaknesses. First, I don't allow the parameters in each equation to vary across the waves which could bias the posterior distribution of counterfactuals outcomes since the changes across waves are only due to the households' attributes and not the way that these attributes affect the agricultural production.

A. Distribution of education, ethnicity and agricultural production in Uganda

Figure 3 – *Agricultural productivity of grain by head's years of education and by average years of education in household*

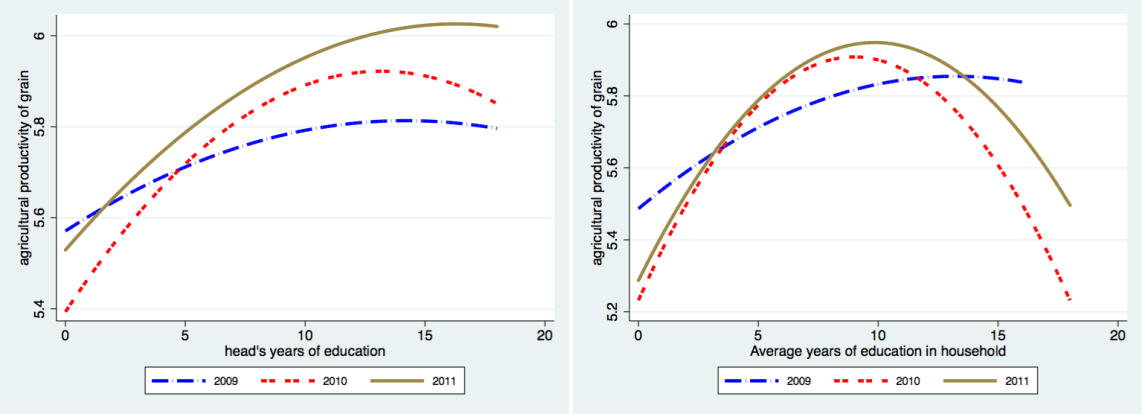


Figure 4 – *Distribution and auto-correlation of the Variance posterior distribution*

(a) *Education distribution in Uganda* **(b)** *Ethnicity and Agricultural productivity*

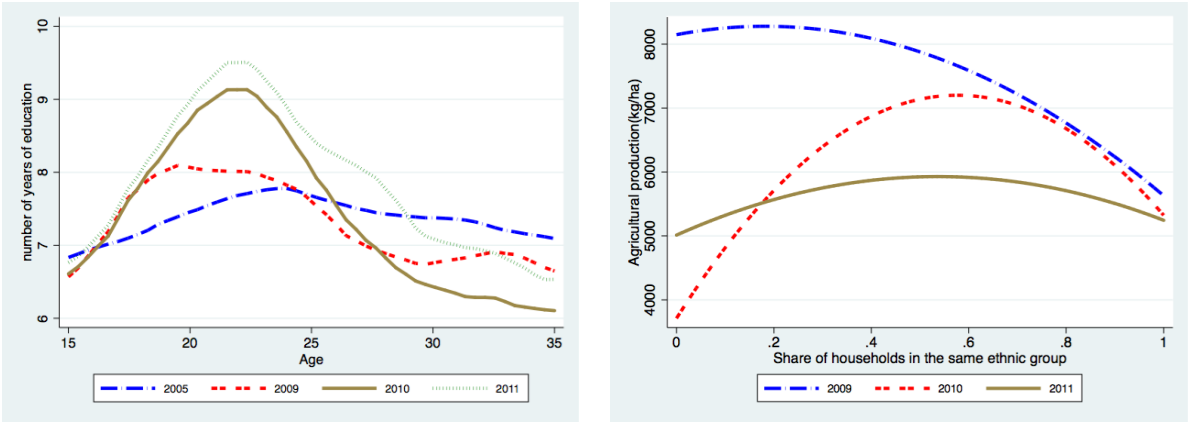


Figure 5 – *Duration of migration and share of household members involved in migration*

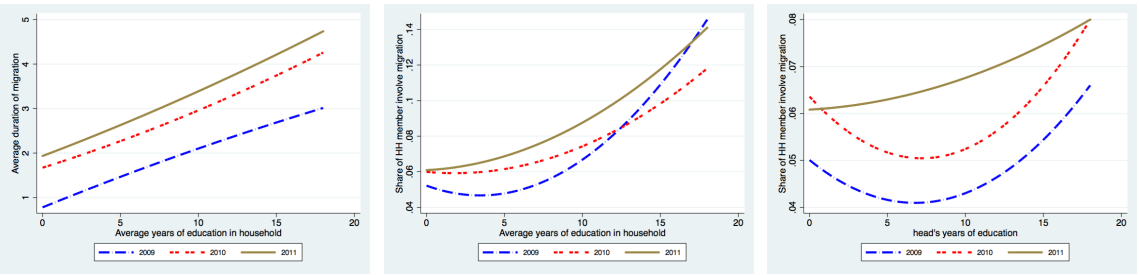


Table 6 – *Household migration status and previous permanent migration status of household head*

	$HH-Migrants$			$HH-Non-Migrants$		
	2009	2010	2011	2009	2010	2011
$Migrant-HHs$	13.1	12.4	13.4	86.9	87.6	86.6
$NonMigrant-HHs$	12.3	11.6	12.0	87.7	88.4	88.0

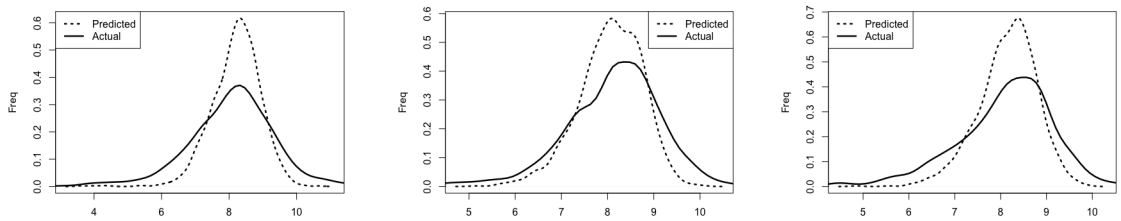
Table 7 – *Posterior distribution of parameters assuming that instruments verify the restriction assumption*

Variables	Migration likelihood		Agr. Production for MIG-hhs		Agr. Production for Non-MIG-hhs	
	mean	SD	mean	SD	mean	SD
Log wealth	-1.04	0.17	0.56	0.14	0.12	0.14
Log wealth sq	0.09	0.02	-0.03	0.01	0.00	0.01
Married poly	-0.01	0.07	-0.06	0.06	0.08	0.07
Div. sep	-0.00	0.12	-0.09	0.09	0.14	0.11
Widow	-0.01	0.11	-0.18	0.08	0.03	0.10
Single	-0.64	0.25	0.04	0.23	-0.08	0.21
Head gender	0.50	0.09	0.09	0.06	-0.26	0.08
Log Head age	0.05	0.10	0.64	0.08	0.07	0.08
Head Educ	0.03	0.01	0.02	0.01	0.00	0.01
Head-MIG	0.06	0.13	-0.11	0.12	-0.01	0.10
Children less than 5	-1.11	0.19	0.15	0.16	-0.43	0.15
Individuals aged more than 65	0.96	0.24	-1.02	0.22	-0.58	0.22
Log HH size	1.36	0.08	0.35	0.08	0.19	0.07
AVE educ in hh	0.04	0.02	-0.01	0.01	0.01	0.01
Adult domes. lab DP.	-0.06	0.04	-0.01	0.03	-0.00	0.02
Children domes.lab. DP.	0.05	0.05	-0.03	0.04	0.02	0.03
Geo-spatial DP.	-0.02	0.03	-0.05	0.02	-0.00	0.02
Hired lab	0.00	0.00	-0.02	0.02	0.01	0.02
Proxy of HH labor	-	-	-0.14	0.07	-0.07	0.07
Area planted	-	-	-0.61	0.04	-0.51	0.03
Nb. crops	-	-	0.85	0.05	0.75	0.04
Agriculture in 1 km radius	-	-	-0.27	0.11	-0.02	0.14
Ethnicity concentration (at country level)	-1.47	0.29	2.61	0.27	0.06	0.30
Ethnicity concentration(at district level)	-0.05	0.10	-0.20	0.08	0.12	0.08
Center	0.04	0.08	-0.21	0.07	-0.06	0.08
East	-0.53	0.08	-0.28	0.07	-0.23	0.08
North	-0.39	0.08	-0.03	0.08	-0.04	0.08
Nb. of migrants in hh(2005)	0.24	0.05	-	-	-	-
Migration rate in the District	1.30	0.31	-	-	-	-
Log Wealth DP.	-0.00	0.01	-	-	-	-
Intercept	-0.76	0.29	2.27	0.28	5.65	0.30
Loading fact.	-0.55	0.06	0.23	0.03	1	-
σ	1	-	0.24	0.02	0.49	0.03

⋆ Head-MIG equals 1 if the current place of living of household is different from the place of household's head place of birth, that is when household's head has migrated permanently in the past (less than 11 years and more than two years).

⋆ In the column of the likelihood migration, I have reported the parameters and they cannot be interpreted as a marginal effect. Only the sign is significant.

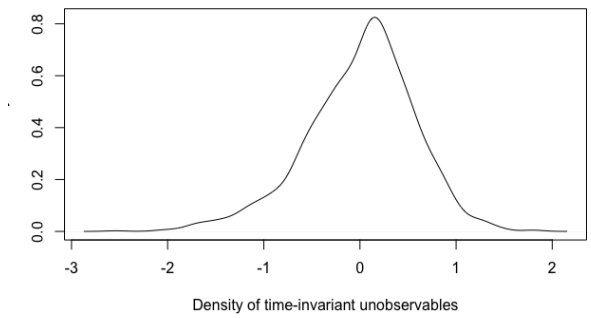
Figure 6 – *Actual and predicted distribution of agricultural production*



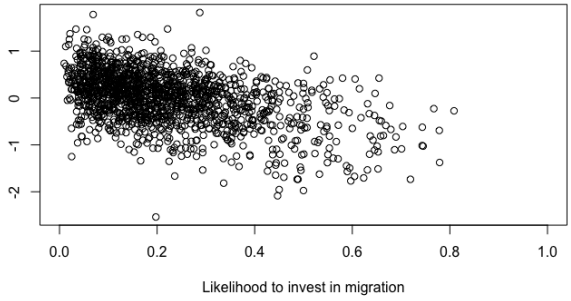
Period 1 *Period 2* *Period 3*

Figure 7

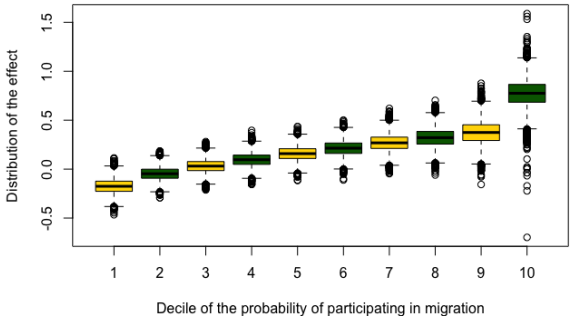
(a) *Density funct. of unobservables, θ_i*



(b) *Inverse demand of migration as a funct. of unobservables θ_i*



(c) *Heterogeneity of the effect by prob. of participating in MIG*



1 : Sampling Algorithm

1. Sample σ from a Metropolis Hastings strategy. The posterior distribution is $h\left(\sigma|\ell_0, L_0, B, \theta, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N\right) = f(\sigma|\ell_0, L_0) \times L(Prod_t, LM_t|B, \sigma, \lambda, \theta)$ and f(.—.) is a multivariate normal distribution of order two.

   To sample σ from function $h(.|.)$, Chib & Greenberg (1998) propose to sample σ from a multivariate-t distribution $q(\nu, V)$ where ν and V are respectively the mode and the inverse of the negative of the hessian of $log(h)$. Therefore, we move from σ to σ' if :

Table 8 – *Percentage of households with positive return to migration*

	Central	Eastern	Northern	Western	Num crops>13	Num crops<=13
			Migrant-HHs			
Period 1	61.46	53.33	60.47	72.94	63.69	62.20
Period 2	75.00	80.00	75.56	76.42	80.35	72.73
Period 3	75.74	58.26	57.66	77.98	72.73	61.93
			Non Migrant-HHs			
Period 1	63.51	62.40	51.95	47.91	56.50	56.67
Period 2	61.21	45.23	37.93	66.22	54.81	48.26
Period 3	40.66	61.25	48.75	54.39	51.00	54.14

$$\min\left\{\frac{h\left(\sigma'|\ell_0,L_0,B,\theta,(\lambda_{1i})_{i=1}^N,(\lambda_{2i})_{i=1}^N,(\lambda_{3i})_{i=1}^N\right)q(\sigma|\nu,V)}{h\left(\sigma|\ell_0,L_0,B,\theta,(\lambda_{1i})_{i=1}^N,(\lambda_{2i})_{i=1}^N,(\lambda_{3i})_{i=1}^N\right)q(\sigma'|\nu,V)},1\right\}=1$$

This strategy enables to reach the convergence of σ more rapidly.

2. Sample the unobserved component of the vector $z_{it}^* = \left(MU_{it}^*, Prod_{1it}^*, Prod_{0it}^*\right)$, $\forall t = 1,2,3$.
   - if $LM_{it} = 1$ then sampled first $MU_{it}^*|B,\sigma,\theta_i,\lambda_{1i},\lambda_{2i},\lambda_{3i}$ a normal distribution truncated to the interval $]0,+\infty[$. Instead, if $LM_{it} = 0$ then sampled $MU_{it}^*|B,\sigma,\theta_i,\lambda_{1i},\lambda_{2i},\lambda_{3i}$ a normal distribution truncated to the interval $]-\infty,0]$.
   - $\forall\, t \in \{1,2,3\}$, $i = 1,\cdots,n$, sample either $Prod_{1it}^*$ or $Prod_{0it}^*$, independently from i and t, from a normal distribution depending on whether LM_{it} is equal to zero or one.

3. Sample the set of parameters $B|z_{it}^*, b_0, B_0, \sigma, (\theta_i)_{i=1}^N, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N$, from the normal distribution $\mathcal{N}(g,G)$, with
   $g = G^{-1}\left(b_0B_0^{-1} + \sum_{i=1}^N R_i'\Omega_i^{-1}(z_i - \Lambda\theta_i)\right)$; $G = \left(B_0^{-1} + \sum_{i=1}^N R_i'\Omega_i^{-1}R_i\right)$ where
   $\Omega_i = diag\,(\lambda_{i1},\lambda_{i2},\lambda_{i3}) \bigotimes diag(1,\sigma_1,\sigma_0)$ [17], $z_i = (z_{i1}, z_{i2}, z_{i3})$
   $R_i = \begin{bmatrix}\Delta_{i1}\\ \Delta_{i2}\\ \Delta_{i3}\end{bmatrix}$ is a matrix, with $\Delta_{it} = \begin{bmatrix} Z_i & W_{it} & 0 & 0\\ 0 & 0 & X_{1it} & 0\\ 0 & 0 & 0 & X_{0it}\end{bmatrix}$ of dimension $(3\times T,k)$;$T = 3$ and k is the length of B and $\Lambda_i = [0,0,1,0,0,1,0,0,1]$.

4. Sample $\theta_i|z_{it}, B, \sigma, (\lambda_{1i})_{i=1}^N, (\lambda_{2i})_{i=1}^N, (\lambda_{3i})_{i=1}^N$ from the normal distribution with mean $\mu_\theta^i = \Sigma_\theta^i D'C^{-1}\bar{z}_i$ and $\Sigma_\theta^i = \left(1/\nu_0 + D'C^{-1}\lambda_i^* D\right)^{-1}$. $D = (\gamma,\gamma_1,1)$ is the vector of loading factors, $\lambda_i^* = \sum_{i=t}^N \lambda_{it}$ and

$$\bar{z}_i = \sum_{t=1}^{3}\lambda_{it}\left(z_{it} - [Z_i\beta + X_{it}\alpha_m, X_{1it}\alpha_1, X_{0it}\alpha_0]\right)$$

5. Sample $\lambda_{it}|z_{it}, B, \sigma, (\theta_i)_{i=1}^N$, $\forall t = 1,2,3$ from a gamma distribution $\mathcal{G}\left(\frac{\lambda_0+3}{2}, \frac{\lambda_0 + \underline{z}_{it}'C^{-1}\underline{z}_{it}}{2}\right)$, with

$$\underline{z}_i = z_{it} - \begin{pmatrix} Z_i\beta + X_{it}\alpha_m + \theta_i\gamma\\ X_{1it}\alpha_1 + \theta_i\gamma_1\\ X_{0it}\alpha_0 + \theta_i\end{pmatrix}$$

6. Complete the sampling procedure by repeating step 1 to step 5.

17. $diag(C)$ represents the diagonal matrix with the elements of vector C on the diagonal and $A \bigotimes B$ stands for the kronecker product of A and B.

B. Some proofs

We demonstrate now the parameters of the posterior distribution of θ_i in the simulation process. Given $z_{it}, B, \sigma, (\lambda_{1i})_{i=1}^{N}, (\lambda_{2i})_{i=1}^{N}, (\lambda_{3i})_{i=}^{N}$ we have from equation 4 that $z_{it} \sim \mathcal{N}\left(\Delta_{it}A + D\theta_i, \lambda_{it}^{-1}\Sigma\right)$, with A be the vector of all parameters of the model except the loading factors.
Thereby, $\lambda_{it}\left(z_{it} - \Delta_{it}A\right) \sim \mathcal{N}\left(D\theta_i, \lambda_{it}\Sigma\right) \Longleftrightarrow \sum_{t=1}^{3}\left[\lambda_{it}\left(z_{it} - \Delta_{it}A\right)\right] \sim \mathcal{N}\left(\lambda_i^* D\theta_i, \lambda_i^*\Sigma\right\}, \lambda_i^* = \sum_{t=1}^{3}\lambda_{it}$

If we set $M_i = \sum_{t=1}^{3}\lambda_{it}\left(z_{it} - \Delta_{it}A\right)$, the posterior distribution of θ_i following the Bayes rules is given by : $P(\theta_i|M_i, \cdot) \propto P(M_i|\theta_i)P(\theta_i)$. In the right hand side of the proportionate sign, the both distribution is a normal distribution meaning they are proportionate to $\exp(-\frac{1}{2}Q)$, with $\theta_i \sim \mathcal{N}(0, \nu_0)$, [18]

$$
\begin{aligned}
Q &= \left[M_i - \lambda_i^* D\theta_i\right]' \left(\lambda_i^*\right)^{-1}\Sigma^{-1}\left[M_i - \lambda_i^* D\theta_i\right] + \theta_i'\nu_0^{-1}\theta_i \\
&= M_i'\lambda_i^*\Sigma^{-1}M_i - M_i'\left(\lambda_i^*\right)^{-1}\Sigma^{-1}\lambda_i^* D\theta_i - \theta_i D'\Sigma^{-1}M_i + \theta_i D'\Sigma^{-1}\lambda_i^* D\theta_i + \theta_i\nu_0^{-1}\theta_i \\
&= \theta_i\left[\lambda_i^* D'\Sigma^{-1}D + \nu_0^{-1}\right]\theta_i - M_i'\Sigma^{-1}D\theta - \theta_i D'\Sigma^{-1}M_i + M_i'\lambda_i^*\Sigma^{-1}M_i \\
&= [\theta_i - AK]' A^{-1}[\theta_i - AK] + M_i'\lambda_i^*\Sigma^{-1}M_i; A^{-1} = \lambda_i^* D'\Sigma^{-1}D + \nu_0^{-1}; \quad K = D'\Sigma^{-1}M_i
\end{aligned}
$$

The last term of the equality is independent from θ_i therefore, the posterior distribution $\theta_i \sim \mathcal{N}(AK, A)$ which ends the demonstration since $A = \Sigma_\theta^i = \lambda_i^* D'\Sigma^{-1}D + \nu_0^{-1}$ and $\mu_\theta^i = AK = \Sigma_\theta^i D'\Sigma^{-1}M_i$.

18. Since θ_i is a constant, it is equal to its transpose.

References

Abbring, J. 2003. Dynamic Econometrics Program Evaluation. *IZA Discussion paper*, **804**.

Adaku, A. 2013. The effect of rural migration on agricultural production in the Northern region of Ghana. *Journal of Agricultural Science and Applications*, **2**(4), 193–201.

Asankha, P., & Takashi, Y. 2011. Impacts of Universal Secondary Education Policy on Secondary School Enrollments in Uganda. *Journal of Accounting, Finance and Economics*, **1**, 16–30. 3

Bandiera, O., & Rasul, I. 2002. Social Network and Technologie Adoption in Northern Mozambique. *DEDPS-London school of Economics and Political Science*, **35**.

Baudassé, T., & Rémi, B. 2014. Gender Inequality and Emigration : Push factor or selection process. *International Economics*, **139**, 19–47.

Bryan, G., Chowdhury, S., & Mobarak, A.M. 2014. Underinvestment in a Profitable Technology : the Case of Seasonal Migration in Bangladesh. *Econometrica*, **82**(5), 1671–1748.

Bukuluki, et al. 2015. Migration in Uganda : A rapid country profile 2013. *International Organization for Migration(IOM)*. 3

Card, D. 2001. Estimating the Return to Schooling : Progress on Some Persistent Econometric Problems. *Econometrica*, **69**(5), 1127–1160. 9

Carneiro, P., Hansen, K.T, & Heckman, J.J. 2003. Estimating Distributions of Treatment Effetcs with an Apllication to Return to schooling and measurenment of the effects of uncertainty on college choice. *International Econometric Reviews*, **44**, 361–422. 1

Chib, S., & Greenberg, E. 1998. Analysis of multivariate probit models. *Biometrika*, **85**(2), 347–361. 13, 24

Chib, S., & Hamilton, B.H. 2002. Semiparametric Bayes analysis of longitudinal data treatment models. *Econometric Reviews*. 12, 13, 14

Conley, T. G., Hansen, C. B., & Rossi, P. E. 2012. Plausibly Exogenous. *The Review of Economics and Statistics*, **94**(1), 260–272. 10

Davis, B., Carletto, G., & Winters, P. 2010. Migration, Transfers and Economic Decision Making among Agricultural Households : an Introduction. *The Journal of Development Studies*, **46**, 1–13. 2

de Brauw, A. 2010. Seasonal Migration and Agricultural Production in Vietnam. *The Journal of Development Studies*, **46**, 114–139. 1, 2

de Haan, A. 1999. Livelihoods and poverty The role of migration a critical review of the migration literature. *The Journal of Development Studies*, **36(2)**, 1–47. 1, 2

Deininger, K. 2003. Does cost of schooling affect enrollment by the poor ? Universal primary education in Uganda. *Economics of Education Review*, **22**, 291–305.

Dinkelman, T., & Mariotti, M. 2016. The long Run Effects of Labor Migration on Human Capital Formation in Communities of Origin. *National Bureau of Economic Research (NBER)*.

Garip, F. 2008. Social Capital and migration : how do similar Ressources Lead to Divergent Outcomes. *Demography*, **45**(3), 591–617. 2, 9

Garlick, J., Leibbrandj, M., & Levinsohm, J. 2016. Individual Migration and Household Incomes. *National Bureau of Economic Research*. 2

Geweke, J. 1992. Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. *In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid, and AFM Smith). Clarendon Press, Oxford, UK*. 15

Glick, P., & Sahn, D. 2006. The Demand for Primary Schooling in Madagascar : Price, Quality, and the choice between Public and Private Providers. *Journal of Development Economics*, **79**(2), 118–145.

Gonzales, R., Aranda, P., & Mendizábal, J. 2016. Is Microfinance Truly Useless for Poverty Reduction and Women Empowerment ? A Bayesian Spatial-Propensity Score Matching Evaluation in Bolivia. *Partnership for Economic Policy (PEP)*, **10**.

Heckman, J.J., Ichimura, H., Smith, J., & Todd, P. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica*, **66**(05), 1017–1098.

Heckman, J.J., Lopes, H.F., & Piatek, R. 2012. Treatment Effects : A Bayesian Perspective. *Econometric Reviews*. 12

Heidelberger, P., & Welch, PD. 1983. Simulation run length control in the presence of an initial transient. *Opns Res.*, **31**, 1109–1144. 15

Huylebroeck, L., & Kristof, T. 2015. Universal Secondary Education(USE) in Uganda : blessing or curse ? The impact of USE on educational attainment and performance. *In : Reyntjens,F., Vandeginste, S. and M. Verpoorten(eds) L'Afrique des Grands Lacs : Annuaire 2014-2015. Antwerp :University Press Antwerp,*, 349–372.

Imbens, G., & Angrist, J. 1994. Identification and Estimation of Local Average Treatment effects . *Econometrica*, **62**(2), 467–475.

Jagger, P., Shively, G., & Arinaitwe, A. 2012. Circular migration, small-scale logging, and household livelihoods in Uganda. *Popul Environ*, **34**, 235–256. 3

Konseiga, A. 2007. Household Migration Decisions as Survival Strategy : The Case of Burkina Faso. *Journal of African Economies*, **16**, 198–233. 9

Kuhn, R. 2015. Internal Migration : Developing Countries. *International Encyclopedia of the Social and Behavioral Sciences*, **12**(2nd edition), 433–442. 2

Larson, D., Savastano, S., Murray, S., & Palacios-López, A. 2015. Are Women Less Productive Farmers ? How Markets and Risk Affect Fertilizer Use, Productivity, and Measured Gender Effects in Uganda. *Policy Research Working Paper,World Bank Group*, **7241**. 11

Lechner, M. 2009. Sequential Causal models for the Evaluation of labor market Programs. *Journal of Business and Economic Statistics*, **27**(01). 9

Lechner, Michael, & Miquel, Ruth. 2001. *Potential Outcomes Approach to Dynamic Programme Evaluation-Part I : Identification* . IZA Discussion Papers. Institute for the Study of Labor (IZA).

Lindley, D.V., & Smith, A.F. 1972. Bayes Estimates for the Linear Model. *Journal of the Royal Statictics Society. Series B (Methodological)*, **34(1)**, 1–41. 13

Lucas, R. 1987. Emigration to south Africa's Mines. *America Economic Review*, **77**(3), 313–330. 9

Mckenzie, D., & Rapoport, H. 2007. Network Effects and the Dynamics of Migration and Inequality : Theory and evidence from Mexico. *Journal of Development Economics*, **84**, 1–24. 9

Mckenzie, D., Gibson, J., & Stillman, S. 2010. How Important is Selection : Experimental Vs Non-experimental Measures of the Income Gains from Migration. *Journal of the European Economic Association*, **8**(4), 913–945.

Mendola, M. 2008. Migration and Technological Change in Rural Household : Complements or Substitutes ? *Journal of Development Economics*, **85**, 150–175. 1, 2, 9

Mu, R., & Van de Walle, D. 2011. Left behind to farm ? Women's labor re-allocation in ruraal china. *Labour Economics*, **18**, 83–97. 1, 2

Muto, M. 2009. The impacts of mobile phone coverage expansion and personal networks on migration : evidence from Uganda. *Prepared for presentation at the international association of Agricultural Economists conference, Beijing, China, August 16-22*. 2, 3

Mwesigye, F, & Matsumoto, T. 2016. The Effect of Population Pressure and Internal Migration on Lands Conflicts : Implications for Agricultural Productivity in Uganda. *World Development*, **79**(01), 25–39. 1, 2, 18

Ndiaye, A. S., Niang, O. K., Ndione, Y. C., & Dedehouanou, S.E. 2016. Migration, Remittances, Labor Market and Human capital in Senegal. *Partnership for Economic Policy (PEP)*, **10**.

Orazem, P, & E., King. 2008. *Schooling in Developing Countries : The Roles of Supply, Demand and Government Policy*. Book chaper. Handbook of Development Ecconomics.

Penny, A., Ward, M., Read, T., & Bines, H. 2008. Education sector reform : The Ugandan experience. *International Journal of Educational Development*, **28**, 268–285.

Rutaremwa, G. 2011. Internal and International Labor Migration in Uganda : The contribution of Remittances to Household Livelihoods. *Makerere University, presented in African conference on Population*. 3

Shannon, C.E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 379–423,623–656.

Sianesi, B. 2004. An Evaluation of the Swedish System of active labor Market Programs in the 1990s. *The Review of Economics and Statistics*, **86**(1), 133–155.

Ssewamala, F., Wang, J., Karimli, L., & Nabunya, P. 2011. Strengthening Universal Primary Education in Uganda : The potential role of an asset-based development policy. *International Journal of Education Development*, **31**, 472–477. 3

Stack, O, & Taylor, J. E. 1991. Migration Incentives, Miration types : Role of relative deprivation. *The Economic Journal*, **101**, 1163–1178. 2, 10

Stark, O., & Bloom, D. 1985. The New Economics of Labor Migration. *American Economics Review*, **75**(2), 173–178. 1

Stark, O., & Taylor, J. E. 1989. Relative Deprivation and International Migration. *Demography*, **26**(1), 1–14.

UBOS, Uganda Bureau of Statistics. 2013. *Uganda Panel National Survey-Wave III report*. Tech. rept. 4

Vasilaky, K., & Leonard, K. 2016. As Good as the Network they Keep ? : Improving the Outcomes Through the Weak Ties in Rural Uganda. *Forthcoming, Economic Development and Cultural Change*.