

THREE ESSAYS ON THE CAUSES AND CONSEQUENCES OF YOUTH MIGRATION IN
TANZANIA

By

Evgeniya Alekseevna Moskaleva

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food, and Resource Economics – Doctor of Philosophy

2022

ABSTRACT

THREE ESSAYS ON THE CAUSES AND CONSEQUENCES OF YOUTH MIGRATION IN TANZANIA

By

Evgeniya Alekseevna Moskaleva

Migration of youth is a prominent phenomenon in Sub-Saharan Africa and in East Africa in particular. International and rural-to-urban migration gained a lot of attention in the older literature, yet internal rural-to-rural migration is the most frequent type. This work revolves around several issues of internal migration of youth in rural Tanzania. First, I determine which factors are associated with destination decisions made by young people. I look at four to six destination types on the rural-urban spectrum and consider various individual, household, and community factors that could affect migration decision. Second, I test how does migration to various destination areas on the rural-urban spectrum contribute to structural transformation through the shifts in main occupation. Although focusing on the shifts from agricultural work to self-employment and wage job, I also consider other employment categories like students, those working mainly in household maintenance, and unemployed people. Third, I estimate the impacts of youth outmigration on the livelihood of non-migrant household members. I consider changes to the labor supplied to the household farm, attraction of new household members, and adjustments to household participation in labor and land markets.

I make contribution to the literature on internal migration of youth in Sub-Saharan Africa, and Tanzania in particular, in four ways. First, I distinguish several migration destinations across the rural-urban spectrum, from low-density rural areas to cities, broadening the conceptualization of migration decision instead of focusing on a specific flow of migrants. I test three categorizations of location types to account for different interpretations of results and to verify

that the main results are not an artifact of the choice of the definition of “rural”. Second, I stress the importance of rural-to-rural migration, which is prevalent in Tanzania, although understudied. I show that even migration to low-density rural areas is associated with a shift towards non-agricultural employment. Third, while looking at occupational shifts, I consider people who are usually excluded from the analysis: students and those employed in household maintenance. I also look at women who state marriage as their main reason for migration. It allows to broaden the view on migration flows and discover employment difficulties for certain groups of people, for example, female rural-to-rural migrants involved mainly in household maintenance and students transitioning into employment. Fourth, I explore the labor adjustment strategies of the households left behind after a young adult migrates, which has rarely been studied in the context of the countries of Sub-Saharan Africa.

This dissertation is dedicated to my grandfather, Anatoly Druzhkov (1944-2019).

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my major professor, Thomas S. Jayne, for his advice, support, and patience over the years. I am also very grateful to my committee members, Songqing Jin, Nicole M. Mason-Wardell, and Amanda Flaim, for their invaluable feedback. I thank Ayala Wineman and Leah Lakdawala who contributed greatly to my work at the early stages of this research project.

I thank all the faculty and students at the Department of Agricultural, Food, and Resource Economics and the Department of Economics, who shared their thoughts during and after my presentations.

I would like to thank students from my cohort who made my first few years at the department precious: Gian Luca Gamberini, Awa Sanou, Ryan Vroegindeway, and Tram Hoang.

I also appreciate the community of students whom I've met at the dissertation support groups. I thank Megumi Moore for organizing these groups, and I thank all the participants. Especially, I would like to thank those who have been there with me every day at the study group meetings: Melissa Dale, Marcela Omans McKeeby, and Jocelyn Dana-Lê. We got this!

Lastly, I want to thank my family and friends. This work wouldn't be possible without the support of my husband, Alexandr; the love of my mother, Nata, and my grandparents; and the encouragement of my friends, Olga, Natasha, Jason, Joanne, and Eddy.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES.....	xv
KEY TO ABBREVIATIONS	xvii
1. INTRODUCTION.....	1
REFERENCES.....	5
2. CAN A REFINED TYPOLOGY OF DESTINATION AREAS IMPROVE OUR UNDERSTANDING OF INTERNAL MIGRATION? EVIDENCE FROM TANZANIA.....	8
Abstract	8
2.1. Introduction	9
2.2. Literature review	12
2.2.1. Migration decisions and migration destination decisions	12
2.2.2. Definitions of “urban” and the continuum of locations on the rural-urban spectrum	16
2.3. Data and definitions	18
2.4. Empirical strategy	26
2.5. Results	29
2.5.1. Summary statistics.....	29
2.5.2. Logistic and multinomial logistic regression results.....	42
2.5.3. Robustness checks.....	55
2.6. Discussion	65
2.7. Conclusion.....	70
APPENDICES.....	73
APPENDIX 1. Data issues related to geospatial information.....	74
APPENDIX 2. Definition of “migrant”	77
APPENDIX 3. Classification of locations on the rural-urban spectrum.....	83
APPENDIX 4. Attrition	113
APPENDIX 5. Geographical zones	115
APPENDIX 6. Additional tables.....	117
REFERENCES.....	158
3. MIGRATION OF YOUTH TO DIFFERENT DESTINATION TYPES IN TANZANIA: HOW DOES THE LEVEL OF URBANIZATION AFFECT EMPLOYMENT SHIFTS?	164
Abstract	164
3.1. Introduction	165
3.2. Literature review	167
3.2.1. Migration destinations.....	167
3.2.2. Employment of youth.....	170
3.3. Data and definitions	173
3.4. Empirical strategy	180

3.5. Results	183
3.5.1. Descriptive analysis.....	183
3.5.2. Regression analysis	194
3.5.3. Additional analysis.....	208
3.6. Discussion	211
3.7. Conclusion.....	217
APPENDIX	220
REFERENCES.....	236
 4. IMPACTS OF YOUTH OUTMIGRATION ON THE LIVELIHOOD OF HOUSEHOLDS LEFT BEHIND: EVIDENCE FROM TANZANIA	 239
Abstract	239
4.1. Introduction	240
4.2. Literature Review	245
4.2.1. The effects of labor withdrawal and remittances on labor outcomes.....	245
4.2.2. Other outcomes of interest	248
4.3. Data and definitions	250
4.4. Empirical strategy	254
4.5. Results	258
4.5.1. Migrant youth at baseline	258
4.5.2. Descriptive results	260
4.5.3. Main results	267
4.6. Discussion	278
4.7. Conclusion.....	280
APPENDICES.....	284
APPENDIX 1. Remittances	285
APPENDIX 2. Additional tables and figures.....	290
REFERENCES.....	301
 5. CONCLUSION	 306
REFERENCES.....	310

LIST OF TABLES

Table 2.1. Main definition constructed for the locations on the rural-urban spectrum.....	20
Table 2.2. Summary statistics for youth living in rural areas at baseline (2,803 observations)....	25
Table 2.3. Variations of the model: values assigned to the dependent variable M for the individual i , $M_i=D$, that indicate a discrete destination type D , for the total number of K migration destination types – for the main constructed definition for locations on the rural-urban spectrum	28
Table 2.4. Migration rates of rural youth, by origin and destination	30
Table 2.5. t-test for the difference in means of key variables between non-migrant rural youth and migrant rural youth (column b); between migrants to rural and urban areas (column d)	33
Table 2.6. Means of key variables for migrants by destination, five destination types	35
Table 2.7. Means of key variables by age group and migration status for youth from rural areas according to the constructed definition	38
Table 2.8. Means of key variables by age group and destination for migrants from rural areas according to the constructed definition	39
Table 2.9. Means of key variables by gender, migration status, and destination for youth from rural areas according to the constructed definition	41
Table 2.10. Regression results (marginal effects, constructed definition of “rural”): binary division, two and three destinations	43
Table 2.11. Regression results (marginal effects, constructed definition of “rural”): four destinations.....	44
Table 2.12. Regression results (marginal effects, constructed definition of “rural”): five destinations.....	45
Table 2.13. Regression results (marginal effects, NBS categorization of “rural”): four destinations.....	52
Table 2.14. Share of households in 2012/2013, by population density, built-up area density, and share of land under agriculture.....	87
Table 2.15. Comparison of the NBS categorization and the constructed definition of “rural”: urban and non-urban households.....	88

Table 2.16. List of coordinates for cities and towns with population of at least 50,000 people in 2012	89
Table 2.17. Mean population density and built-up area density for seven largest urban locations	92
Table 2.18. Comparison of the NBS categorization and the constructed definition: households living in cities and towns.....	95
Table 2.19. Population density and built-up area density for households located further than 30 km from a town with population of at least 50,000; for households with population density below 400 people per sq. km or built-up area density below 8%	100
Table 2.20. Non-urban households located within 30 km of a town with population of at least 50,000.....	105
Table 2.21. Comparison of the NBS categorization and the constructed definition for non-urban households according to the constructed definition	106
Table 2.22. Distribution of population density for rural households: share of households with population density below certain thresholds	106
Table 2.23. Comparison of the NBS categorization and the constructed definition for rural households with high and low population density	107
Table 2.24. Summary statistics for the constructed definition and the NBS categorization for the 2012/2013 survey wave.....	110
Table 2.25. Explanatory variables used in other studies	117
Table 2.26. Migration rates by age group for people from rural areas according to the constructed definition unless stated otherwise	120
Table 2.27. Migration rates by gender for people from rural areas according to the constructed definition unless stated otherwise	121
Table 2.28. Regressions by age groups: logistic regressions and regressions with two destinations; constructed definition of “rural”	122
Table 2.29. Regression results (marginal effects, constructed definition of “rural”) by age group.....	123
Table 2.30. Regressions by gender: logistic regressions and regressions with two destinations; for the constructed definition of “rural”.....	126
Table 2.31. Regression results (marginal effects, constructed definition of “rural”) by gender.	127

Table 2.32. Number of observations of youth by their location at baseline: constructed definition of “rural”, NBS categorization, and their intersection	129
Table 2.33. Number of observations of migrant youth from rural areas by destination: constructed definition of “rural”, NBS categorization, and their intersection	130
Table 2.34. Comparison of characteristics of youth living in rural areas, by definition of “rural”	131
Table 2.35. Regression results (marginal effects): binary division and two destinations; for the NBS categorization of “rural”	132
Table 2.36. Regression results (marginal effects): binary division and two destinations; sample of youth for whom the constructed definition and the NBS categorization agree for all survey waves	133
Table 2.37. Regression results (marginal effects): four destinations; sample of youth for whom the constructed definition and the NBS categorization agree for all survey waves	134
Table 2.38. Number of observations of youth, by the type of their location at baseline according to the cluster analysis definition	135
Table 2.39. Number of observations of youth from rural areas according to the cluster analysis definition, by the type of their location as defined by the constructed definition and the NBS categorization of “rural”	135
Table 2.40. Number of observations of migrant youth from rural areas according to the cluster analysis definition, by destination	136
Table 2.41. Comparison of migration rates among youth living in rural areas, by definition of “rural”	137
Table 2.42. Comparison of characteristics of youth living in rural areas, by definition of “rural”	139
Table 2.43. Regression results (marginal effects) with the cluster analysis definition of “rural”: binary division, two, and three destinations	142
Table 2.44. Regression results (marginal effects) with the cluster analysis definition of “rural”: four destinations	143
Table 2.45. Regression results (marginal effects) with the cluster analysis definition of “rural”: five destinations	144
Table 2.46. Regression results (marginal effects) with the cluster analysis definition of “rural”: six destinations	145

Table 2.47. Migration rates by migration status: comparison of the definitions of “migrant” that are based on distance traveled and self-reports	146
Table 2.48. Comparison of characteristics of youth by their migration status, by definition of “migrant”: for the definitions based on distance traveled and self-reports	147
Table 2.49. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; definition of “migrant” is based on self-reports.....	148
Table 2.50. Regression results (marginal effects): five destinations; definition of “migrant” is based on self-reports.....	149
Table 2.51. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; define “migrant” if an individual is considered to be a migrant by either the definition based on distance traveled or self-reports	150
Table 2.52. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; define “migrant” if an individual is considered to be a migrant by either the definition based on distance traveled or self-reports.....	151
Table 2.53. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; define “migrant” if an individual is considered to be a migrant by both the definition based on distance traveled and by self-reports.....	152
Table 2.54. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; define “migrant” if an individual is considered to be a migrant by both the definition based on distance traveled and by self-reports	153
Table 2.55. Comparison of migration rates by migration status: for the definitions of “migrant” based on distance traveled and administrative change.....	154
Table 2.56. Comparison of characteristics by migration status, by definition of “migrant”: for the definitions based on distance traveled and administrative change.....	155
Table 2.57. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; definition of “migrant” is based on distance traveled and change in administrative area	156
Table 2.58. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; definition of “migrant” is based on distance traveled and change in administrative area	157
Table 3.1. Main occupation of people of age 15-65 in 2008/2009 and 2012/2013, by age group, gender, and location type	176

Table 3.2. Summary statistics for baseline individual, household, and community characteristics of youth living in rural areas according to the constructed definition of “rural” (2,803 observations)	179
Table 3.3. Share of people with main occupation in a certain sector, by migration destination.	184
Table 3.4. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status; each row sums to 100%	187
Table 3.5. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status, for six groups of observations with at least 10 observations in 2008/2009; each row sums to 100%	189
Table 3.6. Selection into migration: marginal values from multinomial logistic regression of indicators to have main occupation in agriculture and non-agricultural wage job or self-employment in 2008/2009 on migration status in 2012/2013	193
Table 3.7. Migration and the probability to stay engaged in work	196
Table 3.8. Migration and the probability to stay engaged in work excluding household maintenance.....	197
Table 3.9. Migration and the probability to become engaged in work	200
Table 3.10. Migration and the probability to become engaged in work excluding household maintenance.....	201
Table 3.11. Migration and the probability to have main occupation in agriculture in the last survey wave; NBS definition of “rural”	203
Table 3.12. Migration and the probability to have main occupation in non-agricultural wage job or self-employment in the last survey wave.....	207
Table 3.13. Contribution of migration to various destinations to the total change in main occupation	213
Table 3.14. Main occupation of people of age 15-65, by age group, gender, and location type; NBS definition of “rural”	221
Table 3.15. Share of people with main occupation in a certain sector, by migration destination; NBS definition of “rural”	223
Table 3.16. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status; each row sums to 100%; NBS definition of “rural”	224

Table 3.17. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status, for six groups of observations with at least 10 observations in 2008/2009; each row sums to 100%; NBS definition of “rural”	225
Table 3.18. Selection into migration: marginal values from multinomial logistic regression of indicators to have main occupation in agriculture and non-agricultural wage job or self-employment in 2008/2009 on migration status in 2012/2013; NBS definition of “rural”	226
Table 3.19. Migration and the probability to stay engaged in work; NBS definition of “rural”	227
Table 3.20. Migration and the probability to stay engaged in work excluding household maintenance; NBS definition of “rural”	228
Table 3.21. Migration and the probability to become engaged in work; NBS definition of “rural”	229
Table 3.22. Migration and the probability to become engaged in work excluding household maintenance; NBS definition of “rural”	230
Table 3.23. Migration and the probability to have main occupation in agriculture in the last survey wave.....	231
Table 3.24. Migration and the probability to have main occupation in non-agricultural wage job or self-employment in the last survey wave; NBS definition of “rural”	232
Table 3.25. Migration and the probability to stay engaged in work including studies	233
Table 3.26. Contribution of migration to various destinations to the total shift in and out of engagement in work	234
Table 3.27. Contribution of migration to various destinations to the total shift from being a student into engaging in certain types of work	235
Table 4.1. Summary statistics for agricultural households with youth at baseline and non-migrant members during the last survey wave	253
Table 4.2. Activities of people of age 15 to 34 who lived in rural areas in 2008/2009 (NBS definition).....	259
Table 4.3. Descriptive results: household-level outcomes in rural households (according to the NBS definition)	261
Table 4.4. Descriptive results: individual-level outcomes, by gender, age, and presence in the household; rural households (NBS definition).....	264

Table 4.5. OLS regressions, household-level outcomes in agricultural households.....	269
Table 4.6. Nearest neighbor matching, household-level outcomes in agricultural households ..	270
Table 4.7. Dependent variable: indicator for working any number of days on the household farm in the first survey wave and not working in the last survey wave (probability to stop working in agriculture); agricultural households.....	275
Table 4.8. Dependent variable: change in the number of days of working on the household farm; subsample of people who participated in agriculture in both the first and the last survey wave; agricultural households	276
Table 4.9. Dependent variable: the number of days of working on the household farm, for new household members; agricultural households	277
Table 4.10. Summary statistics for remittances received in the past 12 months from children living elsewhere in Tanzania, thousand Tanzanian Shilling (TSh); by child.....	285
Table 4.11. Remittances received in the past 12 months, by information about the sender	287
Table 4.12. Characteristics of people of age 15 to 34 who lived in rural areas in 2008/2009	290
Table 4.13. Difference in means between migrant and non-migrant youth from rural areas according to the NBS definition, for people with certain types of main occupation	291
Table 4.14. Activities of youth from rural areas according to the NBS definition – by gender, age, and migration destination	292
Table 4.15. Main occupation by gender, age, and presence in the household; rural areas (according to the NBS definition)	294
Table 4.16. OLS regressions for household-level outcomes in rural households (according to the NBS definition)	296
Table 4.17. Nearest neighbor matching, household-level outcomes in rural households (according to the NBS definition)	297
Table 4.18. Propensity score matching, household-level outcomes in rural households (according to the NBS definition)	298
Table 4.19. Dependent variable: indicator for not working on the household farm in the first survey wave and working any number of days in the last survey wave (probability to start working in agriculture); agricultural households	299
Table 4.20. Dependent variable: change in the number of days of working on the household farm; agricultural households.....	300

LIST OF FIGURES

Figure 2.1. Rural youth, by gender, age in 2008/2009, and destination in 2012/2013 (origin and destination are defined according to the constructed definition of “rural”).....	32
Figure 2.2. Computed and reported distance between the first and the third survey waves	79
Figure 2.3. Computed and reported distance between the first and the third survey waves, for distances below 100 km	79
Figure 2.4. Self-reports for the cases when the difference between the computed and the reported distance is above 10 km: moved if years lived in the community is reported to be equal or below four in the third survey wave.....	81
Figure 2.5. Scatter plots for households’ population density and built-up area density and histograms for built-up area density.....	84
Figure 2.6. Scatter plots for households’ population density and built-up area density and histograms for built-up area density, for built-up area density below 20%	85
Figure 2.7. Scatter plots for households’ population density and built-up area density for those with population density above 400 people per square km and built-up area density above 8%; by distance to the nearest town with population of at least 50,000.....	90
Figure 2.8. Scatter plots for households’ population density and built-up area density for those with population density below 400 people per square km or built-up area density below 8%; by distance to the nearest town with population of at least 50,000.....	91
Figure 2.9. Households’ distance to the nearest town with population of at least 20,000 and distance to the nearest town with population of at least 50,000, for households with population density below 400 people per square km or built-up area density below 8%.....	97
Figure 2.10. Scatter plots for households’ population density and built-up area density for those with population density below 400 people per square km or built-up area density below 8%; by distance to the nearest town with population of at least 20,000.....	99
Figure 2.11. Population density and distance to town for households with population density below 400 people per square km living within 30 km of a town with population of at least 50,000.....	103
Figure 2.12. Geographical zones.....	116
Figure 3.1. Propensity for migration to four location types on the rural-urban spectrum (according to the NBS definition of “rural”).....	222

Figure 4.1. Remittances received from children of the household head, by sender's age group	288
Figure 4.2. Remittances received from children of the household head, by years lived at the host location and age group (colored gray for people of age 15-34).....	289
Figure 4.3. Average number of days spent on agricultural activities, 2008/2009, by age and outmigration experience: for non-migrant household members living in rural areas (according to the NBS definition) in households with youth at baseline	293

KEY TO ABBREVIATIONS

AERC	African Economic Research Consortium
CIRAD	Center de Coopération Internationale en Recherche Agronomique pour le Développement (French) French Agricultural Research Centre for International Development
FAO	Food and Agriculture Organization
GHS	Global Human Settlement
IZA	Forschungsinstitut zur Zukunft der Arbeit (German) Institute of Labor Economics
km	Kilometer(s)
LSMS	Living Standards Measurement Study
NBS	National Bureau of Statistics
RIGA	Rural Income Generating Activities
TLU	Tropical Livestock Units
TSh	Tanzanian Shilling(s)
UN	United Nations
UNCTAD	United Nations Conference on Trade and Development

1. INTRODUCTION

The population of Sub-Saharan Africa is highly mobile, the majority of migrants move within the continent and, in many areas, within their countries of origin (Mercandalli, 2017). Currently, many migrants emerge from rural areas, where the population growth is not expected to slow down before the 2030s, meaning that population pressure and associated employment challenges are likely to continue pushing people to move (Losch, 2017). Most of the African youth lives in rural areas, and the majority of rural youth is employed in agriculture struggling with limited access to resources and hindered productivity (Filmer and Fox, 2014). Hence, the significance of internal migration as a means to improve employment opportunities for youth from rural areas cannot be overestimated. It was shown across various contexts that migration is beneficial for both migrants and their families (Christiaensen and Kanbur, 2017; McKenzie, Gibson, and Stillman, 2010), which makes it one of the possible pathways out of poverty for many of those to whom it is available. The focus on youth, whom I define as people of age from 15 to 34, is justified as they are the most mobile group of people (Dinbabo, Mensah, and Belebema, 2017), while early career choices have a significant impact on the expected lifetime earnings (Bridges et al., 2017).¹ In this study, I look at different aspects of internal migration of youth from rural areas of Tanzania, a country in East Africa with high expected growth rates of rural population and the majority of migrants moving internally (Losch, 2017; Mercandalli, 2017).

This study is built upon several gaps in the current analysis of migration patterns. First, it aims to broaden the conceptualization of the migration decision into a multifaceted one

¹ I use the broad definition of youth including people of age from 15 to 34. In the second and the fourth chapters, I explore the heterogeneity of the observed patterns by age group within this definition of youth. The United Nations (UN) define youth as people of age 15 to 24. African Youth Charter defines youth as people of age 15 to 35.

considering migration destination on a rural-urban spectrum. The literature on the determinants of migration decision could be divided into three categories in respect to its attention to migration destinations. The first one views migration as a binary decision: to move or to stay in place. For example, Ocello et al. (2015) study the effect of environmental shock on the decision to move to another district in Tanzania. The second one looks at a certain migration flow, for example, rural-to-urban migration, but then views migration decision as binary: whether to move from a rural area to an urban area or not. For example, Nguyen, Grote, and Sharma (2017) study the determinants of the length of stay of rural-urban migrants in Vietnam. The third category considers migration as a non-binary decision, distinguishing various destinations on the rural-urban spectrum. For example, Msigwa and Mbongo (2013) study the determinants of the rural-to-urban and town-to-city migration destination decisions for people moving internally in Tanzania. My study falls into the third category and attempts to absorb several advancements of the second one, which allows for a more coherent view on migration that originates from rural areas. I consider two of the previously studied concepts in the urban hierarchy: peri-urban areas (Mueller et al., 2018) and secondary towns (Christiaensen and Todo, 2014) and add to them. This consideration will appear significant in two out of three essays included into this study: I investigate what factors are associated with the destination decision and how does migration to different destinations contribute to occupational shifts. Since working with location types, I pay special attention to the definitions used. Potts (2017) observes that diverging methods to the identification of localities as “rural” and “urban” can potentially lead to very different classifications. I use three categorizations to determine how robust the results are to alternative definitions of location types.

I stress the importance of rural-to-rural migration, which is prevalent for rural youth in Tanzania. Although rural destinations are the most frequent in Sub-Saharan Africa, they are often overlooked by research on migration (Lucas, 2016; Oucho and Gould, 1993). With the focus of attention being shifted towards international and rural-to-urban migration, rural-to-rural migration remains an understudied phenomenon even though it has an impact on millions of people around the world. I verify previously observed features of rural-to-rural migration flows and show the heterogeneity of migration flows to rural areas with high and low population density. Furthermore, I show how rural-to-rural migration could promote occupational shifts, including shifts to non-agricultural jobs.

I also add to the literature on the impacts of youth migration on the labor and other outcomes of the non-migrant household members staying in the origin, which is scarce in Sub-Saharan Africa. There is a vast literature on various aspects of livelihood of the household left behind in China as a result of internal rural-to-urban labor migration. In their review, Ye et al. (2013) show how most studies estimate the impacts of outmigration on the left-behind children and elderly parents. Hence, the most popular areas of study are educational attainment and child labor, and physical and mental health. Studies in other contexts describe impacts on parental health as well, while studies on the impacts on labor are rarer; and the most common migration flows studied refer to international migration (Antman, 2012). Researchers usually come to diverse conclusions on the impacts on labor and/or leisure outcomes (Murard, 2016), which could be explained by diverse channels these effects operate through. The two most common mechanisms are withdrawal of migrant's labor and remittances.

In all three of my essays, I use the 2008/2009 and the 2012/2013 waves of the Living Standards Measurement Study (LSMS) dataset for Tanzania. "Youth" are defined as people of

age 15 to 34. The purpose of my first essay, titled “*Can a refined typology of destination areas improve our understanding of internal migration? Evidence from Tanzania*”, is to describe the existing patterns of youth mobility in Tanzania and test whether one could gain policy-relevant insights about the migration decisions of youth by differentiating destination types. Employing this observation, my second essay, titled “*Migration of youth to different destination types in Tanzania: How does the level of urbanization affect employment shifts?*”, discovers the impact migration to various destination areas has on the changes to main occupation. In my third essay, titled “*Impacts of youth outmigration on the livelihood of households left behind: Evidence from Tanzania*”, the focus is shifted to migrants’ families who stay in the origin area. The purpose of this essay is to estimate the impact of youth outmigration on the labor supply and other choices of the non-migrant members of migrant’s household.

REFERENCES

REFERENCES

- Antman, F. M. 2012. The impact of migration on family left behind. In A. F. Constant and K. F. Zimmerman (eds.) *International Handbook on the Economics of Migration*. Cheltenham, UK, and Northampton, MA, USA: Edward Elgar, pp. 293-308.
- Bridges, S., L. Fox, A. Gaggero, and T. Owens. 2017. Youth unemployment and earnings in Africa: Evidence from Tanzanian retrospective data. *Journal of African Economies* 26(2): 119-139.
- Christiaensen, L., and R. Kanbur. 2017. Secondary towns and poverty reduction: Refocusing the urbanization agenda. *Annual Review of Resource Economics* 9: 405-419.
- Christiaensen, L., and Y. Todo. 2014. Poverty reduction during the rural-urban transformation – The role of the missing middle. *World Development* 63: 43-58.
- Dinbabo, M. F., C. Mensah, and M. N. Belebema. 2017. Diversity of rural migrants' profiles. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in motion. Dynamics and drivers of migration South of the Sahara*, pp. 24-25. Rome: FAO; CIRAD.
- Filmer, D., and L. Fox. 2014. *Youth Employment in Sub-Saharan Africa*. Africa Development Series. Washington, D. C.: World Bank.
- Losch, B. 2017. A lastly booming rural population and the youth employment challenge. In Mercandalli, S., and B. Losch, eds. *Rural Africa in Motion. Dynamics and Drivers of Migration South of the Sahara*. Rome: FAO and CIRAD, pp. 20-21.
- Lucas, R. 2016. Internal migration in developing economies: An overview of recent evidence. *Geopolitics, History and International Relations* 8(2): 159-191.
- McKenzie, D., J. Gibson, and S. Stillman. 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.
- Mercandalli, S. 2017. Prevalent, Contrasted intra-African migration patterns and new territorial dynamics. In Mercandalli, S., and B. Losch, eds. *Rural Africa in Motion. Dynamics and Drivers of Migration South of the Sahara*. Rome: FAO and CIRAD, pp. 22-23.
- Msigwa, R. E., and J. E. Mbongo. 2013. Determinants of internal migration in Tanzania. *Journal of Economics and Sustainable Development* 4(9): 28-35.
- Mueller, V., E. Schmidt, N. Lozano, and S. Murray. 2018. Implication of migration on employment and occupational transitions in Tanzania. *International Regional Science Review*: <https://doi.org/10.1177/0160017617751029>.
- Murard, E. 2016. Consumption and leisure: The welfare impact of migration on family left behind. *IZA Discussion Paper* No. 10305.

Nguyen, L. D., U. Grote, and R. Sharma. 2017. Staying in the cities or returning home? An analysis of the rural-urban migration behavior in Vietnam. *IZA Journal of Development and Migration* 7(1): 1-18.

Ocello, C., A. Petrucci, M. R. Testa, D. Vignoli. 2015. Environmental aspects of internal migration in Tanzania. *Population and Environment* 37(1): 99-108.

Oucho, J. O., and W. T. S. Gould. 1993. Migration, urbanization and population distribution. In K. A. Foote, K. K. Hill, and L. G. Martin, eds. *Demographic Change in Sub-Saharan Africa*. Washington, D. C.: National Academy Press, pp. 256-296.

Potts, D. 2017. Conflict and collisions in Sub-Saharan African urban definitions: Interpreting recent urbanization data from Kenya. *World Development* 97: 67-78.

Ye, J., C. Wang, H. Wu, C. He, J. Liu. 2013. Internal migration and left-behind populations in China. *The Journal of Peasant Studies* 40(6): 1119-1146.

2. CAN A REFINED TYPOLOGY OF DESTINATION AREAS IMPROVE OUR UNDERSTANDING OF INTERNAL MIGRATION? EVIDENCE FROM TANZANIA

Abstract

Older migration literature generally focused on a binary migration decision, to move or to stay in place, or on a decision to move from a rural to an urban area. Recent studies of migration look at a more diverse set of migration destination types, although the bias towards more urbanized locations still persists. This paper contributes to the literature by providing a more refined categorization of migration destinations on the rural-urban spectrum. This differentiation of location types improves our understanding of the dynamics of migration flows and provides the means to more accurately predict future changes in migration patterns. Looking at young adults from rural Tanzania moving internally within the country, I find that there are systemic differences in the characteristics of people migrating to each destination category. In contrast to conventional wisdom, the most frequent destination of young migrants is a relatively sparsely populated rural area. Multinomial logistic regression analysis based on nationally representative survey data shows a highly heterogeneous nature to migration location, which further varies by gender and age. Two distinct migration flows emerge: to low-density rural locations situated further from roads and towns and to more densely populated rural and peri-urban areas near towns. I find that some factors, like prior migration history, to be associated with the decision to migrate but not the choice of destination. Other factors, like gender, education, employment, negative shocks, and remoteness of the origin, are associated with a certain destination choice or can have a more diverse relationship with migration to various destinations.

2.1. Introduction

The number of internal migrants (740 million people in 2009) overwhelms the number of international migrants (221 million people in 2010²), but the literature on migration usually concerns with international migration.³ A similar imbalance between the observed migration patterns and the literature focus is seen in the studies of migration to rural and urban areas: people's destinations are diverse while the focus is set on rural-to-urban migration (Lucas, 1997). In Sub-Saharan Africa, and in East Africa in particular, internal migration is the primary form of relocation, especially for youth from rural areas, and most destinations are rural. In this chapter, I look at the factors associated with the decision to migrate to various destinations on the rural-urban spectrum within Tanzania that youth from rural areas make. In the neighboring Kenya and Uganda, 55% and 79% of migrants moved internally (UNCTAD, 2018), 52% and 85% of migrants originated from rural areas (Mercandalli, 2017), and at least 60% of migrants are between the ages of 15 and 34 (Dinbabo, Mensah, and Belebema, 2017). Hence, internal migration of youth from rural areas is prevalent in the region, and in Tanzania in particular, but still understudied – especially migration to non-urban areas. Migration is an important potential pathway out of poverty for people from rural areas (De Weerd, 2010), and, given the current migration rates and the rates of population growth in rural Sub-Saharan Africa, migration will continue to be a prominent phenomenon affecting the lives of millions (Mercandalli et al., 2017), most of whom are young adults.

Where do these young people move to and what makes them go there? Review studies (Oucho and Gould, 1993; Lucas, 2016) argue that the literature focuses on rural-to-urban and

² In 2020, the estimated number of international migrants is 281 million (McAuliffe and Triandafyllidou, 2021). The estimate for 2010 is here for the comparison with the 2009 estimate of the number of internal migrants, which is the most recent estimate.

³ This chapter is co-authored with Thomas S. Jayne.

international migration, although evidence has long existed that rural-to-rural migration prevails on the African continent.⁴ The most cited reason for migration is wage differential, which is used to explain rural-to-urban migration (Harris and Todaro, 1970). At the same time, search for available land or agricultural work leads farmers from one rural area to another (Bezu and Holden, 2014; Lucas, 2016). Recent studies provide a more nuanced and differentiated description of migration that involves secondary towns, peri-urban areas, and different types of rural areas, in contrast to the conventional binary rural/urban division (e.g., Christiaensen, De Weerdt, and Todo, 2013; Muzzini and Lindeboom, 2008). However, the existent literature does not consider whether the factors affecting individuals' migration decisions are similar regardless of the choice of destination across different destination categories on the rural-urban spectrum. I hypothesize that there may be important differences in the factors driving migration – depending on the destination. For these reasons, my analysis requires a more nuanced differentiation of internal migration destinations and a theoretical framework that guides model specification for these differentiated areas.

The goal of this study is to describe the flows of internal migration of youth in Tanzania, differentiating them by destination type, and to determine whether a more refined typology of destination areas along the rural-urban spectrum can improve our understanding of internal migration. My framework goes beyond the conventional binary rural/urban migration destination models by providing a more differentiated set of destination categories; I then test whether this categorization influences empirical findings regarding the most important and statistically significant drivers of migration to these various destinations. My analysis incorporates multiple

⁴ See Brown and Lawson (1985) for a review of earlier studies that stress the importance of rural-to-rural migration in developing countries.

migration destinations that were studied separately before.⁵ This makes the model more complex but may provide a more refined understanding of migration flows and youth's underlying migration motivations. Differentiation between the types of migration destinations is not uncommon (see, for example, Msigwa and Mbongo, 2013), especially in the studies of the impacts of migration (see, for example, Beegle, De Weerd, and Dercon, 2011), but the set of the destinations I look at is more comprehensive than other studies.

My sample consists of people between the ages of 15 to 34 who lived in rural areas at baseline. I differentiate migration flows by gender and age. The study contributes to the literature by analyzing a wider spectrum of migration destinations. Rural-to-rural migration is commonly under-appreciated, although the literature recognizing its importance is currently growing. I distinguish between types of rural destinations, in addition to other destinations on the rural-urban spectrum, and show that sparsely populated rural areas are a dominant destination among rural youth. Overall, I find that there are significant differences between migrants to various destinations, and I also confirm that rural areas are more accessible for migrants. My work contributes to the plentiful research on certain migration flows in Tanzania that uses the Kagera Health and Development Survey of 1991-2004 and the Living Standards Measurement Study (LSMS) that started in the 2008/2009.

While distinguishing destinations on the rural-urban spectrum, I pay close attention to the definitions of the categories I use. As there is no single “correct” definition of a “rural area”, researchers commonly construct definitions that would be most suitable for their study, although this behavior complicates efforts to compare findings across studies or draw generalizations from the literature. Potts (2017a, 2017b) shows that differences in how migrations destinations are

⁵ For example, in Tanzania, Mueller et al. (2019) focus on the consequences of migration to peri-urban areas; Christiaensen, De Weerd, and Kanbur (2019) focus on the consequences of migration to secondary towns.

defined can promote misleading conclusions. Wineman, Alia, and Anderson (2020) compare how seven alternative definitions of urban and rural areas influence the calculated levels of urbanization and economic indicators. Hence, to examine the robustness of destination definitions, I estimate migration models based on three approaches to the categorization of locations: (1) the one based on the definition used by the Tanzanian National Bureau of Statistics (NBS); (2) the one I construct based on the population density, the built-up area density⁶, and the distance to the nearest town; and (3) the one I construct using cluster analysis based on various household and community characteristics aggregated as a district level.

2.2. Literature review

2.2.1. Migration decisions and migration destination decisions

Two common analytical approaches for examining migration decisions have been widely used in older literature on developing countries. The first one presents migration as a binary decision, to move or to stay in place, without distinguishing destinations. Partly, this approach became popular because of the data limitations: in surveys, unless it directly targeted migrants, people were rarely interviewed on their individual migration history. In most cases, the only information gathered indicated if a person was born in the area of current residence, which is not helpful for migration studies (Lucas, 1997). In some other cases, studies neglected the contextuality of migration decision narrowing it to the simple binary case (Lucas, 1997). This could be related to the second approach to viewing migration decisions. That is, to assume that migration originates in rural areas and that migrants' destinations are urban.

⁶ I use the data on built-up area density from the Global Human Settlement Layer (Corbane et al., 2018). It shows the share of land under buildings in the total size of the cell. I use one km grid for cells.

Rural-to-urban migration has received a lot of attention in the literature due to its contribution to structural transformation through the shifts of labor from agriculture to manufacturing and services that often accompany such movements (de Brauw, Mueller, and Lee, 2014). The two-sector model of rural-to-urban migration introduced by Harris and Todaro (1970) explains how differences in expected earnings between rural and urban areas stimulate migration that “not only continues to exist, but indeed, appears to be accelerating” [p. 126]. Hence, many studies on migration in developing countries strived to explain the patterns of rural-to-urban migration ignoring the continuum of the spectrum of choices and the fact that other forms of migration may have become more important in recent years as Africa’s transformation process has accelerated.

The reasons for an increased interest in the rural-to-rural migration may vary greatly. In recent years, migration studies tend to look beyond the rural-to-urban migration flow. Reed, Andrzejewski, and White (2010), for example, examine whether the drivers of inter-regional rural-to-rural and rural-to-urban migration differ in Ghana. Msigwa and Mbongo (2013) use multinomial logistic regression to distinguish rural-to-urban and town-to-city migration flows. Several studies look at a variety of other destinations. For example, Hirvonen (2016) uses multinomial logistic regressions to distinguish destinations within and outside the district of migrant’s origin. Mueller et al. (2019) distinguish peri-urban and urban destinations.

I observe the majority of rural migrants moving to another rural area, consistent with Lucas (2016), and try to identify the main factors associated with this choice. Ingelaere et al. (2018) discuss the continuum of destination choices that rural people have and stress the importance of a familiar atmosphere at destination that helps migrants to adapt to a new location better. Ingelaere et al. (2018) also focus on urban destinations and find that rural migrants tend to

prefer smaller towns as destinations for the reasons mentioned above. In addition to that, such locations are often closer to the migrants' origin, and it would be easier to return in case they cannot settle in a new location. Rural destinations also provide a familiar atmosphere and some form of a safety net, which might attract youth.

Theory about the determinants of migration decision usually classifies factors affecting this decision into “push” and “pull” factors (Bilsborrow et al., 1987). Some studies distinguish “rural push” and “rural pull”, “urban push” and “urban pull” factors (Jedwab, Christiaensen, and Gindelsky, 2017). The choice of factors depends on the research questions of each particular study: there are examples of environmental factors, land pressure, household composition, and other factors. Some recent studies expand the distinction between push and pull factors by looking at the factors that restrain migration (Dustmann and Okatenko, 2014) and by investigating the core reasons behind the migration decision (Lucas, 2016).

Some studies on the reason for migration separate refugees who flee from wars and violence, land grabs or environmental shocks or consequences of climate change (Sassen, 2016). Weather shocks are extremely important for rural population as they could critically affect income and therefore influence the migration decision (Bohra-Mishra, Oppenheimer, and Hsiang, 2014). Marchiori et al. (2012) describe how weather anomalies could force rural-to-urban migration and then international migration. Gray and Bilsborrow (2013) study how a range of environmental factors, such as access to irrigation, land quality, topographic slope, mean annual rainfall and its seasonality and shocks, affect local, internal, and international migration or make people trapped in place.

In Table 2.25 in Appendix 6, I present a list of explanatory variables employed in a sample of studies on migration decision and migration destination decision that use methods

similar to the ones I do. Among the common individual-level control variables are gender, age, marital status, education, occupation, and migration history (individual's own moves, moves by household members); among the household-level variables are household size, household wealth (asset index), and amount of cultivated land; and among the community-level variables are distance to various facilities (town, road, hospital, primary school). The results showing the impact of these factors on migration decision and destination decision vary across countries and contexts. I further compared them to my results in the Discussion section.

Beegle and Poulin (2013) and Bernard, Bell, and Charles-Edwards (2014) show that relocation decisions relate to education, marriage, employment, and other life events, and may seriously affect young people's livelihoods. Transition to adulthood itself for many is associated with moving away from the community of their parents. Although migration has been a powerful means to improve one's living conditions, sometimes young people could be pushed into migration under duress, for example, by traditional marriage agreements (Kudo, 2015), conflict (Wondimagegnhu and Zeleke, 2017), environmental shocks (Hirvonen, 2016), and deterioration of the local natural resource base (Epule, Peng, and Lepage, 2015).

Access to land is also an important factor that rural youth consider when making their decision to stay or to move. It may be of lesser importance in areas of Tanzania with relatively low population pressure and higher amount of land available for agriculture (Proctor and Lucchesi, 2012). On the other hand, administrative barriers may prevent youth from obtaining land in their home region (Bezu and Holden, 2014). There could be several other issues related to the youth' choice of migration destination, with apparently low number of young people in East Africa aspiring to be predominantly farmers (Proctor and Lucchesi, 2012) as well as employment challenges that youth faces (Fox and Thomas, 2016). Fox and Thomas (2016) suggest the

development of rural employment (both on-farm and off-farm) as a measure to keep rural youth from moving away. Hence, youth can be considering moving to other rural areas for employment as well as for land.

2.2.2. Definitions of “urban” and the continuum of locations on the rural-urban spectrum

Urbanization in Sub-Saharan Africa is a widely discussed topic and a key term in policy initiatives, but the analysis of urbanization patterns is troubled by the lack of a common definition of “urban” (for example, see the list of definitions of “urban” by country provided by the UN, 2008⁷). Although “a universal definition is probably neither possible nor desirable” (Potts, 2017b, p. 967), and the economic dynamism and occupational transition could drive changes in the definitions over time, in some cases researchers aim for comparisons across countries or periods in time. Wenban-Smith (2015) describes urbanization patterns in Tanzania using five censuses. He employs the census definitions of “urban” and notes how inconsistencies between these definitions across years, along with changes in administrative division, affect the observed patterns. Potts (2017a) shows how flaws in the definition of “urban” can have a significant impact on the observations and conclusions I make regarding the urbanization patterns. She also advocates the use of multiple criteria to define “urban” in order to avoid misleading conclusions.⁸

⁷ Some of the definitions listed are based on the amount of people living in the settlement, and the thresholds vary by country, for example: 200 people in Norway, 2,500 people in Bahrain, 20,001 people in Turkey (UN, 2008). With these definitions, there arises a question of identifying the borders of the settlements, which is usually a difficult task in itself, and especially it troubles the analysis of the dynamics of urbanization as settlement borders change over time. Some other definitions employ population density, but, again, the threshold depends on the country: for example, 400 people per square km in Canada, 1500 people per square km in China (UN, 2008).

⁸ The criteria listed by Potts (2017a) are settlement’s form, its function, production, and labor specialization. For the first definition for the locations on the rural-urban spectrum that I construct, I rely on the form of the settlement (using the population density and the share of built-up area). Hence, my results for migration from “rural” to “more urbanized” places should not be interpreted as migration from an area predominantly involved in farming to an area predominantly involved in non-farm activities. In fact, they mostly indicate migration to a more densely populated place. I find that the share of land under agriculture, which could indicate production

Cockx, Colen, and De Weerd (2018) study the effects of migration from rural to urban locations on the diet in Tanzania. They start with a binary division into rural and urban areas and then distinguish migration to secondary towns from migration to cities. In addition, they digress from administrative division into rural/urban locations and use population density to verify the robustness of their results. With the same dataset, Mueller et al. (2019) construct their own definitions of “urban”, “peri-urban”, and “rural” to study occupational transitions of migrants. They find that migration from peri-urban to urban areas is three times as high as migration from rural to urban areas for long-distance moves, which shows how the binary classification might exaggerate rural-to-urban migration if peri-urban areas are counted as rural. The definitions used by the authors are based on population density, distance (in travel time) to the nearest town, and the share of built-up area. I employ a very similar approach to construct my first definition for the locations on the rural-urban spectrum.

While discussing the concepts of rural-urban spectrum and urban hierarchy, Potts (2017b) notes how current analysis tends to use the traditional binary rural-urban relationship due to several reasons, including the aim of a published work to be understood by people “beyond academe” (Potts, 2017b, p. 968, p. 982). Hence, I start my analysis with the rural/urban definition developed by the Tanzanian National Bureau of Statistics, which is commonly used by both researchers and policymakers. On the other hand, the attention to the concepts like “rurban”⁹ is on the rise (Iaquinta and Drescher, 2000; Mercandalli et al., 2017). Models that include a variety of migration destinations are also becoming more common in recent literature.

specialization, is not reliable for classification purposes. For the second definition, I conduct cluster analysis which incorporates a more diverse set of variables that cover not only the form of the settlement but also infrastructure, access to amenities, and labor specialization.

⁹ The term “rurban” is often used to define an area that has many characteristics of a town, although has some portion of its land utilized for farming. With this definition, some peri-urban areas and some urbanized villages are classified as rurban. Some authors use distance to the city center to define rurban, making the new definition identical or close to the one for peri-urban areas.

One of the examples of such differentiation for internal migration is provided by the studies on the benefits of migration.

Beegle, De Weerd, and Dercon (2011) show that people from the Kagera region of Tanzania who migrated internally within the country benefited from their move: a decrease in the poverty rate and an increase in the consumption growth, even among people moving within the same region, exceeded those of non-migrants. The authors also find that the benefits from migration were, on average, higher for people moving to areas closer to urban centers, while even those who moved to remote villages experienced more benefits than those who stayed in place. Growing literature on the role of secondary towns (Christiaensen and Todo, 2014; Christiaensen and Kanbur, 2017) shows that, even though migration is, on average, immensely beneficial for those who migrate to urban areas, the benefits are not distributed equally. People moving to cities experience larger improvements in livelihood, but secondary towns play a very important role in poverty reduction as they are more accessible for migrants from rural areas that constitute the majority of migration's origins (Ingelaere et al., 2018).

2.3. Data and definitions

I use three waves of the Living Standards Measurement Study (LSMS) in Tanzania (World Bank, 2017). The first wave of this individual panel survey was conducted in 2008/2009 and contained a sample of 3,265 households. Subsequent rounds were implemented in 2010/2011 (wave 2) and 2012/2013 (wave 3).¹⁰ People of age 15 and older who moved within the country

¹⁰ An additional wave of survey was conducted in 2014/2015. It extends the panel for 784 households interviewed in 2008/2009-2012/2013 and starts a new sample of 3,352 households. An interesting avenue for future research would be to look at how migration trends change over time, comparing the patterns I observe for the 2008/2009-2012/2013 panel to the patterns seen in the new panel starting in 2014/2015.

between the survey waves were tracked and interviewed in the subsequent waves. International migrants were not tracked.¹¹

I look at people who resided in rural areas at baseline and study their migration decisions and migration destination decisions that realized by the last wave of survey. For the main analysis, I use the NBS definition of “rural” and the first constructed definition of “rural”. The construction of this definition is summarized in Table 2.1 and discussed in more detail in Appendix 3. I use information on population density with one km grid in 2010 from WorldPop Africa Continental Population Databases (Tatem, 2017). The data on the built-up area density with one km grid in 2013/2014 come from the Global Human Settlement Layer (Corbane et al., 2018)¹². As a result, for the constructed definition I identify households as living in a rural area if: (i) they are located further than 30 km away from any town with population of at least 50,000 and have population density below 400 people per sq. km or built-up area density below 8%, or (ii) they are located within 30 km of a town with population of at least 50,000 and have population density below 150 people per sq. km.

¹¹ For the year 2010, the United Nations estimates the international migrant stock to be 247 thousand people (United Nations, Department of Economic and Social Affairs, 2013; Table 7; country of origin: United Republic of Tanzania). The population of Tanzania in 2010 was 44.3 million people, hence the share of international migrants is 0.6%. Based on the LSMS dataset, the share of internal migrants in the sample is 10.5% (calculated for people of all ages who were present both in Wave 1 and Wave 3, using sampling weights from Wave 1).

¹² I use coordinates provided in the LSMS dataset to match households to population density and built-up area density data. In Appendix 1, I discuss how I retrieve missing coordinates and missing population density when the coordinates point to water bodies.

Table 2.1. Main definition constructed for the locations on the rural-urban spectrum

Distinction	Construction of the definition of a city	Construction of the definition of a town	Construction of the definition of a peri-urban area	Construction of the definition of a rural area with high population density	Construction of the definition of a rural area with low population density
How I distinguish urban areas from non-urban areas	<i>Urban area</i> is defined as an area with population density above 400 people per sq. km and built-up area density above 8%		<i>Non-urban area</i> is defined as an area with either population density below 400 people per sq. km or built-up area density below 8%		
Among urban areas, how I distinguish cities from towns	<i>City</i> is defined as an <i>urban area</i> located within 30 km of Dar es Salaam or Mwanza	<i>Town</i> is defined as an <i>urban area</i> located further than 30 km away from Dar es Salaam and Mwanza			
Among non-urban areas, how I distinguish peri-urban areas from rural areas			<i>Peri-urban area</i> is defined as a <i>non-urban area</i> that (i) is located within 30 km of a town with population density of at least 50,000, and (ii) has population density above 150 people per sq. km	<i>Rural area</i> is defined as a <i>non-urban area</i> that (i) is located further than 30 km away from any town with population of at least 50,000, or (ii) is located within 30 km of a town with population of at least 50,000 and has population density below 150 people per sq. km	
How I split rural areas by population density				<i>Rural area with high population density</i> is defined as a <i>rural area</i> with population density above 100 people per sq. km	<i>Rural area with low population density</i> is defined as a <i>rural area</i> with population density below 100 people per sq. km

Note: Data on population density is from WorldPop Africa Continental Population Databases (Tatem, 2017), for 2010; data on built-up area density is from Global Human Settlement Layer (Corbane et al., 2018), for 2013/2014. For both databases, I use the versions with one km grid and match them to the households' coordinates provided in the LSMS. When the coordinates point to a body of water, I replace them with the closest grid cell of land. For each household, I compute distance from the household's location to Dar es Salaam, Mwanza, and other towns with population of at least 50,000 people (based on the 2012 Population and Housing Census) using the households' coordinates provided in the LSMS and the coordinates of town centers that I collected myself from various sources (usually, coordinates point to cross-roads involving the main road(s): see Appendix 3 for the full list of coordinates for towns).

For the majority of observations, I base my definition of “migrant” on the distance between the 2008/2009 and the 2012/2013 waves provided in the dataset: an individual is considered to be a migrant if this distance is over five km.¹³ This definition implies that I observe migration over the four years between the first and the last survey wave.¹⁴ In the dataset, the distance is missing for a small number of individuals, and I compute distance for them with the given coordinates and apply the same threshold of five km for consistency. In Appendix 2, I discuss how distance provided in the dataset corresponds to the computed distance and other parameters that indicate migration. I find the threshold of 0.1 km for the computed distance to indicate that the individual resides in a different place, but I cannot tell if it was a short-distance move or if the observed distance is noisy because of the aggregation of coordinates at the level of enumeration area and the offset.

Among urban areas, I distinguish towns and cities. In addition, the first constructed definition of “rural” allows me to separate peri-urban areas. I distinguish Dar es Salaam and Mwanza as cities, same as Cockx, Colen, and De Weerd (2018). For the constructed definition of “rural”, I use a 30 km threshold for the distance to city to identify cities, towns, and peri-urban areas. The choice of threshold is based on the patterns of distance I observe in the dataset (see Appendix 3, Figure 2.7 and Figure 2.8). To define peri-urban areas, I use data on the distance to

¹³ The threshold of five km is set to follow the survey’s criteria for tracking. People who moved within five km of their original location were not tracked. With the same dataset, Cockx, Colen, and De Weerd (2018) use travel time to the new location to define migrants, with a threshold of one hour. Mueller et al. (2019) use distance to the new location in km and test four thresholds: one km, 10 km, 20 km, and 50 km (only 20 km and 50 km are used for their main analysis). In section 2.5.3, I test whether my results are robust to the definition of “migrant” used. One of the definitions I try is the NBS definition which only considers between-district movements as migration. This definition indirectly eliminates short-distance movements, as the median distance traveled is 13 km for within-district moves (mean is 23 km). Median distance traveled for between-district moves is 138 km (mean is 215 km).

¹⁴ I do not use the 2010/2011 wave for the definition of “migrant”, with the exception for checks described in Appendix 2. This is done due to the concerns of a sharp decrease in the number of observations of migrants to certain destination types. As the main goal of this study is to introduce a more detailed categorization of destination types on the rural-urban spectrum, the loss of precision overwhelms the benefits of using the panel structure of the data.

town, population density, and built-up area density. I use similar variables as Mueller et al. (2019) do, but some of my thresholds are different. I opt for using distance to larger towns and measure it in kilometers instead of travel time.¹⁵ Also, my threshold for the built-up area density is lower: 8% instead of 50%, which classifies more households as urban. Following Mueller et al. (2019), I exclude households that are located within 30 km from a town and have low population density from the list of urban and peri-urban households.

Since the most commonly used definition of “rural” is the one set by the government (Potts, 2017b), I start with the NBS definition but transform it to include more categories. For the differentiation of rural areas into areas with low and high population density, I use the same threshold, 100 people per sq. km, as I use for my constructed definition. To differentiate towns and cities, I look at the district: urban households in all districts of Dar es Salaam and in Nyamagana and Ilemela districts in Mwanza region are categorized as living in cities, and all other urban households are categorized as living in towns. I cannot separate peri-urban areas within the government’s binary division into rural and urban. For the comparison of the constructed definition and the definition based on the NBS categorization, see Table 2.24 in Appendix 3. Locations defined as “rural” under the constructed definition but not the NBS categorization have low average population density and built-up area density; locations defined as “rural” under the NBS categorization but not the constructed definition have high average population density and built-up area density.¹⁶

¹⁵ Mueller et al. (2019) use distance to town with population of at least 20,000 people, I use distance to town with population of at least 50,000 people. In Appendix 3, I describe the difference between these two variables (in particular, see Figure 2.9). Also, Mueller et al. (2019) set a threshold for one hour of travel time for both urban and peri-urban areas, though they do not specify what type of travel does this measure describe. My threshold is 30 km radius from the city center.

¹⁶ For example, the location of 198 households is identified as “town” by the NBS categorization and “low-density rural area” by the main constructed definition. For these households, mean population density is 57 people per sq. km, mean built-up area density is 0.0, and mean share of income coming from farming is 30%. These characteristics are similar to the characteristics of households defined as living in low-density rural areas under both

For a robustness check, I use cluster analysis to build another definition of “rural”. The details of its construction and the comparison to the other two definitions are provided in Appendix 3. I use standardized values of district averages for access to amenities (flooring materials, time to get water), involvement in agriculture (based on the share of household income received from agriculture), and other characteristics (population density, distance to road, and distance to town with population of at least 50,000 people). I run 125 iterations of k-medians algorithm with different random starting points and use adjusted Rand index to select one partition.

For the first wave of survey, cluster analysis provides an optimal division into two groups that I consider to be “rural” and “urban”. I look at the sample of individuals from rural areas according to this definition. For the last survey wave, cluster analysis is not decisive on the number of groups, hence I gradually increase the number of groups that I split the destination regions into. The definitions based on cluster analysis do not distinguish urban areas further than the split into cities, towns, and peri-urban areas. At the same time, I find that rural areas can be differentiated further based on averages of population density, distance to town, distance to road, and share of household income coming from agriculture.

Out of 16,709 individuals surveyed in Wave 1, 14,795 were re-surveyed in Wave 3.¹⁷ The summary on attrition is presented in Appendix 4. I take characteristics from Wave 1 as baseline characteristics. The summary of these variables for youth from rural areas is presented in Table

categorizations. On the other hand, the location of 189 households is identified as “high-density rural area” by the NBS categorization and “town” by the constructed definition. For these households, mean population density is 3,022 people per sq. km, mean built-up area density is 0.33, and mean share of income coming from farming is 27%. These characteristics are similar to the characteristics of households identified as living in towns under both categorizations (except for the average share of income coming from farming, which is lower in areas both definitions agree to identify as “towns”).

¹⁷ 218 of them were surveyed in both Wave 1 and Wave 3 but not in Wave 2. 1,020 individuals out of 16,709 were surveyed in both Wave 1 and Wave 2 but not in Wave 3.

2.2. I define youth as people of age from 15 to 34.¹⁸ The average age in the sample is 23 years, median age is 22 years. The majority of individuals, 64%, completed primary school; many individuals, 45%, are married. They are more likely to be children of a household head (42%) than to be heads of a household themselves (18%). Some individuals have prior migration history: 21% were born outside the village of residence, and 10% were away from the household for at least a month in the past year.

The majority of individuals, 67%, had main occupation in farming or fishing in the past 12 months. Most individuals, 86%, live in households with area under cultivation at or above one acre. With the threshold for land area cultivated by smallholder farms computed by Rapsomanikis (2015) for Tanzania, 5.44 acres, 70% of individuals in the sample live in small farm households. Although the mean number of units of livestock¹⁹ owned by rural households with youth is large (3.47), the median is much smaller (0.23). Median age of the household head is 43. Most heads of the household are male (80%). Average household size is 6.8, median is 6.

¹⁸ For additional analysis, I split the sample into two groups: people of age 15-24, to whom I refer to as a “younger cohort”; and people of age 25-34, to whom I refer to as an “older cohort”.

¹⁹ I use Tropical Livestock Units for the number of livestock owned by the household at the day of the interview. Animals with a coefficient of 0.5: bulls and cows (steers and heifers, male and female calves), horses. Animals with a coefficient of 0.3: donkeys. Animals with a coefficient of 0.2: pigs. Animals with a coefficient of 0.1: goats and sheep. Animals with a coefficient of 0.01: chickens, turkeys, and rabbits.

Table 2.2. Summary statistics for youth living in rural areas at baseline (2,803 observations)

	Mean	Std. dev.	25 th percentile	Median	75 th percentile
Age	22.99	5.91	18.00	22.00	28.00
1 = Male	0.49	0.50			
1 = Completed primary school	0.64	0.48			
1 = Married	0.45	0.50			
1 = Head of the household	0.18	0.38			
1 = Child of household head	0.42	0.49			
1 = Born in this village	0.79	0.41			
1 = Was away from the household for at least one month in the past 12 months	0.10	0.30			
1 = Main occupation in farming or fishing in the past year	0.67	0.47			
Area under cultivation, acres	6.54	18.83	1.50	3.50	6.00
Livestock (TLU)	3.47	13.48	0.03	0.23	2.20
Age of household head	44.66	15.12	32.00	43.00	56.00
1 = Household head is male	0.80	0.40			
Number of working age women	1.81	1.33			
Number of working age men	1.84	1.39			
Number of children of household head living in the household	3.33	2.47			
1 = Household experienced agricultural shock in the past year	0.28	0.45			
1 = Household experienced non-agricultural shock in the past year	0.29	0.45			
Population density, people per square km	100.55	147.52	36.74	72.09	116.17
Distance to road, km	21.35	20.19	6.10	17.50	28.70
Distance to the nearest town with population of at least 50,000, km	67.34	39.45	37.25	61.57	87.03

Note: Rural areas are defined using the constructed definition that is described in Table 2.1. Sampling weights from the 2008/2009 survey wave are applied. Data on population density is from WorldPop Africa Continental Population Databases (Tatem, 2017). Data on distance to road is from the LSMS: it is computed by the survey team using the real coordinates of the households (real coordinates are not provided in the LSMS). Data on the distance to the nearest town with population of at least 50,000 km is computed by the author using households' coordinates provided in the LSMS (the survey team aggregated households' coordinates by enumeration area and added a random offset up to 10 km) and the towns' coordinates listed in Appendix 3.

Agricultural and non-agricultural shocks are self-reported shocks that severely affected the household: three most severe shocks were recorded. I select shocks that occurred in the past year (relative to the interview) and caused a loss of either income or assets. I define the following events as agricultural shocks: drought or floods, crop diseases or crop pests, livestock died or was stolen, large fall in sale price for crops, large rise in agricultural input prices, severe water shortage, and loss of land. On average, 28% of individuals live in households that experienced an agricultural shock in the past year. I define the following events as non-agricultural shocks: household (non-agricultural) business failure, loss of salaried employment or non-payment of salary, large rise in the price of food, chronic or severe illness or accident of household member, death of a member of a household, death of other family member, break-up of the household, household member jailed, fire, hijacking, robbery, burglary, assault, dwelling damaged or destroyed, and shocks reported as “other”. On average, 29% of individuals live in households that experienced a non-agricultural shock in the past year.

2.4. Empirical strategy

I start by looking at the factors associated with a binary migration decision (to migrate or not to migrate) of an individual living in a rural area, and then build a series of more detailed partitions of migration destinations along the rural-urban continuum. I use logistic regression model following the specification similar to Zhang et al. (2018). I distinguish individual, household, and area characteristics that can be associated with migration decision (Bilsborrow et al., 1987). The model takes this form:

$$P(M_i = 1|X_{in}, X_{ih}, X_{ic}) = \Lambda(\beta_0 + X_{in}\beta_n + X_{ih}\beta_h + X_{ic}\beta_c) \quad (1)$$

Here, M_i represents the migration decision made by an individual i . M_i equals one if the individual moved between the first and the last survey waves and equals zero if the individual stayed at the origin:

$$M_i = \begin{cases} 1, & \text{individual } i \text{ moved,} \\ 0, & \text{individual } i \text{ stayed in place} \end{cases}$$

I consider the individual to be a migrant if I observe this individual to settle in a location other than the origin at the last survey wave.²⁰ I consider the individual to be a non-migrant if I observe this individual in the same location at the last survey wave as I did at the first survey wave. Hence, if this individual moved between the first and the last survey waves but returned back to the origin by the last survey wave, I consider this individual to be a non-migrant. I am not able to observe full migration history, so sequential migration is not separated from one-time migration. Using the second wave of survey, I confirm that only 2.5% of young people who moved between the first and the second survey waves returned to their original households by the last survey wave.

In equation 1, Λ is a logistic function, X_{in} is a vector of individual-level characteristics, X_{ih} is a vector of household-level characteristics, and X_{ic} is a vector of community-level characteristics. In addition, I distinguish six geographical zones (Coastal, Northern Highland, Lake, Central, Southern Highland, and Zanzibar: see Appendix 5) and apply zone fixed effects to account for unobserved heterogeneity across zones.

I use multinomial logistic regressions to determine if the observed individual, household, and community characteristics associate with migration to various destination types in different ways. The model could be rewritten as follows:

²⁰ In section “Data and definitions”, migrants are defined based on the distance traveled between the survey waves. Since the survey team treated any distance below five km as zero, “the origin” in the formal definition in this section is any location within five km of the location recorded during the first wave of survey.

$$P(M_i = D|X_{in}, X_{ih}, X_{ic}) = \frac{\exp(\beta_0^D + X_{in}\beta_n^D + X_{ih}\beta_h^D + X_{ic}\beta_c^D)}{1 + \sum_{k=1}^K \exp(\beta_0^k + X_{in}\beta_n^k + X_{ih}\beta_h^k + X_{ic}\beta_c^k)} \quad (2)$$

$$P(M_i = 0|X_{in}, X_{ih}, X_{ic}) = \frac{1}{1 + \sum_{k=1}^K \exp(\beta_0^k + X_{in}\beta_n^k + X_{ih}\beta_h^k + X_{ic}\beta_c^k)}$$

In this equation, D stands for a specific destination along the rural-urban spectrum while staying in place is chosen to be a pivot outcome ($M_i = 0$). The variations of the model with different number of destinations, K , are presented in Table 2.3.

Table 2.3. Variations of the model: values assigned to the dependent variable M for the individual i , $M_i=D$, that indicate a discrete destination type D , for the total number of K migration destination types – for the main constructed definition for locations on the rural-urban spectrum

Model	$D = 0$	$D = 1$	$D = 2$	$D = 3$	$D = 4$	$D = 5$
Logistic regression	Stayed in place	Moved				
Multinomial logistic regression ($K = 2$)	Stayed in place	Moved to a rural area	Moved to an urban area			
Multinomial logistic regression ($K = 3$)	Stayed in place	Moved to a rural area	Moved to a peri-urban area	Moved to a town or a city		
Multinomial logistic regression ($K = 4$)	Stayed in place	Moved to a rural area	Moved to a peri-urban area	Moved to town	Moved to city	
Multinomial logistic regression ($K = 5$)	Stayed in place	Moved to a rural area with low population density	Moved to a rural area with high population density	Moved to a peri-urban area	Moved to town	Moved to city

I keep partitioning migration destinations from the binary rural-urban dichotomy to a broader set of destination types. Depending on the definition of “rural” that I use, the highest

number of destinations range from four (with the NBS definition) to six (with cluster analysis definition).

2.5. Results

2.5.1. Summary statistics

I present migration rates over the observed period, from 2008/2009 to 2012/2013, in panel A of Table 2.4. Using the constructed definition for the location types on the rural-urban spectrum, I categorize both origin (into low-density and high-density rural areas) and destination areas (into five categories). On average, 16.2% of rural youth moved between the first and the last waves of survey, which can be interpreted as an annual migration rate of around 4%. I see that youth from high-density rural areas are, on average, more mobile than youth from low-density rural areas. The choice of destination varies between these two types of origin as well. While rural destinations are pursued by the majority of migrants regardless of their origin, the share of people moving to rural areas is higher among migrants from low-density rural areas (71%) than among migrants from high-density rural areas (62%). Among urban destinations, migrants from low-density rural areas are, on average, more likely to have chosen cities whereas migrants from high-density rural areas are more likely to have chosen peri-urban areas.

Table 2.4. Migration rates of rural youth, by origin and destination

	Stayed in place	Moved to a low-density rural area	Moved to a high-density rural area	Moved to a peri-urban area	Moved to a town	Moved to a city
<i>A. Constructed definition of “rural”</i>						
Youth from low-density rural areas (1,832 obs.)	84.9%	7.9%	2.9%	1.4%	1.1%	1.8%
Youth from high-density rural areas (971 obs.)	81.2%	5.5%	6.2%	2.9%	2.3%	1.9%
Total (2,803 obs.)	83.8%	7.1%	3.9%	1.8%	1.5%	1.8%
	2,364 obs.	194 obs.	103 obs.	48 obs.	40 obs.	54 obs.
<i>B. NBS definition of “rural”, with rural areas split into low- and high-density areas and urban areas split into towns and cities as described in section 3</i>						
Youth from low-density rural areas (1,695 obs.)	84.8%	7.7%	2.8%	undefined	2.7%	2.0%
Youth from high-density rural areas (1,162 obs.)	82.9%	5.3%	5.2%	undefined	4.4%	2.2%
Total (2,857 obs.)	84.1%	6.8%	3.7%	undefined	3.3%	2.1%
	2,423 obs.	183 obs.	100 obs.		87 obs.	64 obs.

Note: Each row sums to 100%. Sampling weights from the 2008/2009 survey wave are applied.

With the NBS definition for the location types, the general patterns are consistent (see panel B of Table 2.4). Notice 69% of migrants from low-density rural areas and 61% of migrants from high-density rural areas chose rural destinations. The results diverge between definitions when I look at a more nuanced distinction of destinations: with the absence of peri-urban category in the NBS definition, migrants are categorized as moving to towns more frequently. Note that it is not as simple as sorting into different destination groups, because the definition for

the location type changes for the origin areas as well as for the destination areas. These changes have several consequences for the sample. For example, some people from peri-urban areas according to the constructed definition are included into the sample as living in rural areas under the NBS definition.²¹ Their choices of migration destinations might differ from the choices of people from areas classified as rural under both definitions: I discuss these issues in the beginning of subsection on the NBS categorization of “rural”.

For the remainder of this subsection, I use the constructed definition for the locations on the rural-urban spectrum. In Figure 2.1, I present the distribution of age by migration status and gender. In graph (a), I show the distribution for non-migrants: there are more men in younger cohorts and more women in older cohorts. In graph (b), I present the distribution for the binary destination choice, and, in graphs (c) and (d), I look at the distribution for all five of the discussed locations on the rural-urban spectrum. There are many more women moving to rural areas than men. This holds for both low-density and high-density rural destinations and is more pronounced among younger cohorts. For additional analysis, I build these graphs distinguishing low-density and high-density rural origins (graphs not provided here). I see that migration flows from the two origin types are alike except for the migration flows from one low-density rural area to another, where the share of migrants of age 15-20 is higher.

²¹ See Appendix 3, and Table 2.23 in particular, for the comparison between the constructed definition and the NBS categorization. The origin areas of 12.4% of migrants from rural areas according to the NBS definition are classified as peri-urban areas under the constructed definition. Hence, they are not included in the main sample when I use the constructed definition but are included in the main sample when I use the NBS categorization.

Figure 2.1. Rural youth, by gender, age in 2008/2009, and destination in 2012/2013 (origin and destination are defined according to the constructed definition of “rural”)

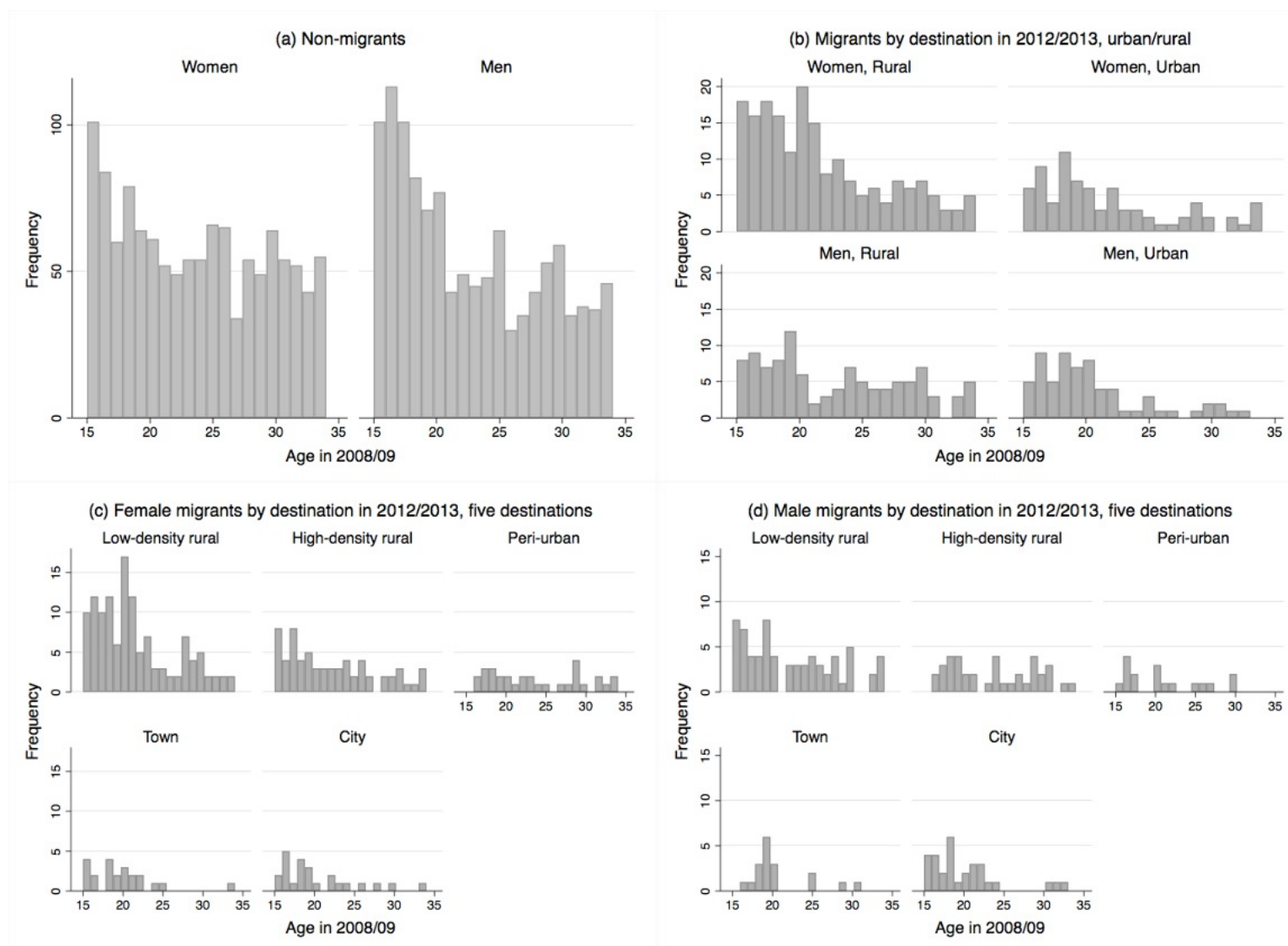


Table 2.5. t-test for the difference in means of key variables between non-migrant rural youth and migrant rural youth (column b); between migrants to rural and urban areas (column d)

	(a) Stayed in place	(b) Moved	(c) Moved to rural	(d) Moved to urban
Age	23.24	21.75***	22.00	21.20
1 = Male	0.51	0.40***	0.37	0.47**
1 = Completed primary school	0.63	0.66	0.60	0.79***
1 = Married	0.48	0.33***	0.36	0.26**
1 = Head of the household	0.19	0.12***	0.13	0.12
1 = Child of household head	0.42	0.42	0.42	0.42
1 = Born in this village	0.82	0.66***	0.64	0.70
1 = Was away from the household for at least one month in the past 12 months	0.09	0.16***	0.14	0.21*
1 = Main occupation in farming or fishing in the past year	0.69	0.57***	0.66	0.37***
Area under cultivation, acres	6.36	7.44	9.16	3.77**
Livestock (TLU)	3.41	3.76	4.65	1.84*
Age of household head	44.39	46.00**	46.05	45.89
1 = Household head is male	0.81	0.78	0.82	0.69***
Number of working age women	1.77	2.04***	2.03	2.06
Number of working age men	1.82	1.90	1.88	1.93
Number of children of household head living in the household	3.35	3.28	3.42	2.99*
1 = Household experienced agricultural shock in the past year	0.29	0.24**	0.22	0.28
1 = Household experienced non-agricultural shock in the past year	0.29	0.31	0.28	0.37**
Population density, people per square km	97.96	113.95**	105.01	132.98*
Distance to road, km	21.48	20.69	22.26	17.34**
Distance to the nearest town with population of at least 50,000, km	67.33	67.40	70.98	59.76***
Number of observations	2364	439	297	142

Note: Rural and urban areas are defined using the constructed definition. Sampling weights from the 2008/2009 survey wave are applied. Column (b): stars indicate significant difference in means between migrants and non-migrants; column (d): stars indicate significant difference in means between migrants to urban areas and migrants to rural areas; *** 0.01; ** 0.05; * 0.1.

A summary of key variables differentiated by binary migration status is shown in Table 2.5. Non-migrants (column a) and migrants (column b) have diverse individual and household characteristics: for example, the share of women among migrants is significantly higher than among non-migrants, and migrants are less likely to have main occupation in farming or fishing than non-migrants. A simple binary distinction of destinations into rural and urban areas presented in columns (c) and (d) of Table 2.5 reveals some information masked by the simplification of migration choice modeling.

From the example above, migrants are, on average, less likely than non-migrants to have main occupation in farming or fishing. Differentiating destination types, I confirm this pattern only for migrants to urban areas. On the other hand, the share of people with main occupation in farming or fishing among those who moved to another rural area is almost as high as that among non-migrants. As for the observed gender patterns, I find the share of women who migrate to rural areas to be much higher than the share of women who do not migrate. There is no significant difference in gender between non-migrants and people who moved to an urban area.

I present summary statistics for migrants to destinations that are further differentiated along the rural-urban spectrum in Table 2.6. I include peri-urban areas, towns, and cities into the “urban” category when I use binary division. Now, I see almost the same share of women among migrants to peri-urban and rural areas, which is higher than the female share of migrants to towns and cities and non-migrants. Also, people moving to cities are, on average, much less likely to have farming or fishing as their main occupation at baseline than people moving to peri-urban areas and towns.

Table 2.6. Means of key variables for migrants by destination, five destination types

	(a) Moved to low- density rural area	(b) Moved to high- density rural area	(c) Moved to peri- urban area	(d) Moved to town	(e) Moved to city
Age	21.67	22.62	23.19	20.12	20.08
1 = Male	0.35	0.39	0.37	0.50	0.53
1 = Completed primary school	0.56	0.67	0.72	0.86	0.79
1 = Married	0.36	0.36	0.29	0.29	0.20
1 = Head of the household	0.13	0.12	0.18	0.14	0.03
1 = Child of household head	0.42	0.42	0.29	0.51	0.47
1 = Born in this village	0.63	0.65	0.64	0.72	0.73
1 = Was away from the household for at least one month in the past 12 months	0.13	0.16	0.17	0.20	0.25
1 = Main occupation in farming or fishing in the past year	0.67	0.63	0.48	0.41	0.23
Area under cultivation, acres	9.66	8.24	3.52	4.79	3.18
Livestock (TLU)	6.30	1.65	1.50	2.05	2.02
Age of household head	45.80	46.52	43.92	46.15	47.67
1 = Household head is male	0.86	0.73	0.67	0.72	0.68
Number of working age women	2.16	1.79	1.81	2.34	2.08
Number of working age men	2.00	1.67	1.48	1.82	2.48
Number of children of household head living in the household	3.61	3.07	2.16	3.73	3.21
1 = Household experienced agricultural shock in the past year	0.27	0.14	0.27	0.20	0.35
1 = Household experienced non-agricultural shock in the past year	0.30	0.24	0.28	0.46	0.40
Population density, people per square km	82.29	146.59	143.65	104.56	145.61
Distance to road, km	24.05	18.98	17.49	17.72	16.88
Distance to the nearest town with population of at least 50,000, km	74.07	65.34	47.61	71.50	62.36
Number of observations	194	103	48	40	54

Note: Locations on the rural-urban spectrum are defined using the constructed definition. Sampling weights from the 2008/2009 survey wave are applied.

One of the other relationships worth noting is between destination types and education. From Table 2.5, there is no statistically significant difference in the proportion of migrants and non-migrants completing primary school. In contrast, there is a large and statistically significant difference in the proportion of people completing primary school between those who moved to rural areas and those who moved to urban areas. In Table 2.6, I see that migrants to all destinations except for low-density rural areas are more likely than non-migrants to have completed primary school. Also, among urban destinations, towns and cities attract the highest share of migrants who completed primary school.

Another interesting relationship is the one between destination types and shocks. In Table 2.5, I see that those who moved to rural areas are less likely to come from households that experienced agricultural shocks. I gain additional information from differentiating location types further: in Table 2.6, the strong negative relationship between agricultural shocks and migration is only present for migration to high-density rural areas. Among those who moved to urban areas, the share of people coming from households that experienced agricultural shocks is low only among migrants to towns. In Table 2.5, those who moved to urban areas are more likely to come from households that experienced non-agricultural shocks. Looking further, in Table 2.6, the strong positive relationship between non-agricultural shocks and migration is only present for migration to towns and cities, while migrants to peri-urban areas are alike rural-destined migrants and non-migrants in that regard.

In Table 2.7 and Table 2.8, I present summary statistics for the two groups distinguished by age (people of age 15-24 and people of age 25-34) and compare them to the cohort of older adults whom I define as people of age 35 and older. In Table 2.26 in Appendix 6, I show migration rates by age group. I see that people of age 15-34 are more likely to move than older

adults, which is consistent with the literature. Migration rate for people of age 15-24 is twice as high as that for people of age 25-34 and four times higher than the migration rate for people of age 35 and older. Migrants of age 25-34 are, on average, more likely to choose high-density rural and peri-urban destinations and less likely to choose urban destinations than migrant of age 15-24. Migrants of age 35 and older are more likely to choose low-density rural destinations and less likely to choose cities.

The share of women among migrants of age 15-24 is higher than that among non-migrants from the same age group, which is driven by a large number of women moving to rural destinations for marriage at a younger age. On the other hand, the shares of women among migrants and non-migrants of age 25-34 are similar. Though, when I look at them by destination, I see more women moving to urban areas and less women moving to rural areas. Interestingly, urban-destined migrants of age 35 and older are more likely to be women than men. Among people of age 15-24, those who moved to an urban area are, on average, more likely to have experienced non-agricultural shock prior to relocating than non-migrants. Among people of age 25-34, this relationship is reversed: migrants to urban areas are, on average, less likely to have experienced non-agricultural shock than non-migrants.

Table 2.7. Means of key variables by age group and migration status for youth from rural areas according to the constructed definition

Age group Migration status	15-24 years of age		25-34 years of age		≥ 35 years of age	
	Non-migrant	Migrant	Non-migrant	Migrant	Non-migrant	Migrant
Age	18.81	18.80	29.38	29.21	51.40	50.87
1 = Male	0.54	0.37	0.46	0.46	0.46	0.36
1 = Completed primary school	0.64	0.68	0.63	0.61	0.43	0.42
1 = Married	0.25	0.19	0.79	0.69	0.77	0.63
1 = Head of the household	0.05	0.01	0.38	0.40	0.59	0.46
1 = Child of household head	0.60	0.51	0.16	0.18	0.02	0.06
1 = Born in this village	0.86	0.71	0.75	0.53	0.65	0.47
1 = Was away from the household for at least one month in the past 12 months	0.10	0.17	0.07	0.13	0.05	0.08
1 = Main occupation in farming or fishing in the past year	0.56	0.50	0.87	0.75	0.88	0.83
Area under cultivation, acres	7.19	8.56	5.21	4.59	6.00	5.05
Livestock (TLU)	3.88	4.46	2.77	1.98	2.52	2.93
Age of household head	47.79	48.97	39.67	38.48	53.51	51.22
1 = Household head is male	0.78	0.77	0.84	0.79	0.80	0.75
Number of working age women	1.98	2.27	1.49	1.44	1.54	1.73
Number of working age men	2.10	2.08	1.44	1.44	1.47	1.69
Number of children of household head living in the household	3.64	3.68	2.94	2.27	2.93	2.66
1 = Household experienced agricultural shock in the past year	0.30	0.24	0.27	0.25	0.26	0.34
1 = Household experienced non-agricultural shock in the past year	0.28	0.32	0.30	0.28	0.27	0.34
Population density, people per square km	100.33	107.11	94.66	131.24	112.62	84.14
Distance to road, km	21.21	19.59	21.84	23.46	21.00	19.38
Distance to the nearest town with population of at least 50,000, km	66.44	69.21	68.57	62.80	66.10	60.73
Number of observations	1,388	316	976	123	2,159	124

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.8. Means of key variables by age group and destination for migrants from rural areas according to the constructed definition

Age group Migration status	15-24 years of age		25-34 years of age		≥ 35 years of age	
	To rural	To urban	To rural	To urban	To rural	To urban
Age	18.90	18.60	29.15	29.39	51.99	46.66
1 = Male	0.32	0.49	0.49	0.39	0.39	0.23
1 = Completed primary school	0.60	0.83	0.60	0.63	0.38	0.56
1 = Married	0.22	0.13	0.69	0.68	0.74	0.23
1 = Head of the household	0.02	0.00	0.38	0.47	0.44	0.53
1 = Child of household head	0.52	0.51	0.19	0.13	0.02	0.24
1 = Born in this village	0.68	0.76	0.54	0.51	0.43	0.58
1 = Was away from the household for at least one month in the past 12 months	0.14	0.24	0.14	0.12	0.09	0.03
1 = Main occupation in farming or fishing in the past year	0.60	0.29	0.79	0.62	0.86	0.72
Area under cultivation, acres	10.87	4.04	5.22	2.92	5.56	3.12
Livestock (TLU)	5.68	2.07	2.29	1.14	3.38	1.25
Age of household head	49.14	48.64	38.94	37.26	50.62	53.46
1 = Household head is male	0.81	0.69	0.84	0.67	0.80	0.57
Number of working age women	2.27	2.28	1.47	1.36	1.65	2.02
Number of working age men	2.01	2.23	1.60	0.99	1.81	1.22
Number of children of household head living in the household	3.87	3.32	2.39	1.95	2.72	2.46
1 = Household experienced agricultural shock in the past year	0.21	0.29	0.25	0.24	0.35	0.34
1 = Household experienced non-agricultural shock in the past year	0.27	0.42	0.30	0.22	0.31	0.45
Population density, people per square km	99.39	122.24	117.97	166.76	84.58	82.50
Distance to road, km	21.31	16.23	24.44	20.84	19.79	17.82
Distance to the nearest town with population of at least 50,000, km	71.65	64.44	69.45	45.02	65.11	44.29
Number of observations	205	111	123	31	97	27

Note: Sampling weights from the 2008/2009 survey wave are applied.

I also look at the characteristics of individuals by gender. In Table 2.27 in Appendix 6, I present migration rates by destination that show that women are more likely to move to rural and peri-urban areas while men are more likely to move to towns and cities. In Table 2.9, I present summary statistics for the sample split by gender. Some characteristics are common for migrants of any gender. For example, those who move to urban areas are, on average, more likely to have completed primary school than non-migrants; their households are more likely to be female-headed, have less land under cultivation, and were more likely to have experienced a non-agricultural shock.

There are characteristics that distinguish women who decided to move from women who decided to stay in place. Migrant women are younger than non-migrant women, they are less likely to be married, and they are more likely to have some migration experience in the previous year. Same patterns are observed for the differences in characteristics of male urban-destined migrants and male non-migrants. Another feature of gender differences among migrants relates to the distances to town and road. Women who live closer to a town are more likely to move to an urban area, while women who live farther from a road are more likely to move to a rural area. Men who live closer to a road are more likely to move to an urban area, while men who live farther from a town are more likely to move to a rural area.

Table 2.9. Means of key variables by gender, migration status, and destination for youth from rural areas according to the constructed definition

Gender	Men				Women			
	Non-migrants	Migrants	Migrants to rural	Migrants to urban	Non-migrants	Migrants	Migrants to rural	Migrants to urban
Migration status / Migration destination								
Age	22.70	22.07	23.08	20.37	23.79	21.54	21.38	21.93
1 = Completed primary school	0.64	0.68	0.62	0.78	0.63	0.64	0.58	0.79
1 = Married	0.36	0.26	0.31	0.17	0.60	0.38	0.39	0.34
1 = Head of the household	0.31	0.22	0.28	0.13	0.07	0.06	0.04	0.10
1 = Child of household head	0.54	0.45	0.45	0.46	0.29	0.40	0.40	0.38
1 = Born in this village	0.88	0.67	0.62	0.77	0.76	0.64	0.65	0.63
1 = Was away from the household for at least one month in the past 12 months	0.11	0.17	0.11	0.27	0.07	0.16	0.16	0.15
1 = Main occupation in farming or fishing in the past year	0.64	0.47	0.58	0.29	0.74	0.63	0.71	0.44
Area under cultivation, acres	6.30	7.52	9.87	3.57	6.43	7.38	8.75	3.94
Livestock (TLU)	3.63	3.06	3.84	1.73	3.18	4.22	5.12	1.94
Age of household head	45.81	46.88	46.00	48.36	42.94	45.42	46.08	43.74
1 = Household head is male	0.81	0.76	0.82	0.67	0.80	0.78	0.81	0.70
Number of working age women	1.60	1.64	1.63	1.65	1.94	2.31	2.26	2.42
Number of working age men	2.19	2.46	2.29	2.74	1.44	1.53	1.65	1.22
Number of children of household head living in the household	3.35	2.94	2.97	2.89	3.34	3.51	3.68	3.08
1 = Household experienced agricultural shock in the past year	0.28	0.25	0.25	0.26	0.29	0.23	0.21	0.29
1 = Household experienced non-agricultural shock in the past year	0.29	0.34	0.29	0.42	0.29	0.29	0.28	0.34
Population density, people per square km	95.50	109.03	100.38	123.57	100.49	117.22	107.71	141.21
Distance to road, km	21.46	17.99	20.40	13.94	21.49	22.48	23.34	20.31
Distance to the nearest town with population of at least 50,000, km	66.24	69.01	71.92	64.10	68.45	66.33	70.44	55.96
Number of observations	1,170	172	107	65	1,194	267	190	77

Note: Sampling weights from the 2008/2009 survey wave are applied.

2.5.2. Logistic and multinomial logistic regression results

Constructed definition of “rural”

First, I run a logistic regression which shows the association between the variables of interest and the probability that the individual moves. I proceed with a series of multinomial logistic regressions for the probability to choose a certain destination. In each subsequent regression out of this series, I split one of the destination types on the rural-urban spectrum to test whether the observed patterns depend on the categorization. The results of the logistic regression for the decision to move or to stay in place and the multinomial logistic regressions for two and three destination choices are presented in Table 2.10. The results of the multinomial logistic regression for four and five destination choices are presented in Table 2.11 and Table 2.12 respectively. I present marginal effects in the tables of regression results. The interpretation of these numbers is the percentage points added to the probability of migration. The base outcome for all regressions is staying in place. For some variables the results align, but for others I see how a narrower set of destination choices can mask important differences in migration decisions.

Table 2.10. Regression results (marginal effects, constructed definition of “rural”): binary division, two and three destinations

	Logistic regression	Multinomial logistic regression		Multinomial logistic regression		
	1 = Migrant	1 = Moved to rural	2 = Moved to urban	1 = Moved to rural	2 = Moved to peri-urban	3 = Moved to town / city
Age	-0.001	-0.001	0.000	-0.001	0.001**	-0.001
Age squared	-0.002**	-0.002	-0.003*	-0.002	-0.004	-0.004*
1 = Male	-0.067***	-0.060***	-0.009	-0.060***	-0.010*	0.000
1 = Completed primary school	0.006	-0.015	0.022**	-0.015	0.003	0.019***
1 = Married	-0.110***	-0.078***	-0.033***	-0.078***	-0.024***	-0.008
1 = Child of household head	-0.050***	-0.016	-0.030***	-0.016	-0.016**	-0.014*
1 = Born in this village	-0.107***	-0.079***	-0.025**	-0.079***	-0.001	-0.026**
1 = Was away from the household for at least one month in the past 12 months	0.078***	0.036*	0.040**	0.036*	0.008	0.028**
1 = Main occupation in farming or fishing in the past year	-0.033*	0.008	-0.039***	0.009	-0.011*	-0.029***
Area under cultivation, acres / 1000	-0.089	0.767	-1.390	0.758	-0.089	-1.237
Squared area under cultivation, acres / 1000000	15.834	-3.090	33.978	-3.102	-3.108	51.895
1 = Household head is male	0.010	0.026*	-0.013	0.026*	-0.003	-0.009
Number of household members	0.001	-0.001	0.002	-0.001	-0.001	0.003**
1 = Household experienced agricultural shock in the past year	-0.026*	-0.029**	0.002	-0.029**	0.000	0.003
1 = Household experienced non-agricultural shock in the past year	0.013	-0.002	0.016*	-0.002	0.000	0.016*
1 = From high-density rural area	0.022	0.009	0.013	0.009	0.005	0.009
Population density, people per square km / 1000	0.019	0.029	-0.002	0.029	0.003	-0.006
Distance to road, km / 1000	-0.140	0.405	-0.547**	0.400	-0.116	-0.375**
Distance to the nearest town with population of at least 50,000, km / 1000	0.147	0.255*	-0.148	0.266*	-0.260***	0.034

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.11. Regression results (marginal effects, constructed definition of “rural”): four destinations

	1 = Moved to rural	2 = Moved to peri-urban	3 = Moved to town	4 = Moved to city
Age	-0.001	0.001**	-0.001	-0.000
Age squared	-0.002	-0.004	-0.009**	-0.002
1 = Male	-0.060***	-0.010*	-0.006	0.004
1 = Completed primary school	-0.015	0.003	0.013***	0.007
1 = Married	-0.078***	-0.024***	-0.001	-0.009
1 = Child of household head	-0.016	-0.016**	-0.000	-0.012*
1 = Born in this village	-0.078***	-0.001	-0.015*	-0.012
1 = Was away from the household for at least one month in the past 12 months	0.036*	0.008	0.009	0.021*
1 = Main occupation in farming or fishing in the past year	0.009	-0.011*	-0.011*	-0.018***
Area under cultivation, acres / 1000	0.767	-0.090	-0.143	-1.153
Squared area under cultivation, acres / 1000000	-3.151	-2.676	10.224	-430.866
1 = Household head is male	0.026*	-0.003	-0.003	-0.004
Number of household members	-0.001	-0.001	0.000	0.002**
1 = Household experienced agricultural shock in the past year	-0.029**	0.000	-0.002	0.005
1 = Household experienced non-agricultural shock in the past year	-0.002	0.000	0.011**	0.004
1 = From high-density rural area	0.007	0.004	0.027**	-0.006
Population density, people per square km / 1000	0.036*	0.005	-0.065*	0.006
Distance to road, km / 1000	0.391	-0.118	-0.024	-0.268**
Distance to the nearest town with population of at least 50,000, km / 1000	0.264*	-0.261***	-0.009	0.059

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.12. Regression results (marginal effects, constructed definition of “rural”): five destinations

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
Age	-0.002	0.001	0.001**	-0.001	-0.000
Age squared	-0.002	-0.002	-0.004	-0.009**	-0.001
1 = Male	-0.045***	-0.015*	-0.010*	-0.006	0.004
1 = Completed primary school	-0.021**	0.007	0.003	0.012***	0.007
1 = Married	-0.058***	-0.022**	-0.024***	-0.001	-0.009
1 = Child of household head	-0.011	-0.005	-0.016**	-0.000	-0.012*
1 = Born in this village	-0.067***	-0.011	-0.001	-0.015*	-0.011
1 = Was away from the household for at least one month in the past 12 months	0.027	0.011	0.008	0.009	0.021*
1 = Main occupation in farming or fishing in the past year	0.008	0.002	-0.011*	-0.011*	-0.018***
Area under cultivation, acres / 1000	1.044*	-0.003	-0.091	-0.150	-1.156
Squared area under cultivation, acres / 1000000	-14.889	16.518	-2.526	10.181	-410.247
1 = Household head is male	0.033***	-0.006	-0.003	-0.003	-0.004
Number of household members	0.000	-0.002	-0.001	0.000	0.002**
1 = Household experienced agricultural shock in the past year	-0.010	-0.018**	0.000	-0.002	0.005
1 = Household experienced non-agricultural shock in the past year	0.002	-0.005	0.000	0.011**	0.004
1 = From high-density rural area	-0.002	0.015*	0.004	0.027**	-0.006
Population density, people per square km / 1000	-0.043	0.021**	0.006	-0.065*	0.008
Distance to road, km / 1000	0.336	0.033	-0.118	-0.023	-0.268**
Distance to the nearest town with population of at least 50,000, km / 1000	0.210*	0.021	-0.260***	-0.008	0.060

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

I find that people who are married at baseline are, on average, less likely to move, regardless of destination type; and people born in the village where I observe them at baseline are, on average, also less likely to move, regardless of destination type. The results with three destinations indicate that these relationships do not hold for all destination types when more types are included in the regression. Among urban destinations, the negative relationship between marriage and migration holds only for peri-urban areas and does not hold for towns and cities, either included together (Table 2.10) or separately (Table 2.11). Among rural destinations, this result holds for both low-density and high-density rural areas (Table 2.12). Similarly, being born in the baseline village reduces the probability of migration for those moving to low-density rural destinations and towns, but not high-density rural and peri-urban areas and cities.

In contrast to the binary migration model, several variables are in fact not significant in the multinomial logistic regressions either for rural or for urban destinations. For example, those whose main occupation was in farming or fishing at baseline were less likely to migrate to an urban destination. Yet, being a farmer had no significant effect on the probability of moving to either low- or high-density rural area. Being a child of the head of the household on average lowers the probability of migrating, but this holds only for some urban destinations: peri-urban areas and cities. Living in a household that experienced an agricultural shock reduces the likelihood of migrating. But this result is an artifact of the conventional way of defining “rural”; the model with more differentiated destination areas reveals that this holds only for the probability to move to a high-density rural area (Table 2.12).

Some factors are not significant in the binary logistic model but are significant in the multinomial logistic model with simple division into rural and urban destinations and in more complex migration destination models. For example, higher distance to road on average lowers

the probability to move to an urban area, but it is significant only for cities. Interesting observations can be made for the indicator of the completion of primary school and the distance to the nearest town with population of at least 50,000 people. Both these factors are not significant in the logistic regression and are positive and significant only for one destination type in the first multinomial logistic regression (urban and rural, respectively). However, the multinomial logistic regression that includes more destination types shows that the sign of the effect of these factors on the probability to migrate differs across destinations.

In particular, primary school completion on average increases the probability to move to an urban area. In further regressions, I see that it holds only for the probability to move to towns. Also, while not being significant for rural destinations when they are combined together, primary school completion on average decreases the probability to move to a low-density rural area. Similarly, higher distance to town on average increases the probability to move to a rural area, although it holds only for low-density rural destinations. In the regression with a diverse set of destination choices, distance to town, on average, turns to be negatively correlated with the probability to move to peri-urban areas; while the regression with the binary rural/urban choice does not show any significant effect for urban destinations.

There are factors that are not significant in either logistic regression with binary decision to migrate or in multinomial logistic regression with rural and urban destination choice but are significant for some destination types when a wider range of choices is considered. One of such factors is an indicator of living in a high-density rural area at baseline: on average, it increases the probability to move to another high-density rural area and to towns. Another variable, population density, has a diverse effect depending on destination. Higher population density at

baseline, on average, increases the probability to move to a high-density rural area but decreases the probability to move to a town.

Overall, there are changes to the significance of different factors depending on the classification of destination types on the rural-urban spectrum. Higher number of destinations considered allows to gain more information about the specific migration flows. I also observe some common migration patterns: higher average probability of migration for women, especially to rural destinations; higher average probability of migration for unmarried people; the importance of prior migration history²² for some destinations; and the positive correlation of the road network and proximity to towns with migration to some destinations.

I run separate regressions by age groups: for people of age 15-24 and for people of age 25-34. The results are presented in Table 2.28 and Table 2.29 in Appendix 6. My model explains the probability to move and the destination choices for younger people better. Most of the main results observed in this section are confirmed for people of age 15-24 and are driven by this part of the sample. Older migrants contribute to the results on migration history, land, and distances to road and to town. From the regressions for adults of age 35 and older (table not presented), I see that the results differ from the results I observe for people of both age cohorts between 15 and 34. For example, gender and marriage effects are not prominent for people older than 35, while being the head of the household is negatively associated with the probability of migration in general and migration to rural areas in particular. Agricultural shocks and being further away from the city are also negatively correlated with the probability of migration among those who are older than 35.

²² As stated in section 2.3, prior migration history at baseline is measured with an indicator to be born outside of the village of residence and an indicator to be away from the household for at least a month in the past year.

Then, I run separate regressions by gender and present the results in Table 2.30 and Table 2.31 in Appendix 6. As I saw from the summary statistics, some factors are associated with the decision to move for everyone, while others are gender specific. For example, position in the household and prior migration history are correlated with both the decision to move and destination decision in a similar way for both men and women. At the same time, some factors are associated with women's decision to migrate but not men's decision. Marriage and an indicator of the completion of primary school are negatively correlated with the probability for women to move to low-density rural areas, while main occupation in farming increases this probability for women; these factors are not associated with men's probability to move. Other factors have diverse effects on men and women. I find that higher distance to road is positively associated with women's decision to move to rural areas regardless of population density category of destination, and negatively associated with men's decision to move to a city. At the same time, higher distance to town is positively associated with men's decision to move to a low-density rural area and a city, and negatively associated with women's decision to move to a peri-urban area.

NBS categorization of "rural"

My sample consists of people who lived in rural areas at baseline, hence the change in the definition of "rural" affects not only the categorization of migrants into destinations but also the selection of people into the sample. I make two changes to the sample when shifting from my constructed definition of "rural" to the NBS definition. First, I need to re-categorize people who lived in areas considered to be rural by the constructed definition but urban according to the NBS definition. Following the NBS definition, I now categorize these individuals as living in urban areas and exclude them from the sample of people living in rural areas. Second, I need to re-

categorize people who lived in areas considered to be non-rural according to the constructed definition but rural according to the NBS definition. Following the NBS definition, I now categorize these individuals as living in rural areas and include them into the sample. In Table 2.32 in Appendix 6, I show how the definition of “rural” affects selection into the sample. For both definitions, the sample of people living in rural areas consists of around 2,800 people, but the sample of people who lived in areas considered to be rural by both definitions consists of only 2,280 people.

In Table 2.33 in Appendix 6, I confirm that the distribution of destination types changes with the definition of “rural”, although the number of migrants is similar between the definitions and is at around 430.²³ Still, if I intersect the definitions and only pick observations that are assigned to the same location type by both definitions in both survey waves, I notice that the numbers of observations and migrants drop significantly.²⁴ It might affect the results in regressions with the intersection of definitions.

I compare sample selection in Table 2.34 in Appendix 6, where I provide migration rates and summary statistics for the characteristics of youth living in rural areas according to the constructed definition and/or NBS categorization. There is a difference in statistics between individuals for whom the definitions of “rural” align and those for whom the definitions diverge. For people living in high-density rural areas according to the constructed definition, the main difference is in the migration rate: it equals 16.9% when NBS definition also categorizes these

²³ From Table 2.32 and Table 2.33 in Appendix 6: for the constructed definition, the total number of observations for youth from rural areas is 2,803, and the number of migrants is 439. For the NBS categorization, the total number of observations is 2,857, and the number of migrants is 434.

²⁴ From Table 2.32 and Table 2.33 in Appendix 6: for the intersection between the constructed definition and the NBS categorization, the total number of observations is 2,280, and the number of migrants is 299. The low number of migrants might affect the results of multinomial logistic regression with several destination choices as, in particular, the number of people whom I observe moving to towns is only 26.

areas as high-density rural; and it equals 28.1% when NBS defines these areas as urban.²⁵ For people living in low-density rural areas according to the constructed definition, migration rates by destination also differ depending on the sample: when definitions align, the rates of migration to rural areas are higher. When definitions diverge and these origin areas are defined as urban by the NBS categorization, the rates of migration to cities are higher.

I also see in this table how some individual and household characteristics differ depending on the definition of “rural” I use to select people into the sample. When definitions diverge and either one of the definitions states that the area is non-rural, the share of youth who completed primary school is higher (77-87%) than that share in low-density (58%) and high-density rural areas (68%) according to both definitions. At the same time, the share of youth with main occupation in farming or fishing is much lower: 20-42% for youth that is categorized as rural according to one criterion but not the other – compared to 62-76% when the two definitions align. Nevertheless, the gap in the share of adults (ages 35-64) with main occupation in farming or fishing is smaller: 45-77% when definitions diverge – compared to 89-93% when definitions align.²⁶

²⁵ In an opposite case of the divergence of definitions, when the constructed definition assigns the areas to be non-rural while the NBS defines them as high-density rural, migration rate is at 17.7%, which is similar to migration rates in other areas; while the share of migrants to peri-urban areas is higher than in other areas and is at 35.9%.

²⁶ A note on “45-77%”, which is a wide range: when definitions diverge, the share of adults with main occupation in farming or fishing equals 45% for areas identified as “high-density rural” by the constructed definition and “urban” by the NBS categorization. For other cases of divergence in definitions, the share of adults with main occupation in farming or fishing is 70-77%.

Table 2.13. Regression results (marginal effects, NBS categorization of “rural”): four destinations

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to town	4 = Moved to city
Age	-0.001	0.000	0.000	0.001
Age squared	-0.002	-0.002	-0.004	-0.009***
1 = Male	-0.049***	-0.027***	-0.007	-0.004
1 = Completed primary school	-0.008	-0.007	0.017***	0.005
1 = Married	-0.044***	-0.021**	-0.041***	-0.011
1 = Child of household head	-0.000	-0.015*	-0.015*	-0.006
1 = Born in this village	-0.041***	-0.016	-0.014	0.003
1 = Was away from the household for at least one month in the past 12 months	0.024	0.018	0.009	0.018
1 = Main occupation in farming or fishing in the past year	0.011	-0.011	-0.005	-0.023***
Area under cultivation, acres / 1000	1.078**	0.100	0.862	1.109
Squared area under cultivation, acres / 1000000	-23.051**	-5.021	-191.660	-12987.580*
1 = Household head is male	0.030***	-0.014	-0.005	-0.003
Number of household members	0.000	-0.002	-0.000	0.001
1 = Household experienced agricultural shock in the past year	-0.007	-0.003	-0.003	-0.004
1 = Household experienced non-agricultural shock in the past year	-0.002	0.002	0.007	0.003
1 = From high-density rural area	0.014	0.009	0.015*	-0.015**
Population density, people per square km / 1000	-0.138**	-0.011	-0.016	0.013**
Distance to road, km / 1000	0.370*	0.071	-0.447**	-0.208
Distance to the nearest town with population of at least 50,000, km / 1000	0.148	-0.115	0.013	-0.065

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

I present the results of logistic regression and multinomial logistic regressions with the NBS definition in Table 2.35 in Appendix 6 (for the decision to move or to stay in place and the categorization of two types of migration destination) and in Table 2.13 (for the categorization of four types of migration destinations) respectively. The results with the intersection of definitions are presented in Table 2.36 and Table 2.37 in Appendix 6. I cannot separate peri-urban areas using the NBS definition, so I compare these results to my main results with the exception of the probability to move to a peri-urban area. The migration patterns I observe in the results with both NBS categorization of “rural” and the intersection of definitions are similar to the ones with the constructed definition. The main conclusion, which is that I gain additional information when I distinguish more destinations on the rural-urban spectrum, is still valid, although, for some factors, the significance for the probability to migrate disappeared or shifted from one destination type to another.

I can separate the factors associated with migration decision and migration destination decision into three groups based on the results. The first group consists of the factors for which the significance for one destination type is lost with the change in the definition of “rural”. For example, being born in the baseline village has no significantly association with the probability of moving to town as it did with the constructed definition, but it still is negatively correlated with the probability to move to a low-density rural area in regressions with both NBS categorization and an intersection of definitions. Similarly, an indicator of the completion of primary school is still positive and significant for the probability to move to a town, but it is no longer negative and significant for the probability to move to a low-density rural area.

The second group consists of factors for which the significance changes only for one definition of “rural”. For example, an indicator of experiencing an agricultural shock is not

significant in the regressions with the NBS categorization of “rural”, but it is negatively correlated with the probability of moving to a high-density rural area when rural is defined with the constructed definition and when I intersect the definitions. Distance to road is found to be negatively associated with the probability of moving to a city in regressions with the constructed definition of “rural” and with the intersection of definitions. In regressions with the NBS categorization, I observe a correlation between distance to road and the probability to move to other destinations but not cities.

The third group consists of factors for which the significance shifts between destination types depending on the definition of “rural”. For example, an indicator of being married is negatively correlated with the probability to move to a peri-urban area according to the constructed definition, the probability to move to a town according to the NBS categorization, and the probability to move to a city according to the intersection of definitions. The results of all three models confirm that being married is negatively correlated with the probability of moving to a low-density rural area, and two models agree that it is negatively correlated with the probability of moving to a high-density rural area. The results for population density are more diverse. It is positive correlated with the probability to move to a city in the regressions with the NBS categorization and the intersection of definitions, while regressions with the constructed definition of “rural” indicate a negative correlation with the probability to move to towns. Population density is positively associated with the probability to move to a high-density rural area under the constructed definition and negatively associated with the probability to move to a low-density rural area under the NBS categorization.

2.5.3. Robustness checks

Cluster analysis definition of “rural”

The construction of the cluster analysis definition of “rural” is described in Appendix 3. For the first wave of survey, two groups are distinguished: rural and urban. Table 2.38 in Appendix 6 shows that the cluster analysis definition mainly agrees with the constructed definition. It places most of the households whose locations were identified as peri-urban by the constructed definition into the “urban” group and leaves only 483 individuals mis-matched (341 individuals if I exclude peri-urban areas according to the constructed definition). At the same time, the NBS definition and the cluster analysis definition do not match for 899 individuals. As for the cluster analysis definition itself, only 125 individuals from areas defined as “rural” by the cluster analysis definitions were not considered to be living in rural areas by either constructed definition or NBS categorization (Table 2.39 in Appendix 6).

In Table 2.40 in Appendix 6, I compare individuals’ destinations according to the cluster analysis and constructed definition. Cluster analysis definition allows me to distinguish several types of rural areas by distance to road and town, although it is not always consistent. For example, when I distinguish five destinations, the majority of individuals categorized as migrants to low-density rural areas by the constructed definition are categorized as migrants to a “rural area close to a road” by the cluster analysis definition. Then, when I distinguish six destinations, it adds separation of rural areas that are close to towns, and the number of individuals defined as migrants who moved to a rural area close to road drops significantly. Overall, the definitions for low-density rural areas close and far from road align with the low-density rural areas according to the constructed definition. Some of other rural locations are categorized as peri-urban areas and towns by the constructed definition. I see that the category “town” of the cluster analysis

definition is very imprecise and sometimes points to areas categorized as “rural” according to the constructed definition. On the other hand, destinations identified as “cities” according to the cluster analysis definition include most of the destinations identified as “cities” by the constructed definition.

Next, I analyze the rates of outmigration from and migration to various location types (see Table 2.41 in Appendix 6). I divide individuals into groups depending on how their baseline location is identified according to the constructed, the NBS, and the cluster analysis definitions. When constructed definition and NBS categorization align, pointing either to low- or high-density rural area, cluster analysis definition classifies at least 82% of observations as rural. When the two main definitions conflict, cluster analysis definition supports the constructed definition. Cluster analysis provides an additional layer to the other definitions: it shows that there are differences within each group even when the other two definitions align. For example, two groups of observations for which only the cluster analysis definition diverge can have drastically different outmigration rates (in some cases, the rate of outmigration in one group is twice as high as in the other group). These differences suggest the existence of an additional layer of complexity within the seemingly alike groups.

Migration rates computed with cluster analysis definition for both the origin and destinations are consistent with those computed using the constructed definition. In some cases, cluster analysis gives more information on the choice of destination. For example, for individuals from high-density rural areas according to both constructed and NBS categorizations that are at the same time identified as urban by the cluster analysis, constructed definition points out that migrants to low-density rural areas constitute 41.9% of all migrants. Cluster analysis shows that the majority of those migrant actually moved to a low-density rural area located far from a road

(34.5% of all migrants). Among people from areas identified as low-density rural by both NBS and the constructed definition but urban by the cluster analysis definition, only 26% of migrants chose low-density rural destinations, all of which are categorized as close to either road or town by the cluster analysis definition.

In Table 2.42 in Appendix 6, I present summary statistics for individual, household, and community characteristics of youth based on the categorization of their location in the first survey wave. I compare the mean values for individuals from rural and urban areas according to the cluster analysis definition who are classified into certain groups according to the constructed and NBS categorizations. The focus is on the groups that are not split too unevenly by cluster analysis definition, so that there are enough observations in each group and the cluster analysis is not decisive. In these groups, at least one definition identifies the location as a high-density rural area (see the last six columns of Table 2.42).

There is a striking difference between individuals from rural and urban areas according to the cluster analysis definition, even when both constructed and NBS categorizations point to the same type of location. On average, 10-16% more young people in rural areas according to the cluster analysis definition report that their main occupation is farming or fishing. Also, they come from households with more livestock. There are characteristics for which the common patterns differ between areas for which the constructed and the NBS definitions align and areas for which they diverge. For example, people from urban areas are, on average, more likely to have completed primary school (but not in areas which are defined as high-density rural by the constructed definition and urban by the NBS categorization); people from rural areas are more likely to come from households with more cultivated land (with the exception of areas which are

defined as non-rural by the constructed definition and high-density rural by the NBS categorization).

With the regression results presented in Table 2.43, Table 2.43, Table 2.45, and Table 2.46 in Appendix 6, I check my two main conclusions. The first one is that distinguishing more destinations on the rural-urban spectrum provides an additional insight about the migration of rural youth. The second inference I test here is that the definitions of “rural” I use in the main part of the paper are robust to changes. I compare the results from the regressions with the origin and destinations defined by the cluster analysis to the results obtained with the constructed and the NBS categorizations. Also, I note when the use of cluster analysis definition provides any additional information about the destinations I describe in the main results section.

The patterns in the results are mostly the same between the main analysis and the cluster analysis definition. There are variables that are not significant in the logistic regression but gain significance in the multinomial logistic regressions. For example, completion of primary school, on average, is positively correlated with migration to urban destinations only. I get more nuanced results for other variables. For example, marriage is negatively correlated with migration in general, and multinomial logistic regression with two destinations shows that it is negatively correlated with the probability to move to both rural and urban areas. A more detailed distinction of destinations shows a strong negative correlation between marriage and the probability to move to a high-density rural area, which is consistent across all models. Finally, some variables have drastically different effect depending on destination. For example, an increase in the distance to the nearest road, on average, increases the probability to move to a rural area that is far from a road but decreases the probability to move to a high-density rural area or a city. All these results are consistent with the results obtained using constructed definition and NBS categorization.

Cluster analysis definition confirms that young adults from rural areas who completed primary school are, on average, more likely to move to urban areas. Regressions with the NBS categorization show that people from households with more livestock are less likely to move to urban areas, while using cluster analysis definition I also see that livestock is positively associated with the probability to move to rural areas that are located far from roads. Results with cluster analysis definition persistently show positive correlation of being away from the household and migration to a high-density rural area, while the results with other definitions are less decisive in regressions with multiple destinations.

In some cases, cluster analysis definition provides more detail on the destination. With an indicator of being born in the baseline village, main analysis concludes that it lowers the probability to move to towns and low-density rural areas. Regressions with cluster analysis definition narrows this set to cities and rural areas that are close to towns, and rural areas that are far from roads. Similarly, people who report their main occupation to be farming are less likely to move to all types of urban areas according to the constructed definition (namely, cities, towns, and peri-urban areas). Cluster analysis definition catches the negative correlation with the probabilities to move to cities, towns, and rural areas that are close to towns.

Agricultural shocks are found to be negative correlated with migration to high-density rural areas according to the constructed definition and rural areas that are close to towns according to the cluster analysis definition. Non-agricultural shocks are positively associated with the probability to move to a town according to the constructed definition and a high-density rural area according to the cluster analysis definition. Regressions with the NBS categorization of “rural” show that youth who live far from roads are more likely to move to low-density rural areas. With cluster analysis definition, distance to road is positively correlated with the

probability to move to a rural area that is far from road. Similarly, with the construction definition I find that increase in the distance to town is positively associated with the chances to migrate to a low-density rural area, which cluster analysis definitions reflects as a higher probability to move to a rural area that is close to a road.

Constructed definition distinguishes peri-urban areas while cluster analysis definition does not. Still, some of the variables associated with migrations to peri-urban areas are also associated with migration to certain types of rural areas under the cluster analysis definition. For example, men are, on average, less likely than women to move to peri-urban areas in regressions with the constructed definition, and men are less likely to move to rural areas that are close to towns in regressions with the cluster analysis definition. Children of the household head are, on average, less likely to move to a peri-urban area, and they are less likely to move to a rural area that is close to a road in regressions with cluster analysis definition. People who live further from towns are less likely to move to a peri-urban area, and they are less likely to move to rural areas that are close to towns and to high-density rural areas according to the cluster analysis definition.

Definitions of “migrant”: self-reports and NBS definition

For the main results above, I use the definition of “migrant” that is based on the reported and the computed distance between the locations of the individuals in the first and the last survey waves. In this subsection, I test several other definitions of “migrant”. First, I use the time spent at the current location reported by individuals during the last survey wave. When a person reports spending four years or less at the current location in the last survey wave, I consider this person to be a migrant. Out of 439 people categorized as migrants according to my main definition, 354 report to have lived at the current location for four years or less, while 85

individuals report to have lived for more than four years in the community at destination.²⁷ In addition to the 354 individuals who report spending little time at the location I observe them at during the last survey wave and who traveled some distance from their location in the first survey wave, there are 139 individuals whom I do not observe to travel but who report being recent migrants. It makes the total number of migrants according to self-reports to be 493.

I present migration rates and summary statistics for individual, household, and community characteristics in Table 2.47 and Table 2.48 in Appendix 6 respectively. Note how, for people who are considered to be migrants only based on self-reports, the type of location does not change from the first to the last survey wave according to the constructed definition of “rural”, while the types defined by the NBS categorization change. It happens because of the changes in administrative division into rural and urban. Among those who report themselves to be migrants but for whom I do not observe a physical move²⁸, there are more women. These people are, on average, more likely to be away from the household for some time during the past year²⁹, and they are likely to come from a household that owns more livestock. Also, their household is less likely to have experienced a negative agricultural shock in the past year.

²⁷ This could happen due to mistakes in self-reports, migration to a familiar place, or return migration: individuals may have traveled back to a well-known community and hence not consider themselves to be migrants. Among those who report to have spent five years or more at their location during the last survey wave, around a third also report to have spent a certain number of years living at the origin that suggests a discontinuity of presence at the location they were at during the first survey wave. An example of such case would be an individual who reports being born in the location I observe them at during the first survey wave and spending eight years at the location I observe them at during the last survey wave. Another example would be an individual who reports spending one year at the location I observe them at during the first survey wave and spending 11 years at the location I observe them at during the last survey wave.

²⁸ Here and further in this subsection I define a physical move the same way I identify migrants in my main analysis: when the distance traveled between the survey waves is at least five km. For most observations, I can use the distance provided in the dataset. When this information is missing, I apply the same threshold of five km to the distance computed using the coordinates provided in the dataset.

²⁹ This observation causes concerns as people might consider these short moves when replying to the question on years lived in the community. I find the difference in averages to be small, although it is significant. In the sample of people whom I can define as migrants only from self-reports, the share of those who was away from the household for at least one month during the last year is 19%. For those whom I can define as migrants only from distance traveled but not self-reports, the share of people with migration history in the past year is 14%. For people whom I can define as migrants from both distance and self-reports, this number is 17%. For non-migrants, this

People who do not consider themselves to be migrants, but for whom I observe a physical move, are more likely to come from a low-density rural area, as well as non-migrants do. Also, they are more likely to be at non-urban destinations during the last survey wave compared to people for whom both self-reports and distance indicate migration. Their characteristics do not differ much from those of non-migrants. The main differences are that they tend to be younger, come from larger households, and live further from towns than non-migrants. Results of regressions where migrants are defined only using self-reports are presented in Table 2.49 and Table 2.50 in Appendix 6.³⁰ They are consistent with my main results.

Then, I test a combination of definitions: an individual is considered to be a migrant if either my main definition based on distance or self-report indicate migration. In comparison to the definition based on self-reports, it shifts 85 people whom I observe traveling between the survey waves to the “migrant” category and makes the total number of migrants 578. I present the results of regressions with this definition in Table 2.51 and Table 2.52 in Appendix 6. I also consider a strict definition: an individual is defined as migrant if both the definition based on distance and self-report indicate migration. It leaves only 354 individuals as migrants. The results of regressions with the strict definition are presented in Table 2.53 and Table 2.54 in Appendix 6.

With the strict definition, both the self-report and tracking by the survey team confirm that the individual moved. The results of the logistic regression in comparison to the results of the multinomial logistic regressions follow the same pattern as I saw with my main definition of

number is 8%. Hence, even if some people consider moves of short duration while replying to the question that I use for self-reports on migration, the share of them in the full sample should be relatively small.

³⁰ I use constructed definition of “rural” for all regressions in this subsection to get results that are comparable to those from the main results with the constructed definition of “rural”, where I use distance-based definition of “migrant”. I present migration rates for both constructed definition and NBS categorization for comparison, as it brings additional insight in some cases.

“migrant”. It supports my key hypothesis that distinguishing a wider range of migration destinations provides new information about youth migration. Though, with the strict definition of “migrant”, I lost some important results: a decrease in the probability to move to a low-density rural area for people who completed primary school, a decrease in the probability to move to a peri-urban area among farmers, and an increase in the probability to move to an urban area if the household experienced non-agricultural shock. On the other hand, I gained an interesting result: young adults who are heads of their households are more likely to move to a low-density rural area and less likely to move to a city. The result with a positive correlation between the amount of land cultivated by the household and the probability to migrate to a low-density rural area is replaced with an analogous connection for the number of livestock the household owns. Similarly, an increase in the probability to move to a low-density rural area as the distance to town increased is replaced with an analogous connection for the distance to road.

I also test the definition of “migrant” used by the NBS. It identifies people who moved between administrative areas as migrants, while people who moved within an administrative area are called “short-distance movers” and pulled together with non-migrants (NBS, 2015). To get an intersection of this definition with my main definition based on distance traveled, I use the district change. District change alone cannot define migrants in my sample as I do not account for administrative changes. Out of 439 people who traveled some distance from the baseline location, 187 did not cross the district borders.³¹ The majority of people, who moved within the districts where they were present in at baseline, came from low-density rural areas and migrated

³¹ Also, I observe 77 individuals with both reported and computed distance being almost zero, which makes them non-migrants according to my definition, but for whom the district changed between the survey waves. I assume this happened due to administrative changes and hence rely on the reported distance, although 13 of these individuals report to have lived in this community for four years or less in the third wave of survey, and only 47 individuals did not have to be tracked by the survey team in both the second and the third waves.

to another low-density rural area. When I use the intersection of definitions, these people are considered to be non-migrants as the definition used by the NBS suggests.

In Table 2.55 and Table 2.56 in Appendix 6, I present migration rates and summary statistics for key variables for non-migrants, people whom I observe physically moving within their district, people for whom I observe district change but not a physical move, and people for whom I observe a physical move between districts. I see that those who moved within their district are almost as likely to come from a low-density rural area as non-migrants do, but those who moved between districts are more likely to come from high-density rural areas. Those who travel within districts are more likely to move to a low-density rural area.

Although the structure of types of origin for people who moved within their district resembles that of non-migrants, other characteristics differ significantly between these two groups. The majority of those who moved within district are women, they are more likely to be a child of the household head, while the head of their household is more likely to be older, and their household is likely to own more livestock. Compared to this group, those who moved between districts are more likely to have completed primary school and to have prior migration history. They are less likely to be farmers and are more likely to come from a more densely populated area located closer to a road or a town.

Regressions results for between-district migrants are presented in Table 2.57 and Table 2.58 in Appendix 6. Some conclusions I made from the main analysis no longer hold, for example, the results for gender that showed that men are less likely to move to rural and peri-urban areas. This is understandable given that many women move to rural areas within their district as shown in Table 2.55 and Table 2.56 in Appendix 6. An indicator of primary school completion is positively associated with the probability of migration to a high-density rural area,

while an indicator of being born at the baseline village is negatively associated with this probability.

I no longer see a negative correlation between being a farmer and the probability to move to a town, neither I see a positive correlation between having more cultivated land in the household and the probability to move to a low-density rural area. On the other hand, I observe a new pattern: youth from households affected by a negative agricultural shock are less likely to move to a low-density rural area. I also see a change in the sign of the effect of distance to town. In my main results, with an increase of the distance, the probability to move to a low-density rural area would increase. When I exclude within-district migration, distance to town is found to be negatively correlated with the probability to move to a low-density rural area. It means that young adults from households located further from towns are less likely to cross the district border in order to move to a low-density rural area. At the same time, as I see in my main results, they are more likely to go to a low-density rural area within their district.

2.6. Discussion

In this section, I propose mechanisms that could explain the observed differences in results as a consequence of how destination locations are defined. I can separate factors that are positively associated with the probability to move to an urban area (urban pull), rural area (rural pull), and factors that can strengthen the decision to move (rural push) or lessen it (stay). In a nutshell, my results confirm that urban areas are more attractive to people who are more likely to find an off-farm job at destination, while people who move to rural areas are more likely to be farmers looking for land. Migration to rural areas, and especially low-density rural areas, is more affordable, and more people pursue it than migration to either peri-urban or urban areas, in contrast to conventional perceptions that rural-to-urban migration is the most common form.

Non-agricultural shocks are associated with the decision to migrate, while stronger connection to the baseline village, marriage, and agricultural shocks encourage youth to stay to place.

First, I look at the factors that are associated with migration to urban areas. I see that people who completed primary school are, on average, more likely to move to an urban area, particularly to a town. At the same time, people who report having main occupation in farming at baseline are, on average, less likely to move to an urban area. Hence, urban areas are attractive to people who are in a relatively advantageous position to find a job there: better educated youth and people with work experience in an off-farm job. Indeed, I observe that a third of urban-destined migrants (almost two thirds of male urban-destined migrants) work in the private sector after their move, and most of them are below age 25.

Then, I discuss factors associated with migration to rural areas. The amount of cultivated land and owned livestock in the household, being the head of the household, and living further from road and town, on average, are positively correlated with the probability to move to a rural area, particularly to a low-density rural area. I see that many rural-destined migrants have main occupation in agriculture after their move. It suggests that the main factor for choosing rural areas as a destination among people who have experience working in farming can be the availability of land and markets for agricultural input and output products.

One of the factors that could be categorized as rural push is distress. I find that people are more likely to move to a high-density urban area or a town if their household experienced a non-agricultural shock in the past year that negatively affected their income or assets. The following three types of shocks are most frequently listed in this category: a large increase in the price of food, death or illness of a household member living in or out of the household, and hijacking, robbery, burglary, or assault (the last four events are grouped into one category in the

questionnaire). These events might indicate a need to send out a migrant specifically for the purpose of employment, which helps to increase the household income and diversify risks.

The frequency of most types of shocks is similar for migrants and non-migrants, although migrants are less likely to have experienced illness of a household member living in the household. It can point to the monetary constraints to migration: funds that the migrant could have used are directed towards the care for the sick household member; as well as to labor constraints: a household with a sick member needs to allocate more time into care, and the sick individuals works less. At the same time, migrants are more likely to have experienced death of a family member living outside of the household. It might make youth move to care for the remaining household members living outside of the household or to work to replace the lost member.

I find that agricultural shocks, on average, are negatively associated with the probability to move, mainly to rural areas. Among agricultural shocks, the most frequently reported ones are drought or flood, death or theft of livestock, drastic change in input or output prices, and severe water shortage. A shock related to livestock is much less frequent among migrants than among non-migrants. People still move to rural destinations if they experienced a shock related to livestock and / or a drastic decrease in the output prices. I observe the frequency of negative weather events to be higher among migrants than among non-migrants, and that people who experienced them are more likely to move to peri-urban areas, towns, and high-density rural areas, suggesting an exit from agriculture. Shocks can be associated with the decision to move through a decrease in income coming from agriculture, which constitutes a large share of income for most rural households. With these negative shocks to income or assets, the household might not be able to send out a migrant, even if it planned to do it before.

Another way to group the results is by how much they differ when I shift from a narrow binary decision to move or to stay in place or from a rural/urban categorization to a more elaborate set of location types. The first group of results covers the split in the location types, when the variables are correlated with the decision to migrate, a decision to migrate to both rural and urban areas, and then they keep being significant only for destination types that stand far from each other on the rural-urban spectrum. Across most models, I see a split in the effects on the decision to migrate to a low-density rural area located further from the road and the decision to migrate to a peri-urban area or a high-density rural area located closer to a town. Being male and being born in the village I observe them at during the first survey wave make individuals less likely to move to the destinations listed above, while completing primary school and living closer to a road or town make individuals less likely to move to low-density rural areas and more likely to move to towns.

The second group of results covers the focus of the location types, when the variables is associated with the decision to migrate and the decision to migrate to either rural or urban area, and then they keep being significant only for one destination. For example, being a farmer has the strongest negative correlation with migration to a city; agricultural shocks have the strongest negative correlation with migration to high-density rural areas located close to a town; and non-agricultural shocks have the strongest positive correlation with migration to a town. The third group of results covers the appearance and the disappearance of the effect as the location types get disaggregated, when the variables show no correlation with the decision to migrate and the decision to migrate to either rural or urban area, but are positively associated with migration to one or two destination types (for example, living in a high-density rural area at baseline, which effects only the probability to move to another high-density rural area); or vice versa, when there

is a strong correlation with the probability to migrate which disappears when more destination types are considered (for example, the effect of migration history in the past year).

For a robustness check, I test variations to the categorization of locations on the rural-urban spectrum. I must note that changes to the categorization do not only affect the results but also have a direct influence on its interpretation. For example, the meaning of moving to a “high-density rural area” is different between my constructed definition, the NBS definition, and the cluster analysis definition. With the constructed definition, the individual moves to an area that I do not consider to be urban by either population density or built-up area density criteria, but, among all such areas, this one has higher population density. With the NBS categorization, the individual moves to an area that has higher population density among areas defined by NBS as rural, which implies that areas defined by the local authorities as urban are excluded. With the cluster analysis definition, the individual moves to an area that differs enough from urban areas and low-density rural areas not only in population density, but also in access to amenities, share of income coming from agriculture, and distance to road and town. One of the limitations of my study is that the definitions I propose are based on the local context. Hence, the interpretation of the results changes when the definitions are applied directly as they are to the data from other time periods or countries.

Another limitation is the decrease in the number of observations as the depth of categorization of migration destinations increases. This causes concerns for the precision of the estimation. With the main constructed definition for the location types I use, the group with the lowest number of observations, which is migrants to towns, counts 40 individuals (see Table 2.4). Cluster analysis categorization provides an even finer division of destination types, with the lowest number of observations per group being 10 for migrants to towns. On the other hand, the

adapted NBS categorization distinguishes four destination types, and the group with the lowest number of observations is migrants to cities, which counts 64 individuals.

2.7. Conclusion

I categorize migration flows of young people from rural Tanzania according to the destination type they chose on the rural-urban spectrum. I observe the majority of young people preferring rural destinations, with low-density rural areas dominating all other destinations, especially for migrants from other low-density rural areas. I use multinomial logistic regressions to show that factors associated with migration destination decision vary between destinations and that some relationships might be hidden by overgeneralization of destinations. I find that the probability to move to a rural area increases if the individual is the head of the household, lives in a household with more livestock, or lives further from towns. This probability decreases if the individual was born in the baseline village or lives in a household that recently experienced a negative agricultural shock. Characteristics of people moving to low-density and more remote rural areas differ from the characteristics of people moving to more densely populated rural areas, especially to rural areas located closer to road or town. Migration to urban areas, which contributes to structural transformation, is more likely to be observed among youth from high-density rural areas. I find that people who completed primary school or live in the households that recently experienced a negative shock not related to agriculture are more likely to move to an urban area. At the same time, children of the head of the household, people with main occupation in farming, and people who live further from roads are, on average, less likely to move to an urban area. There are two distinct flows of migrants: young women moving to rural and peri-urban areas and well-educated and very young men and women moving to urban areas, especially cities.

I distinguish four to six types of destinations on the rural-urban spectrum. I employ two main definitions of “rural”: the modification of the NBS administrative division and the definition based on population density, the built-up area density, and the distance to the nearest town.³² I also build a cluster analysis definition that is based on the access to amenities, involvement in agriculture, population density, and distance to roads and towns.³³ I find that a group of people who were categorized into one location type by the two main definitions can be split further into two groups using the cluster analysis definition of “rural”, and that these groups will differ significantly in their characteristics. Hence, while the results are often quite similar across the definitions, in some cases by using the more differentiated migration categories I obtain novel findings or a more nuanced interpretation of the results. For example, I find more urban-destined migrants to be living in high-density rural areas, closer to roads and towns, and having main occupation in an area not related to agriculture, which suggests that rural-to-urban migration can have smaller effect on structural transformation than presumed before.

I also test the robustness of my results to the definition of “migrant”. The main definition I use is based on the distance the individual traveled, with a threshold of five km. The other two definitions are based on respondents’ self-reports and the fact of crossing the border of an administrative unit (a district). Migration flows between districts are very different from migration flows within one district. Relocation within a district is more common among women³⁴ and people from low-density rural areas. These migrants are more likely to be children

³² For the constructed definition, I use the data on population density and built-up area density with one km grid (Tatem, 2017; Corbane et al., 2018) and on distance to the nearest town with population of at least 50,000 people. To get this data, I use the household coordinates provided in the LSMS dataset, which are aggregated at the level of enumeration area with a random offset of up to two km for urban households (according to the NBS definition of “rural/urban”) and up to five km for rural households (with an additional offset of up to 10 km for 1% of rural households).

³³ For this definition, I use district averages.

³⁴ Women who move within the district are more likely to report moving for marriage and are less likely to report moving to get access to better housing or services.

of the household head and live in the households with more livestock. For those who traveled between districts, education, prior employment history, and migration history³⁵ were the most important drivers of the decision to migrate. These people, on average, are less likely to be farmers at baseline and are more likely to come from an area that is located closer to a road or a town.

The definitions I employ to identify locations on the rural-urban spectrum are based on the survey data. I look at the distributions of population density, built-up area density, and distance to town to set the thresholds for these variables. Hence, both the definitions and the interpretation of the results are specific to the context. The main conclusion of this essay is that there is inevitable subjectivity to the categorization of migration destination areas and that our understanding of the drivers of migration will be influenced by how many categories are used along the rural-urban spectrum and how these categories are defined. While simplicity certainly has an advantage of being able to convey findings more easily, this essay shows that our understanding of the labor flows associated with structural transformation may depend on a more nuanced categorization of migration destination areas. There are factors, for example, migration history, which are associated with the decision to migrate in general but not with the choice of destination. Other factors, for example, living in a more densely populated area, are correlated with migration to a certain destination (another densely populated place) but not with the decision to move in general. Finally, factors like remoteness and unfinished education are positively associated with migration to a low-density rural area but negatively associated with migration to cities.

³⁵ Prior migration history, both as a child (an indicator for the individual to be born in the village where I observe this individual to reside at baseline) and more recent one (an indicator for the individual to be away from the household for at least one month at baseline), is an important predictor for current migration.

APPENDICES

APPENDIX 1. Data issues related to geospatial information

I use the following specifications of the Living Standards Measurement Study (LSMS) datasets for Tanzania (World Bank, 2017): March 2019 version of the 2008/2009 and 2010/2011 survey waves and June 2017 version of the 2012/2013 survey wave. Two major issues I have to tackle are missing coordinates and missing data on population density and built-up area density when coordinates point to water bodies. In this section, I describe how I recover or approximate most of this information.

The coordinates are provided in separate files called “HH.Geovariables_Y1”, “HH.Geovariables_Y2”, and “HouseholdGeovars_Y3” for the first, the second, and the third survey waves respectively. Coordinates are provided at the level of enumeration area to maintain the confidentiality of respondents. To achieve that, households’ coordinates were averaged across the enumeration area and a random offset was applied by the survey team. The files are organized at the household level, and each household is linked to the identification for its enumeration area, “ea_id”. In the first wave, households within one enumeration area are assigned the same coordinates, but in further waves, with administrative changes and migration, households within one enumeration area can have different coordinates, although this difference is usually small³⁶.

For the first survey wave, “HH.Geovariables_Y1” has information for 2,990 out of 3,265 interviewed households. Another file, “EA.Offsets”, is available only for this wave and contains coordinates at the level of enumeration area for all 409 enumeration areas listed in the survey. I

³⁶ Average distance from the household’s coordinates to the coordinates averaged across the enumeration area in the second and the third survey waves is 1.5 km. The number of households with the difference above 100 km (10 km) is 20 (44) in the second wave and 13 (73) in the third wave of survey. For most outliers, it happens because migrant households got assigned the enumeration area of their origin.

can link this file to the main dataset through “ea_id”, but the linking file, “SEC_A”, lists only 407 enumeration areas. For some households in the remaining two enumeration areas, “ea_id” is provided in “HH.Geovariables_Y1”, and I can transmit it to the rest of the households matching by enumeration area³⁷. Finally, I can pull the coordinates from “EA.Offsets” leaving no missing information for the 2008/2009 wave.

“HH.Geovariables_Y2” and “HH.Geovariables_Y3” are missing coordinates for seven and 22 households in the second and the third survey waves respectively. If it is possible to compute an average across the enumeration area and the average is the same for all households, then I replace missing information with enumeration area average. For households that are the sole household in that enumeration area, I check if it is possible to recover coordinates using their migration history. For example, if tracking information and self-reports indicate that the household moved between the first and the second survey waves and stayed in place between the second and the third waves, then I replace missing information in the third wave with the coordinates from the second wave. When migration history is inconclusive (tracking and self-reports point to different directions), I replace missing information with ward average.

As a result of this procedure for the second and the third survey waves, I replace missing coordinates with enumeration area average for 11 households, with household’s own coordinates from a different wave for eight households, and with ward average for eight households³⁸, leaving two households with missing coordinates. These two are the only households in its wards and are missing the coordinates for the second wave of survey. One household moved between

³⁷ I can match by region name, district number, and enumeration area number. This enumeration area number is not unique: enumeration areas in different districts have the same number, therefore, I need information on region and district. This information is available for all households, so I can recover the missing ea_id, which is unique for each enumeration area.

³⁸ Average distance from the household’s coordinates to the coordinates averaged across the ward is 1.4 km for households in wards where these eight households are located. Maximum distance does not exceed eight km.

the first and the second survey waves and was lost due to attrition after; another household moved both between the first and the second and between the second and the third survey waves.

I use these coordinates to get data on population density for 2010 from WorldPop Africa Continental Population Databases (Tatem, 2017) and on the built-up area density for 2013/2014 from Global Human Settlement Layer (Corbane et al., 2018). For both datasets, I use one km grid, while a denser grid is available (100 m for population density and 250 m for built-up area density). As mentioned above, the coordinates provided in the LSMS dataset were averaged across the enumeration area by the survey team, and a random offset was applied. Its range for urban areas is 0 – 2 km, its range for rural areas is 0 – 5 km, and an additional offset of 0 – 10 km is applied for 1% of rural households. Hence, I opt for a less dense grid when I use external datasets for population density and built-up areas density.

Some coordinates point to water bodies, and the chances of that are higher when a map with a less dense grid is used. For these cases, I replace the coordinates with a point in the closest grid cell of land. As a result, I use coordinates that are different from the ones provided in the dataset for four locations in the first survey wave, seven locations in the second wave, and 15 locations in the third wave, replacing 26 coordinates in total. In 20 of these cases, the distance between the original point and the substitute is below one km. In four cases, the distance is 1 km – 1.25 km, and the distance is 1.25 km – 2.5 km in the remaining two cases.

APPENDIX 2. Definition of “migrant”

For the LSMS dataset for Tanzania, I found four ways to get information on whether the individual moved between the survey waves: distance between the waves that is provided in the data, distance I can compute with the given coordinates, self-report on migration, and an indication of tracking by the survey team. In this section, I discuss how I use both reported and computed distance between the waves to identify migrants and how this data aligns with other available information. I identify individuals as migrants if between the first and the last survey waves: (i) the reported distance is over five km; or (ii) the reported distance is missing and the computed distance is at least five km.

For youth from rural areas, both according to my constructed definition of “rural” (see Appendix 3) and the NBS definition, this definition of “migrant” works as follows. I am able to identify migrants with the use of the reported distance for the majority of the sample. For four observations the reported distance is missing: (i) for one of them, the computed distance is below 0.1 km and I identify this individual as non-migrant; (ii) for three of them, the computed distance is over 400 km and I identify these individuals as migrants. I must note that, although it did not affect my subsample of youth, in the full sample there are cases when the reported distance is zero and the computed distance significantly differs from zero. An offset applied to the coordinates reported in the dataset or inaccuracies in data recording might be the reason for this, as I discuss below.

Reported and computed distance

The dataset contains information on the distance between the first and the second and between the first and the third survey waves. It is computed with the original coordinates³⁹ and

³⁹ Distance is computed by the data processing team using the coordinates without the offset.

reported in the set of geospatial information. As I discussed in Appendix 1, this information, along with the information on coordinates, is missing for seven households in the second wave and 22 households in the third wave. In addition to this, the way of reporting information is different between the waves. In the second wave, the distance is either zero (for 3,497 households) or a number above five (for 420 households). In the third wave, the distance is either zero (for 37 households), a number above five (for 1,253 households), or missing (for 3,698 households). I assume that in the second wave any distance below five km was discarded and replaced with zero. Then, I suspect that in the third wave the distance was above zero but below five km for 37 households, and it was zero for 3,698 households. Hence, I replace missing values with zeroes for the third wave, except for the 22 households for which the information was missing in the set of geospatial information.

As described in Appendix 1, I am able to retrieve the missing coordinates for all households except for two in the second wave. Therefore, I am able to compute the distance between the first and the third survey waves for all individuals present in these waves (14,740 individuals). In Figure 2.2 and Figure 2.3, I show how the computed distance aligns with the reported distance. Most dissimilarities between the distances can be explained with the 0 – 10 km offset applied to the coordinates provided in the dataset (see Appendix 1). I cannot explain the large difference between the computed and the reported distance for 66 individuals, and for four of them the reported distance is zero. I further investigate these cases in the next subsection.

Figure 2.2. Computed and reported distance between the first and the third survey waves

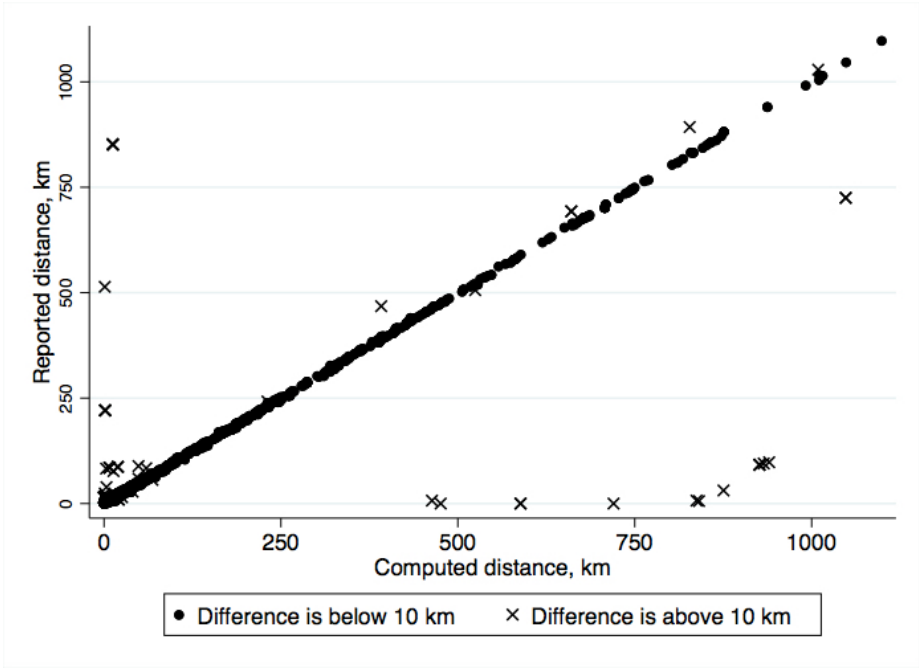
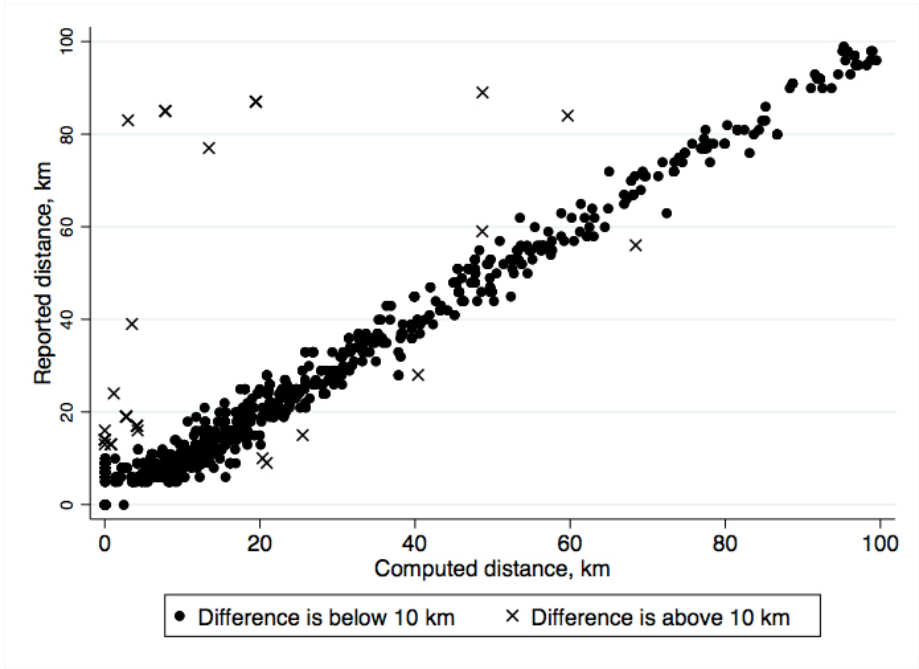


Figure 2.3. Computed and reported distance between the first and the third survey waves, for distances below 100 km



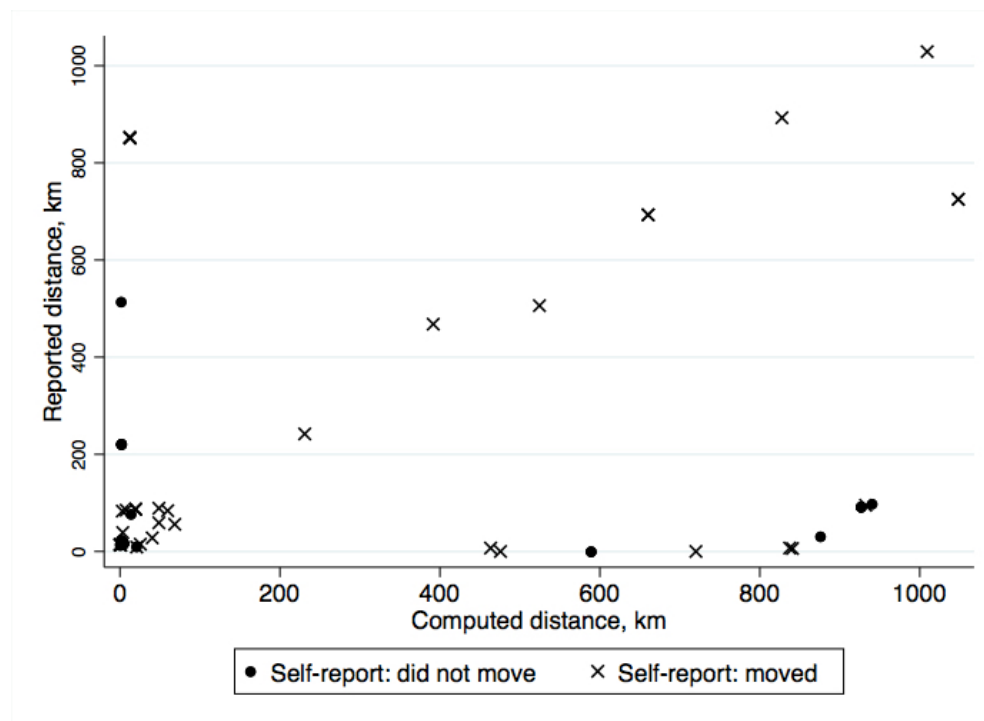
Cases with large difference between the reported and the computed distance

The dataset contains information on the self-reported years the individual lived in this community and on whether the household or individual had to be tracked between the survey waves (between the first and the second wave and between the second and the third wave). From this, I state that the individual reports to be a migrant if the years spent in the community in the third wave is equal to four or is below four. The way of recording tracking information is different between the waves: an indication of a split-off households is recorded separately in the third survey wave. I state that the household was tracked if it is indicated to be tracked locally⁴⁰ or over distance, or if it is indicated as a split-off household.

All 66 individuals with large differences between the reported and the computed distance are tracked between the waves. For 62 of them there is information on years lived in the community in the last survey wave, but only 35 of them report to be migrants. Although the individuals who report to be non-migrants have either reported or computed distance closer to zero (see Figure 2.4), since they all have been tracked, I believe they are migrants. Also, 20% of individuals for whom the computed distance between the first and the third waves matches the reported distance, and both distances are above five km, report to be non-migrants. With tracking, this is true only for 3% individuals, so I should not rely profoundly on self-reports to identify migrants. Overall, I can identify the individuals for whom the computed and the reported distances do not match as migrants since for all of them at least one distance is above five km and all of them were tracked between the waves either locally or over distance.

⁴⁰ Local tracking indicates that the new locations was within one hour of travel from the original location.

Figure 2.4. Self-reports for the cases when the difference between the computed and the reported distance is above 10 km: moved if years lived in the community is reported to be equal or below four in the third survey wave



Cases with missing reported distance

I use the computed distance between the waves to identify the migration status for individuals with missing reported distance. Since the offset distorts the computation of distance, I expect the threshold for the computed distance in identifying migrants to be different from that for the reported distance – although I would not be able to separate moves longer and shorter than five km (which is the threshold used in the survey data for the reported distance) when they are around five km. I find that for 99.33% of observations with computed distance below five km, the distance is also below 0.1 km. All individuals with computed distance from 0.1 km to five km were tracked either locally or over distance, and their reported distance is non-zero (except for one observation for which the reported distance equals zero and the computed distance equals 2.4 km).

Among 55 observations with missing reported distance, the computed distance is below 0.1 km for 37 observations, so I can identify these individuals as non-migrants with high level of confidence. For 11 observations of these 55, the computed distance is above five km, so I can identify them as migrants for the definition to be consistent with the one used by the survey team in the reported distance for other observations. It leaves me with seven observations out of 55 with missing reported distance, for whom the computed distance is from 0.1 km to five km and tracking information indicates migration. I decide to identify these observations as non-migrant for the consistency across definitions.

Cases with different results on migration status from the reported and the computed distance

When I use the threshold of 0.1 km for the computed distance from the previous subsection, I find that for 106 observations the results on migration status based on the reported and the computed distances do not match. Nine of these observations are covered in the second subsection since the difference in the reported and the computed distance is above 10 km, and one observation is covered in the third subsection (the household with reported distance equals zero and computed distance equals 2.4 km that was tracked). All of the other individuals are tracked, although they have computed distance below 0.1 km. The reported distance for them is from five km to 16 km, so for the majority of them the offset could be the reason for the results on migration status to not match. Since the reported distance for them indicates migration, I identify them as migrants.

APPENDIX 3. Classification of locations on the rural-urban spectrum

The main constructed definition for locations on the rural-urban spectrum I construct employs population density, built-up area density, and distance to the nearest urban location. It distinguishes cities, towns, peri-urban areas, and rural areas with high and low population density. I also expand the binary categorization into rural and urban areas used by the NBS. I split urban areas into towns and cities, and I split rural areas into areas with high and low population density. I cannot separate peri-urban areas for the NBS definition. Finally, I use cluster analysis with a wider set of variables. It allows me to isolate up to six location types.

Define “urban”

I construct the definition of “urban”, that would include cities and towns, based on the population density and the built-up area density, and then I use distance to an urban location to distinguish cities from towns. From the definitions of “urban” listed in the description of Table 6 of the Demographic Yearbook 2005 (UN, 2008), three, in Canada, China, and India, use population density as one of several criteria. Two definitions, in Canada and India, use a threshold of 400 people per square km, and I adopt it too. To account for other criteria, I use built-up area density with a threshold of eight percent.

As shown in Figure 2.5, the majority of households from areas with population density below 400 people per sq. km have built-up area density below 10%. Some households from areas with population density above 400 people per sq. km also have low built-up areas density. From Figure 2.6, I see the distribution in more detail bounding the built-up area density at 20%. I set a threshold for the built-up area density at 8% based on the distribution for households from areas with population density above 400 (pictures b, d, and f on Figure 2.6).

Figure 2.5. Scatter plots for households' population density and built-up area density and histograms for built-up area density

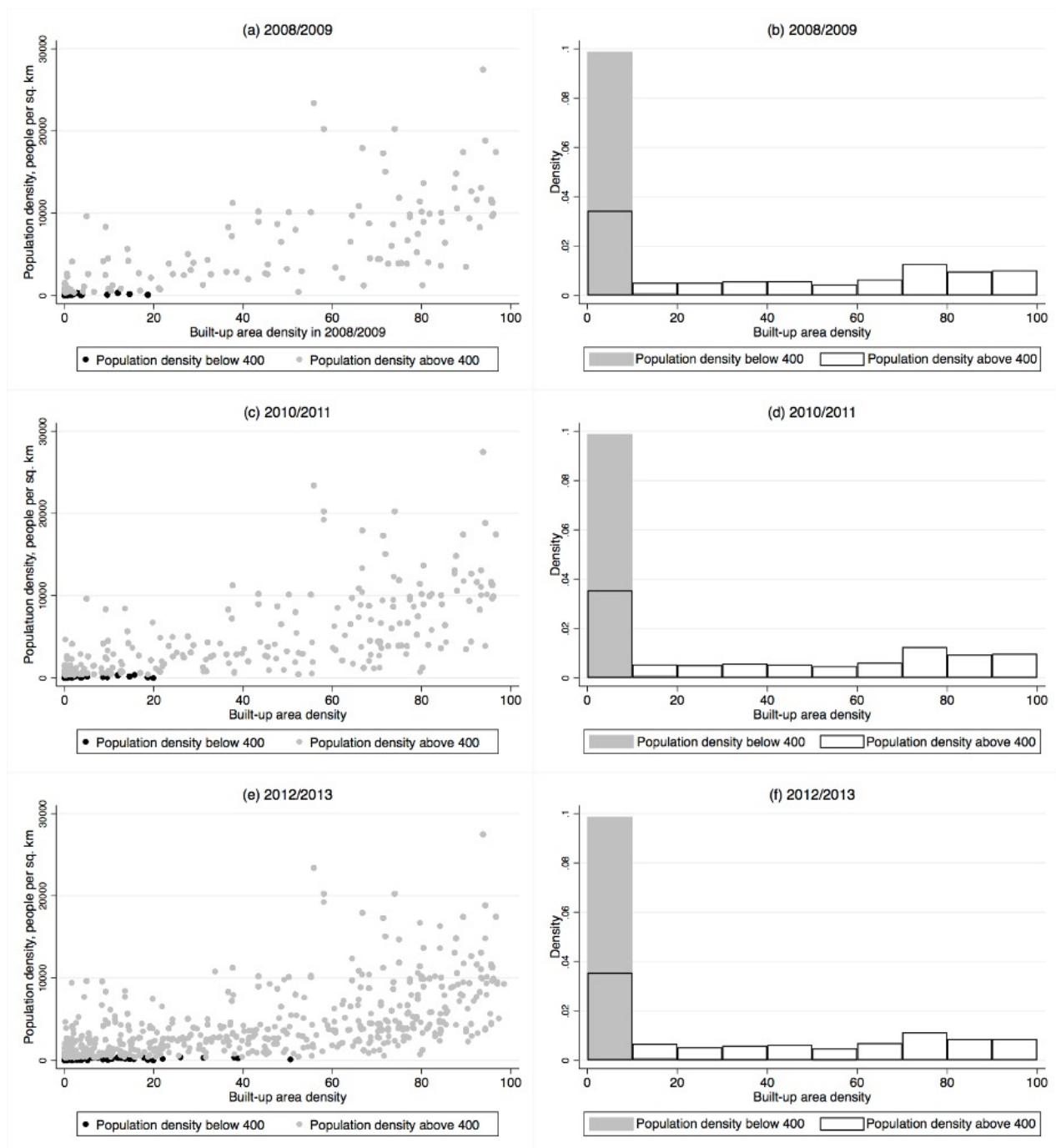
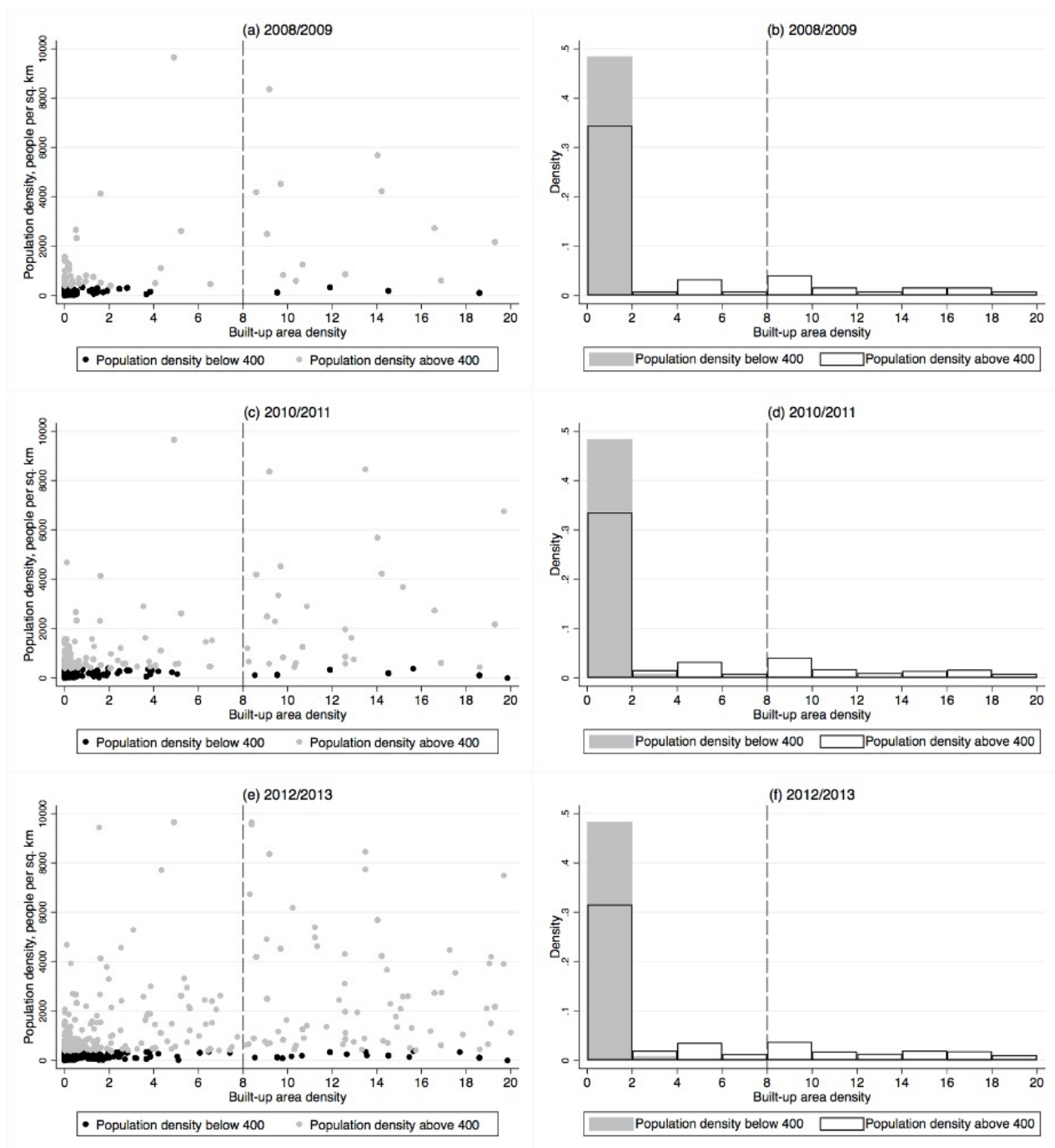


Figure 2.6. Scatter plots for households' population density and built-up area density and histograms for built-up area density, for built-up area density below 20%



Compared to a stricter threshold for the built-up area density, for example, 50% used by Mueller et al. (2019), the definition with the 8% threshold categorizes 311 new households as urban in 2008/2009 in addition to 539 categorized with 50% threshold; 369 new households in addition to 625 in 2010/2011; and 494 new households in addition to 730 in 2012/2013. These additional households categorized as urban with the 8% threshold have average population density of 3,500 people per square km (it ranges from 416 to 11,272 people per sq. km) and average built-up area density of 25% (it ranges from 8.1% to 49.8%). NBS definition categorizes around 89% of them as urban.

There is another variable that I could use to define “urban”, share of land under agriculture in one km radius of the household, but I opt for not using it for two reasons. First, it is provided in the set of geospatial information, hence it is missing for some households (275 households in the first survey wave, see Appendix 1). Second, the available information does not align well with the data on population density and built-up area density. In Table Table 2.14, I show how an increase in the built-up area density is associated with an increase in population density, while the share of land under agriculture does not change much. Another problem is outliers, for example, cases with population density of at least 3,000 people per square km, built-up area density above 90%, and other household characteristics suggesting that the household is urban, while the share of land under agriculture being 40-60%.

Table 2.14 also provides a better understanding of the threshold for population density. The majority of households living in areas with population density below 800 people per sq. km also have population density below 200 people per sq. km. A threshold that I use, 400 people per sq. km, leads to an additional 424 households to be classified as urban if they meet the criteria for built-up area density, compared to the number of households I would classify as urban if I

used a threshold of 200 people per sq. km. I see that the distribution of built-up area density for these 424 households does not differ much from the distribution among households with population density below 200 people per sq. km, but differs from the distribution among households with population density above 400 people per sq. km.

Table 2.14. Share of households in 2012/2013, by population density, built-up area density, and share of land under agriculture

Population density, people per square km	0 – 200 (2,778 households)		200 – 400 (424 households)		400 – 600 (275 households)		600 – 800 (139 households)	
Variable for quantiles	Built- up	Ag. land	Built- up	Ag. land	Built- up	Ag. land	Built- up	Ag. land
0% – 20%	100%	40%	99%	36%	93%	35%	91%	36%
20% – 40%	0%	22%	1%	17%	1%	28%	6%	22%
40% – 60%	0%	24%		23%	5%	22%	1%	15%
60% – 80%		12%		17%	0%	11%	1%	17%
80% – 100%		3%		6%		4%		10%

Note: “Built-up” stands for the built-up area density. “Ag. land” stands for the share of land under agriculture in one km radius from the household; this data comes from the LSMS and is computed by the survey team using the real coordinates of the household.

As a result, constructed definition of “urban” classifies more households as non-urban than the NBS definition does (see Table 2.15). Average population density for households classified as urban under the NBS categorization and as non-urban under the constructed definition is relatively high, ranging from 516 to 620 people per sq. km depending on the survey wave, but the average built-up area density is low, ranging from 1.0% to 1.2%. It could also be seen from Figure 2.6: the majority of households with population density above 400 people per square km and built-up area density below 20% fall into the group with built-up area density below 2%.

Households classified as rural under the NBS categorization and as urban under the constructed definition have high average population density, ranging from 874 to 6,072 people per square km, and high built-up area density, ranging from 14.0% to 42.7%. For households

classified as non-urban under both definitions, average population density is low, ranging from 139 to 145 people per sq. km, and average built-up area density is low, ranging from 0.1% to 0.2%. For households classified as urban under both definitions, average population density is high, ranging from 6,141 to 6,715 people per sq. km, and average built-up area density is also high, ranging from 52.3% to 54.7%.

Table 2.15. Comparison of the NBS categorization and the constructed definition of “rural”: urban and non-urban households

NBS categorization	Rural	Urban	Rural	Urban
Constructed definition	Non-urban	Non-urban	Urban	Urban
<i>2008/2009</i>				
Number of households	2039	376	24	826
Mean population density	139.3	607.9	874.1	6714.7
Mean built-up area	0.1	1.1	14.0	54.2
<i>2010/2011</i>				
Number of households	2401	529	227	766
Mean population density	144.9	516.1	6071.8	6145.6
Mean built-up area	0.2	1.0	42.7	52.3
<i>2012/2013</i>				
Number of households	3022	764	197	1027
Mean population density	144.1	620.2	3316.4	6140.9
Mean built-up area	0.2	1.2	36.9	54.7

Note: Constructed definition: “urban” if the population density is above 400 people per sq. km and the built-up area density is above 8%, “non-urban” otherwise. Sampling weights from the respective survey waves are applied.

Cities and towns

I set a threshold for distance to an “urban location” at 30 km to determine whether an urban area is a city or a town and to distinguish peri-urban areas (see the next sub-section). An “urban location” is defined as a town with population of at least 50,000 people according to the 2012 Population and Housing Census. I compute the distance using households’ coordinates provided in the dataset and the coordinates of town centers which I collect myself (they mostly point to crossroads; see the list of coordinates in Table 2.16).

Table 2.16. List of coordinates for cities and towns with population of at least 50,000 people in 2012

Name of the city/town	Latitude	Longitude	Name of the city/town	Latitude	Longitude
Dar es Salaam	-6.8	39.283333	Mtwara	-10.273611	40.183742
Mwanza	-2.51716	32.9	Kahama	-3.8375	32.595981
Zanzibar	-6.165	39.199	Kasulu	-4.573991	30.10804
Arusha	-3.374341	36.684006	Singida	-4.8154	34.75
Mbeya	-8.910475	33.455466	Njombe	-9.34747	34.769643
Morogoro	-6.824305	37.6633	Mpanda	-6.34289	31.071446
Tanga	-5.079678	39.098297	Masasi	-10.729718	38.805577
Kigoma	-4.882097	29.648458	Tunduma	-9.3	32.766667
Dodoma	-6.170821	35.741944	Makambako	-8.85	34.833333
Songea	-10.676357	35.64475	Babati	-4.208585	35.744697
Moshi	-3.34627	37.336203	Geita	-2.871757	32.230391
Tabora	-5.024937	32.807621	Handeni	-5.421927	38.025032
Iringa	-7.781029	35.693025	Lindi	-9.996728	39.714852
Musoma	-1.514314	33.800515	Sengerema	-2.650254	32.64347
Bukoba	-1.325727	31.810996	Bunda	-2.02189	33.872417
Kibaha	-6.784039	38.993489	Korogwe	-5.155833	38.450278
Sumbawanga	-7.95399	31.617671	Vwawa	-9.10998	32.941358
Shinyanga	-3.670218	33.426546	Mafinga	-8.300033	35.296458

I show the distribution of population density and built-up area density in relation to distance to the nearest urban location for both urban and non-urban areas in Figure 2.7 and Figure 2.8. I see how, for both urban and non-urban households, population density and built-up area density decrease as distance to the nearest urban location increases, though the effect for built-up area density in urban areas is less pronounced. From these observations, I set a threshold at 30 km, where the effect of proximity to an urban location on the two variables of interest seems to dissipate.

Figure 2.7. Scatter plots for households' population density and built-up area density for those with population density above 400 people per square km and built-up area density above 8%; by distance to the nearest town with population of at least 50,000

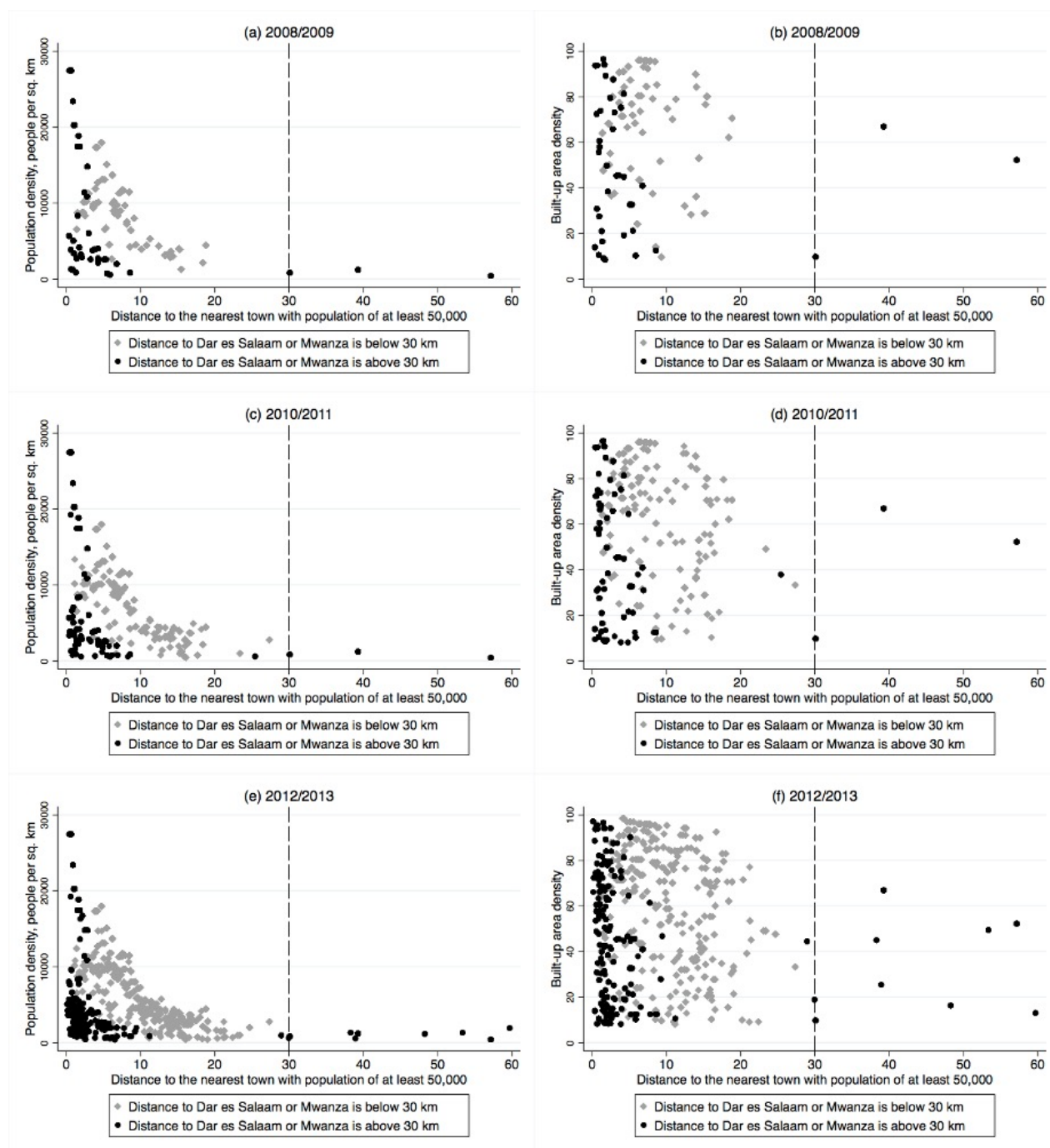


Figure 2.8. Scatter plots for households' population density and built-up area density for those with population density below 400 people per square km or built-up area density below 8%; by distance to the nearest town with population of at least 50,000

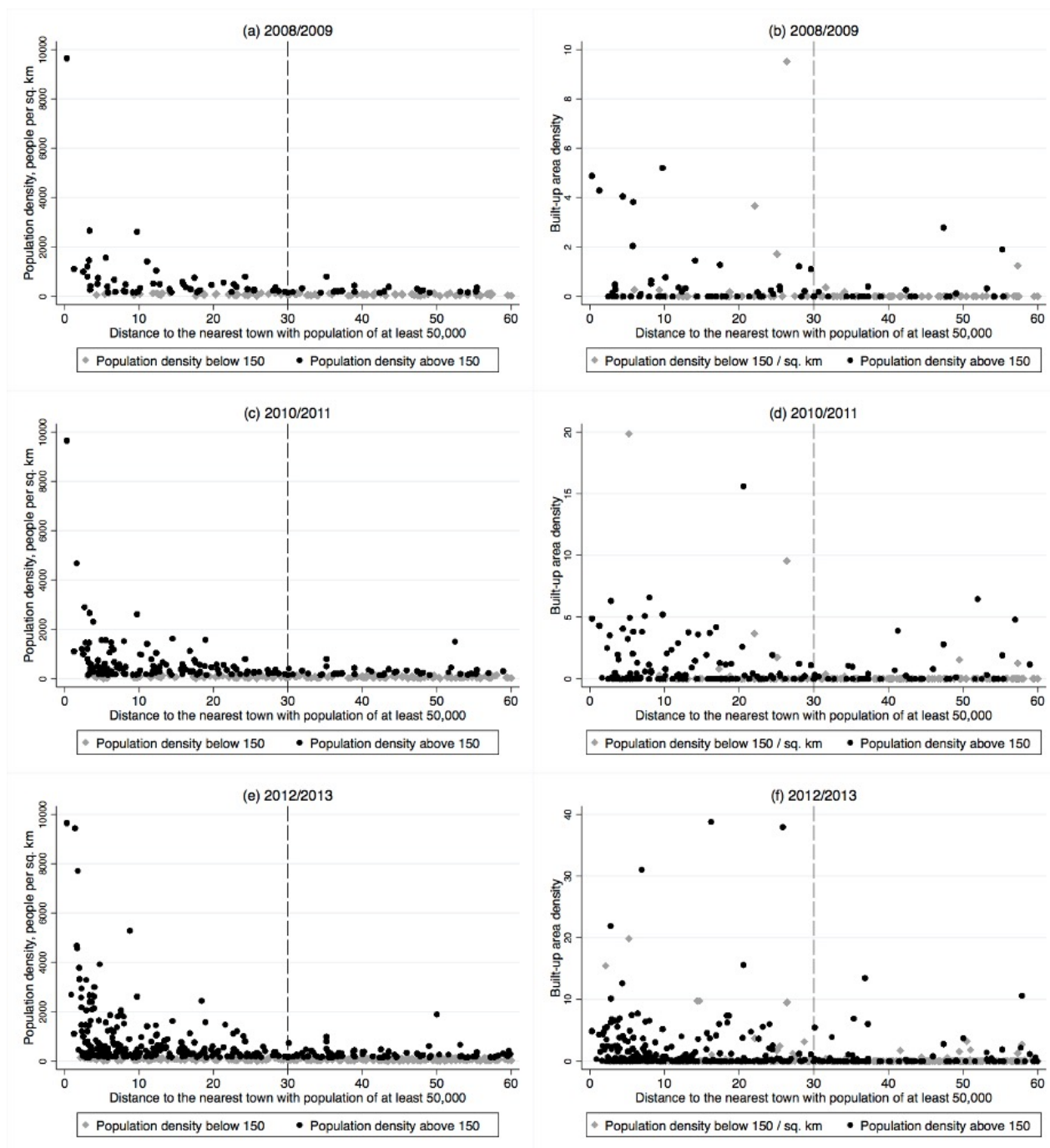


Table 2.17. Mean population density and built-up area density for seven largest urban locations

	NBS categorization			Constructed definition		
	Number of households in the sample	Mean population density	Mean built-up area density	Number of households in the sample	Mean population density	Mean built-up area density
<i>2008/2009</i>						
Dar es Salaam	483	8626	70	499	8339	68
Mwanza	24	4861	40	8	10625	88
Zanzibar	175	10764	60	167	11398	64
Arusha	24	7047	5	8	8369	9
Mbeya	8	768	21	8	768	21
Morogoro	24	3111	21	16	4207	29
Tanga	no data	no data	no data	no data	no data	no data
<i>2010/2011</i>						
Dar es Salaam	541	8116	68	578	7954	67
Mwanza	34	5201	40	14	10278	79
Zanzibar	71	2142	27	185	10693	61
Arusha	28	6931	6	9	8374	9
Mbeya	15	820	17	9	907	22
Morogoro	30	2899	18	16	4302	30
Tanga	3	2877	35	2	6229	79
<i>2012/2013</i>						
Dar es Salaam	770	7368	63	676	7613	65
Mwanza	49	4958	38	29	7696	61
Zanzibar	95	4074	40	195	6163	53
Arusha	37	6444	6	14	7596	10
Mbeya	17	1283	23	15	1588	32
Morogoro	45	2679	18	26	4036	32
Tanga	9	3334	43	8	5000	66

Note: Locations are listed in the order of total urban population based on the 2012 Population and Housing Census. For the constructed definition, households are considered living in the named urban location if they live within 30 km from its center. Sampling weights from the respective survey waves are applied.

To distinguish cities from towns, I compare average population density and built-up area density for seven cities with the largest population for both the NBS categorization and the constructed definition (see Table 2.17). Four cities with the largest population (Dar es Salaam,

Mwanza, Zanzibar, and Arusha) also have the highest average population density that is far above the average population density of the other three cities. All four of these locations except Arusha have high average built-up area density. Note that Tanga also have high average built-up area density, although I do not have that many observations for Tanga. The information on Zanzibar is inconsistent across the waves when I look at the NBS definition. Hence, I categorize Dar es Salaam and Mwanza as cities, and identify other urban locations as towns. Cockx, Colen, and De Weerd (2018) use the same categorization. From Figure 2.7, I see that the relationship between distance to city or town and population density and built-up area density is different for cities and towns: cities' effect spreads over larger area.

For the NBS definition, I distinguish cities from towns using the region and district identifiers. I categorize urban households in all districts of Dar es Salaam and in Nyamagana and Ilelela districts in Mwanza region as living in a "city", and all other urban households as living in a "town". For the constructed definition, I define an urban area (an area with population density above 400 people per sq. km and built-up area density above 8%) as a "city" if it is located within 30 km from Dar es Salaam or Mwanza, and I define it as a "town" otherwise. The results from the two definitions are compared in Table 2.18.

From Table 2.18, I see that for many households the distinction between cities and towns aligns between the two definitions. Average population density and built-up area density of households living in cities is almost twice that of households living in towns. Nevertheless, there are differences among households categorized as "non-urban" under one definition and "urban" by another. I could compare these results to Table 2.15: households are more likely to be classified as living in a town on this step if they were classified into living in an urban area by one definition and into living in a non-urban area by the other definition on the previous step.

These households have much higher average population density and built-up area density if they are categorized as living in a town by the constructed definition and as living in a rural area by the NBS categorization, than if the definitions are inverted.

Overall, the number of households living in cities aligns between the two definitions, while the number of households categorized as living in cities relative to the number of households living in towns is much higher for the constructed definition. As I noted from Table 2.15, the NBS categorization classifies more households as “urban” than the constructed definition does. Here, I see that most of these households are further classified into living in towns according to the NBS categorization. Under the constructed definition, they are classified as living in “non-urban” areas, and most of them will be classified as living in “peri-urban” areas on the next step.

Table 2.18. Comparison of the NBS categorization and the constructed definition: households living in cities and towns

NBS categorization	Town	City	Town	City	Rural	Rural	Town	City
Constructed definition	Town	Town	City	City	Town	City	Non-urban	Non-urban
<i>2008/2009</i>								
Number of households	327	0	0	499	16	8	368	8
Mean population density	4196.1	-	-	8480.1	731.9	2630.6	610.6	525.4
Mean built-up area	32.7	-	-	69.3	13.2	24.2	1.1	0.4
<i>2010/2011</i>								
Number of households	207	3	6	550	191	36	507	22
Mean population density	3108.4	1774.4	6907.9	8149.2	6135.4	5677.8	515.5	537.1
Mean built-up area	28.2	19.5	73.1	68.0	40.1	58.9	1.0	0.6
<i>2012/2013</i>								
Number of households	325	5	6	691	189	8	641	123
Mean population density	3159.3	1751.1	6022.4	7650.9	3022.0	5661.3	613.7	711.5
Mean built-up area	34.0	19.1	51.4	65.2	33.2	66.6	1.1	2.7

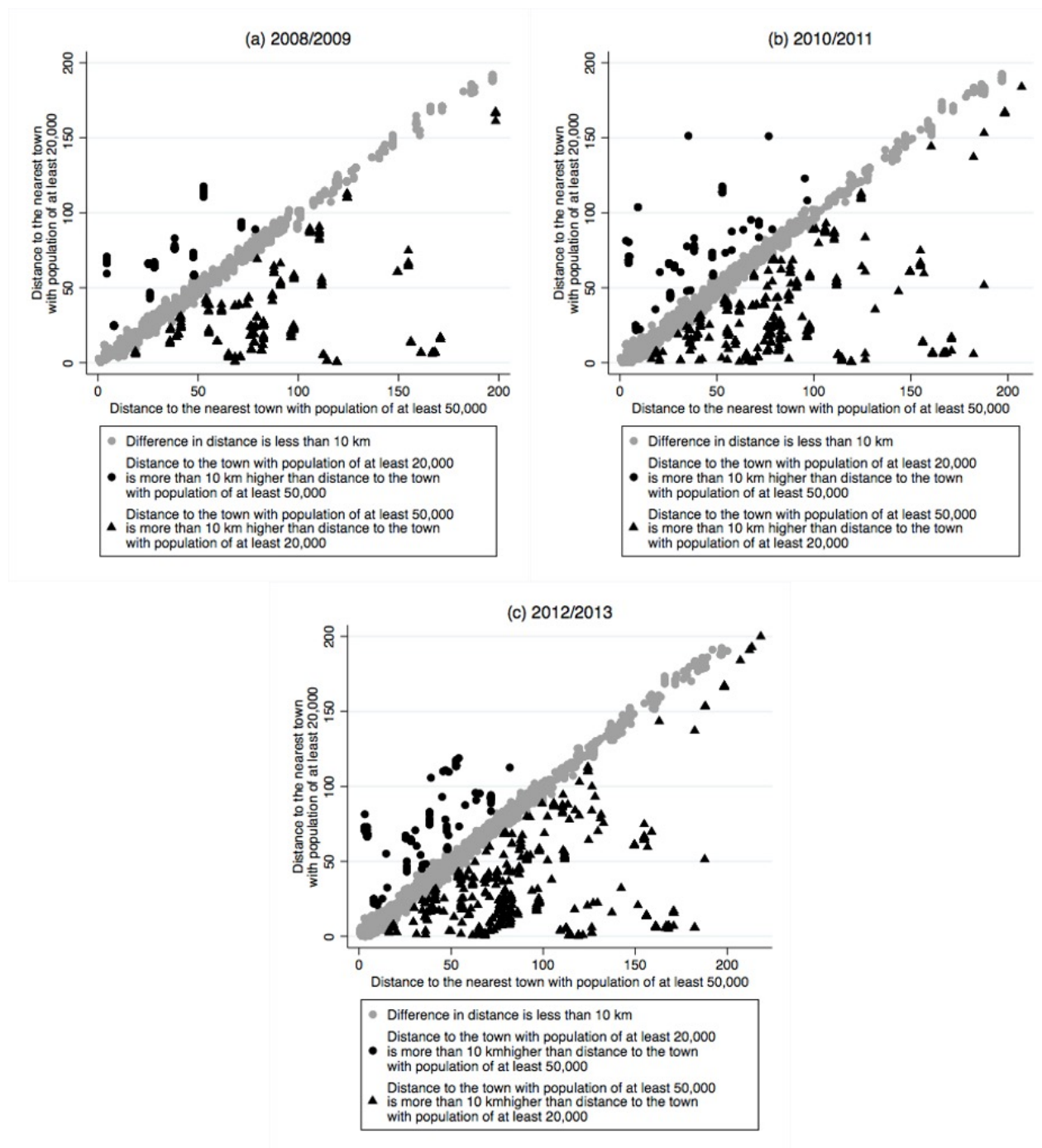
Note: NBS categorization: “city” if identified as “urban” in all districts of Dar es Salaam or in Nyamagana and Ilemela districts of Mwanza; “town” if identified as “urban” in any other district. Constructed definition: “urban” if population density is above 400 people per sq. km and built-up area density is above 8%; “city” if urban and within 30 km of Dar es Salaam or Mwanza; “town” if urban and not within 30 km of Dar es Salaam or Mwanza. Sampling weights from the respective survey waves are applied.

Define “peri-urban”

I define peri-urban areas for the constructed definition as non-urban areas with population density above 150 people per sq. km located within 30 km of towns with population of at least 50,000. The choice of this threshold for distance is described in the previous sub-section (see Figure 2.8), but I discuss it in more details related to the definition of “peri-urban” in this sub-section. The distinction by population density is shown on Figure 2.8: many non-urban households with relatively high population density and built-up area density are located within 30 km from an urban location with at least 50,000 inhabitants, although there are outliers (see graphs b, d, f). In this sub-section, I first compare my measure of distance with the distance to the nearest town with population of at least 20,000 that is provided in the dataset, and then discuss the threshold for distance and explain the choice of the threshold for population density.

The geographical coordinates are provided at the level of enumeration areas with a random offset that could be up to 10 km for rural households (though it is up to five km for 99% of them) and up to two km for urban households, and it includes many households identified as living in peri-urban areas by the constructed definition. The dataset also includes information on the distance to the nearest population center with at least 20,000 inhabitants which was computed by the survey team for each household using the original coordinates, but, as I discussed in Appendix 1, it is missing for some households (275 households in the first wave of survey, seven in the second, and 22 in the third survey wave respectively).

Figure 2.9. Households' distance to the nearest town with population of at least 20,000 and distance to the nearest town with population of at least 50,000, for households with population density below 400 people per square km or built-up area density below 8%

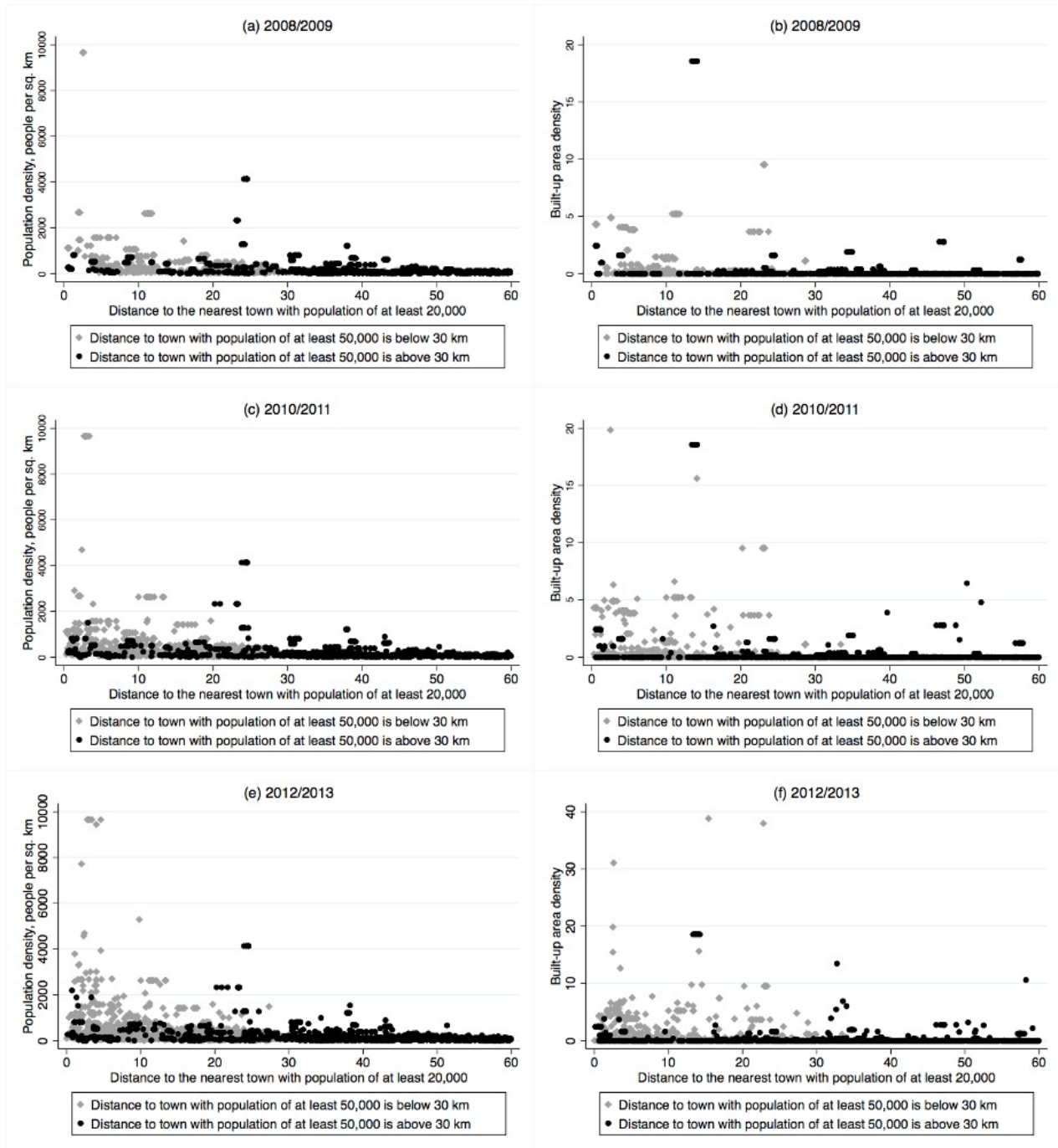


On Figure 2.9, I show how my measure of distance to towns with population of at least 50,000 people relates to the one provided in the dataset for the distance to towns with populations of at least 20,000 people. For most households, the difference in distance is less than 10 km.⁴¹ For 425 non-urban (according to the constructed definition) households in the first wave, 522 non-urban households in the second wave, and 642 non-urban households in the third wave the distance to town with population of at least 50,000 is more than 10 km higher than the distance to town with population of at least 20,000. So, for these households there is a town with population of 20,000 – 50,000 that is at least 10 km closer than a town with population of at least 50,000. For most of these household, the town with population of at least 50,000 is further than 30 km away, which means they cannot be defined as “peri-urban” under my definition even if the population density in that area is above 150 people per sq. km.

On Figure 2.10, I show the same graphs as on Figure 2.8, where I set the threshold for distance to distinguish towns and cities and peri-urban areas to be 30 km, but for the distance to the nearest town with population of 20,000 instead of 50,000. Non-urban households that are located within 30 km distance from towns with population of at least 50,000 (gray rhombi on the graphs) exhibit the same patterns as on Figure 2.8 for the distance to the nearest town with population of 20,000: population density and built-up area density decline as the distance to town increases. For households that are located further than 30 km from towns with population of at least 50,000 but are near towns with population from 20,000 to 50,000 the relationship between density and distance is not that clear.

⁴¹ I use a threshold of 10 km here as it is the highest possible offset, as I cannot estimate the average distortion introduced by averaging households’ coordinates at the enumeration area level.

Figure 2.10. Scatter plots for households' population density and built-up area density for those with population density below 400 people per square km or built-up area density below 8%; by distance to the nearest town with population of at least 20,000



I examine the distance to the nearest town with population of at least 20,000 in more detail in Table 2.19, looking at households that are further than 30 km away from any town with population of at least 50,000. Population density and built-up area density are, on average, higher for households that are located within five km from the town with population of at least 20,000, though there are outliers (at 20-25 km and further for population density and at 10-15 km and further for built-up area density), which also could be seen on Figure 2.9. Hence, I decide to use distance to the nearest town with population of at least 50,000 to define “peri-urban”.

Table 2.19. Population density and built-up area density for households located further than 30 km from a town with population of at least 50,000; for households with population density below 400 people per sq. km or built-up area density below 8%

Distance to the nearest town with population of at least 20,000	2008/2009			2010/2011			2012/2013		
	Number of households	Mean population density	Mean built-up area dens.	Number of households	Mean population density	Mean built-up area dens.	Number of households	Mean population density	Mean built-up area dens.
0 - 5 km	36	371.8	0.8	53	337.9	0.7	68	371.7	0.7
5 - 10 km	52	133.3	0.0	65	145.8	0.0	79	135.4	0.0
10 - 15 km	38	156.7	6.4	42	158.7	6.2	49	153.0	6.5
15 - 20 km	50	136.0	0.1	57	155.3	0.1	74	135.4	0.1
20 - 25 km	59	280.3	0.1	71	312.2	0.1	79	275.8	0.1
25 - 30 km	40	177.2	0.0	56	145.4	0.0	81	142.4	0.0
over 30 km	1326	90.7	0.2	1718	95.6	0.2	2256	93.9	0.2

Note: Sampling weights from the respective survey waves are applied. Data on distance to the nearest town with population of at least 20,000 people is from the LSMS: it is computed by the survey team using the real coordinates of the households.

On Figure 2.10, for 78 non-urban household in the first wave, 108 non-urban households in the second wave, and 128 non-urban households in the third wave the distance to town with population of at least 20,000 is more than 10 km higher than the distance to town with population of at least 50,000. For these households, the median of population density is 60-80 people per sq.

km depending on the survey wave, the 75th percentile of population density is 140-165 people per sq. km, the 95th percentile of population density is 245 people per sq. km, and the 95th percentile for built-up area density is 1.72%. Hence, most of these households would have been categorized as rural even if I used distance to a town with population of at least 20,000. NBS categorization cannot help either: among these households, those categorized as urban are located further than 45 km away from any town with population of at least 20,000.

Allen (2018) defines peri-urban districts as those located within 10 km of any urban district, which results in the inclusion of households located further than 10 km away from cities. In the 1998 Tanzania Peri-Urban Survey, the interviewed households are located within 20 km from the city perimeter for six cities, five of which had the population of at least 140,000 (2002 Census) and one had the population of less than 30,000 (2002 Census). Using this dataset, Lanjouw, Quizon, and Sparrow (2001) find that the share of income coming from agriculture is different for rural households and peri-urban households near “relatively dynamic cities”, Dar es Salaam and Arusha. Kombe (2005) and Mapunda, Chen, and Yu (2018) look at peri-urban areas near Dar es Salaam: the studied locations are within 18-20 km and 24-28 km of the city respectively. The peri-urban settlement studied by Msigala et al. (2017) is located within 25 km of Morogoro, a city with population of at least 305,000 according to the 2012 Census. Hence, my threshold of 30 km is slightly higher than those used in the literature for Tanzania.

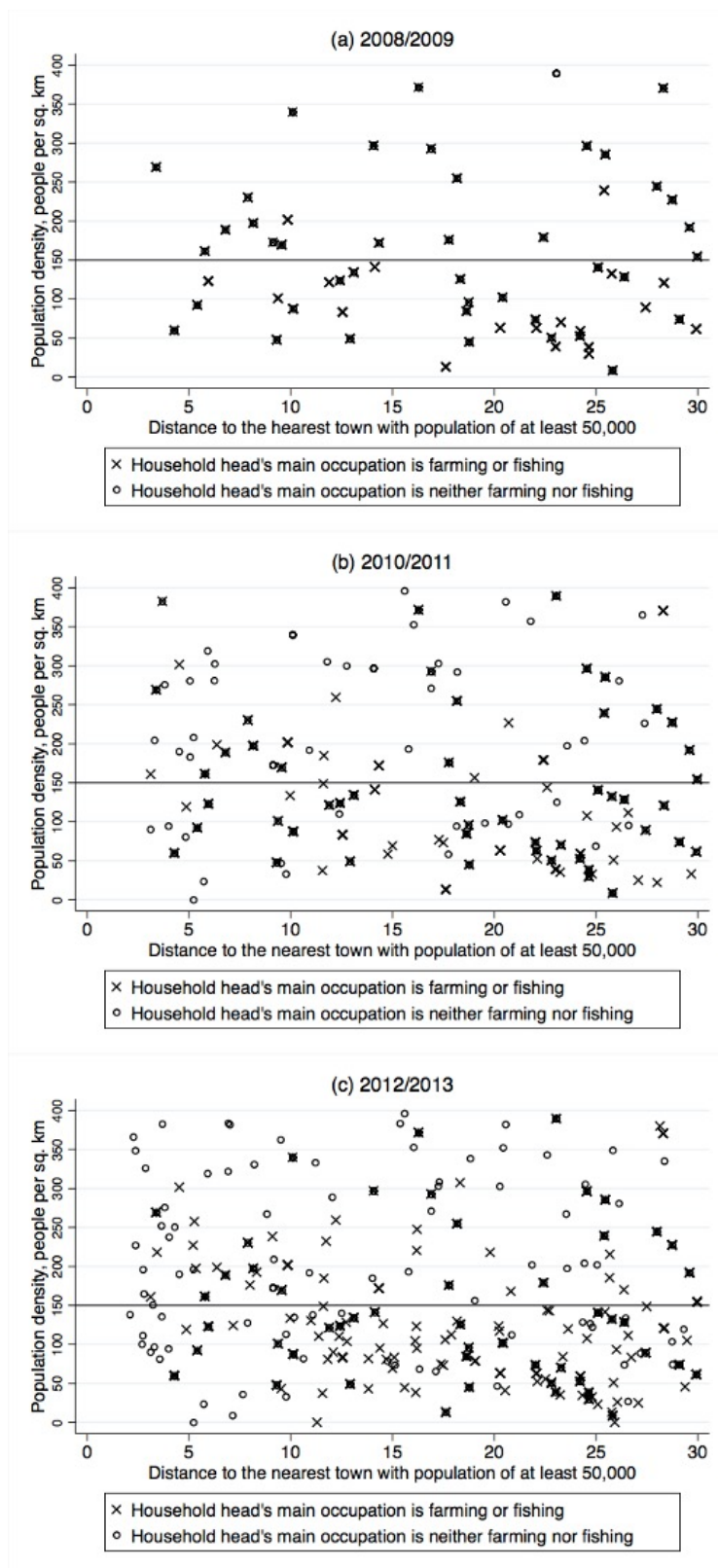
I set a threshold for population density at 150 people per sq. km. Mueller et al. (2019) use the same threshold along with distance to town and built-up area density. For the distance to town, they use an indicator for the travel time to a town with population of at least 20,000 to be less than one hour. For the built-up area density, they use the threshold of 50%. Muzzini and

Lindeboom (2008) use the same threshold of 150 people per sq. km to distinguish rural and urban areas for the 1998 and 2002 census data for Tanzania.

In Figure 2.11, I show the distribution of population density and distance to town for non-urban households that are located near towns and hence could be considered peri-urban.

Reference lines (in black) represent the threshold for population density. I separate households where the head's main occupation is farming or fishing as I expect an average share of household heads with main occupation in farming or fishing to be higher in rural areas. While households where head's main occupation is neither farming nor fishing are scattered almost uniformly across the presented range of population density and distance to town, I observe more households where the head's main occupation is farming or fishing to live in areas with lower population density that are located further from towns.

Figure 2.11. Population density and distance to town for households with population density below 400 people per square km living within 30 km of a town with population of at least 50,000



In Table 2.20, I compare the average built-up area density, the share of households in which the head's main occupation is farming or fishing, and the average share of income coming from agriculture by population density bins for non-urban households located within 30 km from towns with population of at least 50,000. The results for the built-up area density are not conclusive: it increases with population density when population density is below 200 people per sq. km, and it decreases with population density when population density is from 200 to 300 people per sq. km. Both the share of households in which the head's main occupation is farming or fishing and the average share of income coming from agriculture remain high when population density is below 150 people per square km. Then they decrease drastically when the population density is 150-200 people per sq. km but increase again when the population density is 200-250 people per sq. km (for some survey waves, even to the same levels as when population density is below 150 people per sq. km). They decrease again when the population density is above 250 people per sq. km. Hence, I can use a threshold of 250 people per sq. km for a robustness check. It will lead to a recategorization of 120-176 households (depending on the survey wave) as rural instead of peri-urban.

Overall, around 17% of non-urban households are categorized as peri-urban. I define rural households as non-urban households that are either located further than 30 km away from any town with population of at least 50,000 or located within 30 km of such town and have population density lower than 150 people per sq. km. Among rural households, 86% are defined as "rural" by the NBS categorization (see Table 2.21). Among peri-urban households, 52-71% are defined as "rural" by the NBS categorization (depending on the survey wave). The average population density of peri-urban households is higher than the average population density of rural households.

Table 2.20. Non-urban households located within 30 km of a town with population of at least 50,000

Population density, people per sq. km	2008/2009	2010/2011	2012/2013
<i>Number of households</i>			
0 - 50	64	75	101
50 - 100	128	151	180
100 - 150	96	120	160
150 - 200	80	101	116
200 - 250	40	49	60
250 - 300	48	57	66
300 - 350	8	12	24
350 and above	232	303	391
<i>Average built-up area density</i>			
0 - 50	0.05	0.21	0.23
50 - 100	0.13	0.16	0.15
100 - 150	0.36	0.31	0.41
150 - 200	0.49	0.52	0.66
200 - 250	0.39	0.37	0.43
250 - 300	0.23	0.18	0.32
300 - 350	0.78	0.28	4.67
350 and above	0.56	0.63	0.91
<i>Share of households in which the main occupation of the household head is farming or fishing</i>			
0 - 50	87%	81%	80%
50 - 100	88%	79%	77%
100 - 150	90%	82%	71%
150 - 200	73%	70%	73%
200 - 250	86%	80%	81%
250 - 300	65%	55%	56%
300 - 350	13%	42%	34%
350 and above	44%	42%	39%
<i>Average share of income coming from crops and livestock</i>			
0 - 50	61%	68%	57%
50 - 100	69%	64%	50%
100 - 150	66%	57%	51%
150 - 200	46%	42%	44%
200 - 250	58%	58%	43%
250 - 300	45%	40%	27%
300 - 350	5%	21%	29%
350 and above	32%	27%	17%

Note: Households are considered to be non-urban when they live in an area with population density below 400 people per sq. km or built-up area density below 8%. Sampling weights from the respective survey waves are applied. Share of households in which the head's main occupation is farming or fishing is computed using the LSMS data. Data on the average share of income coming from crops and livestock is from the Rural Income Generating Activities dataset.

Table 2.21. Comparison of the NBS categorization and the constructed definition for non-urban households according to the constructed definition

NBS categorization	Rural	Urban	Rural	Urban
Constructed definition	Rural	Rural	Peri-urban	Peri-urban
<i>2008/2009</i>				
Number of households	1751	256	288	120
Mean population density	90.7	206.8	508.3	1395.0
Mean built-up area	0.1	1.2	0.3	0.9
<i>2010/2011</i>				
Number of households	2069	339	332	190
Mean population density	98.1	183.6	500.0	1126.4
Mean built-up area	0.2	1.0	0.3	0.9
<i>2012/2013</i>				
Number of households	2680	449	342	315
Mean population density	94.5	182.7	546.9	1246.9
Mean built-up area	0.1	1.0	0.3	1.5

Note: Constructed definition: “non-urban” if the population density is below 400 people per sq. km or the built-up area density is below 8%; “peri-urban” if non-urban and located within 30 km of a town with population of at least 50,000. Sampling weights from the respective survey waves are applied.

Table 2.22. Distribution of population density for rural households: share of households with population density below certain thresholds

Thresholds for population density, people per sq. km	2008/2009		2010/2011		2012/2013	
	Constr.	NBS	Constr.	NBS	Constr.	NBS
50	33%	30%	32%	30%	34%	31%
75	53%	48%	51%	46%	53%	47%
100	70%	64%	68%	61%	69%	62%
125	79%	72%	78%	70%	78%	70%
150	85%	78%	84%	76%	85%	77%
175	88%	82%	86%	80%	88%	81%
200	91%	84%	89%	82%	90%	83%

Note: “Constr.” stands for the constructed definition. “NBS” stands for the NBS categorization. Constructed definition: “rural” if the population density is below 400 people per sq. km or the built-up area density is below 8% and the distance to the nearest town with population of at least 50,000 is above 30 km, or if the distance is less than 30 km and the population density is below 150 people per sq. km. Sampling weights from the respective survey waves are applied.

Rural households with high and low population density

I distinguish rural areas with high and low population density for both the constructed definition and the NBS categorization. To connect the definitions, I set the same threshold for population density, at 100 people per sq. km, for both of them. I compare the distribution of population density in Table 2.22. With the threshold of 100 people per sq. km, 61-64% of rural households according to the NBS categorization and 68-70% of rural households according to the constructed definition are identified as living in areas with low population density.

Table 2.23. Comparison of the NBS categorization and the constructed definition for rural households with high and low population density

NBS categorization	Low	High	High	Urban	Urban
Constructed definition	Low	High	Non-rural	Low	High
<i>2008/2009</i>					
Number of households	1215	536	312	104	152
Mean population density	52.4	193.7	538.8	58.5	324.5
Mean built-up area	0.0	0.3	1.4	0.0	2.1
Mean share of income from crops and livestock	67.8%	68.6%	51.7%	20.8%	27.2%
<i>2010/2011</i>					
Number of households	1397	672	560	174	165
Mean population density	52.4	211.6	1536.1	58.9	291.7
Mean built-up area	0.1	0.4	8.2	0.0	1.9
Mean share of income from crops and livestock	66.0%	60.3%	41.5%	35.6%	21.3%
<i>2012/2013</i>					
Number of households	1811	869	539	207	242
Mean population density	51.1	204.0	1005.4	56.2	320.3
Mean built-up area	0.1	0.4	6.4	0.0	2.1
Mean share of income from crops and livestock	58.8%	53.0%	33.4%	30.1%	27.1%

Note: “Low” stands for low population density (below 100 people per sq. km). “High” stands for high population density (above 100 people per sq. km). There are no observations of low-density areas under the NBS categorization that would be identified as non-rural areas under the constructed definition. Constructed definition: “rural” if the population density is below 400 people per sq. km or the built-up area density is below 8% and the distance to the nearest town with population of at least 50,000 is above 30 km, or if the distance is less than 30 km and the population density is below 150 people per sq. km; “non-rural” if the population density is above 400 people per sq. km and the built-up area density is above 8%, or if the population density is below 400 people per sq. km or the built-up area density is below 8% and the distance to the nearest town with population of at least 50,000 is below 30 km and the population density is above 150 people per sq. km. Sampling weights from the respective survey waves are applied.

In Table 2.23, I report the average population density, the built-up area density, and the share of income coming from agriculture for different types of rural households and compare the results for the constructed definition and the NBS categorization. For the households for which the types match between the two definitions, the reported variables meet the expectations for rural areas: the average built-up area density is below 0.5%, the average share of income coming from crops and livestock is above 50%, and the average population density for low-density areas is 51-52 people per sq. km, and for high-density areas it is 194-212 people per sq. km.

There are no observations defined as low-density rural areas with the NBS categorization that would be identified as non-rural areas with the constructed definition, but there are observations defined as high-density rural areas with the NBS categorization that are identified as non-rural areas with the constructed definition. These households have high average population density (539-1536 people per sq. km) and high average built-up area density (1-8%), while the average income coming from agriculture for them is relatively low (33-53%). The majority (81-92%) of these households are defined as living in peri-urban areas by the constructed definition.

Households defined as living in urban areas with the NBS categorization and living in rural areas with the constructed definition have the same trends in the average population density and the built-up area density as households defined as rural with both definitions, although the average population density is higher for the former. At the same time, the average share of income coming from agricultural activities is drastically different: it ranges from 21% to 36% for households whose location type varies depending on the definition, and it is much smaller than for the households identified as rural with both definitions. I also find that these households

differ from the households with matched identifications in the distance from their location to the nearest town.

Households defined as urban with the NBS categorization and high-density rural under the constructed definition are more likely to be located near towns with population of at least 20,000 (for the available data, see the description of data limitations in Appendix 1): 39-52% of the households with mixed identifications are located within 30 km of a town while only 25-27% of households with matched identifications (and defined as high-density rural) are located within 30 km of a town. On the other hand, only 0-9% of households with mixed identifications are located within 30 km of a town with population of at least 50,000 while 17-19% of households with matched identifications are located within 30 km of a town. For the households defined as low-density rural with the constructed definition the difference is smaller: 27% (4-11%) of households with mixed identifications and 22% (16-17%) of households with matched identifications are located within 30 km of a town with population of at least 20,000 (50,000). Hence, it seems like some households with mixed identifications are located near or in towns with population of 20,000-50,000 that are not a part of my definition of “peri-urban areas” as I only look at proximity to larger towns.

Constructed definition and NBS categorization

I compare the average population density, the built-up area density, and the share of income coming from crops and livestock for the 2012/2013 survey wave for all categories for the constructed definition and the NBS categorization in Table 2.24. Peri-urban areas differ from rural and urban areas: the average population density there is higher than in rural areas but not as high as in urban areas, while the average share of income coming from agriculture there is lower than in rural areas but not as low as in urban areas.

Table 2.24. Summary statistics for the constructed definition and the NBS categorization for the 2012/2013 survey wave

NBS categorization	Low-density rural	High-density rural	Town	City
<i>Constructed definition: low-density rural</i>				
Number of households	1811		198	9
Mean population density	51.1		56.5	20.3
Mean built-up area density	0.1		0.0	0.0
Mean share of income coming from crops and livestock	58.8%		30.1%	25.5%
<i>Constructed definition: high-density rural</i>				
Number of households		869	223	19
Mean population density		204.0	317.3	403.7
Mean built-up area density		0.4	2.2	0.7
Mean share of income coming from crops and livestock		53.0%	27.8%	8.6%
<i>Constructed definition: peri-urban</i>				
Number of households		342	220	95
Mean population density		546.9	1316.2	795.4
Mean built-up area density		0.3	1.3	3.2
Mean share of income coming from crops and livestock		35.0%	14.9%	1.5%
<i>Constructed definition: town</i>				
Number of households		189	325	5
Mean population density		3022.0	3159.3	1751.1
Mean built-up area density		33.2	34.0	19.1
Mean share of income coming from crops and livestock		27.4%	5.9%	1.4%
<i>Constructed definition: city</i>				
Number of households		8	6	691
Mean population density		5661.3	6022.4	7650.9
Mean built-up area density		66.6	51.4	65.2
Mean share of income coming from crops and livestock		10.2%	3.0%	2.8%

Note: Sampling weights from the 2012/2013 survey wave are applied.

Cluster analysis definition of “rural”

Cluster analysis is suggested by Potts (2017a) to be an alternative to the binary local classification (NBS definition, in my case). I use the following variables for cluster analysis: (i)

share of households in the district⁴² with the floor made of concrete, cement, tiles, or timber, (ii) district average time to get water from the source of drinking water to the dwelling in minutes, (iii) district average share of household income coming from farm activities (from the Rural Income Generating Activities dataset⁴³), (iv) district average of the logarithm of population density⁴⁴ (v) district average distance to the nearest road, and (vi) district average distance to the nearest town with the population of at least 50,000 people.

I use averages across households for the variables reflecting access to amenities because amenities might not only point to a more urbanized place, but also be an indication of wealth of a particular household. Higher share of land under agriculture might be both a sign of a rural area and an indication of the remoteness of the household even if it's located in a more densely populated area, hence I use district average for this variable as well. I opt for using district averages instead of ward or enumeration area averages. When people move and are being tracked in the subsequent survey waves, there appear cases when migrants are the only representatives of their enumeration area or ward. Hence, a ward average for a variable simply equals to a variable measured for a particular household in those cases, and I want to avoid that.

I focus on the following observations to prove the point of using district averages instead of ward averages. In the first wave of survey, there are at least seven households in each ward, and there are eight households in most wards. In the third wave of survey, 432 households are the only households in their respective wards. Looking at my constructed definition of “rural”, I see that, among these 432 households, high-density rural and peri-urban areas are over-

⁴² For this and other variables used for cluster analysis, I apply sample weights when computing district averages.

⁴³ <https://www.fao.org/economic/riga>

⁴⁴ For observations with zero population density, the value of the logarithm is replaced with 0.01 after a check for observations with population density close to zero.

represented in comparison to the overall sample, while cities are under-represented. In the last wave of survey, there is only one household that is the only household in its district.⁴⁵ I do not see any peculiar connection between the position of the district on the rural-urban spectrum and the number of households I observe there. Also, the differences between the district average and the ward and enumeration area averages computed for the first survey wave are small. On the other hand, one of the disadvantages of using district averages instead of ward or enumeration area averages is the loss of precision.

Following the process outlined by Brusco et al. (2017), I run multiple iterations of k-medians clustering using random starting points.⁴⁶ To find the best partitions, I apply adjusted Rand index that measures the agreement between partitions correcting for the chance of random partition (Halpin, 2017). I perform k-medians clustering to separate two groups, “rural” and “urban”, for the first survey wave. Separation into three groups adds a group which is more urbanized than “rural” but much less urbanized than “urban” according to the variables I look at. Cluster analysis performed for four groups does not produce results with clear differences between the two categories situated in the middle, although it still produces a clear distinction of “the most rural” and “the most urban” locations. Hence, even though the results are enough to distinguish between “rural”, “urban”, and the in-between location, I use the partition into two groups for the sake of consistency with my previous definitions. For the last survey wave, I can produce a partition into up to six groups. A more detailed partition, which would group locations into seven groups, appears to be much weaker than the partition into six groups.

⁴⁵ This household entered the survey sample because of a migrant who was 11 years old at baseline, hence, this household is not included in my sample.

⁴⁶ One of the limitations of the cluster analysis definition I use is the low number of iterations ran, which is only 125 iterations. Among this low number of restarts, in general, I see some agreement between partitions: the adjusted Rand index for the partition into 6 groups is 0.64.

APPENDIX 4. Attrition

The issue of attrition is very important for studies of migration as the individuals not found by the survey team could potentially be migrants. In my study, I define an individual to be lost due to attrition if this individual was listed as a household member in the first wave of survey and had information but was not listed in the third wave of survey or had no information. “No information” cases occur when the individual is listed, but no information other than age, gender, relationship to the household head, and time spent in the household is provided⁴⁷. Under the assumption that all the deaths in the household are recorded in the respective section of the questionnaire, the four other potential sources of individuals’ attrition remaining in the LSMS for Tanzania are: (1) movements of individuals or households that resulted in the loss of track to the new location, (2) migration abroad, since such individuals were not subject to tracking, (3) temporary migration of an individual: an individual is listed as a household member but has not been present in the household for a significant amount of time in the past year (and hence has no information recorded) and was not tracked by the survey team as a migrant, and (4) migration of children below age 15 both to other locations in Tanzania and abroad since individuals below age 15 were not subject to tracking.

In 2008/2009, there were 3,196 individuals of age 15-34 living in rural areas according to the NBS definition of “rural”. Out of them, 88 had no information available. Out of the remaining 3,108 individuals, 2,888 were listed as household members in 2012/2013, out of

⁴⁷ In the first survey wave, the information was not collected for individuals listed as household members and having the answer “No” to the question “For the last 12 months has [Individual’s name] stayed in this household for three months or more?”, and for whom there is missing information on how many months the individual was away from the household during the past 12 months. For 80 individuals there is a discrepancy in the answers: although the answer to the first question was “No”, the information was collected, and the number of months spent away from the household ranges from zero to 12 months (for 53 individuals the number of months is at least ten). In the third survey wave, there is no discrepancy between the answers: the information was not collected for individuals listed as household members and having the answer “No” to the question “For the last 12 months has [Individual’s name] stayed in this household for three months or more?”.

whom 2,857 individuals had information in 2012/2013. It means that 220 individuals had information in 2008/2009 but were not listed in 2012/2013, and 31 individuals had information in 2008/2009 and were listed in 2012/2013 but had no information then. Hence, the total number of people lost due to attrition is 251: they had information in 2008/2009 but not in 2012/2013.

Individuals lost due to attrition differ in their characteristics from individuals re-interviewed in the last survey wave. They are on average younger and more educated, and there are more women among them. They are less likely to be married and more likely to have been away from the household for at least a month in the past year. They live in more densely populated areas, closer to roads and towns. When I compare the characteristics of individuals lost due to attrition to the characteristics of six groups of youth distinguished by their migration status (non-migrants, migrants to low-density rural areas, migrants to high-density rural areas, migrants to peri-urban areas, migrants to towns, and migrants to cities), individuals lost due to attrition differ from non-migrants and migrants regardless of destination. I cannot conclude that individuals lost due to attrition resemble a single group more than all other groups, but I observe the largest differences with the non-migrant youth. Still, there is a possibility that individuals lost due to attrition for different reasons (mainly, international migration versus non-tracked internal migrants) have different characteristics, hence aggregating them in one group can be uninformative.

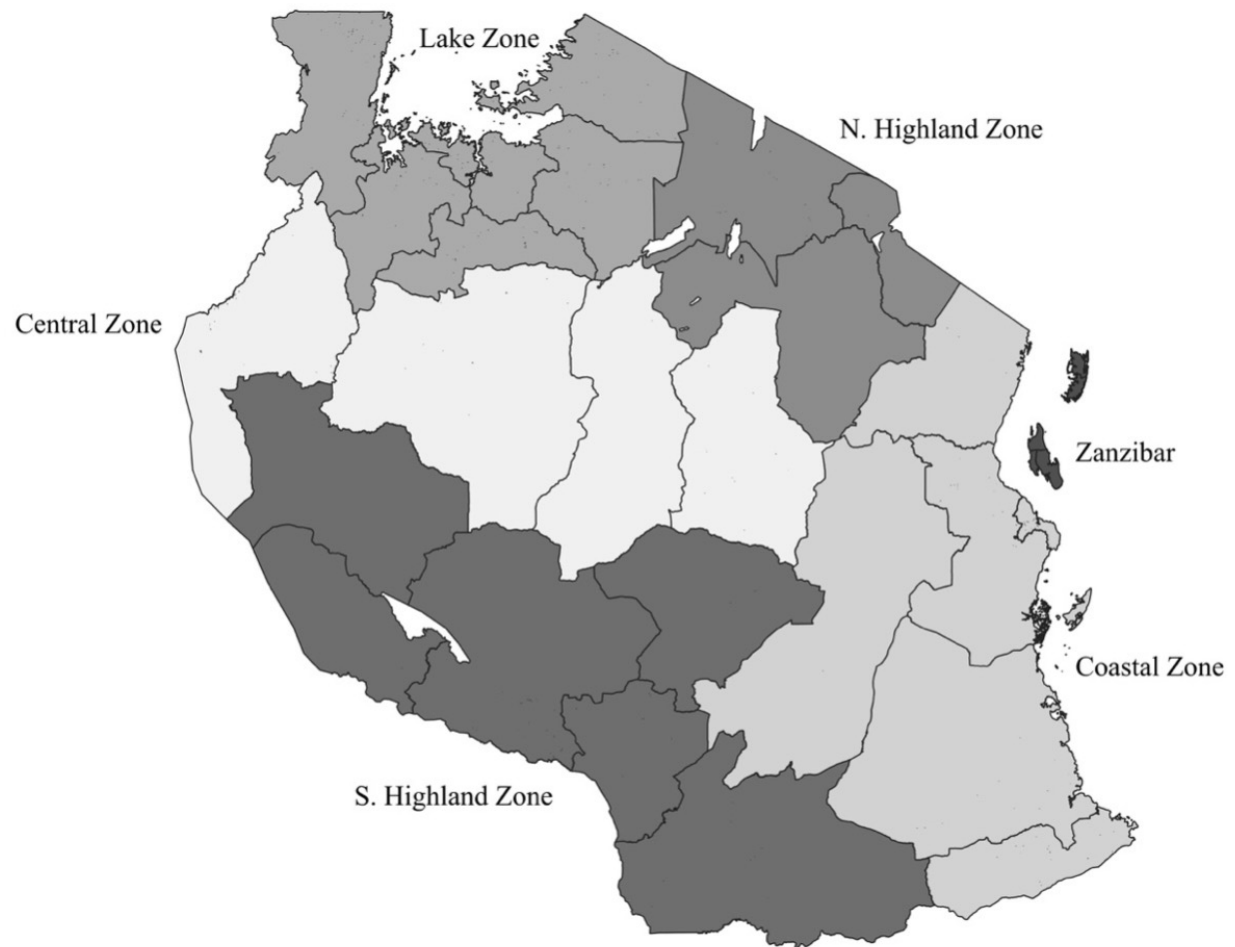
APPENDIX 5. Geographical zones

I distinguish six geographical zones depicted in Figure 2.12. Based on the classification into regions provided in the dataset, I distinguish the following zones: (1) Coastal Zone: Dar es Salaam, Morogoro, Pwani, Tanga, Lindi, and Mtwara – with 662 individuals of age 15-34 living in rural⁴⁸ areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 8.1%⁴⁹); (2) Northern Highland Zone: Arusha, Kilimanjaro, and Manyara – with 275 individuals of age 15-34 living in rural areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 7.6%); (3) Lake Zone: Kagera, Mara, Shinyanga, and Mwanza – with 619 individuals of age 15-34 living in rural areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 7.3%); (4) Central Zone: Dodoma, Singida, Tabora, and Kigoma – with 484 individuals of age 15-34 living in rural areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 7.2%); (5) Southern Highland Zone: Iringa, Mbeya, Rukwa, and Ruvuma – with 484 individuals of age 15-34 living in rural areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 9.8%); (6) Zanzibar – with 333 individuals of age 15-34 living in rural areas in 2008/2009 and re-surveyed in 2012/2013 (attrition rate is 8.0%).

⁴⁸ Here and elsewhere in Appendix 5 I use the NBS definition of “rural”.

⁴⁹ Household weights from the first survey wave are applied when computing attrition rates in Appendix 5.

Figure 2.12. Geographical zones



APPENDIX 6. Additional tables

Table 2.25. Explanatory variables used in other studies

Study	Herrera and Sahn (2020)	Mueller et al. (2019)	Reed, Andrzejewski, and White (2010)	Beegle and Poulin (2013)	Koubi et al. (2016)	Zhang et al. (2018)
Country	Senegal	Tanzania	Ghana	Malawi	Vietnam	China
Destinations	rural and urban	peri-urban and urban	rural and urban, for inter-regional moves	not distinguished	not distinguished	not distinguished
Age group	21-35 years of age	15-65 years of age	15 years of age and above	15-24 years of age	18-64 years of age	15-59 years of age
<i>Explanatory variables</i>						
Gender	women are more likely to move to rural areas	women are more likely to move to peri-urban areas	women are less likely to move	women are more likely to move	women are less likely to move	women are less likely to move
Age	older people are more likely to move to a rural area and less likely to move to an urban area	descriptive: older people are less likely to move; migrant to peri-urban areas are older than migrants to urban areas	binary: older people are less likely to move; multinomial: stays significant only for female rural-to-rural migrants	not significant	older people are less likely to move	older people are less likely to move
Marital status		married people are less likely to move to urban areas	married men are more likely to move to a rural area			married people are more likely to move; single women are more likely to move
Education	better educated men are more likely to move; better educated women are less likely to move to rural areas	descriptive: better educated people are more likely to move to urban areas	binary: better educated people are more likely to move; multinomial: stays significant for female migrants to both destinations		better educated people are more likely to move	better educated people are more likely to move

Table 2.25 (cont'd)

Study	Herrera and Sahn (2020)	Mueller et al. (2019)	Reed, Andrzejewski, and White (2010)	Beegle and Poulin (2013)	Koubi et al. (2016)	Zhang et al. (2018)
Occupation			binary: being employed or in school in the previous year is associated with lower probability of migration; multinomial: stays significant for male rural-to-rural migrants and rural-to-urban migrants of both genders	being in school is associated with lower probability of migration; descriptive: the main activity of male migrants is less likely to be farming and more likely to be domestic chores than of male non-migrants	relative to agricultural workers, people with other professions (civil servants, entrepreneurs, wage workers, people with “elementary profession”, people living from remittances) have lower probability to move	
Migration history			binary: higher number of previous moves is associated with higher probability of migration; multinomial: stays significant for both destinations and both genders		having a family member who migrated is associated with higher probability to move (in some models)	indicator of having a family member with a history of migration is associated with higher probability of migration
Household size	people from households with higher number of younger siblings are more likely to move to a rural area	descriptive: rural-to-urban migrants come from households with more adults	binary: people with two or more children are less likely to move; multinomial: stays significant for both destinations for both genders	people from larger households are less likely to move		not significant

Table 2.25 (cont'd)

Study	Herrera and Sahn (2020)	Mueller et al. (2019)	Reed, Andrzejewski, and White (2010)	Beegle and Poulin (2013)	Koubi et al. (2016)	Zhang et al. (2018)
Household wealth	people from wealthier households (based on the asset index at the time when the individual was 10 years old) are more likely to move to an urban area	descriptive: rural-to-urban migrants come from households with less land and higher asset index		people from wealthier households (based on the asset index) are more likely to move	people who report having economic reasons to move are more likely to move	people from households with higher amount of cultivated land are less likely to move
Community characteristics	people living within five km of a primary school are less likely to move to an urban area; people living closer to a hospital are less likely to move to a rural area			distance to town: not significant		people living closer to a hospital are less likely to move
Other variables	ethnicity; education of parents; indicators of parent's death by the time individual was 10 years old		interaction terms: sex*education, sex*employment	indicator of being a child of household head; indicator of having both parents alive	indicators of sudden weather events and gradual events	indicator of being a single woman; gender, age, education, and marital status of the household head; house elevation; compensations received from the programs studied

Table 2.26. Migration rates by age group for people from rural areas according to the constructed definition unless stated otherwise

Age	15-24	25-34	35 and above
Number of observations	1,704	1,099	2,283
Share of people living in low-density rural areas at baseline	69.2%	70.2%	68.4%
Share of migrants	19.3%	11.6%	5.4%
Migrants: to low-density rural	44.4%	42.9%	48.1%
Migrants: to high-density rural	21.8%	29.9%	30.9%
Migrants: to peri-urban	9.6%	16.0%	13.4%
Migrants: to town	11.0%	5.1%	6.3%
Migrants: to city	13.3%	6.1%	1.4%
Migrants from rural areas according to the NBS definition: to low-density rural	42.0%	45.1%	44.9%
Migrants from rural areas according to the NBS definition: to high-density rural	20.1%	31.4%	36.5%
Migrants from rural areas according to the NBS definition: to town	22.9%	15.8%	13.4%
Migrants from rural areas according to the NBS definition: to city	15.0%	7.7%	5.2%

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.27. Migration rates by gender for people from rural areas according to the constructed definition unless stated otherwise

Gender	Men	Women
Number of observations	1,342	1,461
Share of people living in low-density rural areas at baseline	70.6%	68.6%
Share of migrants	13.2%	19.1%
Migrants: to low-density rural	39.0%	47.4%
Migrants: to high-density rural	23.7%	24.3%
Migrants: to peri-urban	10.6%	11.9%
Migrants: to town	11.7%	7.7%
Migrants: to city	15.0%	8.8%
Migrants from rural areas according to the NBS definition: to low-density rural	38.6%	45.5%
Migrants from rural areas according to the NBS definition: to high-density rural	21.2%	24.5%
Migrants from rural areas according to the NBS definition: to town	24.8%	18.6%
Migrants from rural areas according to the NBS definition: to city	15.4%	11.5%

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.28. Regressions by age groups: logistic regressions and regressions with two destinations; constructed definition of “rural”

	People of age 15-24			People of age 25-34			People of age 35 and above		
	Logistic regression	Multinomial logistic regression		Logistic regression	Multinomial logistic regression		Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban	1 = Migrant	1 = Moved to rural	2 = Moved to urban	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	0.009**	0.008**	0.002	-0.000	-0.001	0.001	0.001	0.002***	-0.001*
1 = Male	-0.117***	-0.112***	-0.012	-0.049	-0.020	-0.031*	-0.003	0.004	-0.009
1 = Completed primary school	-0.028	-0.039*	0.011	-0.030	-0.016	-0.017	0.025	0.023	0.001
1 = Married	-0.183***	-0.161***	-0.028	-0.013	-0.020	-0.004	-0.034	0.020	-0.132***
1 = Child of household head	-0.120***	-0.089***	-0.027	0.030	0.009	0.012	0.067	0.006	-0.010
1 = Born in this village	-0.123***	-0.116***	0.001	-0.049*	-0.048*	-0.004	-0.040***	-0.036***	0.000
1 = Was away from the household for at least one month in the past 12 months	0.145***	0.068*	0.067**	0.080	0.047	0.026	-0.002	-0.001	-0.007
1 = Main occupation in farming or fishing in the past year	-0.011	0.020	-0.027*	-0.017	0.009	-0.022	-0.015	-0.002	-0.010
Area under cultivation, acres / 1000	0.764	2.566*	-2.265	1.065	1.838	2.128	0.653	1.011	-0.714
1 = Household head is male	0.040	0.059***	-0.011	0.027	0.045*	-0.010	-0.009	-0.023	0.023
Number of household members	0.001	-0.004	0.006***	-0.002	0.001	-0.005	-0.002	-0.002	0.001
1 = Household experienced agricultural shock in the past year	-0.021	-0.008	-0.013	0.010	0.003	0.007	0.035**	0.026*	0.009
1 = Household experienced non-agricultural shock in the past year	0.020	0.001	0.015	0.034	0.028	0.003	0.014	0.005	0.009
Population density, people per square km / 1000	0.665	0.794*	0.014	0.276	0.011	0.356	-0.727***	-0.530**	-0.128
Distance to road, km / 1000	-0.611	0.845*	-1.602***	0.670	0.467	0.164	-0.078	-0.175	0.042
Distance to the nearest town with population of at least 50,000, km / 1000	0.424	0.164	0.219	-0.212	0.261	-0.685**	-0.355**	-0.112	-0.261**

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.29. Regression results (marginal effects, constructed definition of “rural”) by age group

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
<i>A. People of age 15-24</i>					
Age	0.005	0.003	0.002	0.001	-0.001
1 = Male	-0.089***	-0.022*	-0.010	-0.002	-0.002
1 = Completed primary school	-0.050***	0.009	0.002	0.009	0.003
1 = Married	-0.110***	-0.050***	-0.009	-0.009	-0.012
1 = Child of household head	-0.042*	-0.045***	-0.012	-0.007	-0.009
1 = Born in this village	-0.113***	-0.002	-0.003	-0.000	0.008
1 = Was away from the household for at least one month in the past 12 months	0.044	0.028	0.013	-0.001	0.064**
1 = Main occupation in farming or fishing in the past year	0.023	-0.000	-0.015	-0.003	-0.009
Area under cultivation, acres / 1000	2.264**	0.679	-0.329	-0.524	-1.684
1 = Household head is male	0.053***	0.007	0.003	-0.009	-0.007
Number of household members	-0.002	-0.001	-0.001	0.001	0.006***
1 = Household experienced agricultural shock in the past year	0.011	-0.019*	-0.000	-0.018***	0.005
1 = Household experienced non-agricultural shock in the past year	0.013	-0.014	-0.008	0.016	0.005
Population density, people per square km / 1000	0.300	0.511**	0.075	-0.046	-0.041
Distance to road, km / 1000	0.838**	-0.060	-0.243	-0.098	-1.160***
Distance to the nearest town with population of at least 50,000, km / 1000	0.098	0.053	-0.127	0.057	0.255**

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.29 (cont'd)

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
<i>B. People of age 25-34</i>					
Age	-0.003	0.002	0.002	-0.000	-0.000
1 = Male	-0.032	0.002	-0.033	0.000	-0.000
1 = Completed primary school	-0.007	-0.007	-0.019	-0.000	-0.000
1 = Married	-0.019	-0.006	-0.004	0.000	-0.007
1 = Child of household head	-0.017	0.023	0.011	0.000	-0.000
1 = Born in this village	-0.054**	0.006	-0.001	0.000	-0.000
1 = Was away from the household for at least one month in the past 12 months	0.014	0.047	0.033	-0.000	-0.008
1 = Main occupation in farming or fishing in the past year	0.002	0.006	-0.001	-0.000	-0.061
Area under cultivation, acres / 1000	1.093	1.146	2.465**	0.000	0.000
1 = Household head is male	0.048**	0.001	-0.006	-0.000	-0.000
Number of household members	0.005	-0.003	-0.005	-0.000	-0.000
1 = Household experienced agricultural shock in the past year	0.002	-0.004	0.018	0.000	-0.008
1 = Household experienced non-agricultural shock in the past year	0.014	0.013	0.006	-0.000	0.000
Population density, people per square km / 1000	-0.032	0.045	0.224	-0.000	0.000
Distance to road, km / 1000	0.093	0.610*	-0.202	0.000	0.000
Distance to the nearest town with population of at least 50,000, km / 1000	0.302	-0.086	-0.478**	-0.000	-0.000

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.29 (cont'd)

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
<i>C. People of age 35 and above</i>					
Age	0.001**	0.001*	-0.000	-0.000	-0.000
1 = Male	-0.002	0.008	-0.004	-0.007***	-0.007
1 = Completed primary school	0.014	0.013	0.004	-0.006	-0.008
1 = Married	0.018	0.006	-0.059	-0.052	-0.057
1 = Child of household head	0.024	-0.013***	-0.007	-0.003	-0.000
1 = Born in this village	-0.022	-0.015	-0.000	0.003	-0.014
1 = Was away from the household for at least one month in the past 12 months	0.003	-0.005	-0.007***	-0.000	-0.002***
1 = Main occupation in farming or fishing in the past year	-0.007	0.005	-0.006	0.004	0.004
Area under cultivation, acres / 1000	1.684	0.484	-2.083*	2.149**	0.000
1 = Household head is male	-0.023	-0.002	0.022	-0.013	0.011
Number of household members	-0.000	-0.002	-0.000	0.002**	-0.000
1 = Household experienced agricultural shock in the past year	0.021	0.005	-0.003	0.011*	-0.000
1 = Household experienced non-agricultural shock in the past year	-0.005	0.009	0.010	0.001	-0.003
Population density, people per square km / 1000	-0.335	-0.188	-0.048	-0.114	0.000
Distance to road, km / 1000	-0.180	0.072	0.132	-0.202	0.000
Distance to the nearest town with population of at least 50,000, km / 1000	-0.003	-0.146	-0.193*	-0.120*	-0.000

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.30. Regressions by gender: logistic regressions and regressions with two destinations; for the constructed definition of “rural”

	Logistic regression 1 = Migrant	Men Multinomial logistic regression 1 = Moved to rural 2 = Moved to urban		Logistic regression 1 = Migrant	Women Multinomial logistic regression 1 = Moved to rural 2 = Moved to urban	
Age	0.001	0.002	0.001	-0.005	-0.002	-0.003
1 = Completed primary school	-0.032	-0.019	-0.016	-0.017	-0.031	0.016
1 = Married	-0.035	-0.023	-0.005	-0.185***	-0.201***	-0.002
1 = Child of household head	-0.069	-0.069*	-0.013	0.048	-0.045	0.122
1 = Born in this village	-0.146***	-0.140***	0.011	-0.053	-0.044	-0.007
1 = Was away from the household for at least one month in the past 12 months	0.070*	0.003	0.058**	0.170***	0.106**	0.058*
1 = Main occupation in farming or fishing in the past year	-0.036	-0.013	-0.025	0.041	0.065**	-0.024
Area under cultivation, acres / 1000	-1.741	-0.050	-1.568	2.470	2.642	-0.486
1 = Household head is male	-1.221	-0.323	-0.782	0.288	0.955	-0.954
Number of household members	0.002	-0.002	0.003*	-0.002	-0.004	0.003
1 = Household experienced agricultural shock in the past year	0.008	0.016	-0.009	-0.022	-0.012	-0.009
1 = Household experienced non-agricultural shock in the past year	0.024	0.005	0.016	0.033	0.025	0.007
Population density, people per square km / 1000	0.408	0.422	0.117	0.528	0.534	-0.003
Distance to road, km / 1000	-1.171**	-0.062	-1.452***	0.980*	1.398***	-0.527
Distance to the nearest town with population of at least 50,000, km / 1000	0.566**	0.281	0.229	-0.364	-0.089	-0.267

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.31. Regression results (marginal effects, constructed definition of “rural”) by gender

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
<i>A. Men</i>					
Age	-0.000	0.002*	0.001	0.002	-0.003
1 = Completed primary school	-0.023	-0.000	-0.015	-0.006	-0.002
1 = Married	-0.002	-0.024	0.004	-0.011	0.009
1 = Child of household head	-0.065*	-0.004	-0.008	0.039	-0.027***
1 = Born in this village	-0.121***	-0.007	-0.003	-0.002	0.014
1 = Was away from the household for at least one month in the past 12 months	-0.007	0.008	0.025	0.018	0.019
1 = Main occupation in farming or fishing in the past year	-0.006	-0.009	-0.009	-0.009	-0.012
Area under cultivation, acres / 1000	3.462*	-0.447	0.551	0.563	2.001
1 = Household head is male	0.429	-3.657	-1.224	-4.218	0.659
Number of household members	-0.000	-0.002	-0.002	0.001	0.003
1 = Household experienced agricultural shock in the past year	0.015	-0.006	0.002	-0.010	-0.007
1 = Household experienced non-agricultural shock in the past year	0.024	-0.023**	-0.002	0.005	0.003
Population density, people per square km / 1000	0.348	-0.000	0.105	-0.162	0.072
Distance to road, km / 1000	-0.119	-0.053	-0.461	0.051	-1.165***
Distance to the nearest town with population of at least 50,000, km / 1000	0.345*	0.008	-0.031	-0.063	0.223**

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.31 (cont'd)

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
<i>B. Women</i>					
Age	-0.002	-0.000	0.001	-0.002	-0.001
1 = Completed primary school	-0.037*	0.007	0.001	0.008	0.005
1 = Married	-0.131***	-0.066**	-0.006	0.011	0.002
1 = Child of household head	-0.012	-0.019	0.055	-0.010***	0.050
1 = Born in this village	-0.053*	0.008	0.001	-0.005	-0.002
1 = Was away from the household for at least one month in the past 12 months	0.059	0.047	0.033	-0.012***	0.050
1 = Main occupation in farming or fishing in the past year	0.051**	0.016	-0.022	-0.003	-0.010
Area under cultivation, acres / 1000	1.256	2.904*	0.592	0.217	0.042
1 = Household head is male	1.164	-1.157	0.131	0.361	-3.013
Number of household members	-0.002	-0.002	-0.001	0.000	0.004**
1 = Household experienced agricultural shock in the past year	0.005	-0.018	0.006	-0.016**	-0.001
1 = Household experienced non-agricultural shock in the past year	0.010	0.014	-0.004	0.010	0.000
Population density, people per square km / 1000	0.072	0.438	0.022	0.014	0.017
Distance to road, km / 1000	0.882**	0.516*	-0.035	-0.009	-0.293
Distance to the nearest town with population of at least 50,000, km / 1000	0.037	-0.094	-0.599**	0.053	0.033

Note: All regressions contain squared age, squared area under cultivation, indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.32. Number of observations of youth by their location at baseline: constructed definition of “rural”, NBS categorization, and their intersection

		Origin according to the constructed definition of “rural”					
		Low-density rural	High-density rural	Peri-urban	Town	City	Total
<i>A. Youth from areas defined as rural by the constructed definition (in the first survey wave)</i>							
Origin according to the NBS categorization of “rural”	Low-density rural	1,695					1,695
	High-density rural		768				768
	Town	137	203				340
	City						
	Total	1,832	971				2,803
<i>B. Youth from areas defined as rural by the NBS categorization (in the first survey wave)</i>							
Origin according to the NBS categorization of “rural”	Low-density rural	1,695					1,695
	High-density rural		768	358	27	9	1,162
	Town						
	City						
	Total	1,695	768	358	27	9	2,857
<i>C. Youth from rural areas, for whom the defined locations on the rural-urban spectrum are always consistent between the constructed definition and the NBS categorization (for both the first and the third survey waves)</i>							
Origin according to the NBS categorization of “rural”	Low-density rural	1,580					1,580
	High-density rural		700				700
	Town						
	City						
	Total	1,580	700				2,280

Table 2.33. Number of observations of migrant youth from rural areas by destination: constructed definition of “rural”, NBS categorization, and their intersection

		Destination according to the constructed definition of “rural”					
		Low-density rural	High-density rural	Peri-urban	Town	City	Total
<i>A. Migrant youth from areas defined as rural by the constructed definition (in the first survey wave)</i>							
Destination according to the NBS categorization of “rural”	Low-density rural	179					179
	High-density rural		75	18	6		99
	Town	14	27	24	34		99
	City	1	1	6		54	62
	Total	194	103	48	40	54	439
<i>B. Migrant youth from areas defined as rural by the NBS categorization (in the first survey wave)</i>							
Destination according to the NBS categorization of “rural”	Low-density rural	183					183
	High-density rural		66	26	7	1	100
	Town	8	26	21	32		87
	City	1	1	9		53	64
	Total	192	93	56	39	54	434
<i>C. Migrant youth from rural areas, for whom the defined locations on the rural-urban spectrum are always consistent between the constructed definition and the NBS categorization (for both the first and the third survey waves)</i>							
Destination according to the NBS categorization of “rural”	Low-density rural	169					169
	High-density rural		63				63
	Town				26		26
	City					41	41
	Total	169	63		26	41	299

Table 2.34. Comparison of characteristics of youth living in rural areas, by definition of “rural”

Constructed definition	LD rural	LD rural	HD rural	HD rural	Non-rural
NBS categorization	LD rural	Urban	HD rural	Urban	HD rural
Share of migrants	15.2%	14.1%	16.9%	28.1%	17.7%
Migrants: moved to low-density rural area	53.0%	41.0%	34.3%	13.5%	25.9%
Migrants: moved to high-density rural area	20.0%	6.0%	33.1%	33.1%	10.5%
Migrants: moved to peri-urban area	8.9%	13.2%	13.7%	20.6%	35.9%
Migrants: moved to town	7.6%	6.6%	11.4%	15.3%	7.7%
Migrants: moved to city	10.5%	33.2%	7.4%	17.4%	19.9%
Age	23.06	22.07	22.7	24.4	22.98
1 = Male	0.49	0.54	0.48	0.42	0.49
1 = Completed primary school	0.58	0.87	0.68	0.87	0.77
1 = Married	0.49	0.25	0.42	0.38	0.36
1 = Head of the household	0.19	0.2	0.13	0.26	0.14
1 = Child of household head	0.42	0.35	0.45	0.28	0.45
1 = Born in this village	0.81	0.74	0.8	0.55	0.79
1 = Was away from the household for at least one month in the past 12 months	0.09	0.14	0.13	0.11	0.11
1 = Main occupation in farming or fishing in the past year	0.76	0.20	0.62	0.28	0.42
Area under cultivation, acres	7.27	2.82	6.21	2.46	12.52
Livestock (TLU)	4.16	1.31	2.71	0.53	1.04
Age of household head	44.12	46	46.25	42.47	46.49
1 = Household head is male	0.83	0.67	0.77	0.74	0.79
Number of working age women	1.81	2.01	1.83	1.56	1.89
Number of working age men	1.82	2.24	1.88	1.45	1.99
Number of children of household head living in the household	3.53	2.05	3.36	1.99	3.04
1 = Household experienced agricultural shock in the past year	0.28	0.33	0.28	0.27	0.33
1 = Household experienced non-agricultural shock in the past year	0.28	0.43	0.27	0.36	0.38
Population density, people per square km	52.15	60.95	182.18	342.23	609.49
Distance to road, km	23.53	11.24	17.77	20.61	5.77
Distance to the nearest town with population of at least 50,000, km	70.51	88.9	53.25	76.09	18.48
Number of observations	1,695	137	768	203	394

Note: Sampling weights from the 2008/2009 survey wave are applied. “LD rural” stands for “low-density rural area”; “HD rural” stands for “high-density rural area”.

Table 2.35. Regression results (marginal effects): binary division and two destinations; for the NBS categorization of “rural”

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	-0.000	-0.001	0.001
Age squared	-0.002**	-0.002	-0.006***
1 = Male	-0.084***	-0.074***	-0.011
1 = Completed primary school	0.008	-0.014	0.023***
1 = Married	-0.113***	-0.065***	-0.051***
1 = Child of household head	-0.040**	-0.017	-0.020*
1 = Born in this village	-0.071***	-0.058***	-0.009
1 = Was away from the household for at least one month in the past 12 months	0.071***	0.040*	0.028*
1 = Main occupation in farming or fishing in the past year	-0.030*	-0.000	-0.027***
Area under cultivation, acres / 1000	1.015	0.940	-0.012
Squared area under cultivation, acres / 1000000	-28.373	-7.939	-39.948
1 = Household head is male	0.000	0.014	-0.007
Number of household members	-0.000	-0.001	0.001
1 = Household experienced agricultural shock in the past year	-0.018	-0.011	-0.006
1 = Household experienced non-agricultural shock in the past year	0.008	-0.000	0.009
1 = From high-density rural area	0.012	0.017	-0.001
Population density, people per square km / 1000	-0.044	-0.083**	0.002
Distance to road, km / 1000	-0.120	0.425	-0.609**
Distance to the nearest town with population of at least 50,000, km / 1000	0.083	0.099	-0.049

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.36. Regression results (marginal effects): binary division and two destinations; sample of youth for whom the constructed definition and the NBS categorization agree for all survey waves

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	-0.001	-0.001	0.001
Age squared	-0.003***	-0.002	-0.009**
1 = Male	-0.073***	-0.074***	-0.004
1 = Completed primary school	-0.005	-0.017	0.013*
1 = Married	-0.092***	-0.077***	-0.017*
1 = Child of household head	-0.025	-0.012	-0.010
1 = Born in this village	-0.102***	-0.078***	-0.021
1 = Was away from the household for at least one month in the past 12 months	0.055**	0.024	0.025*
1 = Main occupation in farming or fishing in the past year	-0.036*	-0.003	-0.026***
Area under cultivation, acres / 1000	-0.119	0.405	-0.699
Squared area under cultivation, acres / 1000000	17.434	-0.638	32.797
1 = Household head is male	0.010	0.028*	-0.012
Number of household members	0.000	-0.002	0.002
1 = Household experienced agricultural shock in the past year	-0.029**	-0.026*	-0.004
1 = Household experienced non-agricultural shock in the past year	0.013	-0.002	0.014*
1 = From high-density rural area	0.006	0.017	-0.006
Population density, people per square km / 1000	0.037	-0.066	0.050*
Distance to road, km / 1000	-0.026	0.452	-0.494**
Distance to the nearest town with population of at least 50,000, km / 1000	0.294*	0.203	0.061

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.37. Regression results (marginal effects): four destinations; sample of youth for whom the constructed definition and the NBS categorization agree for all survey waves

	1 = Moved to low- density rural	2 = Moved to high- density rural	3 = Moved to town	4 = Moved to city
Age	-0.002	0.000	0.001	-0.000
Age squared	-0.002	-0.003	-0.033**	-0.004
1 = Male	-0.062***	-0.013*	-0.004	-0.001
1 = Completed primary school	-0.010	-0.007	0.010**	0.005
1 = Married	-0.062***	-0.016	-0.004	-0.013*
1 = Child of household head	-0.002	-0.010	0.003	-0.011
1 = Born in this village	-0.063***	-0.013	-0.007	-0.013
1 = Was away from the household for at least one month in the past 12 months	0.024	0.001	0.003	0.024*
1 = Main occupation in farming or fishing in the past year	0.008	-0.011	-0.010*	-0.014*
Area under cultivation, acres / 1000	1.028	-0.751	0.330	-1.502
Squared area under cultivation, acres / 1000000	-12.179	59.377	-99.656	-3572.407
1 = Household head is male	0.046***	-0.017	-0.004	-0.006
Number of household members	-0.000	-0.003*	-0.001	0.003**
1 = Household experienced agricultural shock in the past year	-0.012	-0.013*	-0.006	0.002
1 = Household experienced non-agricultural shock in the past year	-0.005	0.003	0.014**	-0.001
1 = From high-density rural area	0.010	0.008	0.018	-0.015**
Population density, people per square km / 1000	-0.120	0.008	-0.063	0.060***
Distance to road, km / 1000	0.298	0.152	-0.039	-0.403**
Distance to the nearest town with population of at least 50,000, km / 1000	0.204	-0.046	0.027	0.027

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.38. Number of observations of youth, by the type of their location at baseline according to the cluster analysis definition

		Low-density rural	High-density rural	Peri-urban	Town	City	Total
<i>A. Columns: constructed definition of “rural”</i>							
Cluster analysis definition	Rural	1750	801	142	89	0	2782
	Urban	82	170	416	445	772	1885
<i>B. Columns: NBS categorization of “rural”</i>							
Cluster analysis definition	Rural	1628	742	-	412	0	2782
	Urban	67	420	-	662	776	1885

Table 2.39. Number of observations of youth from rural areas according to the cluster analysis definition, by the type of their location as defined by the constructed definition and the NBS categorization of “rural”

		Origin according to the constructed definition of “rural”					
		Low-density rural	High-density rural	Peri-urban	Town	City	Total
Origin according to the NBS categorization of “rural”	Low-density rural	1628	0	0	0	0	1628
	High-density rural	0	636	96	10	0	742
	Town	122	165	46	79	0	412
	City	0	0	0	0	0	0
	Total	1750	801	142	89	0	2782

Table 2.40. Number of observations of migrant youth from rural areas according to the cluster analysis definition, by destination

	Destination according to the constructed definition of “rural”					
	Low-density rural	High-density rural	Peri-urban	Town	City	Total
<i>Destination according to the cluster analysis definition: two groups</i>						
Rural	191	88	14	22	2	317
Urban	6	15	35	18	51	125
<i>Destination according to the cluster analysis definition: three groups</i>						
Rural	144	46	4	3		197
High-density rural	52	56	19	28	2	157
Urban	1	1	26	9	51	88
<i>Destination according to the cluster analysis definition: four groups</i>						
Rural	129	38	4	3		174
High-density rural	50	47	19	27	2	145
Town	17	17				34
City	1	1	26	10	51	89
<i>Destination according to the cluster analysis definition: five groups</i>						
Rural & far from road	17	13				30
Rural & close to road	129	33	4	3		169
High-density rural	49	47	19	27	2	144
Town	1	9				10
City	1	1	26	10	51	89
<i>Destination according to the cluster analysis definition: six groups</i>						
Rural & far from road	62	21		1		84
Rural & close to road	73	30		1		104
Rural & close to town	49	25	10	17	2	103
High-density rural	11	18	14	12		55
Town	1	9				10
City	1		25	9	51	86

Table 2.41. Comparison of migration rates among youth living in rural areas, by definition of “rural”

Constructed definition	Low-density rural		Low-density rural		High-density rural		High-density rural		Non-rural	
NBS categorization	Low-density rural		Urban		High-density rural		Urban		High-density rural	
Cluster analysis definition	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Number of obs.	1628	67	122	15	636	132	165	38	106	288
Share of migrants	15.0%	19.2%	12.0%	26.7%	17.2%	12.8%	28.9%	6.8%	25.6%	12.5%
<i>Destinations according to the constructed definition of “rural”</i>										
Migrants: moved to low-density rural area	54.3%	26.0%	46.8%	25.0%	33.9%	41.9%	13.6%	0.0%	24.6%	8.3%
Migrants: moved to high-density rural area	20.3%	14.4%	8.3%	0.0%	34.7%	3.9%	32.8%	70.1%	9.2%	4.1%
Migrants: moved to peri-urban area	7.7%	32.5%	18.1%	0.0%	12.7%	33.3%	20.8%	0.0%	25.0%	20.6%
Migrants: moved to town	8.0%	0.0%	9.0%	0.0%	11.4%	11.0%	15.5%	0.0%	31.7%	47.6%
Migrants: moved to city	9.8%	27.1%	17.8%	75.0%	7.2%	9.9%	17.3%	29.9%	9.5%	19.3%
<i>Destinations according to the cluster analysis definition of “rural” with two groups</i>										
Rural	78.8%	24.4%	64.9%	25.0%	68.6%	41.9%	61.8%	0.0%	69.9%	14.5%
Urban	21.2%	75.6%	35.1%	75.0%	31.4%	58.1%	38.2%	100.0%	30.1%	85.5%
<i>Destinations according to the cluster analysis definition of “rural” with three groups</i>										
Rural	52.4%	24.4%	46.8%	0.0%	34.4%	37.9%	30.5%	0.0%	19.0%	9.5%
High-density rural	31.8%	12.7%	26.3%	25.0%	46.5%	33.3%	31.7%	70.1%	68.1%	59.9%
Urban	15.8%	62.9%	26.9%	75.0%	19.0%	28.7%	37.8%	29.9%	12.9%	30.7%

Table 2.41 (cont'd)

Constructed	Low-density rural		Low-density rural		High-density rural		High-density rural		Non-rural	
NBS	Low-density rural		Urban		High-density rural		Urban		High-density rural	
Cluster analysis	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
<i>Destinations according to the cluster analysis definition of “rural” with four groups</i>										
Rural	49.8%	24.4%	46.8%	0.0%	26.1%	20.7%	27.1%	0.0%	19.0%	9.5%
High-density rural	31.1%	12.7%	26.3%	25.0%	45.6%	33.3%	31.3%	0.0%	65.5%	18.1%
Town	3.3%	0.0%	0.0%	0.0%	9.2%	17.3%	3.8%	70.1%	0.0%	0.0%
City	15.8%	62.9%	26.9%	75.0%	19.0%	28.7%	37.8%	29.9%	15.5%	72.4%
<i>Destinations according to the cluster analysis definition of “rural” with five groups</i>										
Rural & far from road	6.6%	0.0%	0.0%	0.0%	8.3%	34.5%	3.4%	0.0%	0.0%	4.1%
Rural & close to road	46.7%	24.4%	46.8%	0.0%	26.1%	3.4%	27.1%	0.0%	19.0%	5.4%
High-density rural	30.8%	12.7%	26.3%	25.0%	45.6%	33.3%	31.3%	0.0%	65.5%	18.1%
Town	0.1%	0.0%	0.0%	0.0%	0.9%	0.0%	0.4%	70.1%	0.0%	0.0%
City	15.8%	62.9%	26.9%	75.0%	19.0%	28.7%	37.8%	29.9%	15.5%	72.4%
<i>Destinations according to the cluster analysis definition of “rural” with six groups</i>										
Rural & far from road	22.0%	0.0%	9.0%	0.0%	16.7%	34.5%	10.8%	0.0%	9.3%	4.1%
Rural & close to road	29.8%	11.6%	37.8%	0.0%	17.9%	3.4%	20.1%	0.0%	3.1%	2.9%
Rural & close to town	22.2%	12.7%	18.1%	25.0%	27.4%	0.0%	16.6%	0.0%	47.7%	7.1%
High-density rural	10.6%	24.4%	8.3%	0.0%	18.0%	33.3%	17.1%	0.0%	27.0%	55.7%
Town	0.1%	0.0%	0.0%	0.0%	0.9%	0.0%	0.4%	70.1%	0.0%	0.0%
City	15.4%	51.3%	26.9%	75.0%	19.0%	28.7%	35.0%	29.9%	12.9%	30.2%

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.42. Comparison of characteristics of youth living in rural areas, by definition of “rural”

Constructed definition	Low-density rural		Low-density rural		High-density rural		High-density rural		Non-rural	
NBS categorization	Low-density rural		Urban		High-density rural		Urban		High-density rural	
Cluster analysis definition	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Age	23.09	22.18	22.3	20.67	22.67	23.12	24.46	22.78	22.79	23.11
1 = Male	0.49	0.60	0.52	0.67	0.49	0.46	0.42	0.46	0.49	0.49
1 = Completed primary school	0.59	0.56	0.85	1.00	0.68	0.77	0.87	0.81	0.71	0.81
1 = Married	0.49	0.39	0.30	0.00	0.42	0.43	0.39	0.25	0.34	0.38
1 = Head of the household	0.19	0.10	0.22	0.07	0.13	0.21	0.27	0.04	0.11	0.16
1 = Child of household head	0.42	0.43	0.32	0.53	0.46	0.41	0.28	0.42	0.50	0.41
1 = Born in this village	0.81	0.83	0.72	0.87	0.80	0.75	0.54	0.85	0.83	0.76
1 = Was away from the household for at least one month in the past 12 months	0.09	0.14	0.13	0.20	0.13	0.08	0.12	0.00	0.07	0.14
1 = Main occupation in farming or fishing in the past year	0.76	0.57	0.22	0.07	0.63	0.47	0.28	0.15	0.48	0.38

Table 2.42 (cont'd)

Constructed definition	Low-density rural		Low-density rural		High-density rural		High-density rural		Non-rural	
NBS categorization	Low-density rural		Urban		High-density rural		Urban		High-density rural	
Cluster analysis definition	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Area under cultivation, acres	7.36	4.67	2.74	3.25	6.55	1.64	2.51	1.10	5.97	16.77
Livestock (TLU)	4.18	3.60	0.29	7.50	2.77	1.94	0.54	0.20	1.69	0.62
Age of household head	44.03	46.44	44.94	52.40	46.55	42.14	42.15	51.84	49.50	44.53
1 = Household head is male	0.83	0.84	0.65	0.80	0.77	0.78	0.74	0.87	0.73	0.83
Number of working age women	1.82	1.50	1.94	2.47	1.85	1.60	1.53	2.35	1.82	1.94
Number of working age men	1.80	2.39	1.69	5.60	1.90	1.52	1.41	2.55	1.89	2.05
Number of children of household head living in the household	3.54	3.32	1.95	2.67	3.39	2.94	1.95	3.05	2.92	3.12
1 = Household experienced agricultural shock in the past year	0.28	0.18	0.26	0.73	0.28	0.28	0.28	0.00	0.24	0.38

Table 2.42 (cont'd)

Constructed definition	Low-density rural		Low-density rural		High-density rural		High-density rural		Non-rural	
NBS categorization	Low-density rural		Urban		High-density rural		Urban		High-density rural	
Cluster analysis definition	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
1 = Household experienced non-agricultural shock in the past year	0.29	0.23	0.40	0.60	0.28	0.17	0.36	0.12	0.27	0.45
Population density, people per square km	51.55	68.56	58.78	74.05	179.25	222.16	331.84	640.91	362.19	769.74
Distance to road, km	24.24	4.25	13.03	0.40	17.99	14.76	20.05	36.65	9.57	3.31
Distance to the nearest town with population of at least 50,000, km	72.37	20.14	85.00	112.45	53.70	47.18	76.71	58.27	16.51	19.76
Number of observations	1628	67	122	15	636	132	165	38	106	288

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.43. Regression results (marginal effects) with the cluster analysis definition of “rural”: binary division, two, and three destinations

	1 = Migrant	1 = Moved to rural	2 = Moved to urban	1 = Moved to rural	2 = Moved to high- density rural	3 = Moved to urban
Age	-0.001	-0.001	0.000	-0.001	-0.001	0.001
Age squared	-0.002*	-0.003**	0.000	-0.002	-0.001	-0.003
1 = Male	-0.073***	-0.052***	-0.018**	-0.041***	-0.023**	-0.008
1 = Completed primary school	0.009	-0.010	0.020**	-0.012	0.006	0.017***
1 = Married	-0.098***	-0.065***	-0.037***	-0.052***	-0.030**	-0.017*
1 = Child of household head	-0.025	-0.018	-0.006	-0.015	-0.006	-0.004
1 = Born in this village	-0.098***	-0.083***	-0.015	-0.066***	-0.010	-0.021**
1 = Was away from the household for at least one month in the past 12 months	0.082***	0.056**	0.026*	0.023	0.039**	0.019
1 = Main occupation in farming or fishing in the past year	-0.031*	-0.000	-0.031***	0.005	-0.005	-0.031***
Area under cultivation, acres / 1000	0.633	0.210	0.858	0.966	0.704	-0.060
Squared area under cultivation, acres / 1000000	-10.207	-0.901	-36.349	-114.021	-7.605	-109.371
Livestock (TLU) / 1000	-0.090	0.472	-3.371*	0.240	-0.331	-2.006
1 = Household head is male	0.004	0.003	0.004	0.009	-0.001	-0.002
Number of household members	-0.000	0.001	-0.002	0.002	-0.003*	0.001
1 = Household experienced agricultural shock in the past year	-0.020	-0.022*	0.004	-0.015	-0.012	0.008
1 = Household experienced non-agricultural shock in the past year	0.007	0.003	0.004	-0.005	0.014	-0.000
Population density, people per square km / 1000	-0.010	-0.033	-0.002	-0.007	-0.003	-0.007
Distance to road, km / 1000	-0.191	0.368	-0.672***	0.525**	-0.733**	-0.373*
Distance to the nearest town with population of at least 50,000, km / 1000	0.184	0.272*	-0.141	0.411***	-0.273**	-0.043

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.44. Regression results (marginal effects) with the cluster analysis definition of “rural”: four destinations

	1 = Moved to rural	2 = Moved to high-density rural	3 = Moved to town	4 = Moved to city
Age	-0.001	-0.001	0.000	0.001
Age squared	-0.002	-0.001	-0.001	-0.003
1 = Male	-0.032***	-0.016*	-0.015***	-0.009
1 = Completed primary school	-0.013	0.006	0.001	0.016**
1 = Married	-0.045***	-0.032**	-0.007	-0.017*
1 = Child of household head	-0.016	-0.005	0.001	-0.003
1 = Born in this village	-0.052***	-0.010	-0.017*	-0.021**
1 = Was away from the household for at least one month in the past 12 months	0.029*	0.037**	-0.010**	0.020
1 = Main occupation in farming or fishing in the past year	0.018	-0.003	-0.014**	-0.031***
Area under cultivation, acres / 1000	0.962	0.672	-0.116	-0.193
Squared area under cultivation, acres / 1000000	-128.262	-7.929	-16.389	-45.309
Livestock (TLU) / 1000	0.189	-0.271	-0.393	-2.012
1 = Household head is male	0.004	0.003	0.004	-0.003
Number of household members	0.002	-0.003*	-0.001	0.001
1 = Household experienced agricultural shock in the past year	-0.013	-0.011	-0.004	0.010
1 = Household experienced non-agricultural shock in the past year	-0.003	0.015	-0.001	-0.001
Population density, people per square km / 1000	-0.003	-0.015	-0.002	-0.006
Distance to road, km / 1000	0.078	-0.663**	0.475***	-0.395**
Distance to the nearest town with population of at least 50,000, km / 1000	0.448***	-0.270**	-0.142*	-0.048

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.45. Regression results (marginal effects) with the cluster analysis definition of “rural”: five destinations

	1 = Moved to rural & far from road	2 = Moved to rural & close to road	3 = Moved to high- density rural	4 = Moved to town	5 = Moved to city
Age	0.000	-0.001	-0.001	-0.000	0.001
Age squared	-0.001	-0.002	-0.001	0.007	0.008
1 = Male	-0.010**	-0.031***	-0.016*	-0.006	-0.009
1 = Completed primary school	-0.002	-0.011	0.006	0.002	0.016**
1 = Married	-0.008	-0.047***	-0.031**	0.004	-0.017*
1 = Child of household head	0.005	-0.023*	-0.004	0.001	-0.003
1 = Born in this village	-0.015**	-0.050***	-0.011	-0.001	-0.021**
1 = Was away from the household for at least one month in the past 12 months	-0.005	0.027	0.037**	-0.004	0.020
1 = Main occupation in farming or fishing in the past year	-0.007	0.013	-0.002	-0.003	-0.031***
Area under cultivation, acres / 1000	-0.219	1.100	0.646	1.548	-0.241
Squared area under cultivation, acres / 1000000	-37.073	-140.677	-7.452	-171247	-159449.9
Livestock (TLU) / 1000	-0.366	0.233	-0.253	-0.656	-2.015
1 = Household head is male	0.000	0.009	0.002	-0.003	-0.003
Number of household members	-0.000	0.002	-0.003*	-0.000	0.001
1 = Household experienced agricultural shock in the past year	-0.001	-0.014	-0.010	-0.004	0.010
1 = Household experienced non-agricultural shock in the past year	0.002	-0.002	0.014	-0.004	-0.001
Population density, people per square km / 1000	-0.011	-0.002	-0.014	0.000	-0.006
Distance to road, km / 1000	0.401***	-0.003	-0.677**	-0.088	-0.386**
Distance to the nearest town with population of at least 50,000, km / 1000	-0.091	0.438***	-0.284**	0.011	-0.051

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.46. Regression results (marginal effects) with the cluster analysis definition of “rural”: six destinations

	1 = Moved to rural & far from road	2 = Moved to rural & close to road	3 = Moved to rural & close to town	4 = Moved to high- density rural	5 = Moved to town	6 = Moved to city
Age	0.000	-0.001	-0.000	-0.000	-0.000	0.001
Age squared	-0.004*	-0.002	-0.003	0.005**	0.005**	0.006**
1 = Male	-0.025***	-0.009	-0.015*	-0.007	-0.006	-0.009
1 = Completed primary school	-0.001	-0.004	-0.007	0.005	0.002	0.017
1 = Married	-0.016*	-0.033	-0.010	-0.026***	0.004	-0.019
1 = Child of household head	0.002	-0.025	-0.000	0.000	0.001	-0.005
1 = Born in this village	-0.033***	-0.015	-0.023**	-0.001	-0.001	-0.018
1 = Was away from the household for at least one month in the past 12 months	-0.003	0.013	0.022	0.024**	-0.004	0.016
1 = Main occupation in farming or fishing in the past year	-0.004	0.016	-0.019*	0.009	-0.003	-0.030
Area under cultivation, acres / 1000	0.107	0.204	0.593	0.570	1.550	-0.177
Squared area under cultivation, acres / 1000000	-127.831	-6.347	-15.034	-95.374	-101.455	-95.174
Livestock (TLU) / 1000	0.370*	-0.642	0.020	-0.508	-0.656	-1.761
1 = Household head is male	-0.003	0.012	-0.003	0.006	-0.003	-0.002
Number of household members	0.000	0.004	-0.002	-0.003**	-0.000	0.001
1 = Household experienced agricultural shock in the past year	-0.005	-0.002	-0.014*	-0.006	-0.004	0.008
1 = Household experienced non-agricultural shock in the past year	0.005	-0.007	0.005	0.011*	-0.004	-0.002
Population density, people per square km / 1000	-0.024	0.008	-0.025	0.012	0.000	-0.007
Distance to road, km / 1000	0.740***	-0.279	-0.274	-0.184	-0.089	-0.425
Distance to the nearest town with population of at least 50,000, km / 1000	-0.194*	0.513	-0.317***	-0.042	0.012	-0.049

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.47. Migration rates by migration status: comparison of the definitions of “migrant” that are based on distance traveled and self-reports

		Non-migrants (both definitions)	Migrants (distance only)	Migrants (self-reports only)	Migrants (both definitions)
Constructed definition of “rural”	Located in a low-density rural area in 2008/2009	70.8%	72.2%	66.2%	63.0%
	Located in a high-density rural area in 2008/2009	29.2%	27.8%	33.8%	37.0%
	Located in a low-density rural area in 2012/2013	70.8%	55.8%	66.2%	41.3%
	Located in a high-density rural area in 2012/2013	29.2%	29.1%	33.8%	22.9%
	Located in a peri-urban area in 2012/2013		6.9%		12.4%
	Located in a town in 2012/2013		3.6%		10.6%
	Located in a city in 2012/2013		4.5%		12.8%
	Located in a low-density rural area in 2008/2009	66.0%	67.3%	60.2%	59.1%
	Located in a high-density rural area in 2008/2009	24.7%	23.7%	28.6%	26.7%
	Located in a town in 2008/2009	9.3%	9.0%	11.2%	14.3%
NBS definition of “rural”	Located in a low-density rural area in 2012/2013	63.9%	51.7%	56.4%	38.4%
	Located in a high-density rural area in 2012/2013	24.4%	28.9%	27.9%	20.9%
	Located in a town in 2012/2013	11.6%	14.7%	13.6%	25.5%
	Located in a city in 2012/2013	0.1%	4.7%	2.1%	15.2%
	Number of observations	2,222	85	139	354

Note: Sampling weights from the 2008/2009 survey wave are applied.

Table 2.48. Comparison of characteristics of youth by their migration status, by definition of “migrant”: for the definitions based on distance traveled and self-reports

	Non-migrants (both definitions)	Migrants (distance only)	Migrants (self-reports only)	Migrants (both definitions)
Age	23.29	21.93**	22.41*	21.71***
1 = Male	0.52	0.43*	0.29***	0.40***
1 = Completed primary school	0.64	0.53**	0.58	0.69*
1 = Married	0.48	0.40	0.44	0.32***
1 = Head of the household	0.19	0.10**	0.13*	0.13***
1 = Child of household head	0.42	0.38	0.36	0.43
1 = Born in this village	0.82	0.75*	0.79	0.64***
1 = Was away from the household for at least one month in the past 12 months	0.08	0.14*	0.19***	0.17***
1 = Main occupation in farming or fishing in the past year	0.69	0.65	0.68	0.55***
Area under cultivation, acres	6.24	7.86	8.37	7.35
Livestock (TLU)	3.27	4.63	5.73**	3.56
Age of household head	44.33	46.48	45.50	45.89*
1 = Household head is male	0.81	0.80	0.79	0.78
Number of working age women	1.76	2.00*	2.00**	2.05***
Number of working age men	1.83	2.11*	1.69	1.86
Number of children of household head living in the household	3.36	2.97	3.13	3.36
1 = Household experienced agricultural shock in the past year	0.29	0.29	0.19***	0.24**
1 = Household experienced non- agricultural shock in the past year	0.29	0.32	0.22*	0.31
Population density, people per square km	97.79	88.94	100.80	119.80***
Distance to road, km	21.63	20.21	19.11	20.80
Distance to the nearest town with population of at least 50,000, km	67.55	76.33**	64.63	65.32
Number of observations	2,222	85	139	354

Note: Sampling weights from the 2008/2009 survey wave are applied. Stars indicate significant differences in means between migrants and non-migrants: *** 0.01; ** 0.05; * 0.1.

Table 2.49. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; definition of “migrant” is based on self-reports

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	-0.000	-0.001	0.001
Age squared	-0.003***	-0.004***	-0.005**
1 = Male	-0.111***	-0.103***	-0.012
1 = Completed primary school	0.011	-0.011	0.024***
1 = Married	-0.122***	-0.092***	-0.032***
1 = Head of the household	0.047	0.059*	-0.001
1 = Child of household head	-0.048***	-0.020	-0.024***
1 = Born in this village	-0.091***	-0.070***	-0.019*
1 = Was away from the household for at least one month in the past 12 months	0.120***	0.082***	0.035**
1 = Main occupation in farming or fishing in the past year	-0.031*	0.009	-0.036***
Area under cultivation, acres / 1000	-1.753	-0.289	-1.923*
Squared area under cultivation, acres / 1000000	65.907**	7.162	64.130
Livestock (TLU) / 1000	1.055*	1.548***	-1.092
1 = Household head is male	0.023	0.043***	-0.016
Number of household members	-0.002	-0.004**	0.002
1 = Household experienced agricultural shock in the past year	-0.034**	-0.037***	0.002
1 = Household experienced non-agricultural shock in the past year	0.003	-0.007	0.011
1 = From high-density rural area	0.012	-0.003	0.016*
Population density, people per square km / 1000	0.043**	0.043***	0.000
Distance to road, km / 1000	-0.026	0.494*	-0.539**
Distance to the nearest town with population of at least 50,000, km / 1000	-0.131	-0.003	-0.172

Note: All regressions contain age of the household head and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.50. Regression results (marginal effects): five destinations; definition of “migrant” is based on self-reports

	1 = Moved to low-density rural	2 = Moved to high-density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
Age	-0.000	-0.001	0.001**	-0.001	-0.000
Age squared	-0.004***	-0.003	-0.007**	-0.008**	-0.002
1 = Male	-0.069***	-0.034***	-0.010**	-0.007	0.002
1 = Completed primary school	-0.008	-0.002	0.007	0.012***	0.006
1 = Married	-0.079***	-0.013	-0.025***	-0.000	-0.008
1 = Head of the household	0.048*	0.005	-0.004	0.028	-0.012*
1 = Child of household head	-0.019	-0.001	-0.015**	0.001	-0.009
1 = Born in this village	-0.043***	-0.025**	0.005	-0.016**	-0.011
1 = Was away from the household for at least one month in the past 12 months	0.052**	0.031*	0.006	0.010	0.016
1 = Main occupation in farming or fishing in the past year	0.007	0.001	-0.007	-0.012**	-0.019***
Area under cultivation, acres / 1000	0.248	-0.191	-0.987	0.006	-0.303
Squared area under cultivation, acres / 1000000	1.853	20.615	93.785	0.915	-6447.317
Livestock (TLU) / 1000	1.517***	-1.376	-0.404	-0.441	-0.304
1 = Household head is male	0.040***	0.002	-0.002	-0.004	-0.007
Number of household members	-0.004**	-0.000	-0.000	0.000	0.002*
1 = Household experienced agricultural shock in the past year	-0.022**	-0.015*	0.000	-0.002	0.005
1 = Household experienced non-agricultural shock in the past year	0.005	-0.013	-0.003	0.009*	0.005
1 = From high-density rural area	-0.058***	0.061***	0.006	0.026**	-0.004
Population density, people per square km / 1000	-0.046	0.027***	0.007	-0.060*	0.009
Distance to road, km / 1000	0.300	0.183	-0.149	-0.005	-0.236*
Distance to the nearest town with population of at least 50,000, km / 1000	0.041	-0.094	-0.268***	0.013	0.025

Note: Constructed definition of “rural” is used. All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.51. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; define “migrant” if an individual is considered to be a migrant by either the definition based on distance traveled or self-reports

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	-0.001	-0.001	0.000
Age squared	-0.003***	-0.003**	-0.004**
1 = Male	-0.114***	-0.108***	-0.010
1 = Completed primary school	-0.002	-0.022	0.022**
1 = Married	-0.137***	-0.106***	-0.033***
1 = Child of household head	-0.067***	-0.032*	-0.030***
1 = Born in this village	-0.102***	-0.075***	-0.026**
1 = Was away from the household for at least one month in the past 12 months	0.139***	0.097***	0.040**
1 = Main occupation in farming or fishing in the past year	-0.032	0.011	-0.039***
Area under cultivation, acres / 1000	-0.982	0.107	-1.410
Squared area under cultivation, acres / 1000000	39.581	2.537	36.992
Livestock (TLU) / 1000	0.979	1.404**	-0.701
1 = Household head is male	0.027	0.043***	-0.013
Number of household members	-0.001	-0.003	0.002
1 = Household experienced agricultural shock in the past year	-0.042***	-0.045***	0.002
1 = Household experienced non-agricultural shock in the past year	0.005	-0.010	0.016*
1 = From high-density rural area	0.014	0.001	0.013
Population density, people per square km / 1000	0.042*	0.045**	-0.001
Distance to road, km / 1000	-0.160	0.385	-0.546**
Distance to the nearest town with population of at least 50,000, km / 1000	0.093	0.202	-0.149

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.52. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; define “migrant” if an individual is considered to be a migrant by either the definition based on distance traveled or self-reports

	1 = Moved to low-density rural	2 = Moved to high-density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
Age	-0.001	-0.001	0.001**	-0.001	-0.000
Age squared	-0.003**	-0.002	-0.004	-0.009**	-0.002
1 = Male	-0.074***	-0.035***	-0.010*	-0.006	0.004
1 = Completed primary school	-0.021*	0.000	0.003	0.013***	0.007
1 = Married	-0.080***	-0.026**	-0.024***	-0.001	-0.009
1 = Child of household head	-0.023	-0.009	-0.016**	-0.000	-0.012*
1 = Born in this village	-0.055***	-0.018	-0.001	-0.015*	-0.012
1 = Was away from the household for at least one month in the past 12 months	0.064***	0.035**	0.008	0.009	0.021*
1 = Main occupation in farming or fishing in the past year	0.006	0.004	-0.011	-0.011*	-0.018***
Area under cultivation, acres / 1000	0.718	-0.263	-0.092	-0.159	-1.161
Squared area under cultivation, acres / 1000000	-2.985	19.207	-0.478	13.489	-412.603
Livestock (TLU) / 1000	1.344***	-1.522	-0.335	-0.134	-0.293
1 = Household head is male	0.045***	-0.003	-0.003	-0.003	-0.004
Number of household members	-0.002	-0.001	-0.001	0.000	0.002**
1 = Household experienced agricultural shock in the past year	-0.026**	-0.019**	0.000	-0.002	0.005
1 = Household experienced non-agricultural shock in the past year	0.005	-0.017*	-0.000	0.011**	0.003
1 = From high-density rural area	-0.053***	0.062***	0.004	0.027**	-0.006
Population density, people per square km / 1000	-0.077	0.033***	0.007	-0.064*	0.010
Distance to road, km / 1000	0.244	0.110	-0.119	-0.023	-0.269**
Distance to the nearest town with population of at least 50,000, km / 1000	0.201	-0.059	-0.259***	-0.007	0.061

Note: All regressions contain indicator of being the head of the household, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.53. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; define “migrant” if an individual is considered to be a migrant by both the definition based on distance traveled and by self-reports

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	-0.000	-0.001	0.001
Age squared	-0.002**	-0.003**	-0.005**
1 = Male	-0.078***	-0.066***	-0.016*
1 = Completed primary school	0.014	-0.008	0.024***
1 = Married	-0.119***	-0.079***	-0.040***
1 = Head of the household	0.032	0.046	-0.004
1 = Child of household head	-0.046***	-0.011	-0.030***
1 = Born in this village	-0.114***	-0.085***	-0.024*
1 = Was away from the household for at least one month in the past 12 months	0.091***	0.041*	0.047***
1 = Main occupation in farming or fishing in the past year	-0.030*	0.010	-0.037***
Area under cultivation, acres / 1000	-0.984	0.391	-2.118*
Squared area under cultivation, acres / 1000000	49.573	2.843	64.446
Livestock (TLU) / 1000	0.187	0.731	-1.156
1 = Household head is male	0.010	0.031**	-0.016
Number of household members	0.000	-0.002	0.002
1 = Household experienced agricultural shock in the past year	-0.021	-0.024**	0.003
1 = Household experienced non-agricultural shock in the past year	0.009	-0.001	0.012
1 = From high-density rural area	0.017	0.001	0.017
Population density, people per square km / 1000	0.031	0.036**	0.002
Distance to road, km / 1000	-0.109	0.493*	-0.623***
Distance to the nearest town with population of at least 50,000, km / 1000	-0.060	0.068	-0.174

Note: All regressions contain age of the household head and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.54. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; define “migrant” if an individual is considered to be a migrant by both the definition based on distance traveled and by self-reports

	1 = Moved to low-density rural	2 = Moved to high-density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
Age	-0.001	0.000	0.001**	-0.001	-0.000
Age squared	-0.004**	-0.002	-0.007**	-0.008**	-0.002
1 = Male	-0.049***	-0.017**	-0.013**	-0.008	0.002
1 = Completed primary school	-0.013	0.005	0.007	0.013***	0.005
1 = Married	-0.067***	-0.013	-0.029***	-0.002	-0.011
1 = Head of the household	0.055*	-0.005	-0.006	0.027	-0.014**
1 = Child of household head	-0.011	0.001	-0.017***	0.000	-0.012*
1 = Born in this village	-0.063***	-0.021**	0.005	-0.018**	-0.014
1 = Was away from the household for at least one month in the past 12 months	0.030	0.014	0.010	0.012	0.022*
1 = Main occupation in farming or fishing in the past year	0.011	-0.001	-0.008	-0.013**	-0.019***
Area under cultivation, acres / 1000	0.615	0.087	-1.120	0.012	-0.471
Squared area under cultivation, acres / 1000000	-8.744	17.520	99.961	0.332	-5730.438
Livestock (TLU) / 1000	0.798**	-2.274	-0.449	-0.476	-0.268
1 = Household head is male	0.032***	-0.001	-0.002	-0.005	-0.006
Number of household members	-0.001	-0.001	-0.000	0.001	0.002*
1 = Household experienced agricultural shock in the past year	-0.008	-0.016**	0.000	-0.002	0.005
1 = Household experienced non-agricultural shock in the past year	0.001	-0.002	-0.004	0.009	0.005
1 = From high-density rural area	-0.010	0.014	0.006	0.028**	-0.004
Population density, people per square km / 1000	-0.022	0.022***	0.007	-0.065*	0.010
Distance to road, km / 1000	0.383*	0.088	-0.179	-0.012	-0.268*
Distance to the nearest town with population of at least 50,000, km / 1000	0.056	-0.006	-0.292***	0.022	0.032

Note: All regressions contain age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.55. Comparison of migration rates by migration status: for the definitions of “migrant” based on distance traveled and administrative change

		Non-migrants (both definitions)	Migrants (distance only)	Migrants (administr. change only)	Migrants (both definitions)
Constructed definition of “rural”	Located in a low-density rural area in 2008/2009	70.5%	70.9%	70.7%	60.5%
	Located in a high-density rural area in 2008/2009	29.5%	29.1%	29.3%	39.5%
	Located in a low-density rural area in 2012/2013	70.5%	65.4%	70.7%	29.2%
	Located in a high-density rural area in 2012/2013	29.5%	26.8%	29.3%	22.1%
	Located in a peri-urban area in 2012/2013		5.1%		15.7%
	Located in a town in 2012/2013		2.6%		13.9%
	Located in a city in 2012/2013		0.1%		19.0%
	Located in a low-density rural area in 2008/2009	65.5%	68.0%	70.7%	55.5%
	Located in a high-density rural area in 2008/2009	24.8%	27.1%	27.7%	25.4%
	Located in a town in 2008/2009	9.7%	4.8%	1.6%	19.1%
NBS definition of “rural”	Located in a low-density rural area in 2012/2013	63.4%	60.9%	66.7%	27.0%
	Located in a high-density rural area in 2012/2013	24.6%	21.6%	24.5%	22.9%
	Located in a town in 2012/2013	11.9%	17.1%	7.4%	27.9%
	Located in a city in 2012/2013	0.2%	0.3%	1.4%	22.1%
	Number of observations	2,287	187	77	252

Note: For people of age 15-34 living in rural areas in 2008/2009 according to the constructed definition. For migrants: “distance only” stands for the sample of people who traveled some distance but did not cross the district border; “administrative change only” stands for the sample of people for whom no travel is observed, but the district changed; “both definitions” stands for the sample of people who traveled some distance and crossed the district border. Sampling weights from the 2008/2009 survey wave are applied.

Table 2.56. Comparison of characteristics by migration status, by definition of “migrant”: for the definitions based on distance traveled and administrative change

	Non-migrants (both definitions)	Migrants (distance only)	Migrants (administrative change only)	Migrants (both definitions)
Age	23.3	21.14***	21.78**	22.17***
1 = Male	0.50	0.32***	0.61**	0.46
1 = Completed primary school	0.65	0.60	0.37***	0.71*
1 = Married	0.48	0.37***	0.42	0.31***
1 = Head of the household	0.19	0.10***	0.19	0.15*
1 = Child of household head	0.41	0.48*	0.47	0.38
1 = Born in this village	0.82	0.73***	0.74**	0.61***
1 = Was away from the household for at least one month in the past 12 months	0.09	0.11	0.07	0.2***
1 = Main occupation in farming or fishing in the past year	0.69	0.68	0.68	0.5***
Area under cultivation, acres	6.36	7.51	6.44	7.39
Livestock (TLU)	3.42	6.24***	3.11	2.04*
Age of household head	44.41	47.79***	44.02	44.76
1 = Household head is male	0.80	0.79	0.88*	0.78
Number of working age women	1.77	2.17***	1.74	1.95**
Number of working age men	1.81	1.88	2.08*	1.92
Number of children of household head living in the household	3.31	3.50	4.09***	3.14
1 = Household experienced agricultural shock in the past year	0.29	0.30	0.26	0.21***
1 = Household experienced non- agricultural shock in the past year	0.29	0.29	0.20**	0.33
Population density, people per square km	98.58	92.94	82.92	128.5***
Distance to road, km	21.51	22.96	20.58	19.11*
Distance to the nearest town with population of at least 50,000, km	67.53	76***	62.62	61.43**
Number of observations	2,287	187	77	252

Note: For migrants: “distance only” stands for the sample of people who traveled some distance but did not cross the district border; “administrative change only” stands for the sample of people for whom no travel is observed, but the district changed; “both definitions” stands for the sample of people who traveled some distance and crossed the district border. Sampling weights from the 2008/2009 survey wave are applied. Stars indicate significant differences in means between migrants and non-migrants: *** 0.01; ** 0.05; * 0.1.

Table 2.57. Regression results (marginal effects) with the constructed definition of “rural”: binary division and two destinations; definition of “migrant” is based on distance traveled and change in administrative area

	Logistic regression	Multinomial logistic regression	
	1 = Migrant	1 = Moved to rural	2 = Moved to urban
Age	0.001	0.002*	0.000
Age squared	0.000	-0.002	-0.003
1 = Male	-0.015	-0.014	-0.003
1 = Completed primary school	0.014	-0.002	0.018**
1 = Married	-0.079***	-0.050***	-0.028**
1 = Child of household head	-0.040***	-0.011	-0.026***
1 = Born in this village	-0.079***	-0.051***	-0.026**
1 = Was away from the household for at least one month in the past 12 months	0.066***	0.030**	0.035**
1 = Main occupation in farming or fishing in the past year	-0.045***	-0.007	-0.038***
Area under cultivation, acres / 1000	-0.960	-0.238	-0.679
Squared area under cultivation, acres / 1000000	57.262	20.922	16.648
1 = Household head is male	0.009	0.014	-0.006
Number of household members	0.001	-0.000	0.001
1 = Household experienced agricultural shock in the past year	-0.024**	-0.028***	0.003
1 = Household experienced non-agricultural shock in the past year	0.015	0.002	0.013
1 = From high-density rural area	0.016	0.005	0.012
Population density, people per square km / 1000	0.005	0.003	0.004
Distance to road, km / 1000	-0.113	0.246	-0.343*
Distance to the nearest town with population of at least 50,000, km / 1000	-0.176	-0.156	-0.039

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

Table 2.58. Regression results (marginal effects) with the constructed definition of “rural”: five destinations; definition of “migrant” is based on distance traveled and change in administrative area

	1 = Moved to low-density rural	2 = Moved to high-density rural	3 = Moved to peri- urban area	4 = Moved to town	5 = Moved to city
Age	0.001	0.001	0.001**	-0.001	-0.000
Age squared	-0.002	-0.001	-0.005	-0.007*	-0.001
1 = Male	-0.008	-0.006	-0.005	-0.005	0.005
1 = Completed primary school	-0.010	0.009*	0.001	0.010**	0.007
1 = Married	-0.039***	-0.013	-0.019**	-0.005	-0.008
1 = Child of household head	-0.010	-0.001	-0.010*	-0.002	-0.011*
1 = Born in this village	-0.032***	-0.019**	-0.002	-0.014*	-0.010
1 = Was away from the household for at least one month in the past 12 months	0.018	0.013	0.000	0.011	0.022*
1 = Main occupation in farming or fishing in the past year	-0.004	-0.003	-0.015**	-0.008	-0.017***
Area under cultivation, acres / 1000	0.142	-0.435	0.311	0.096	-0.903
Squared area under cultivation, acres / 1000000	4.787	55.889	-116.946	-18.035	-694.703
1 = Household head is male	0.015**	-0.001	0.006	-0.003	-0.005
Number of household members	0.001	-0.001	-0.002*	0.001	0.002*
1 = Household experienced agricultural shock in the past year	-0.012*	-0.017***	0.000	-0.001	0.004
1 = Household experienced non-agricultural shock in the past year	0.002	0.000	0.002	0.008	0.003
1 = From high-density rural area	-0.001	0.009	0.003	0.027**	-0.007
Population density, people per square km / 1000	-0.029	0.006	0.006	-0.051	0.007
Distance to road, km / 1000	0.191	0.060	0.037	-0.023	-0.260*
Distance to the nearest town with population of at least 50,000, km / 1000	-0.181*	0.026	-0.156**	0.012	0.064

Note: All regressions contain indicator of being the head of the household, livestock units owned by the households, age of the household head, and geographical zone fixed effects. The marginal effect of age squared is calculated separately using a formula for partial effect. *** 0.01; ** 0.05; * 0.1.

REFERENCES

REFERENCES

- Allen IV, J. E. 2018. Are agricultural markets more developed around cities? Testing for urban heterogeneity in separability in Tanzania. *Food Policy*, 79, 199-212.
- Beegle, K., J. De Weerdt, and S. Dercon. 2011. Migration and economic mobility in Tanzania: Evidence from a tracking survey. *The Review of Economics and Statistics*, 93(3), 1010-1033.
- Beegle, K., and M. Poulin. 2013. Migration and the transition to adulthood in contemporary Malawi. *ANNALS of the American Academy of Political and Social Science*, 648(1), 38-51.
- Bernard, A., M. Bell, and E. Charles-Edwards. 2014. Life-course transitions and the age profile of internal migration. *Population and Development Review* 40(2): 213-239.
- Bezu, S., and S. Holden. 2014. Are rural youth in Ethiopia abandoning agriculture? *World Development*, 64, 259-272.
- Bilsborrow, R. E., T. M. McDevitt, S. Kossoudji, and R. Fuller. 1987. The impact of origin community characteristics on rural-urban out-migration in a developing country. *Demography*, 24(2), 191-210.
- Bohra-Mishra, P., M. Oppenheimer, and S. M. Hsiang. 2014. Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences*, 111(27), 9780–9785.
- Brown, L. A., and V. A. Lawson. 1985. Rural-destined migration in Third World setting: A neglected phenomenon? *Regional Studies* 19(5): 415-432.
- Brusco, M. J., R. Singh, J. D. Cradit, and D. Steinley. 2017. Cluster analysis in empirical OM research: survey and recommendations. *International Journal of Operations & Production Management*, 37(3), 300-320.
- Christiaensen, L., J. De Weerdt, and R. Kanbur. 2019. Decomposing the contribution of migration to poverty reduction: methodology and application to Tanzania. *Applied Economics Letters*, 26(12), 978-982.
- Christiaensen, L., J. De Weerdt, and Y. Todo. 2013. Urbanization and poverty reduction: the role of rural diversification and secondary towns. *Agricultural Economics*, 44(4-5), 435-447.
- Christiaensen, L., and R. Kanbur. 2017. Secondary towns and poverty reduction: refocusing the urbanization agenda. *Annual Review of Resource Economics*, 9, 405-419.
- Christiaensen, L., and Y. Todo. 2014. Poverty reduction during the rural-urban transformation – The role of the missing middle. *World Development*, 63, 43-58.

Cockx, L., L. Colen, and J. De Weerd. 2018. From corn to popcorn? Urbanization and dietary change: Evidence from rural-urban migrants in Tanzania. *World Development*, 110, 140-159.

Corbane, C., A. Florczyk, M. Pesaresi, P. Politis, and V. Syrris. 2018. GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A. *European Commission, Joint Research Centre (JRC)* doi: 10.2905/jrc-ghsl-10007 PID: <http://data.europa.eu/89h/jrc-ghsl-10007>

De Brauw, A., V. Mueller, and H. L. Lee. 2014. The role of rural-urban migration in the structural transformation of Sub-Saharan Africa. *World Development*, 63, 33-42.

De Weerd, J. 2010. Moving out of poverty in Tanzania: Evidence from Kagera. *Journal of Development Studies*, 46(2), 331-349.

Dinbabo, M. F., C. Mensah, and M. N. Belebema. 2017. Diversity of rural migrants' profiles. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in motion. Dynamics and drivers of migration South of the Sahara*, pp. 24-25. Rome: FAO; CIRAD.

Dustmann, C., and A. Okatenko. 2014. Out-migration, wealth constraints, and the quality of local amenities. *Journal of Development Economics*, 110, 52-63.

Epule, T. E., C. Peng, and L. Lepage. 2015. Environmental refugees in Sub-Saharan Africa: A review of perspectives on the trends, causes, challenges and way forward. *GeoJournal*, 80(1), 79-92.

Fox, L., and A. Thomas. 2016. Africa's got work to do: A diagnostic of youth employment challenges in Sub-Saharan Africa. *Journal of African Economies*, 25(AERC supplement 1), i16-i36.

Gray, C., and R. Bilsborrow. 2013. Environmental influences on human migration in rural Ecuador. *Demography*, 50(4), 1217-1241.

Halpin, B. 2017. SADI: Sequence analysis tools for Stata. *The Stata Journal*, 17(3), 546-572.

Harris, J. R., and M. P. Todaro. 1970. Migration, unemployment and development: a two-sector analysis. *American Economic Review*, 60(1), 126-142.

Herrera-Almanza, C., and D. E. Sahn. 2020. Childhood determinants of internal youth migration in Senegal. *IZA Discussion Papers*, No. 12988.

Hirvonen, K. 2016. Temperature changes, household consumption, and internal migration: Evidence from Tanzania. *American Journal of Agricultural Economics*, 98(4), 1230-1249.

Iaquinta, D. L., and A. W. Drescher. 2000. Defining the peri-urban: Rural-urban linkages and institutional connections. *Land Reform, Land Settlement and Cooperatives*, 2, 8-27.

Ingelaere, B., L. Christiaensen, J. De Weerd, and R. Kanbur, R. 2018. Why secondary towns can be important for poverty reduction – A migrant perspective. *World Development*, 105, 273-282.

- Jedwab, R., L. Christiaensen, and M. Gindelsky, M. 2017. Demography, urbanization and development: rural push, urban pull and... urban push? *Journal of Urban Economics*, 98, 6-16.
- Kombe, W. J. 2005. Land use dynamics in peri-urban areas and their implications on the urban growth and form: The case of Dar es Salaam, Tanzania. *Habitat International*, 29(1), 113-135.
- Koubi, V., G. Spilker, L. Schaffer, and T. Bernauer. 2016. Environmental stressors and migration: Evidence from Vietnam. *World Development*, 79, 197-210.
- Kudo, Y. 2015. Female migration for marriage: Implications from the land reform in rural Tanzania. *World Development*, 65, 41-61.
- Lanjouw, P., J. Quizon, and R. Sparrow. 2001. Non-agricultural earnings in peri-urban areas of Tanzania: Evidence from household survey data. *Food Policy*, 26(4), 385-403.
- Lucas, R. E. B. 1997. Internal migration in developing countries. In Rosenzweig, M. R., and O. Stark., (eds.) *Handbook on population and family economics*. Amsterdam: Elsevier, pp. 721-798.
- Lucas, R. E. B. 2016. Internal migration in developing economies: An overview of recent evidence. *Geopolitics, History, and International Relations*, 8(2), 159-191.
- Mapunda, D. W., S. S. Chen, and C. Yu. 2018. The role of informal small-scale water supply system in resolving drinking water shortages in peri-urban Dar es Salaam, Tanzania. *Applied Geography*, 92, 112-122.
- Marchiori, L., J. F. Maystadt, and I. Schumacher. 2012. The impact of weather anomalies on migration in Sub-Saharan Africa. *Journal of Environmental Economics and Management*, 63(3), 355-374.
- McAuliffe, M., and A. Triandafyllidou (eds.). 2021. *World Migration Report 2022*. International Organization for Migration (IOM), Geneva.
- Mercandalli, S. 2017. Prevalent, contrasted intra-African migration patterns and new territorial dynamics. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in motion. Dynamics and drivers of migration South of the Sahara*, pp. 22-23. Rome: FAO and CIRAD.
- Mercandalli, S., B. Losch, C. Rapone, R. Bourgeois, and C. A. Khalil. 2017. Rural migration and the new dynamics of structural transformation in Sub-Saharan Africa. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in motion. Dynamics and drivers of migration South of the Sahara*, pp. 14-17. Rome: FAO and CIRAD.
- Msigala, S. C., F. P. Mabiki, B. Stylishave, and R. H Mdegela. 2017. Performance of wastewater stabilization ponds in treatment of endocrine disrupting estrogens in Morogoro urban and peri-urban, Tanzania. *International Journal of Public Health and Epidemiology*, 6(1), 305-317.
- Msigwa, R. E., and J. E. Mbongo. 2013. Determinants of internal migration in Tanzania. *Journal of Economics and Sustainable Development*, 4(9), 28-35.

- Mueller, V., E. Schmidt, N. Lozano, and S. Murray. 2019. Implications of migration on employment and occupational transitions in Tanzania. *International Regional Science Review*, 42(2), 181-206.
- Muzzini, E., and W. Lindeboom. 2008. The urban transition in Tanzania: Building the empirical base for policy dialogue. *Working Paper* 44972. Washington, DC: World Bank.
- National Bureau of Statistics (NBS), The United Republic of Tanzania. 2015. *Migration and urbanization monograph. Migration and urbanization report: 2012 population and housing census*, Volume IV. Dar es Salaam and Zanzibar.
- Oucho, J. O., and W. T. S. Gould. 1993. Internal migration, urbanization, and population distribution. In Foote, K. A., K. H. Hill, and L. G. Martin. (Eds.) *Demographic change in Sub-Saharan Africa*, pp. 256-296. Washington, D. C.: National Academy Press.
- Potts, D. 2017. Conflict and collisions in Sub-Saharan African urban definitions: Interpreting recent urbanization data from Kenya. *World Development*, 97, 67-78.
- Potts, D. 2017. Urban data and definitions in Sub-Saharan Africa: Mismatches between the pace of urbanization and employment and livelihood change. *Urban Studies*, 55(5), 965-986.
- Proctor, F. J., and V. Lucchesi. 2012. *Small-Scale Farming and Youth in an Era of Rapid Rural Change*. London/The Hague: International Institute for Environment and Development / Humanist Institute for Development Cooperation.
- Rapsomanikis, G. 2015. *The Economic Lives of Smallholder Farmers: An Analysis Based on Household Data from Nine Countries*. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Reed, H. E., C. S. Andrzejewski, and M. J. White. 2010. Men's and women's migration in coastal Ghana: An event history analysis. *Demographic Research*, 22, 771-812.
- Sassen, S. 2016. A massive loss of habitat: New drivers for migration. *Sociology of Development*, 2(2), 204-233.
- Tatem, A. J. 2017. WorldPop: Open data for spatial demography. *Scientific Data*, 4:170004.
- Tegegne, A. D., and M. Penker. 2017. Determinants of rural out-migration in Ethiopia: Who stays and who goes? *Demographic Research*, 35, 1011-1044.
- United Nations (UN). 2008. Table 6. In *United Nations Demographic Yearbook 2005: Fifty-Seventh Issue*, pp. 103-107. New York: United Nations.
- United Nations Conference on Trade and Development (UNCTAD). 2018. *Economic Development in Africa Report 2018: Migration for Structural Transformation*. New York and Geneva: United Nations.

United Nations, Department of Economic and Social Affairs. 2013. *Trends in International Migrant Stock: Migrants by Destination and Origin* (United Nations database, POP/DB/MIG/Stock/Rev.2013).

Wenban-Smith, H. 2015. Population growth, internal migration and urbanization in Tanzania, 1967-2012. Phase 2 (Final Report). *Working Paper C-40211-TZA-1*, International Growth Centre.

Wineman, A., D. Y. Alia, and C. L. Anderson. 2020. Definitions of “rural” and “urban” and understandings of economic transformation: Evidence from Tanzania. *Journal of Rural Studies*, 79, 254-268.

Wondimagegnhu, B. A., and M. E. Zeleke. 2017. Determinants of rural out-migration in Habru district of Northeast Ethiopia. *International Journal of Population Research*, 2018, 1-8.

World Bank. 2017. *Living Standards Measurement Study – Integrated Surveys on Agriculture, Tanzania*.

Wave 1 (2008-2009) retrieved from <http://microdata.worldbank.org/index.php/catalog/76>.

Wave 2 (2010-2011) retrieved from <http://microdata.worldbank.org/index.php/catalog/1050>.

Wave 3 (2012-2013) retrieved from <http://microdata.worldbank.org/index.php/catalog/2252>.

Zhang, Q., R. E. Bilsborrow, C. Song, S. Tao, S., and Q. Huang. 2018. Determinants of out-migration in rural China: Effects of payments for ecosystem services. *Population and Environment*, 40, 182-203.

3. MIGRATION OF YOUTH TO DIFFERENT DESTINATION TYPES IN TANZANIA: HOW DOES THE LEVEL OF URBANIZATION AFFECT EMPLOYMENT SHIFTS?

Abstract

This paper investigates how different migration destination categories on the rural-urban spectrum facilitate shifts in main occupation among rural youth. The study is motivated by the arising debate on the role of rural areas and rural non-farm economy in structural transformation, as well as by the recent evidence on the differences that various destinations on the rural-urban spectrum could uphold. Using the data from the Living Standards Measurement Study in Tanzania, I describe migration trends for various destinations and the associated occupational shifts. My analysis distinguishes low- and high-density rural areas, peri-urban areas, small towns, and large cities, which enables a more nuanced understanding of which destination types involve the most drastic shifts in employment associated with structural transformation. I account for selection into migration using matching techniques and compare employment outcomes of migrants to those of non-migrants with similar initial characteristics. I show that the majority of migration in Tanzania is rural-to-rural, not rural-to-urban as is often presumed, and that even migration to low-density rural areas promotes structural transformation through an increase in the probability to shift main occupation to non-farm wage job or self-employment. People who move to more urbanized areas are less likely to have main occupation in agriculture and more likely to have a main occupation in an off-farm sector at destination, but those who move to the most urbanized places already leaned towards off-farm employment at baseline compared to non-migrants. Migration to peri-urban areas is associated with underemployment, while migration to cities is associated with unemployment at destination.

3.1. Introduction

Structural transformation, the deep change to the structure of the economy from the agricultural sector to manufacturing and services, is a central and essential part of economic growth which encompasses many spheres of economic life, from employment and labor productivity to consumption (Herrendorf, Rogerson, and Valentinyi, 2014).⁵⁰ Shift in occupation from agricultural to non-agricultural activities is an integral part of structural transformation that developing countries undergo, and migration could facilitate this shift.⁵¹ Classical models associate transition from agriculture to manufacturing solely with rural-to-urban migration (Lewis, 1954; Harris and Todaro, 1970). Due to this, urban destinations received superior coverage in the literature, although rural destinations commonly prevail in migrants' choices (Lucas, 2016). At the same time, constant growth of the rural non-farm economy and the overall economic development provide better off-farm employment opportunities in less urbanized locations (Diao, Magalhaes, and McMillan, 2018). The effect of this is two-fold: on the one hand, people who move to rural areas can get better access to non-agricultural employment; on the other hand, people moving from rural areas are less likely to be farmers prior to their move. Recent studies expanded the perspective on migration destinations from the binary rural-urban case to looking at the role particular destinations, like secondary towns, peri-urban areas, and rural areas, play in structural transformation (Emran and Shilpi, 2018; Mueller et al., 2019; De Brauw, Mueller, and Lee, 2014). Overall, with stable and high rates of migration in Sub-Saharan Africa (Mercandalli et al., 2017), and a rapidly increasing rural population (Losch, 2017),

⁵⁰ This chapter is co-authored with Thomas S. Jayne.

⁵¹ According to the World Development Indicators database (World Bank, <https://databank.worldbank.org/source/world-development-indicators>), in 2008 in Tanzania the value added of agriculture, forestry, and fishing was 24.8% of GDP, while 71.3% of employed population were employed in agriculture. By 2019, the share of value added of the agricultural sector in GDP increased to 26.5% while the share of employment decreased to 65.1%. At the same time, the share of employment in industry and services increased.

migration should remain an important means of occupational shift affecting millions of people. This study aims to provide a comprehensive view on the shifts occurring with migration of rural youth to different destinations on the rural-urban spectrum and assess the ability of different destination types to help people transition to working in a new sector.

The impact of migration on the employment choice at destination can be viewed as an outcome of two forces: migrant's will to shift and the structure of employment at destination. For many migrants, their reason for migration is tied to employment – although it does not necessarily translate to a sectoral shift. More urbanized destinations could offer a wider range of non-farm employment opportunities, and the actual structure of employment and the welfare outcomes could differ a lot depending on the destination's type (Christiaensen and Kanbur, 2017). Self-selection into migration from rural population is not random (McKenzie, Stillman, and Gibson, 2010), and it can be related to selection into occupation at destinations and thus should be accounted for. In this paper, I employ matching techniques to build a counterfactual for migrants' employment at the origin and estimate how youth' migration to various destinations on the rural-urban spectrum promote shifts in main occupation. I look at the internal movements of people of age 15-34 from rural areas in Tanzania using the 2008/2009 and the 2012/2013 waves of the Living Standards Measurement Study (LSMS; World Bank, 2017) dataset.

I start with the description of migration trends and the associated occupational shifts. Then, I estimate how migration to certain destination types contributes to the shifts from employment to unemployment, from unemployment to employment, into agricultural sector, and into non-agricultural sector. Transition into unemployment associated with migration to urban areas is a major concern for the ability of rural-to-urban migration to provide the means to

sectoral shifts in employment (Harris and Todaro, 1970). On the other hand, migration can provide employment opportunities for people who were underemployed or unemployed at baseline. Shifts away from farming and into the non-farm sector contribute directly to the structural transformation of the economy. As the majority of youth in the sample reports to move for reasons not related to work, I analyze certain types of migrants separately: women moving for marriage and students. I contribute to the literature on employment transitions and employment challenges of youth in Sub-Saharan Africa and consider the impacts of migration on these issues. By studying destinations on the rural-urban spectrum, I enhance the knowledge on these destinations in particular and on the spectrum itself. I expand the analysis of Mueller et al. (2019) done for the first two waves of the LSMS by looking at a wider range of migration destinations and by increasing the time scope. I also contribute to the growing literature on the importance of secondary towns as migration destinations for rural youth (Christiaensen and Kanbur, 2017; Ingelaere et al., 2018) and stress the importance of rural destinations.

3.2. Literature review

3.2.1. Migration destinations

The classic dual-sector model is built on the assumption of the existence of rural agricultural and urban manufacturing sectors with different productivities and, consequently, different wages (Lewis, 1954). Therefore, rural-to-urban migration is viewed as a way to shift labor from the less productive agricultural to the more productive manufacturing sector and foster structural transformation. In particular, urban destinations allow youth to transition from the main occupation of their parents, which is mostly farming, to non-farm activities (Fox and Thomas, 2016). With this positive view on rural-to-urban migration, it has long been a focus of

many researchers, although urban areas are not the most prominent destination among people moving internally in developing countries (Lucas, 2016). At the same time, concerns rise regarding unemployment among migrants to urban areas.

Harris and Todaro (1970) incorporated urban unemployment into the two-sector model and showed that people still move to urban areas because of higher expected earnings there. Recent evidence suggests that young people migrating from rural to urban areas are often employed, but they are likely to be underemployed (Filmer and Fox, 2014). One of the examples of underemployment in cities could be a shoe-shine industry. Elkan, Ryan, and Mukui (1982) describe the industry in Nairobi, Kenya, as the one with low barriers to entry, with average worker's age at 25, and the one that allows to generate income soon after entrance. In a more recent study from Ghana, Tanle (2018) describes workers in the industry to be young, many of them are migrant from rural areas. Although most workers don't see their job as a long-term position as it is physically and mentally exhausting, many people stay to earn enough money to settle in an urban area.

The perspective in the literature has long been focused on big cities, but it is now shifting to other non-rural destinations. Emran and Shilpi (2018) discuss to the role that the changes in employment in secondary towns plays in structural transformation. Mueller et al. (2019) look at the migration to and from peri-urban areas and provide arguments for the importance of these areas to employment shifts, and, consequently, structural transformation. Filmer and Fox (2014) argue that young people settled in peri-urban areas and secondary towns could use agriculture as a "steppingstone" before transitioning into self-employment. Hence, migrant farmers might choose to be employed in agriculture at destination as it is a familiar activity for them and, right after migration, they may rely on it to provide higher expected returns than activities that are new

for them. Smaller towns and peri-urban areas give more opportunities for agricultural activities, either as wage work or self-employment, than cities do.

On the other hand, migration is not necessarily associated with an intent to shift one's occupation: some farmers may want to stay in agriculture. Though, if migrants needed time to collect enough capital to start their own farm at destination, I would observe a gap between their move and their employment transition back to agriculture. Masvaure (2016) shows that the majority of urban farmers in Harare, Zimbabwe, originate from rural areas, although they are not recent migrants. In my sample, I only observe people who spent up to 4 years at destination, so I might not be able to see this transition back to agriculture among migrants with less starting capital. This argument is also correct for people who lack capital to start a non-farm business, and, unfortunately, I have no way of knowing the intentions of a migrant.

In the past, with less non-farm employment opportunities in rural areas, migration definitely played a huge role in the shift from agricultural to non-agricultural activities when people moved from rural to urban areas (hence, providing empirical evidence for dual-sector models). De Brauw, Mueller, and Lee (2014) also argue that rural-to-urban migration is tied to structural transformation, but they briefly discuss the role of rural-to-rural migration in this process as well. Nowadays, rural areas become more attractive with the rise of the rural non-farm economy (Nagler and Naudé, 2017; Davis, Di Giuseppe, and Zezza, 2017). Hence, people willing to shift to non-farm employment do not require to move to a city anymore. At the same time, it means that I might observe non-farmers as well as farmers moving to rural areas.

An increasing number of studies on the outcomes of migration distinguishes migration destinations. Ingelaere et al. (2018) argue that it is easier for migrants from rural areas to adjust to their new community when they move to another rural area or a smaller town rather than when

they move to a big city. While studying the impacts of migration in Tanzania on the composition of migrants' diets, Cockx, Colen, and De Weerd (2018) distinguish migration to rural areas, secondary, and cities. One of the focus points of their analysis is the movement from an agricultural household to a non-agricultural household (based on the occupation of the household head). Christiaensen and Kanbur (2017) look at the benefits of migration to rural areas, towns, and cities in Tanzania and conclude that gains from migration increase with the movement across the rural-urban spectrum towards more urbanized locations – although they do not specify whether the welfare benefits are associated with shifts in the type of employment. Mueller et al. (2019) show that migration to various destinations across the rural-urban spectrum in Tanzania leads to diverse pattern in both industrial shifts and shifts in and out of unemployment depending on destination.

3.2.2. Employment of youth

Decisions made early in life have a huge impact on the future career path and earnings. With the data for people of age 20-35 living in urban Tanzania, Bridges et al. (2017) find that early career choices greatly impact future earnings. The four jobs considered in their study are wage job, self-employment, participation in family business, and job seeking; and the authors find wage job to be the most favorable early position in terms of the earning prospects. Whenever youth struggles to secure a job, migration may serve as a pathway to improved livelihood (Filmer and Fox, 2014). Beegle, De Weerd, and Dercon (2011) find that, for migrants, the move itself is correlated with a growth of consumption beyond improved opportunities coming from better connection to markets or a more urbanized environment at the new location. For youth, when they transition from school to work, it may be easier to make the decision to enter a particular sector or have a certain type of job than the decision to shift from

one sector to another later in life. Rural youth in particular could be eager to shift away from agriculture that their parents and grandparents pursue (Fox and Thomas, 2016). On the other hand, those who are not willing to shift away from farming could experience hardships at their village of residence, for example, due to the lack of available agricultural land (Bezu and Holden, 2014).

It might be hard for rural youth to find a job in the formal sector, which would provide more stability⁵², after they move to an unfamiliar place. Beauchemin and Bocquier (2004) find that in West Africa migrants, especially younger migrants, are more likely to start their employment in the formal sector than non-migrants do once education is controlled for. Hence, they deem lack of education and not migration itself to be the reason why migrants are employed in the informal sector. Fox, Senbet, and Simbanegavi (2016) explain African youth entering informal sector instead of formal wage jobs by the fact that many young people struggle to get a set of cognitive and non-cognitive skills that is necessary to start formal employment, behavioral skills in particular (for example, they list perseverance, risk aversion, and self-esteem). They argue that poor rural youth have to get these skills at school, from parents, and in the community they live in. From the observations of Elkan, Ryan, and Mukui (1982), it means that even work in the informal sector then could be associated with a higher probability of failure.

Decisions to shift occupations or to move away from the household of origin are made by most young people as they transition into adulthood. Klasen and Woolard (2008) find that in areas with high rural unemployment in South Africa young people could delay splitting up with

⁵² Blekking et al. (2020a) find a negative relationship between informal employment and food security in Lusaka, Zambia. Crush (2013) finds that non-migrants at urban destinations, through better employment opportunities and urban agriculture, are more likely to be food secure than migrants. Blekking et al. (2020b) find that recent migrants in a small city actually tend to have better access to food through more household assets and more members earning wage.

their parents and starting a household on their own. On the other hand, rural unemployment could stimulate rural out-migration as it increases the gap in expected wages between rural and urban areas (Harris and Todaro, 1970). In Ethiopia, Bezu and Holden (2014) find that many parents are willing to transfer land to their children once they get married, hence children who want to delay marriage are more likely to move, as at their origin the probability of them getting land is low. Filmer and Fox (2014) show that the majority of people of age from 15 to 34 in Sub-Saharan Africa transitions from working for their household to self-employment around the age of 20-25, regardless of gender and the location of origin on the rural-urban spectrum. On the other hand, they also find that people from rural areas tend to work longer for their families, with women continuing working for their family of origin even after they transition into adulthood.

In my main analysis in this chapter, I look at the shifts in main occupation, leaving aside secondary occupations. It could be a major drawback of my study: Filmer and Fox (2014) show that about half of youth in Tanzania is employed in more than one activity, with the share being much higher in rural than in urban areas. For agrarian households, involvement in non-agricultural activities could be seasonal and tied to the agricultural cycle (Bryceson, 2010; Burnod, Rakotomalala, and Bélières, 2017), hence some households may consider sending out a migrant, which induces repetitive, seasonal migration (Radel et al., 2018). I will look at the contribution of semi-permanent⁵³ migration into occupational shifts, as seasonal migrants are unidentifiable with the dataset I use. On the one hand, it allows to track more permanent changes and avoid bias with observed rural-to-urban labor migration among farmers during the lean season (Bryan, Chowdhury, and Mobarak, 2014). On the other hand, I will not be able to capture the variety of occupations migrants have at their origin prior to movement.

⁵³ I look at people who moved to a new location at most 4 years ago.

3.3. Data and definitions

I use two waves of the LSMS data for Tanzania that were conducted in 2008/2009 and 2012/2013 (World Bank, 2017). The sample is narrowed to individuals within the age range from 15 to 34 years old who lived in rural areas in 2008/2009. Mueller et al. (2019) use the first two waves (2008/2009 and 2010/2011) of the LSMS dataset for Tanzania and look at people of working age (15-65). People of age 15 and older who moved internally within the country were tracked by the survey team and interviewed in the subsequent survey waves, while younger people and international migrants were not tracked. I distinguish several types of migrants⁵⁴ destinations on the rural-urban spectrum: low-density rural areas, high-density rural areas, peri-urban areas, towns, and cities.

I use the definition of “rural” constructed in the first essay as the main definition for location types on the rural-urban spectrum. It is based on population density, built-up area density, and distance to the nearest town with population of at least 50,000 people.⁵⁵ Areas with population density above 400 people per sq. km and built-up area density above 8% located within 30 km radius of Dar es Salaam or Mwanza are considered to be cities, while such areas located elsewhere are considered to be towns. For all other areas, location within 30 km radius of a town with population of at least 50,000 people and population density of at least 150 people per sq. km places the area into the “peri-urban” category. All the remaining locations are split into

⁵⁴ To determine if an individual is a migrant, I use distance between survey waves provided in the dataset. The threshold for migration with this distance is set to five km by the survey team. For some observations, the information on distance is missing, and for them I check the distance computed using the coordinates provided in the dataset. These coordinates are aggregated across enumeration areas by the survey team, and a random offset is applied, which can be up to 10 km. For consistency, I apply the same threshold of five km to the computed distance traveled to define if the individual is a “migrant”.

⁵⁵ Data on the population density for 2010 comes from the WorldPop Africa Continental Population Databases (Tatem, 2017). Data on the built-up area density for 2013/2014 comes from the Global Human Settlement Layer (Corbane et al., 2018). A grid of one km is used for both datasets. I adjust coordinates pointing to water bodies to point at the nearest land instead.

high- and low-density rural areas using a threshold of 100 people per sq. km. Mueller et al. (2019) use different thresholds: a location with population density above 150 people per sq. km within an hour travel of a town with population of at least 20,000 people is defined as urban if its built-up area density is above 50% and peri-urban otherwise, and all other locations are considered to be rural. The two main differences between my definition and the definition employed by Mueller et al. (2019) are in (i) the use of towns with population of at least 50,000 people instead of 20,000, and (ii) the use of threshold of 8% for the built-up area density instead of 50%.⁵⁶

For a robustness check, I use the definition of “rural” from the Tanzanian National Bureau of Statistics (NBS) that is employed by the survey team. With this definition, locations are divided into rural and urban. I split rural areas further into low- and high-density areas using the same threshold of 100 people per sq. km that is used in the main definition. I split urban areas into towns and cities based on the district: all districts of Dar es Salaam and Nyamagana and Ilemela districts of Mwanza region are considered to be cities, all other urban districts are considered to be towns. With this definition, peri-urban areas are not distinguished.

Transition of main occupation observed after four years is of main interest for this chapter. I group self-reported main occupations in the following way: (1) “farming or fishing” when individuals state agriculture / livestock or fishing to be their main occupation; (2) “self-employment” when individuals state to be self-employed not in agriculture (with or without employees); (3) “wage job” when individuals state to be employed in a private sector, mining, or

⁵⁶ The choice of these thresholds is described in Appendix 3. See Figure 2.9 for the difference between the distance to the nearest town with population of at least 50,000 people and the distance to the nearest town with population of at least 20,000 people. Figure 2.5 and Figure 2.6 motivate the choice of threshold for the built-up area density. With the threshold of 50%, the number of people living in areas defined as urban is extremely low, while some locations re-categorized as “rural” have very high population density, low share of people employed in agriculture and low share of household income coming from agriculture.

tourism, or in a government, parastatal, or NGO/religious organization; (4) “student” when individuals state studies to be their main occupation; (5) “household maintenance” when individuals state paid or unpaid family work; (6) “no occupation” when individuals states to be unemployed (having no job or be job-seeking – more details are provided below) or disabled. In Table 3.1, I present the frequency of main occupations within each group by gender, age, and location type. I compare young adults of age 15-34 to adults of age 35-65 to see occupation trajectories by gender and location type. Table 3.14 in the Appendix is a re-calculation of Table 3.1 with the NBS definition of “rural”. The comparison of patterns of main occupation and key characteristics between the definitions of “rural” is given in the end of this section.

Prevalence of farming as main occupation depends on age and gender, although location type is still the most important factor. In 2008/2009, men of age 15-34 have the lowest share of people with main occupation in farming among them: 60% in rural and 13% in urban areas. Women of age 35-64 have the highest share: 92% in rural and 48% in urban areas state farming to be their main occupation. Age is also an important predictor of having main occupation in farming as among men of age 35-65 the share of people with main occupation in farming is higher than among women of age 15-34. People in urban areas are more likely to be self-employed without employees (the rates in 2008/2009 are 16-29% in urban and 2-5% in rural areas). In urban areas, older people are more likely to be self-employed alone. People in rural areas are, on average, less likely to work at private enterprises (0-2% in rural and 3-15% in urban areas) or governmental organizations (0-3% in rural and 1-9% in urban areas).

Table 3.1. Main occupation of people of age 15-65 in 2008/2009 and 2012/2013, by age group, gender, and location type

	Rural								Urban							
	Of age 15-34 in 2008/09				Of age 35-65 in 2008/09				Of age 15-34 in 2008/09				Of age 35-65 in 2008/09			
	Men		Women		Men		Women		Men		Women		Men		Women	
	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013
<i>A. Agriculture</i>	62%	70%	72%	77%	87%	85%	92%	87%	15%	13%	16%	16%	37%	32%	48%	47%
Agriculture/livestock	60%	68%	72%	77%	85%	84%	92%	87%	13%	12%	16%	16%	34%	31%	48%	47%
Fishing	2%	1%	0%	0%	1%	1%	0%	0%	2%	1%	0%	0%	2%	1%	0%	0%
<i>B. Self-employment</i>	4%	7%	3%	5%	6%	6%	4%	5%	18%	22%	18%	24%	33%	31%	25%	26%
Self-employed alone	4%	5%	2%	5%	5%	5%	3%	5%	16%	19%	17%	22%	29%	25%	24%	23%
Self-employed with employees	0%	1%	0%	0%	1%	0%	0%	0%	2%	3%	1%	2%	4%	6%	1%	2%
<i>C. Wage job</i>	3%	8%	1%	3%	7%	8%	2%	3%	18%	35%	8%	18%	28%	33%	9%	10%
Private enterprise	2%	6%	0%	2%	2%	2%	1%	2%	15%	30%	6%	13%	15%	21%	3%	5%
Government	1%	2%	0%	1%	3%	3%	1%	1%	2%	4%	1%	4%	9%	7%	5%	4%
Parastatal	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	2%	3%	0%	0%
Mining	0%	0%	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tourism	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NGO/religious	0%	0%	0%	0%	1%	1%	0%	0%	0%	1%	0%	1%	1%	2%	1%	0%
<i>D. Student</i>	26%	9%	18%	6%	0%	0%	0%	0%	37%	16%	26%	11%	0%	0%	0%	0%
<i>E. HH maintenance</i>	4%	4%	5%	7%	0%	0%	1%	2%	5%	4%	25%	24%	0%	0%	14%	13%
Family work without pay	3%	4%	5%	7%	0%	0%	1%	2%	4%	3%	22%	23%	0%	0%	13%	13%
Family work with pay	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	1%	0%	0%	0%	0%
<i>F. Other</i>	1%	2%	2%	2%	1%	2%	1%	3%	8%	10%	7%	8%	2%	4%	4%	5%
No job	1%	2%	1%	2%	0%	1%	0%	1%	5%	7%	6%	8%	1%	3%	3%	4%
Job seeker	1%	1%	0%	0%	0%	0%	0%	0%	2%	2%	1%	0%	0%	0%	0%	0%
Disabled	0%	0%	0%	0%	1%	1%	1%	2%	0%	1%	0%	1%	1%	1%	1%	1%
Number of observations	1342	1301	1461	1437	873	877	1022	1012	889	930	974	998	532	528	563	573

Note: "HH maintenance" stands for "household maintenance". Constructed definition of "rural" is used. Sample weights from each respective wave (2008/2009 or 2012/2013) are applied.

Around a quarter of people of age 15-34 are occupied without salary: they are students or work on household maintenance. Mueller et al. (2019) exclude these people from their main analysis⁵⁷. People of age 15-34 living in urban areas are more likely to be students, although a significant share of youth living in rural areas are students too: 26% of women and 37% of men living in urban areas in 2008/2009 list studies as their main occupation, while 18% of women and 26% of men living in rural areas do. In addition to that, more people of age 15-34 are students in 2012/2013 in urban areas (11% of women and 16% of men) than in rural areas (6% of women and 9% of men). Women living in urban areas are much more likely to state family work with no pay to be their main occupation: 22% (13%) of women of age 15-34 (of age 35-65) living in urban areas do so, while only 5% (1%) of women living in rural areas do.

The share of people without a job⁵⁸ is slightly higher in urban areas, especially among youth: it is 1-6% there, in contrast to 0-1% in rural areas. Between the answers “job seeker” and “no job”, men are, on average, more likely to list “job seeker” as their main occupation, and younger people are, on average, more likely to list “no job” and to state that they have never worked in their life yet. Ideally, “job seekers” should have no job, be available for work and looking for a job, while people without a job should have no job, not be available for work and not be looking for a job (otherwise they should have been listed as “job seekers”), but the data seems to paint a different picture.

⁵⁷ They include students into the category “employed” for the analysis of transition in and out of unemployment, while people involved mainly in household maintenance are excluded. They exclude both of these groups from the analysis of sectoral (agricultural and non-agricultural sectors) transitions. They consider self-employment in non-farm enterprises and wage labor to be non-agricultural employment, and I do the same in this chapter.

⁵⁸ I include disabled people in the category of individuals without a job. In my sample of rural youth, there are four disabled people in the first wave of survey. In the third survey wave, three of them had a job; none of them migrated.

First of all, with the data for 2012/2013, which is more nuanced than for 2008/2009, many of those who listed either “job seeker” or “no job” as their main occupation actually were involved in unpaid apprenticeship, paid employment, and agricultural and non-agricultural unpaid family work as their primary or secondary activity. Then, I can look at job seeking behavior and availability for work for the past week. This information is consistent: people who chose “job seeker” as their main occupation in the past 12 months in 2008/2009, compared to people who listed “no job”, are indeed more likely to be available for work, to take steps to find a job, and to do some work for pay, barter, or home use, although some people who listed “no job” as their main occupation answered the same way. Among people who listed “job seeker” and “no job” as their main occupation and who were not available for work in the past 7 days, most state household duties to be the main reason for not being available, while some people state being sick or disabled, and some youth state being busy with school.

For the main analysis, I use the sample of 2,803 individuals who were of age 15-34 and lived in rural areas (according to the main definition if not specified otherwise) in 2008/2009. Among them, 439 are migrants, of whom 142 individuals moved to an urban area and 297 individuals moved to another rural area.⁵⁹ As seen in Table 3.1, among youth who lived in rural areas in 2008/2009, the share of students and people with wage job as their main occupation is higher for men than for women, while women are, on average, more likely to have main occupation in farming. I use two sets of control variables, a small one and a large one.⁶⁰ Summary statistics for them are presented in Table 3.2.

⁵⁹ With the NBS definition of “rural”, the sample consists of 2,857 people. Of them, 151 moved to an urban area and 283 moved to a rural area by the last survey wave.

⁶⁰ Both sets are listed in section 3.4. In rare cases, land area and asset index are replaced with indicators of living in a household that was above median in land area and asset index respectively. In rare cases, an indicator of being a household head or an indicator of being married is dropped due to low number of observations.

Table 3.2. Summary statistics for baseline individual, household, and community characteristics of youth living in rural areas according to the constructed definition of “rural” (2,803 observations)

	Mean	Std. dev.	25th percentile	Median	75th percentile
<i>Small set of controls</i>					
Age	22.99	5.91	18	22	28
1 = Male	0.49	0.50			
1 = Married	0.40	0.49			
1 = Completed primary school	0.59	0.49			
1 = Born in this village	0.79	0.41			
Household size	6.81	4.27	4	6	8
Land area under cultivation, acres	6.54	18.83	1.5	3.5	6.0
Asset index	0.56	2.76	-1.31	-0.19	1.48
<i>Large set of controls (includes the small set)</i>					
1 = Head of the household	0.18	0.38			
1 = Child of the household head	0.42	0.49			
1 = Was away from the household for at least a month in the past year	0.10	0.30			
Age of the household head	44.66	15.12	32	43	56
1 = Household head is male	0.80	0.40			
Livestock units (TLU)	3.47	13.48	0.03	0.23	2.20
1 = Household experienced an agricultural shock in the past year	0.28	0.45			
1 = Household experienced a non-agricultural shock in the past year	0.29	0.45			
Population density, people per sq. km	100.55	147.52	36.74	72.09	116.17
Distance to the nearest road, km	21.35	20.19	6.10	17.50	28.70
Distance to the nearest town with population of at least 50,000 people, km	67.34	39.45	37.25	61.57	87.03

Note: Sample weights from the 2008/2009 survey wave are applied. Data on population density is from WorldPop Africa Continental Population Databases (Tatem, 2017). Data on distance to road is from the LSMS: it is computed by the survey team using the real coordinates of the households (real coordinates are not provided in the LSMS). Data on distance to the nearest town with population of at least 50,000 km is computed using households' coordinates provided in the LSMS (the survey team aggregated households' coordinates by enumeration area and added a random offset up to 10 km) and the towns' coordinates.

People living in rural areas according to the main definition of “rural” are less likely to have main occupation in farming and more likely to be self-employed without employees compared to people living in rural areas according to the NBS definition of “rural”. At the same time, for people living in urban areas the pattern is reversed. Also, older people living in urban areas according to the main definition of “rural” are less likely to have a wage job than people within the same age group living in urban areas according to the NBS definition, and women are less likely to have main occupation in household maintenance. With the main definition of “rural”, the sample of rural youth lives in households with lower average amount of land under cultivation, younger household head. The areas where they live, on average, have higher population density and are located closer to roads and towns. All these differences stem from the re-categorization of the households living in rural areas near towns.

3.4. Empirical strategy

My study is split into two parts. In the first part, I build tables of occupational shifts for non-migrants and people who moved to different destination types on the rural-urban spectrum. These tables support the descriptive analysis of the outcomes of migration. For migrants, they provide a single difference estimate: I compare changes in outcomes within one group of people between 2008/2009 and 2012/2013. This estimate is very likely to be biased, although it provides some insight into the dynamism at different destination types. For the unbiasedness, I need to assume that migrants’ occupation would have been the same in 2012/2013 as it was in 2008/2009, if they had not chosen to move. Since I look at youth, I expect to see people in my sample to change their occupation type quite often. The main concern is with students who are likely to finish their studies and progress to some type of employment in the span of four years between the survey waves.

Hence, I take a second difference and compare the change in occupation from 2008/2009 to 2012/2013 between migrants and non-migrants. This estimate adjusts the time bias but also introduces a new one based on the difference between migrants and non-migrants. For this estimate to be unbiased, I need to assume that migrants would have had the same shift in occupation as non-migrants did, if they had not chosen to move. As shown in the previous chapter, migration is not random: migrants to various destinations on the rural-urban spectrum differ from each other and from non-migrants, which might affect occupational shifts. Also, a decision to change occupation and a decision to migrate to a certain destination can be related. For example, a person who wants to improve their education can choose an urban destination expecting the quality of education to be higher there. Therefore, I need to account for selection into migration. I compare the outcomes of migrants to the outcomes of non-migrants with similar characteristics, this way I account for observable differences between migrants and non-migrants. But the bias emerging from non-observable characteristics, like skills, ambitions, and aspirations, might still remain if these unobservable characteristics are not captured by differences in observable characteristics. This is the main concern of non-experimental methods to the estimation of migration outcomes (McKenzie, Stillman, and Gibson, 2010).

In this paper, I show results of different approaches to accounting for selectivity in migration. Busso, DiNardo, and McCrary (2014) compare the performance of different reweighting and matching estimators of the treatment effect on the treated (that I aspire to estimate for migrants) and suggest using a set of estimators, since their properties depend on the data and the specification used. McKenzie, Stillman, and Gibson (2010) show that, when correcting for selection into migration to estimate income gains from migration, bias-adjusted nearest neighbor matching and difference-in-differences specification perform better than other

non-experimental methods. Though, any method they try overstates the experimental estimate by at least 20%. Mueller et al. (2019) estimate the impact of migration on employment outcomes using both propensity score and bias-adjusted nearest neighbor matching to account for selection, emphasizing the latter due to poor overlap in some of their data.

For the main analysis, for migrants to each destination type (and often, depending on the model specification, for migrants to a certain destination type who had a certain main occupation at baseline), I find matches among non-migrants. Following the literature discussed above, I use several matching strategies to limit the possibility for the results to be a feature of a specific estimation method. I use logistic regressions with controls, propensity score matching, and bias-adjusted nearest neighbor matching. Two sets of controls are employed: a smaller set⁶¹ and a larger set⁶². An attempt to find matches living in the same administrative area failed due to low number of observations in some categories. Selection into migration is further discussed in section 3.5.1: I am not able to account for selection into migration to the most urbanized locations using the observable characteristics. The quality of propensity score matching worsens with the level of urbanization as well (see Figure 3.1 in the Appendix).

⁶¹ It includes age, gender, indicator for being married, indicator for having completed primary school, indicator for being born in the village of residence, land area that the household cultivates, and asset index that compares the household's assets to assets of other rural households.

⁶² Along with variables from the smaller set of controls, it includes an indicator for being away from the household in the past year, indicator for being a head of the household, indicator for being a child of the household head, household head's age, household head's gender, units of livestock owned by the household, indicator for the household to experience an agricultural shock in the past year that negatively affected either household's income or assets, indicator for the household to experience a non-agricultural shock in the past year that negatively affected either household's income or assets, population density, distance to the nearest road, and distance to the nearest town with population of at least 50,000 people (computed using the coordinates provided in the dataset that were aggregated and adjusted by the survey team).

3.5. Results

3.5.1. Descriptive analysis

First, I aggregate the types of main occupation into the following groups: (1) agriculture, (2) wage job and self-employment, (3) studies, and (4) household maintenance, unemployment, and disability. In Table 3.3, I present the structure of employment across these four groups, migration status, and survey wave.⁶³ At baseline (years 2008/2009), migrants to rural areas are similar to non-migrants in the structure of their main occupation: 66-69% of youth who will choose to stay in their home village or move to another village by the last survey wave (years 2012/2013) were employed in agriculture, 21% were students, 5-7% were mainly involved in household maintenance, were unemployed or disabled, and 5-6% had a non-agricultural wage job or were self-employed. Migrants to urban areas, on the other hand, are different: only 37% of them had main occupation in agriculture at baseline, 39% were students, 16% worked in household maintenance, were unemployed or disabled, and 8% had a wage job or were self-employed.

⁶³ Table 3.15 in the Appendix is a re-calculation of Table 3.3 with the NBS definition of “rural”. The differences that occur due to changes in the definition of “rural” are discussed towards the end of this sub-section, after the description of Table 3.4 and Table 3.5.

Table 3.3. Share of people with main occupation in a certain sector, by migration destination

	Non-migrants (2,423 observations)	Rural-destined migrants (283 observations)	Urban- destined migrants (151 observations)	Difference between rural- destined migrants and non-migrants	Difference between urban- destined migrants and non-migrants
<i>Panel A. Agriculture</i>					
2008/2009	0.690 (0.010)	0.658 (0.029)	0.370 (0.045)	-0.031 (0.028)	-0.319 (0.040)
2012/2013	0.750 (0.010)	0.690 (0.029)	0.081 (0.024)	-0.060 (0.026)	-0.669 (0.036)
Difference between 2012/2013 and 2008/2009	0.060 (0.013)	0.032 (0.038)	-0.289 (0.047)	-0.028 (0.038)	-0.350 (0.054)
<i>Panel B. Wage job or self-employment in a non-agricultural sector</i>					
2008/2009	0.050 (0.005)	0.063 (0.016)	0.080 (0.024)	0.014 (0.013)	0.030 (0.019)
2012/2013	0.104 (0.007)	0.175 (0.023)	0.555 (0.046)	0.072 (0.019)	0.451 (0.027)
Difference between 2012/2013 and 2008/2009	0.054 (0.008)	0.112 (0.026)	0.475 (0.048)	0.058 (0.023)	0.421 (0.033)
<i>Panel C. Student</i>					
2008/2009	0.207 (0.009)	0.213 (0.025)	0.393 (0.045)	0.006 (0.025)	0.186 (0.035)
2012/2013	0.075 (0.006)	0.032 (0.011)	0.059 (0.022)	-0.043 (0.015)	-0.016 (0.022)
Difference between 2012/2013 and 2008/2009	-0.132 (0.010)	-0.181 (0.026)	-0.334 (0.046)	-0.049 (0.029)	-0.202 (0.042)
<i>Panel D. Household maintenance, unemployment, disability</i>					
2008/2009	0.054 (0.005)	0.065 (0.015)	0.157 (0.035)	0.012 (0.014)	0.104 (0.020)
2012/2013	0.071 (0.005)	0.102 (0.019)	0.305 (0.043)	0.031 (0.016)	0.234 (0.023)
Difference between 2012/2013 and 2008/2009	0.018 (0.007)	0.037 (0.023)	0.148 (0.049)	0.019 (0.021)	0.131 (0.031)

Note: Constructed definition of “rural” is used. Standard errors are in parentheses. Sample weights from the 2008/2009 survey wave are applied.

Over time, the structure of main occupation of migrants to rural areas diverges from that of non-migrants. The biggest difference between them is in the changes to the shares of people in agriculture and in wage job or self-employment. The share of people with main occupation in agriculture increased among non-migrants by 6% and increased among migrants to rural areas by 3% (and this change is not significantly different from zero). The share of people with main occupation in a non-agricultural wage job or self-employment increase among non-migrants by 5%, while among migrants to rural areas the increase was 11%. The share of students among non-migrants decreased by 13% by the last survey wave, and the share of people in other categories increased by 2%. Among migrants to rural areas, the decline in the share of students was 18% and the increase in the share of people in household maintenance and unemployment, or with disability was 4%.

Employment outcomes observed during the third survey wave for migrants to urban areas differ a lot from those of both non-migrants and migrants to rural areas. Among migrants to urban areas, there is a 48% increase in the share of people with a wage job or self-employed, a 33% drop in the share of students, a 29% drop in the share of people with main occupation in agriculture, and a 15% increase in the share of people employed in household maintenance, unemployed, or disabled. Overall, regardless of migration status, there is an increase in both the share of people with main occupation in a non-agricultural wage job or self-employment and the share of people with main occupation in household maintenance, unemployment, or disabled people, while the share of students decreases.

For each panel of Table 3.3, its bottom right corner shows how the change of main occupation over time differs between non-migrants and migrants to rural and urban areas. For example, migrants to urban areas have 35% lower change in the share of people with main

occupation in agriculture than non-migrants do. This outcome stems from the fact that the share of people with main occupation in agriculture among non-migrants increased between the survey waves by 6%, while the share among migrants to urban areas decreased by 29%. One can come to the same conclusion knowing that the share for non-migrants was 32% higher than for migrants to urban areas in the first survey wave and 67% higher in the last survey wave. When comparing differences in changes, I see that migrants to rural areas, compared to non-migrants, have a significantly higher increase in the share of people with main occupation in a non-agricultural wage job or self-employment. At the same time, the changes to the occupational structure of migrants to urban areas are significantly different from that of non-migrants for all four groups of occupational categories.

Table 3.4. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status; each row sums to 100%

		Main occupation in 2012/2013				Number of observations: 2008/2009
		Agriculture	Wage job and self-employment	Student	Household maintenance, unemployment, disability	
<i>Panel A. Non-migrants</i>						
Main occupation in 2008/2009	Agriculture	89%	7%	1%	3%	1539
	Wage job and self-employment	35%	58%	1%	7%	125
	Student	45%	10%	31%	14%	519
	Household maintenance, unemployment, disability	43%	17%	6%	34%	181
	Number of observations: 2012/2013	1665	262	191	246	2364
<i>Panel B. Rural-destined migrants</i>						
Main occupation in 2008/2009	Agriculture	76%	12%	1%	11%	186
	Wage job and self-employment	69%	31%	0%	1%	20
	Student	52%	26%	12%	11%	69
	Household maintenance, unemployment, disability	51%	30%	6%	13%	22
	Number of observations: 2012/2013	200	54	10	33	297
<i>Panel C. Urban-destined migrants</i>						
Main occupation in 2008/2009	Agriculture	14%	54%	2%	30%	49
	Wage job and self-employment	9%	86%	0%	5%	14
	Student	2%	53%	13%	33%	59
	Household maintenance, unemployment, disability	9%	51%	0%	40%	20
	Number of observations: 2012/2013	12	76	7	47	142

Note: Constructed definition of “rural” is used. Sample weights from the 2008/2009 survey wave are applied.

Aside from the aggregates, it is also important to look at the transition of people between these occupational groups. In Table 3.4, I present the shares of youth who had a certain main occupation in the first survey wave and shifted (or not) to a different occupation type by the last survey wave.⁶⁴ For example, 89% of non-migrants with main occupation in agriculture at baseline – the most numerous group – stayed in agriculture, while 7% shifted to wage job or self-employment. Among migrants to rural (urban) areas with main occupation in agriculture at baseline, 76% (14%) stayed in agriculture and 12% (54%) shifted to non-agricultural wage job or self-employment. In Table 3.5, I expand the number of groups back to six and choose the groups with the highest number of observations in 2008/2009, which allows me to look at the differences within the aggregated categories.⁶⁵ In this table, wage job is distinguished from self-employment, and household maintenance is distinguished from unemployment and disability. Table 3.5 shows that non-migrants and migrants to rural areas who had main occupation in agriculture at baseline are, on average, more likely to shift to self-employment rather than wage job, while migrants to urban areas are more likely to shift from agriculture to wage job rather than self-employment.

⁶⁴ Table 3.16 in the Appendix is a re-calculation of Table 3.4 with the NBS definition of “rural”. The differences that occur due to changes in the definition of “rural” are discussed towards the end of this sub-section.

⁶⁵ Table 3.17 in the Appendix is a re-calculation of Table 3.5 with the NBS definition of “rural”. The differences that occur due to changes in the definition of “rural” are discussed towards the end of this sub-section.

Table 3.5. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status, for six groups of observations with at least 10 observations in 2008/2009; each row sums to 100%

		Main occupation in 2012/2013						Number of observations: 2008/2009
		Agriculture	Wage job	Self-employment	Student	Household maintenance	Unemployed or disabled	
<i>Panel A. Non-migrants</i>								
Main occupation in 2008/2009	Agriculture	89%	3%	4%	1%	2%	1%	1539
	Wage job	22%	48%	25%	0%	1%	5%	43
	Self-employment	41%	15%	35%	2%	5%	2%	82
	Student	45%	5%	4%	31%	9%	5%	519
	Household maintenance	43%	10%	7%	4%	30%	7%	103
	Unemployed or disabled	43%	5%	13%	14%	11%	14%	78
<i>Panel B. Rural-destined migrants</i>								
Main occupation in 2008/2009	Agriculture	76%	5%	7%	1%	8%	3%	186
	Self-employment	63%	19%	18%	0%	0%	0%	14
	Student	52%	25%	1%	12%	10%	0%	69
	Household maintenance	48%	29%	0%	8%	15%	0%	16
<i>Panel C. Urban-destined migrants</i>								
Main occupation in 2008/2009	Agriculture	14%	35%	19%	2%	21%	9%	49
	Wage job	12%	66%	18%	0%	4%	0%	10
	Student	2%	40%	13%	13%	26%	6%	59
	Household maintenance	10%	36%	12%	0%	27%	15%	18

Note: Constructed definition of “rural” is used. Sample weights from the 2008/2009 survey wave are applied.

From Table 3.5, the transition patterns among non-migrants are different for people who were mainly employed at a wage job and those who were mainly self-employed at baseline. First, among non-migrants, more people with main occupation in non-agricultural wage job kept the same occupation type over time (48%) than people with main occupation in non-agricultural self-employment (35%). Many people shifted from wage job and self-employment to agriculture: 22% of people with a wage job and 41% of self-employed. Transitions between these two groups themselves are somewhat limited: only 25% of people with wage job shifted to self-employment and only 15% of self-employed shifted to wage job. Self-employed people are slightly more likely to shift to household maintenance (5% of them did), while people with a wage job are slightly more likely to shift to disability (5 % of them did).

Table 3.4 and Table 3.5 also show differences in transition patterns between non-migrants and migrants. More students and people with main occupation in household maintenance, unemployed, or disabled shifted to agriculture among migrants to rural areas (51-52%) than among non-migrants (43-45%). Also, people from these occupational groups who moved to rural areas are more likely to shift to a non-agricultural wage job or self-employment: 26% of students do (10% among non-migrants), 30% of migrants with main occupation in household maintenance, unemployed, or disabled at baseline do (17% among non-migrants). Table 3.5 shows that people with main occupation in household maintenance at baseline are, on average, more likely to shift to a wage job when they move to another rural area (29% of them do) than when they stay (10% of them do), while shift to self-employment is less common among rural-destined migrants (0%) than among non- migrants (7%).

For migrants to urban areas, there is a drastic shift away from agriculture: among people whose main occupation was farming or fishing at baseline, 35% shifted to a non-agricultural

wage job at destination, 21% shifted to household maintenance, 19% shifted to a non-agricultural self-employment, and 9% shifted to unemployment or disability, leaving only 14% in agriculture. Surprisingly high share, 26%, of students shift to household maintenance, compared to 9-10% of students among non-migrants and migrants to rural areas. Migrants to urban areas have the highest rates of shifting to unemployment or disability among people with main occupation in agriculture (9% - compared to 1% among non-migrants and 3% among migrants to rural areas) and students (6% - compared to 5% among non-migrants and 0% among migrants to rural areas), although the rates of shifting from other occupational categories into disability are the highest among non-migrants.

The use of the NBS definition of “rural” instead of the constructed one introduces several differences to the patterns of employment (see Table 3.15, Table 3.16, and Table 3.17 in the Appendix). It happens both due to changes in the categorization of destination areas and due to changes to the sample which is restricted to youth who lived in rural areas at baseline. With the constructed definition of “rural”, the average share of people with main occupation in agriculture at baseline among migrants is lower than with the NBS definition. On the other hand, there is a lower chance for migrants to rural areas to maintain their occupation in non-agricultural wage job or self-employment and a higher chance to shift into agriculture when the constructed definition is used. The shift into wage job and self-employment is more pronounced among migrants to urban areas when the constructed definition is used. At the same time, the share of people with main occupation in household maintenance, unemployment, and disability at baseline is higher among urban-destined migrants under the constructed definition, which makes the shift into this group less pronounced. Overall, there are lower chances for migrants with main

occupation in household maintenance, unemployment, or disability to shift into agriculture, wage job, or self-employment under the constructed definition.

As was evident from Table 3.3, baseline (2008/2009) occupational structure of non-migrants differs from that of migrants, especially urban-destined migrants. This observation emphasizes selection into migration, which I additionally test for. I run multinomial logistic regressions for the impact of future migration on the probability to have main occupation in agriculture in the first wave of survey and the probability to have main occupation in a non-agricultural wage job or self-employment in the first wave of survey. The base outcome is being a student, have main occupation in household maintenance, being unemployed or disabled in the first wave of survey. I run regressions with each of the following indicators of migration status: being a migrant, moving to a rural area, moving to an urban area, moving to a low-density rural area, moving to a high-density rural area, moving to a peri-urban area, moving to a town, and moving to a city. For each of these regressions, I try three specifications: without controls, with a smaller set of controls, and with a larger set of controls.⁶⁶

⁶⁶ The smaller and the larger set of controls are the same as the ones used for the main specification (they are listed in section 3.4).

Table 3.6. Selection into migration: marginal values from multinomial logistic regression of indicators to have main occupation in agriculture and non-agricultural wage job or self-employment in 2008/2009 on migration status in 2012/2013

Outcome variable	Indicator for migration	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a peri-urban area	Indicator for migration to a town	Indicator for migration to a city
<i>A. Multinomial logistic regression without controls</i>								
Main occupation in farming or fishing	-0.111*** (0.024)	-0.024 (0.029)	-0.287*** (0.040)	-0.007 (0.035)	-0.052 (0.047)	-0.233*** (0.074)	-0.399*** (0.071)	-0.216*** (0.067)
Main occupation in wage job or self-employment	0.023** (0.011)	0.013 (0.013)	0.043*** (0.016)	-0.001 (0.017)	0.034* (0.018)	0.025 (0.031)	0.075*** (0.021)	0.015 (0.030)
Number of observations	2,803	2,661	2,506	2,558	2,467	2,412	2,404	2,418
<i>B. Multinomial logistic regression with controls for age, gender, marital status, primary school completion, being born in the village of residence, household size, land area the household cultivates, and asset index</i>								
Main occupation in farming or fishing	-0.032 (0.020)	0.013 (0.024)	-0.127*** (0.033)	0.029 (0.029)	-0.011 (0.038)	-0.074 (0.058)	-0.167*** (0.056)	-0.142** (0.059)
Main occupation in wage job or self-employment	0.021** (0.011)	0.020 (0.013)	0.030* (0.016)	0.009 (0.017)	0.031* (0.017)	0.028 (0.030)	0.059** (0.024)	0.002 (0.029)
Number of observations	2,801	2,659	2,504	2,556	2,465	2,410	2,402	2,416
<i>C. Multinomial logistic regression with controls from Panel B and controls for being away from the household in the past year, being a head of the household, being a child of the household head, household head's age and gender, units of livestock owned by the household, agricultural and non-agricultural shocks experienced by the household, population density, distance to the nearest road, and distance to the nearest town with population of at least 50,000 people</i>								
Main occupation in farming or fishing	-0.039* (0.020)	0.005 (0.024)	-0.131*** (0.033)	0.006 (0.029)	0.002 (0.038)	-0.077 (0.057)	-0.180*** (0.055)	-0.136** (0.058)
Main occupation in wage job or self-employment	0.018* (0.011)	0.017 (0.013)	0.028* (0.016)	0.012 (0.017)	0.023 (0.018)	0.015 (0.030)	0.058** (0.024)	0.009 (0.027)
Number of observations	2,801	2,659	2,504	2,556	2,465	2,410	2,402	2,416

Note: Base outcome is to list one of these four categories as main occupation: studies, household maintenance, unemployment, or disability. Constructed definition of "rural" is used. Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Results are presented in Table 3.6 for the constructed definition of “rural” and in Table 3.18 in the Appendix for the NBS definition. Main occupation in agriculture at baseline is less likely to be taken by people who will move to urban areas, especially towns (under the constructed definition) and cities (under the NBS definition). Main occupation in a non-agricultural wage job or self-employment at baseline is more likely to be taken by people who will move to high-density rural areas or towns (constructed definition) and cities (NBS definition). For all indicators, the inclusion of controls for observable characteristics leads to a weaker relationship between the probability to have main occupation in a certain sector at baseline and future migration. Still, even with the largest set of controls, the indicator of migration to the most urbanized areas is negative and significant for the probability to have main occupation in agriculture at baseline. Hence, there are significant differences between non-migrants and migrants to towns and cities that cannot be explained by the observable characteristics I employ. The results for the difference between non-migrant and migrants to rural and peri-urban areas are more promising and show that the inclusion of controls helps to account to selection into migration to these destinations.

3.5.2. Regression analysis

Probability to stay engaged in work

First, I look at the impact of migration on the probability to stay engaged in work between the survey waves. This is an important concern, as becoming unemployed at destination can be associated with the need to receive remittances from the household of origin and the worsening of career options in the future (which, in turn, leads to lower lifetime earnings). For people who had main occupation in agriculture, wage job, self-employment, or household maintenance at baseline, I check if they stayed in one of these sectors by the last survey wave or

shifted into studies, unemployment, or disability. With the results presented in Table 3.7, I conclude that migration to peri-urban areas and cities might have some negative effects on the probability to stay employed, which is concerning. This effect does not disappear in some models once I account for selection into migration. Migration to rural areas and towns has no significant effect on the probability to stay engaged in work.

If I exclude household maintenance from the definition of “work”, the results strongly indicate that migration is associated with a shift away from agricultural employment, wage work, and self-employment into household maintenance, studies, unemployment, and disability (see Table 3.8). This effect is smaller and weaker for migration to rural areas (simple difference in means between non-migrants and migrants to rural areas is 6%) and larger and stronger for migration to urban areas (simple difference in means between non-migrants and migrants to urban areas is 23%). Once I control for observable characteristics, the negative effect of migration mostly disappears for migration to low-density rural areas, while some effect is preserved for migration to high-density rural areas and towns. Migration to peri-urban areas and towns has a strong negative effect on the probability to stay engaged in work: my estimate for the share of migrants who shift away from work ranges from 9% to 26% for peri-urban destinations and from 9% to 29% for cities.

Table 3.7. Migration and the probability to stay engaged in work

	Migration to a rural area	Migration to an urban area	Migration to a low- density rural area	Migration to a high- density rural area	Migration to a peri- urban area	Migration to a town	Migration to a city
Difference in means between non-migrants and migrants	-0.012 (0.011)	-0.087*** (0.018)	-0.008 (0.012)	-0.020 (0.017)	-0.092*** (0.026)	-0.019 (0.029)	-0.158*** (0.032)
<i>Logistic regression</i>							
Without controls	-0.008 (0.011)	-0.035*** (0.013)	0.001 (0.014)	-0.019 (0.015)	-0.039** (0.018)	-0.009 (0.028)	-0.043** (0.018)
With a small set of controls	-0.004 (0.011)	-0.014 (0.013)	0.003 (0.014)	-0.016 (0.015)	-0.024 (0.018)	0.024 (0.030)	-0.023 (0.018)
With a large set of controls	-0.007 (0.011)	-0.021 (0.013)	-0.000 (0.014)	-0.017 (0.015)	-0.031 (0.019)	0.029 (0.032)	-0.035** (0.018)
<i>Propensity score matching</i>							
With a small set of variables	-0.023 (0.016)	0.000 (0.042)	0.007 (0.022)	-0.041 (0.029)	-0.103** (0.041)	-0.038** (0.018)	-0.038 (0.074)
With a large set of variables	-0.018 (0.016)	-0.074** (0.034)	0.000 (0.019)	0.027 (0.041)	-0.069** (0.030)	0.038 (0.066)	-0.115* (0.066)
<i>Nearest neighbor matching</i>							
With a small set of variables	-0.008 (0.018)	-0.024 (0.043)	0.008 (0.022)	-0.054** (0.026)	-0.103* (0.057)	0.089 (0.078)	-0.050 (0.087)
With a large set of variables	-0.017 (0.015)	-0.060 (0.041)	-0.012 (0.014)	-0.036 (0.029)	-0.103* (0.057)	-0.031 (0.073)	-0.188 (0.162)

Note: “Engaged in work” is defined as having main occupation in agriculture, non-agricultural wage job, non-agricultural self-employment, and household maintenance. Students, unemployed people, and disabled people are considered to not be engaged in work. For people engaged in work in the first survey wave, I estimate the probability to stay engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. Constructed definition of “rural” is used.

Table 3.8. Migration and the probability to stay engaged in work excluding household maintenance

	Migration to a rural area	Migration to an urban area	Migration to a low- density rural area	Migration to a high- density rural area	Migration to a peri- urban area	Migration to a town	Migration to a city
Difference in means between non-migrants and migrants	-0.061*** (0.016)	-0.230*** (0.028)	-0.057*** (0.019)	-0.069*** (0.024)	-0.216*** (0.040)	-0.139*** (0.047)	-0.358*** (0.051)
<i>Logistic regression</i>							
Without controls	-0.034** (0.015)	-0.106*** (0.018)	-0.027 (0.018)	-0.042* (0.022)	-0.098*** (0.027)	-0.081*** (0.031)	-0.117*** (0.027)
With a small set of controls	-0.024 (0.015)	-0.085*** (0.018)	-0.019 (0.018)	-0.037* (0.021)	-0.089*** (0.027)	-0.052 (0.033)	-0.091*** (0.026)
With a large set of controls	-0.029* (0.015)	-0.093*** (0.019)	-0.027 (0.018)	-0.034 (0.021)	-0.095*** (0.028)	-0.060* (0.034)	-0.100*** (0.027)
<i>Propensity score matching</i>							
With a small set of variables	-0.049* (0.028)	-0.175*** (0.060)	-0.037 (0.026)	-0.042 (0.045)	-0.261*** (0.092)	-0.211** (0.089)	-0.095 (0.119)
With a large set of variables	-0.063*** (0.022)	-0.190*** (0.056)	-0.074*** (0.026)	-0.042 (0.048)	-0.217** (0.106)	-0.105 (0.102)	-0.286*** (0.109)
<i>Nearest neighbor matching</i>							
With a small set of variables	-0.030 (0.028)	-0.138** (0.063)	-0.025 (0.035)	-0.055 (0.039)	-0.172** (0.081)	-0.053 (0.111)	-0.244* (0.134)
With a large set of variables	-0.036 (0.026)	-0.175*** (0.068)	-0.023 (0.031)	-0.055 (0.047)	-0.140 (0.137)	-0.125 (0.203)	0.365 (0.441)

Note: “Engaged in work excluding household maintenance” is defined as having main occupation in agriculture, non-agricultural wage job, and non-agricultural self-employment. People with main occupation in household maintenance, students, unemployed people, and disabled people are considered to not be engaged in work. For people engaged in work in the first survey wave, I estimate the probability to stay engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. Constructed definition of “rural” is used.

Interestingly, the share of people with main occupation in household maintenance during the last survey wave among migrants to towns is comparable to that in other urban destinations, but many of these people had main occupation in household maintenance at baseline. It makes the share of people shifting from agriculture, wage job, or self-employment into household maintenance to be lower among migrants to towns compared to migrants to peri-urban areas and cities. A shift into household maintenance at destination can indicate underemployment and be temporary for someone who is looking for a job. As seen in the descriptive results, migrants to peri-urban areas are more likely to shift into household maintenance while migrants to cities are more likely to become unemployed.

A robustness check with the NBS definition of “rural” (Table 3.19 and Table 3.20 in the Appendix) confirms that urban destinations, and especially cities, are more likely to be associated with falling out of labor force. Migration to cities is associated with a 10-36% lower chance to stay engaged from work excluding household maintenance (Table 3.20). A strong negative effect of migration to peri-urban areas evident from the models with constructed definition of “rural” is now present for migration to towns and high-density rural areas: the estimated effect is 7-13%. When selection into migration is accounted for, migration to low-density rural areas has the smallest, if any, negative effect on the probability to stay engaged in work, for both the definition that includes and excludes household maintenance: the estimates range from 2% to 5%.

Probability to become engaged in work

Next, I look at the impact of migration on the probability to become engaged in work by the last survey wave following disengagement from work at baseline. Migration can provide a new set of employment options that were not available at the origin, attracting underemployed

and unemployed people and improving their livelihood. Again, I start with a definition of “work” that includes main occupation in agriculture, wage job, self-employment, and household maintenance. So, I estimate the impact of migration on the probability for students, unemployed and disabled people (at baseline) to shift into one of the categories labeled as “work” by the last survey wave. The results presented in Table 3.9 show migration in general to have a positive and significant effect on employment among people who were not engaged in work at baseline. Simple difference in means shows that the share of people who became engaged in work is 25% higher among migrant to rural areas and 17% higher among migrants to urban areas than among non-migrants. In larger models that account for selection into migration, the estimates are pretty consistent between low- and high-density rural and peri-urban destinations: the probability to become engaged in work increases with migration to these regions by 35% on average. The results are weaker and less consistent for migration to towns and cities.

When I exclude household maintenance from the definition of “work”, I estimate the probability for people with main occupation in household maintenance, studies, unemployment, or disability at baseline to shift into agriculture, wage job, or self-employment by the last survey wave (see the results in Table 3.10). Now, migration to low-density rural areas has a consistently positive and significant effect on the probability to become engaged in work. Migration to high-density rural areas shows up as positive and significant in some models. Migration to other destinations has no significant effect.

Table 3.9. Migration and the probability to become engaged in work

	Migration to a rural area	Migration to an urban area	Migration to a low- density rural area	Migration to a high- density rural area	Migration to a peri- urban area	Migration to a town	Migration to a city
Difference in means between non-migrants and migrants	0.249*** (0.055)	0.172*** (0.061)	0.227*** (0.070)	0.286*** (0.088)	0.350*** (0.112)	0.252** (0.118)	0.026 (0.088)
<i>Logistic regression</i>							
Without controls	0.340*** (0.076)	0.265*** (0.078)	0.353*** (0.101)	0.343*** (0.126)	0.592** (0.240)	0.338* (0.182)	0.129 (0.102)
With a small set of controls	0.327*** (0.075)	0.271*** (0.076)	0.336*** (0.098)	0.337*** (0.123)	0.556** (0.228)	0.338* (0.176)	0.160 (0.101)
With a large set of controls	0.335*** (0.075)	0.248*** (0.075)	0.326*** (0.098)	0.379*** (0.127)	0.546** (0.223)	0.294* (0.171)	0.138 (0.099)
<i>Propensity score matching</i>							
With a small set of variables	0.213*** (0.068)	0.410*** (0.084)	0.315*** (0.091)	0.345*** (0.094)	0.368*** (0.126)	0.286* (0.172)	0.161 (0.125)
With a large set of variables	0.253*** (0.072)	0.131* (0.076)	0.348*** (0.085)	0.448*** (0.094)	0.368*** (0.106)	0.143** (0.063)	0.357*** (0.118)
<i>Nearest neighbor matching</i>							
With a small set of variables	0.258*** (0.075)	0.296*** (0.085)	0.230*** (0.088)	0.258* (0.132)	0.460*** (0.136)	0.117 (0.181)	0.258** (0.127)
With a large set of variables	0.228*** (0.077)	0.249*** (0.075)	0.204* (0.115)	0.218 (0.147)	-0.158 (0.440)	-0.289 (0.326)	0.241* (0.124)

Note: “Engaged in work” is defined as having main occupation in agriculture, non-agricultural wage job, non-agricultural self-employment, and household maintenance. Students, unemployed people, and disabled people are considered to not be engaged in work. For people not engaged in work in the first survey wave, I estimate the probability to become engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. For the propensity score matching, marital status is excluded due to low number of observations. Constructed definition of “rural” is used.

Table 3.10. Migration and the probability to become engaged in work excluding household maintenance

	Migration to a rural area	Migration to an urban area	Migration to a low- density rural area	Migration to a high- density rural area	Migration to a peri- urban area	Migration to a town	Migration to a city
Difference in means between non-migrants and migrants	0.227*** (0.053)	0.002 (0.056)	0.244*** (0.064)	0.194** (0.086)	-0.027 (0.099)	0.085 (0.100)	-0.034 (0.082)
<i>Logistic regression</i>							
Without controls	0.255*** (0.059)	0.023 (0.059)	0.349*** (0.080)	0.119 (0.093)	-0.029 (0.102)	0.063 (0.112)	0.037 (0.089)
With a small set of controls	0.250*** (0.058)	0.024 (0.059)	0.338*** (0.078)	0.121 (0.090)	-0.043 (0.100)	0.099 (0.109)	0.036 (0.087)
With a large set of controls	0.253*** (0.059)	0.005 (0.058)	0.328*** (0.078)	0.154* (0.091)	-0.040 (0.099)	0.057 (0.107)	0.017 (0.086)
<i>Propensity score matching</i>							
With a small set of variables	0.132* (0.080)	0.127 (0.087)	0.220** (0.088)	0.266** (0.119)	-0.020 (0.140)	0.095 (0.150)	0.182 (0.135)
With a large set of variables	0.264*** (0.082)	0.101 (0.082)	0.237*** (0.087)	0.125 (0.085)	-0.080** (0.033)	0.238* (0.133)	0.061 (0.111)
<i>Nearest neighbor matching</i>							
With a small set of variables	0.185** (0.074)	0.158* (0.083)	0.215** (0.087)	0.107 (0.126)	0.158 (0.129)	0.140 (0.144)	0.153 (0.138)
With a large set of variables	0.179** (0.076)	0.078 (0.090)	0.168* (0.098)	0.238* (0.129)	-0.022 (0.138)	-0.239 (0.238)	0.205* (0.121)

Note: “Engaged in work excluding household maintenance” is defined as having main occupation in agriculture, non-agricultural wage job, and non-agricultural self-employment. People with main occupation in household maintenance, students, unemployed people, and disabled people are considered to not be engaged in work. For people not engaged in work in the first survey wave, I estimate the probability to become engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. For the propensity score matching, marital status and the indicator of being the head of the household are excluded due to low number of observations. Constructed definition of “rural” is used.

The results with the NBS definition of “rural” are presented in Table 3.21 and Table 3.22 in the Appendix. For the definition of “work” that includes household maintenance, the positive effect of migration I find is stronger with the NBS definition than with the constructed definition of “rural”, and it is more consistent across destinations when they are defined using the NBS definition. Under this definition, migration to cities is associated with 27-33% higher chance to become engaged in work. Though, after the exclusion of household maintenance from the definition of “work”, the impact of migration to a city become insignificant. With this definition of “work”, the results between the NBS and the constructed definition of “rural” align and point to the positive and significant impact of migration to rural areas, especially low-density rural areas, on the probability to become engaged in work.

Probability to be employed in a certain sector during the last survey wave

Finally, I study whether migration to various destination types is associated with the sectoral transitions, looking at transition into the agricultural sector and into the non-farm wage job or self-employment. I start with estimating the probability to have main occupation in agriculture during the last survey wave. I run estimations with and without an indicator for having main occupation in agriculture at baseline. For this estimation only, the main specification features the NBS definition of “rural”, while the robustness check is done with the constructed definition. The results are presented in Table 3.11 (with NBS definition) and Table 3.23 (see Appendix; with constructed definition).

Table 3.11. Migration and the probability to have main occupation in agriculture in the last survey wave; NBS definition of “rural”

	I. for migration to a rural area	I. for migration to an urban area	I. for migration to a low- density rural area	I. for migration to a high- density rural area	I. for migration to a town	I. for migration to a city
Difference in means between non-migrants and migrants in the last survey wave	-0.074*** (0.026)	-0.642*** (0.035)	-0.034 (0.032)	-0.149*** (0.043)	-0.591*** (0.044)	-0.725*** (0.055)
Difference in means between non-migrants' and migrants' differences between the last and the first survey waves	-0.033 (0.039)	-0.376*** (0.052)	-0.028 (0.048)	-0.037 (0.064)	-0.401*** (0.067)	-0.311*** (0.084)
<i>Logistic regression</i>						
Without controls	-0.034 (0.028)	-0.557*** (0.045)	0.033 (0.036)	-0.141*** (0.043)	-0.486*** (0.055)	-0.730*** (0.103)
With I(ag.)	-0.028 (0.025)	-0.441*** (0.037)	0.016 (0.031)	-0.098*** (0.038)	-0.411*** (0.045)	-0.532*** (0.081)
With a small set of controls	-0.025 (0.027)	-0.467*** (0.042)	0.032 (0.034)	-0.118*** (0.041)	-0.407*** (0.051)	-0.616*** (0.093)
With a small set of controls and I(ag.)	-0.032 (0.025)	-0.425*** (0.037)	0.010 (0.031)	-0.094** (0.038)	-0.390*** (0.045)	-0.530*** (0.081)
With a large set of controls	-0.054** (0.025)	-0.430*** (0.038)	-0.023 (0.031)	-0.103*** (0.038)	-0.384*** (0.046)	-0.554*** (0.083)
With a large set of controls and I(ag.)	-0.051** (0.024)	-0.402*** (0.035)	-0.025 (0.029)	-0.088** (0.036)	-0.374*** (0.041)	-0.495*** (0.075)
<i>Propensity score matching</i>						
With a small set of variables	-0.057 (0.041)	-0.530*** (0.045)	0.038 (0.048)	-0.150** (0.059)	-0.483*** (0.066)	-0.562*** (0.069)
With a small set of variables and I(ag.)	0.019 (0.041)	-0.434*** (0.048)	-0.087** (0.044)	-0.080 (0.071)	-0.494*** (0.074)	-0.375*** (0.062)
With a large set of variables	-0.025 (0.042)	-0.530*** (0.052)	0.027 (0.044)	-0.170** (0.069)	-0.345*** (0.080)	-0.484*** (0.074)
With a large set of variables and I(ag.)	-0.064 (0.040)	-0.490*** (0.053)	-0.022 (0.047)	-0.060 (0.069)	-0.471*** (0.074)	-0.375*** (0.075)
<i>Nearest neighbor matching</i>						
With a small set of variables	-0.061 (0.038)	-0.400*** (0.050)	-0.005 (0.044)	-0.166** (0.070)	-0.344*** (0.068)	-0.492*** (0.069)
With a small set of variables and I(ag.)	-0.029 (0.037)	-0.409*** (0.050)	0.021 (0.045)	-0.131** (0.058)	-0.405*** (0.069)	-0.419*** (0.072)
With a large set of variables	-0.056 (0.038)	-0.463*** (0.053)	-0.044 (0.046)	-0.102 (0.064)	-0.456*** (0.070)	-0.482*** (0.075)
With a large set of variables and I(ag.)	-0.059* (0.035)	-0.416*** (0.049)	-0.085** (0.042)	-0.027 (0.060)	-0.442*** (0.069)	-0.425*** (0.069)

Note: “I.” stands for “indicator”. I(ag.) is an indicator for having main occupation in agriculture at baseline. For the computation of the differences in means, sample weights from the 2008/2009 survey wave were applied.

Migration to rural areas has a small negative effect on the probability to have main occupation in agriculture at destination. After controlling for selection into migration and for main occupation at baseline, I find that the probability to have main occupation in agriculture is 5% smaller among migrants to rural areas and 40-49% smaller among migrants to urban areas. Among rural destinations, low-density rural areas rarely have any significant effect on agricultural occupation. Migration to high-density rural areas is associated with 9-13% lower probability to have main occupation in agriculture at destination.

A simple difference in means between migrants and non-migrant shows that the share of people with main occupation in agriculture in 2012/2013 is 13% smaller among migrants to cities than among migrants to towns. But after taking the difference with the 2008/2009 shares, the relationship reverses: the share of people with main occupation in agriculture is 9% higher among migrants to cities than among migrants to towns. Regressions results diverge too. The results of logistic regressions with the full set of controls show that migration to towns is associated with 37% lower chance to have main occupation in agriculture, while migration to cities is associated with 50% lower chance. The results of propensity score matching are the opposite: migration to towns is associated with 47% lower chance, while migration to cities is associated with 38% lower chance. The results of nearest neighbor matching are closer to each other: migration to towns is associated with 44% lower chance, while migration to cities is associated with 43% lower chance.

NBS definition of “rural” does not distinguish peri-urban areas, but a larger model with five destinations according to the constructed definition is not converging even with the small set of controls. Hence, for this definition, I had to cut the model to include less controls. The models comparable to the ones presented in Table 3.11 that I could run (not presented here) – namely,

for the rural destinations, show no significant difference between migrants and non-migrants. All models in Table 3.23 use a smaller set of controls and are comparable within this table only.

Migration to peri-urban destinations is associated with a 32-40% lower probability to have main occupation in agriculture during the last survey wave. This result is much closer to the impact of migration to towns and cities than to the impact of migration to high-density rural areas.

Between towns and cities, the results with the constructed definition of “rural” are inconclusive, same as with the NBS definition. Here, the results of logistic regressions are similar between migration to towns and migration to cities (56% lower probability). The results of propensity score matching are more drastic for migration to cities (42% lower probability for migration to towns and 53% lower probability for migration to cities), while the results of nearest neighbor matching are more drastic for migration to towns (52% lower probability for migration to towns and 36% lower probability for migration to cities).

In Table 3.12, I estimate the impact of migration on the probability to have main occupation in non-agricultural wage job or self-employment during the last survey wave. The main definition of “rural” is the one I constructed. NBS definition is used as a robustness check; results are presented in Table 3.24 in the Appendix. The results of both logistic regressions and propensity score matching point to migration having a positive impact on the probability to have a non-agricultural main occupation during the last survey wave, regardless of the destination migrants chose. The same conclusion is made from the regressions that use the NBS definition. With nearest neighbor matching, the largest model (with the full set of controls and the indicator of having main occupation in non-agricultural wage job or self-employment at baseline) picks up no impact of migration to low-density rural and peri-urban areas with the constructed definition

and high-density rural areas with the NBS definition of “rural”. I will focus on the results of logistic regressions and propensity score matching which are more consistent.

I estimate migration to low-density rural areas to be associated with a 7% higher chance to have main occupation in non-agricultural wage job or self-employment at destination. This result holds across both definitions of “rural” in most models once selection into migration and baseline occupation are controlled for. The impact of migration to high-density rural areas is around 6-13%. With the constructed definition, it is possible to look at migrants to peri-urban areas. I find that for them the probability to have a non-farm occupation at destination is comparable to that of migrants to cities and is 16-35% higher than that of non-migrants. Migration to towns is associated with the highest probability to have main occupation in non-agricultural wage job or self-employment: for migrants, the probability is 20-48% higher. In general, the estimates I got from the logistic regressions are smaller than the results of propensity score matching and nearest neighbor matching for all destinations except for low-density rural areas.

Table 3.12. Migration and the probability to have main occupation in non-agricultural wage job or self-employment in the last survey wave

	I. for migration to a rural area	I. for migration to an urban area	I. for migration to a low- density rural area	I. for migration to a high- density rural area	I. for migration to a peri- urban area	I. for migration to a town	I. for migration to a city
Difference in means between non-migrants and migrants in 2012/13	0.072*** (0.019)	0.451*** (0.027)	0.057** (0.023)	0.099*** (0.030)	0.392*** (0.043)	0.578*** (0.048)	0.406*** (0.044)
Difference in means between non- migrants' and migrants' differences between 2008/09 and 2012/13	0.034 (0.025)	0.414*** (0.034)	0.035 (0.030)	0.027 (0.040)	0.346*** (0.057)	0.550*** (0.062)	0.330*** (0.057)
<i>Logistic regression</i>							
Without controls	0.060*** (0.017)	0.238*** (0.018)	0.043** (0.021)	0.085*** (0.025)	0.186*** (0.030)	0.261*** (0.033)	0.235*** (0.028)
With I(NA)	0.055*** (0.016)	0.223*** (0.017)	0.044** (0.020)	0.070*** (0.024)	0.179*** (0.027)	0.247*** (0.030)	0.210*** (0.026)
With a small set of controls	0.076*** (0.017)	0.224*** (0.018)	0.066*** (0.021)	0.082*** (0.024)	0.169*** (0.030)	0.259*** (0.034)	0.213*** (0.029)
With a small set of controls and I(NA)	0.067*** (0.017)	0.210*** (0.017)	0.060*** (0.020)	0.069*** (0.024)	0.166*** (0.027)	0.241*** (0.031)	0.189*** (0.027)
With a large set of controls	0.073*** (0.017)	0.216*** (0.019)	0.071*** (0.021)	0.072*** (0.025)	0.159*** (0.030)	0.246*** (0.034)	0.211*** (0.029)
With a large set of controls and I(NA)	0.066*** (0.017)	0.205*** (0.018)	0.065*** (0.020)	0.061** (0.024)	0.158*** (0.027)	0.233*** (0.031)	0.189*** (0.027)
<i>Propensity score matching</i>							
With a small set of variables	0.037 (0.031)	0.394*** (0.043)	0.093*** (0.031)	0.126*** (0.045)	0.333*** (0.059)	0.525*** (0.073)	0.463*** (0.068)
With a small set of variables and I(NA)	0.047 (0.032)	0.380*** (0.047)	0.036 (0.036)	0.097* (0.054)	0.271*** (0.068)	0.525*** (0.055)	0.389*** (0.071)
With a large set of variables	0.091*** (0.028)	0.345*** (0.044)	0.052 (0.034)	0.078 (0.048)	0.281*** (0.061)	0.450*** (0.070)	0.389*** (0.066)
With a large set of variables and I(NA)	0.088*** (0.029)	0.366*** (0.048)	0.077*** (0.029)	0.126*** (0.042)	0.271*** (0.075)	0.475*** (0.082)	0.352*** (0.076)
<i>Nearest neighbor matching</i>							
With a small set of variables	0.055* (0.033)	0.438*** (0.050)	0.037 (0.038)	0.089* (0.054)	0.325*** (0.079)	0.568*** (0.079)	0.434*** (0.087)
With a small set of variables and I(NA)	0.038 (0.032)	0.380*** (0.051)	0.027 (0.038)	0.061 (0.052)	0.291*** (0.077)	0.527*** (0.084)	0.344*** (0.088)
With a large set of variables	0.073** (0.030)	0.396*** (0.052)	0.042 (0.035)	0.143*** (0.052)	0.251*** (0.090)	0.558*** (0.088)	0.386*** (0.086)
With a large set of variables and I(NA)	0.044 (0.029)	0.355*** (0.052)	0.018 (0.034)	0.104** (0.051)	0.154 (0.097)	0.514*** (0.086)	0.394*** (0.085)

Note: "I." stands for "indicator". I(NA) is an indicator for having main occupation in non-agricultural wage job or self-employment at baseline. For the computation of the differences in means, sample weights from the 2008/2009 survey wave were applied. Constructed definition of "rural" is used.

Within non-agricultural occupations, there are some differences in transition patterns by destination. From Table 3.4, migrants to rural areas who had main occupation in agriculture, studies, household maintenance, were unemployed or disabled at baseline are almost twice more likely to have main occupation in wage job or self-employment at destination than non-migrants do. Migrants to urban areas are much more likely to shift to wage job or self-employment than non-migrants do regardless of their occupation at baseline. More detailed sectoral transitions presented in Table 3.5 show that students who moved are more likely to shift to wage job: 25% of students who moved to rural areas and 40% of students who moved to urban areas shifted to a wage job; while only 5% of non-migrant students did. Only in urban destinations students shift to self-employment: 13% of them did. In rural destinations, only 1% of students shifted to self-employment, while 4% of non-migrant students did. People with main occupation in agriculture at baseline are more actively switching to self-employment if they move to a rural area and to wage job if they move to an urban area. People with main occupation in household maintenance at baseline are more likely to shift to self-employment than wage job if they stay in place. If they move, they are more likely to shift to wage job than self-employment, especially if they move to a rural area.

3.5.3. Additional analysis

Some people can use household maintenance as an alternative to paid employment when they cannot find another occupation at destination. Hence, I check if migration impacts the probability to stay employed in a sector other than household maintenance. In the main analysis, I used two definitions of employment: (i) main occupation in agriculture or non-agricultural wage job or self-employment (Table 3.8); (ii) main occupation as in (i) or main occupation in household maintenance (Table 3.7). In this subsection, I add main occupation in studies to (i)

rather than household maintenance. For people with this main occupation at baseline, I see if they list household maintenance, unemployment, or disability as their main occupation during the last survey wave. As seen in Table 3.3, there is no significant difference between non-migrants and migrants to rural areas in the change to the share of people with main occupation in household maintenance, unemployed, or disabled, once the share in the first survey wave is accounted for. Among migrants to urban areas, on the other hand, the change in the share is 20% higher than among non-migrants.

The results with this new definition are presented in Table 3.25. They are comparable to the main results from Table 3.8 with a stricter definition of “work”, although there are a few exceptions. Low-density rural areas now are the only destination with no significant effect to the probability to stay engaged in work. The negative effect of migration to high-density rural areas, peri-urban areas, and towns became larger, while the negative effect of migration to cities became smaller. The difference shows that students who move to cities are more likely to stay engaged in work. It mostly happens because students moving to cities are more likely to have main occupation in wage job and to stay in school. In fact, among 14 migrant students who continued their studies, 5 moved to a city and 6 moved to a low-density rural area.

The occupational transition of students differs not only by destination type, but also by gender. More female students move to low-density rural areas (41% of migrant female students chose this destination type while only 25% of migrant male students did), and more male students move to cities (33% of migrant male students moved to a city while only 12% of migrant female students did). Among migrant male students, the share of people shifting to a non-farm wage job or self-employment is even both within rural destinations: almost a third of them shift when they move to either low- or high-density rural areas; and within urban

destinations: almost two thirds of them shift when they move to either peri-urban areas, town, or city. Among migrant female students, the share of people shifting to a non-farm wage job or self-employment is high only among those who moved to towns (51%) and high-density rural areas (28%); among migrants to other destinations the share is at or below 20%.

Reason for migration can be one of the factors explaining these gender differences, as a third of migrant female students listed marriage as their main reason for migration. Next, I check whether the patterns of occupational shifts among women migrating for marriage (among all women, not just students) differs from that among other female migrants and among male migrants. The information on the main reason for migration is missing for 12% of the sample of migrants, equally so for male and female migrants. The share of women who moved for marriage is 20% among all migrants who reported the reason for migration. Among women moving for marriage, 47% are of age 15-19. The migration rate to low-density rural areas is higher and the migration rate to towns is lower among women who moved for marriage.

At baseline, women who will move for marriage are, on average, slightly less likely to have main occupation in agriculture and slightly more likely to have main occupation in studies than other female migrants. At destination, however, women who moved for marriage are more likely to have main occupation in agriculture, self-employment, and household maintenance than other migrants. They are much less likely to have a wage job or be students. They are more likely to be unemployed than male migrants but less likely than other female migrants. Taking into account occupational category at baseline, women who moved for marriage make the largest shift into household maintenance and self-employment than other groups of migrants. Same as all other groups of migrants, women who moved for marriage, on average, tend to shift away from agriculture and studies.

3.6. Discussion

The center of attention for this study is the contribution of various migration destinations to the employment shifts associated with structural transformation. To accompany the main analysis which takes into account the non-randomness of migration destination decision, I conduct a simple calculation of every destination's contribution to the occupational shifts I observe in the total population. The results are presented in Table 3.13, and Table 3.26 and Table 3.27 in the Appendix and explained below.

First, let's look at the share of people who had main occupation in non-agricultural wage job or self-employment during the last survey wave. In the full sample, this share is 13.5% (row A*B in Table 3.13). One can also compute this number knowing the share of people with this occupation by their migration status and location type observed during the last survey wave. In the full sample, using the constructed definition of "rural", I can categorize every individual into one of the following groups based on their migration status: non-migrant, migrant to a low-density rural area, migrant to a high-density rural area, migrant to a peri-urban area, migrant to a town, or migrant to a city. Then, for the share of people with main occupation in non-farm sector (wage job or self-employment) during the last survey wave, the following is true:

$$S^{non-farm} = \sum_X C_X^{non-farm} = \sum_X S_X * S_X^{non-farm}$$

Here, $S^{non-farm}$ is the share of people with main occupation in a non-farm sector during the last survey wave in the full sample; X is the migration status I observe during the last survey wave; $C_X^{non-farm}$ is the contribution of migration status X to the total share ($S^{non-farm}$); S_X is the share of people with migration status X in the full sample; and $S_X^{non-farm}$ is the share of people with main occupation in non-farm sector during the last survey wave among people with

migration status X . So, $S^{non-farm} = 13.5\%$ is a sum of $C_{non-migrant}^{non-farm} = 8.7\%$,

$C_{to low-dens.rural}^{non-farm} = 1.2\%$, $C_{to high-dens.rural}^{non-farm} = 0.8\%$, $C_{to peri-urban}^{non-farm} = 0.9\%$, $C_{to town}^{non-farm} =$

1.0% , and $C_{to city}^{non-farm} = 0.9\%$ (from Table 3.13). The share $S_X^{non-farm}$ differs widely by

migration status X , from 10.4% among non-migrant to 68.2% among migrant to towns, but it

loses its importance once the share of people with certain migration status in the sample, S_X , is

taken into account, because the share of non-migrants in the sample is very high at 83.7%. On

the other hand, even with the lowest share of people in the sample, migration to towns (1.5% of

the sample) contributes more to the total share of people with main occupation in a non-

agricultural sector during the last survey wave than migration to peri-urban areas and towns (a

bit over 1.8% of the sample each).

For all subsequent calculations, the formula has an additional component:

$$S^{A to B} = \sum_X C_X^{A to B} = \sum_X S_X * S_X^A * S_X^{A to B}$$

Here, $S^{A to B}$ is the share of people in the full sample who shifted their main occupation from A to B ; $C_X^{A to B}$ is the contribution of people with migration status X to this shift; S_X^A is the share of people who had main occupation A at baseline among people with migration status X ; and $S_X^{A to B}$ is the share of people who shifted their main occupation to B among people with migration status X who had main occupation A at baseline. In Table 3.13, I calculate the contributions of each migration status to the shift from agriculture to non-farm wage job and self-employment (see rows C, D, and A*C*D).

Table 3.13. Contribution of migration to various destinations to the total change in main occupation

	Non-migrants	Migrants to low-density rural areas	Migrants to high-density rural areas	Migrants to peri-urban areas	Migrants to towns	Migrants to cities
A = Share of population that this group represents	83.76%	7.15%	3.91%	1.85%	1.51%	1.83%
B = Among people in this group, share of those in the non-farm sector during the last survey wave	10.37%	16.04%	20.25%	49.53%	68.16%	50.96%
A * B = Contribution of this group to the total share of people in the non-farm sector during the last survey wave (13.50%)	8.68%	1.15%	0.79%	0.92%	1.03%	0.93%
C = Among people in this group, share of those in farm sector at baseline	68.96%	67.35%	63.04%	47.94%	40.68%	23.03%
D = Among people in this group in farm sector at baseline, share of those who shifted to non-farm sector	6.65%	10.74%	15.39%	47.38%	70.63%	43.78%
A * C * D = Contribution of this group to the total shift from farm to non-farm sector (5.77%)	3.84%	0.52%	0.38%	0.42%	0.43%	0.18%
E = Among people in this group, share of those in the non-farm sector at baseline	4.95%	4.61%	9.48%	7.34%	5.80%	10.35%
A * E = Contribution of this group to the total share of people in the non-farm sector at baseline (5.26%)	4.15%	0.33%	0.37%	0.14%	0.09%	0.19%

Note: “Non-farm sector” refers to main occupation in non-agricultural wage job or self-employment. “Farm sector” refers to main occupation in agriculture. Sample weights from 2008/2009 are applied.

Although only 6.7% of non-migrants with main occupation in agriculture shift to non-farm wage job or self-employment by the last survey wave, they contribute 3.8% out of 5.8% in the total share of people who shift, simply because non-migrants are the most populous group.

Migration to low-density rural areas contributes 0.5%, again mainly because of high amount of people with this migration status. The contribution of migration to high-density rural areas, peri-urban areas, and towns are comparable at 0.4%. Migration to cities contributes only 0.2% due to a low share of people with main occupation in farming at baseline and a relatively low share of people shifting to non-farm occupations among farmers.

This last point raises a concern for migration draining rural areas of people who already work in a non-agricultural sector at baseline. In Table 3.13, rows E and A*E show the share of people with main occupation in non-agricultural wage job or self-employment at baseline by their migration status. With the total share of 5.3%, the contribution of non-migrants is 4.2%, meaning that almost 80% of people with main occupation in a non-farm sector stayed in the origin. Migration to high-density rural areas is the most draining with the contribution of 0.4%, while migration to towns is the least draining with the contribution of 0.1%. Among urban destinations, migration to cities is the most draining with the contribution of 0.2%.

I showed that migration in general is associated with sectoral shifts in employment (for example, see Table 3.4), while non-migrants are more likely to maintain the same type of main occupation. Transition of workers from the agricultural sector is of the utmost importance for the structural transformation, hence looking at the share of people who keep their main occupation in farming after migration is of interest. Among non-migrants with main occupation in farming at baseline, 89.4% maintained their main occupation type by the last survey wave. Migration to rural areas is associated with lower probability for farmers to maintain main occupation in agriculture: 78.6% farmers who moved to low-density rural areas and 72.0% of farmers who moved to high-density rural areas maintain their main occupation in agriculture. Migration to urban areas does not lead to a full withdrawal of labor from agriculture: 22.9% of farmers who

moved to peri-urban areas and 10.7% of farmers who moved to towns maintain their main occupation in agriculture, while none of the farmers who moved to cities did.⁶⁷

I use the same approach to estimate the contribution of migration status to the shift from being engaged in work to not being engaged, and from not being engaged in work to being engaged. As in the parts of section 5.2, “engaged in work” is defined as having main occupation in agriculture, or non-agricultural wage job or self-employment. “Not engaged in work” is defined as having main occupation in studies, household maintenance, unemployment, or disability. The results are presented in Table 3.26. Migration contributes significantly to both the share of people who stopped being engaged in work (36% of those who stopped being engaged in work by the last survey wave are migrants, 20% are migrants to rural areas) and the share of people who became engaged in work (25% of those who became engaged in work are migrants, 15% are migrants to rural areas).

The share of people who were engaged (and not engaged) in work at baseline among migrants to rural areas of both types is comparable to that share among non-migrants. At the same time, the share of people shifting from being engaged in work to not being engaged is more than twice as high among migrants (10-11% among migrants to rural areas and 4% among non-migrants). The share of people shifting from not being engaged in work to being engaged is also much higher among migrants (75-80% among migrants to rural areas and 56% among non-migrants). The share of people engaged (not engaged) at work at baseline decreases (increases) with the level of urbanization of migration destination. Peri-urban areas and cities have the highest rates of shifting to not being engaged in work among people engaged at baseline: 26% and 40% respectively. Also, these two destinations have the lowest rates of shifting to being

⁶⁷ There are people with main occupation in farming during the last survey wave among migrants to cities, but they had main occupation in other sectors at baseline.

engaged in work among people not engaged at baseline, 52-53%. Towns are the best urban destination in this regard: the share of people shifting to not being engaged in work among those who were engaged is 18%, and the share of people shifting to being engaged among those who were not engaged is 64%.

Finally, I look at the contribution of migration status to the shift from studies to being engaged in work, and in work in a non-agricultural sector in particular (see results in Table 3.27). Migration contributes significantly to the shift from studies to being engaged in work: 24% of students who made this shift are migrants. But the contribution of migration to the shift of students to the non-farm sector is tremendous: 50% of students who made this shift are migrants. Rural destinations contribute more to the shift of students to work in general, while students who moved to peri-urban areas and cities have similar (or lower) rates of shifting to being engaged in work as (than) non-migrants do. At the same time, urban destinations contribute more to the shift of students to non-agricultural work, although all destination have the rates of students shifting to non-farm work over twice as high as among non-migrants.

As discussed in section 2.6, the main limitation for the study with a more detailed set of destination types is a decrease in the number of observations with an increase in the number of the distinguished location types. For this chapter in particular it is crucial, as I look at subsamples of migrants moving to a certain destination: people who had a certain main occupation at baseline. The group with the lowest number of observations is people who were not engaged in work at baseline, which is why I use a modified definition of “being engaged in work” to include people who had main occupation in household maintenance. Also, the use of the NBS categorization for locations on the rural-urban spectrum allows to increase the number of observations, especially for migrants to urban areas. An important limitation of the matching

strategy is the decrease in the quality of matching when the sample is small, as well as the inability to observe (and then match) some important characteristics like ability and aspirations.

3.7. Conclusion

This study estimates the impact that various types of migration destinations have on the employment outcomes of youth from rural areas of Tanzania. I confirm that the impact of migration on occupational shifts differs drastically by destination type. The four main outcomes I look at are: (i) the probability to stay engaged in work – for people who were engaged in work at baseline; (ii) the probability to become engaged in work – for people who were not engaged in work at baseline; (iii) the probability to have main occupation in the agricultural sector; and (iv) the probability to have main occupation in the non-agricultural sector. I find the evidence of higher unemployment among migrants to peri-urban areas and cities. At the same time, migration to low-density rural areas is associated with a higher probability to become employed. Among all destinations, only low-density rural areas are not tied to a decrease in the share of people with main occupation in agriculture; while they still – along with other destinations – are tied to an increase in the share of people with main occupation in non-agricultural wage job and self-employment. Some farmers moving to peri-urban areas and towns maintain their main occupation in agriculture.

Throughout the study, two destination types stand out: low-density rural areas and towns. These two location types are the most important for structural transformation, although they promote it through different channels. The rates of migration to low-density rural areas are almost as high as the rates of migration to all other destinations combined. People moving to these areas do not differ much in their initial characteristics from non-migrants. Yet, the observed occupational outcomes of migrants to low-density rural areas differ significantly from

that of non-migrants. Low-density rural destinations allow migrants to shift to non-farm employment or begin working. Youth who choose urban destinations, on the other hand, differs significantly from non-migrants and migrants to rural areas. But, compared to cities, towns are more likely to attract farmers and less likely to attract people employed in a non-farm sector. Also, unemployment rates in towns are smaller than in cities, which allows people to stay engaged in work after migration.

The use of a more complex set of location types on the rural-urban spectrum allows to see the nuances in the employment patterns. Migrants to high-density rural areas are already different from non-migrants and migrants to low-density rural areas both in their initial characteristics and in their employment choices at baseline. These differences intensify with the level of urbanization of the destination type. Yet, high-density rural areas still provide relatively low unemployment rates and high rates of agricultural employment. Peri-urban areas, on the other hand, already have a significant share of migrants not being employed in agriculture at baseline; and unemployment rates there are comparable to those in towns and cities.

The highest share of people shifting from farming to employment in non-agricultural sector do not migrate, yet migration still contributes to structural transformation through the shift from other sectors, like studies and household maintenance, into non-agricultural wage job and self-employment. I separate two groups of people to look at their employment shifts: students and women moving for marriage. I find that, as these groups are populous among rural youth, the patterns of their occupational transition are crucial to understanding the total picture of youth employment. I find that migration of students to any destination leads to almost twofold increase in the probability to have main occupation in non-agricultural wage job or self-employment at destination. The share of students among migrants to cities is the highest, and cities contribute

the most to their shift into non-farm employment. Movements of women who state marriage to be their main reason for migration is associated with a shift towards main occupation in household maintenance; but at the same time these women are much more likely to shift into non-agricultural self-employment at destination than women migrating for reasons other than marriage and migrant men.

The focus of this study is the self-reported main occupation of youth. The caveat to this approach is the inability to observe whether the occupation is in the formal or informal sector (which is especially important for wage jobs). It does not undermine the results on the sectoral shifts contributing to structural transformation, although it prevents from making certain conclusions about the expected lifetime earnings of youth. Also, self-reports do not necessarily reflect the full scope of the work people do. It is more common for people living in rural areas to maintain several occupations throughout the year: depending on the agricultural season, access to school, availability of non-farm jobs, etc. Hence, the answer about one's "main occupation" may differ throughout the year. Ideally, the self-reports would reflect person's time spent on various activities. Alternatively, the self-reports can capture the priorities people set for the allocation of their time, their aspiration. In that regard, migration can allow people to change the set of opportunities they are exposed to, and to reallocate their time accordingly. This is what the models used in this study aim to explain.

APPENDIX

Table 3.14. Main occupation of people of age 15-65, by age group, gender, and location type; NBS definition of “rural”

	Rural								Urban							
	Of age 15-34 in 2008/09				Of age 35-65 in 2008/09				Of age 15-34 in 2008/09				Of age 35-65 in 2008/09			
	Men		Women		Men		Women		Men		Women		Men		Women	
	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013
<i>A. Agriculture</i>																
Agriculture/livestock	61%	70%	73%	78%	86%	84%	93%	89%	6%	10%	12%	14%	21%	22%	34%	35%
Fishing	2%	1%	0%	0%	2%	1%	0%	0%	1%	1%	0%	0%	1%	1%	0%	0%
<i>B. Self-employment</i>																
Self-employed alone	3%	4%	2%	4%	4%	4%	3%	3%	21%	22%	18%	23%	36%	29%	31%	30%
Self-employed with employees	1%	1%	0%	0%	1%	1%	0%	0%	2%	4%	1%	1%	4%	7%	2%	2%
<i>C. Wage job</i>																
Private enterprise	2%	7%	0%	2%	2%	3%	0%	2%	15%	27%	6%	13%	19%	21%	4%	5%
Government	1%	2%	0%	1%	3%	3%	1%	1%	3%	4%	1%	3%	11%	8%	6%	4%
Parastatal	0%	0%	0%	0%	0%	1%	0%	0%	0%	1%	0%	0%	2%	3%	1%	1%
Mining	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1%	1%	0%	0%
Tourism	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NGO/religious	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	1%	2%	3%	1%	0%
<i>D. Student</i>																
	26%	9%	18%	5%	0%	0%	0%	0%	38%	16%	26%	12%	0%	0%	0%	0%
<i>E. Household maintenance</i>																
Family work without pay	3%	3%	4%	7%	0%	0%	1%	2%	5%	5%	25%	23%	0%	0%	16%	15%
Family work with pay	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	1%	0%	0%	0%	0%
<i>F. Unemployed or disabled</i>																
No job	1%	2%	1%	2%	0%	1%	1%	1%	6%	7%	7%	8%	2%	3%	3%	5%
Job seeker	1%	0%	0%	0%	0%	0%	0%	0%	2%	2%	1%	0%	0%	0%	0%	0%
Disabled	0%	0%	0%	0%	1%	1%	1%	2%	0%	1%	0%	0%	1%	1%	2%	2%
Number of observations	1377	1384	1480	1494	925	965	1063	1101	854	847	955	941	480	440	522	484

Note: Sample weights from each respective wave (2008/2009 or 2012/2013) are applied.

Figure 3.1. Propensity for migration to four location types on the rural-urban spectrum (according to the NBS definition of “rural”)

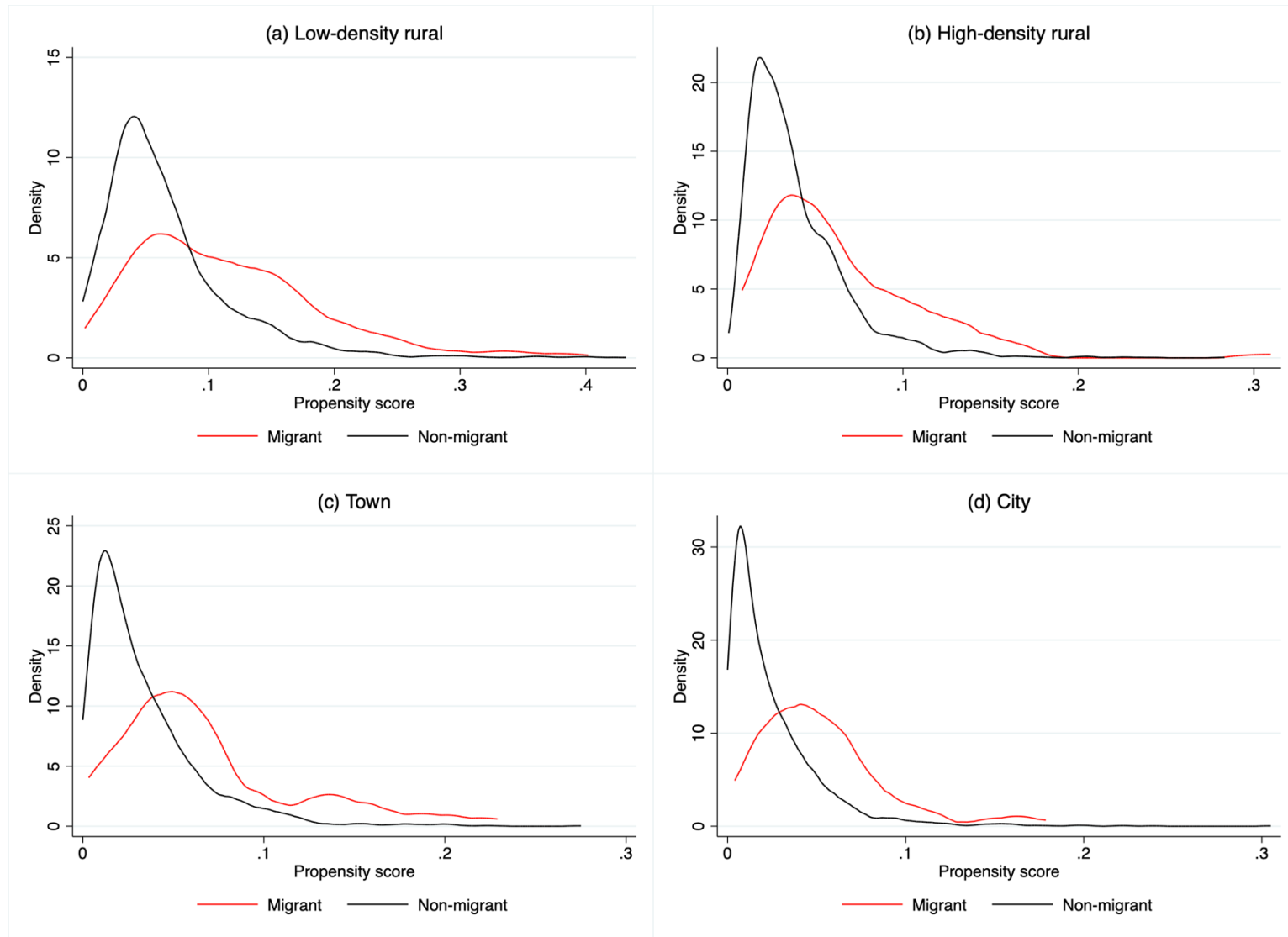


Table 3.15. Share of people with main occupation in a certain sector, by migration destination;
NBS definition of “rural”

	Non-migrants (2,423 observations)	Migrants to rural areas (283 observations)	Migrants to urban areas (151 observations)	Difference between rural- destined migrants and non-migrants	Difference between urban- destined migrants and non-migrants
<i>Panel A. Agriculture</i>					
2008/2009	0.699 (0.010)	0.681 (0.030)	0.439 (0.045)	-0.018 (0.028)	-0.260 (0.038)
2012/2013	0.761 (0.009)	0.687 (0.030)	0.119 (0.028)	-0.074 (0.026)	-0.642 (0.035)
Difference between 2012/2013 and 2008/2009	0.062 (0.013)	0.006 (0.039)	-0.320 (0.048)	-0.056 (0.039)	-0.382 (0.052)
<i>Panel B. Wage job or self-employment in a non-agricultural sector</i>					
2008/2009	0.043 (0.004)	0.038 (0.012)	0.078 (0.023)	-0.005 (0.012)	0.034 (0.017)
2012/2013	0.103 (0.007)	0.185 (0.025)	0.519 (0.045)	0.082 (0.019)	0.416 (0.026)
Difference between 2012/2013 and 2008/2009	0.060 (0.007)	0.146 (0.026)	0.441 (0.046)	0.087 (0.023)	0.381 (0.032)
<i>Panel C. Student</i>					
2008/2009	0.209 (0.009)	0.210 (0.026)	0.390 (0.044)	0.001 (0.025)	0.181 (0.034)
2012/2013	0.073 (0.006)	0.028 (0.011)	0.051 (0.020)	-0.044 (0.015)	-0.022 (0.021)
Difference between 2012/2013 and 2008/2009	-0.136 (0.010)	-0.182 (0.026)	-0.339 (0.044)	-0.045 (0.029)	-0.203 (0.040)
<i>Panel D. Household maintenance, unemployment, disability</i>					
2008/2009	0.049 (0.005)	0.070 (0.016)	0.093 (0.025)	0.022 (0.013)	0.044 (0.018)
2012/2013	0.063 (0.005)	0.100 (0.019)	0.312 (0.042)	0.037 (0.015)	0.248 (0.022)
Difference between 2012/2013 and 2008/2009	0.015 (0.007)	0.030 (0.023)	0.219 (0.045)	0.015 (0.020)	0.204 (0.028)

Note: Standard errors are in parentheses. Sample weights from the 2008/2009 survey wave are applied.

Table 3.16. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status; each row sums to 100%; NBS definition of “rural”

		Main occupation in 2012/2013				Number of observations: 2008/2009
		Agriculture	Wage job and self-employment	Student	Household maintenance, unemployment, disability	
<i>Panel A. Non-migrants</i>						
Main occupation in 2008/2009	Agriculture	89%	7%	1%	3%	1597
	Wage job and self-employment	42%	54%	0%	4%	132
	Student	47%	10%	30%	12%	525
	Household maintenance, unemployment, disability	45%	15%	9%	31%	169
	Number of observations: 2012/2013	1710	278	182	253	2423
<i>Panel B. Migrants to rural areas</i>						
Main occupation in 2008/2009	Agriculture	76%	14%	1%	10%	183
	Wage job and self-employment	31%	53%	0%	16%	10
	Student	57%	23%	12%	9%	67
	Household maintenance, unemployment, disability	56%	31%	0%	14%	23
	Number of observations: 2012/2013	190	52	8	33	283
<i>Panel C. Migrants to urban areas</i>						
Main occupation in 2008/2009	Agriculture	20%	47%	2%	31%	62
	Wage job and self-employment	21%	75%	0%	4%	15
	Student	3%	50%	7%	40%	58
	Household maintenance, unemployment, disability	4%	63%	17%	17%	16
	Number of observations: 2012/2013	20	76	7	48	151

Note: Sample weights from the 2008/2009 survey wave are applied.

Table 3.17. Share of people with main occupation in a certain sector in 2008/2009, by their main occupation in 2012/2013 and migration status, for six groups of observations with at least 10 observations in 2008/2009; each row sums to 100%; NBS definition of “rural”

		Main occupation in 2012/2013						Number of observations: 2008/2009
		Agriculture	Wage job	Self-employment	Student	Household maintenance	Unemployed or disabled	
<i>Panel A. Non-migrants</i>								
Main occupation in 2008/2009	Agriculture	89%	4%	3%	1%	2%	1%	1597
	Wage job	33%	50%	16%	0%	1%	1%	51
	Self-employment	48%	10%	37%	0%	2%	4%	81
	Student	47%	6%	4%	30%	8%	4%	525
	Household maintenance	47%	4%	10%	6%	30%	4%	93
	Unemployed or disabled	42%	6%	11%	17%	10%	14%	76
<i>Panel B. Migrants to rural areas</i>								
Main occupation in 2008/2009	Agriculture	76%	5%	9%	1%	8%	2%	183
	Student	57%	19%	3%	12%	8%	0%	67
	Household maintenance	55%	27%	3%	0%	16%	0%	18
<i>Panel C. Migrants to urban areas</i>								
Main occupation in 2008/2009	Agriculture	20%	30%	17%	2%	24%	8%	62
	Student	3%	37%	14%	7%	36%	4%	58
	Household maintenance	4%	36%	25%	19%	16%	0%	13

Note: Sample weights from the 2008/2009 survey wave are applied.

Table 3.18. Selection into migration: marginal values from multinomial logistic regression of indicators to have main occupation in agriculture and non-agricultural wage job or self-employment in 2008/2009 on migration status in 2012/2013; NBS definition of “rural”

Outcome variable	Indicator for migration	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a town	Indicator for migration to a city
<i>A. Multinomial logistic regression without controls</i>							
Main occupation in farming or fishing	-0.091*** (0.024)	-0.009 (0.030)	-0.230*** (0.038)	0.041 (0.038)	-0.092** (0.047)	-0.143*** (0.049)	-0.357*** (0.062)
Main occupation in wage job or self-employment	0.004 (0.012)	-0.022 (0.017)	0.040*** (0.015)	-0.036 (0.023)	-0.003 (0.024)	0.015 (0.022)	0.069*** (0.020)
Number of observations	2,857	2,706	2,574	2,606	2,523	2,510	2,487
<i>B. Multinomial logistic regression with controls for age, gender, marital status, primary school completion, being born in the village of residence, household size, land area the household cultivates, and asset index</i>							
Main occupation in farming or fishing	-0.021 (0.020)	0.021 (0.025)	-0.092*** (0.031)	0.064** (0.032)	-0.056 (0.040)	-0.027 (0.040)	-0.186*** (0.050)
Main occupation in wage job or self-employment	0.016 (0.011)	-0.004 (0.016)	0.046*** (0.016)	-0.014 (0.021)	0.009 (0.023)	0.023 (0.022)	0.076*** (0.022)
Number of observations	2,855	2,704	2,572	2,604	2,521	2,508	2,485
<i>C. Multinomial logistic regression with controls from Panel B and controls for being away from the household in the past year, being a head of the household, being a child of the household head, household head's age and gender, units of livestock owned by the household, agricultural and non-agricultural shocks experienced by the household, population density, distance to the nearest road, and distance to the nearest town with population of at least 50,000 people</i>							
Main occupation in farming or fishing	-0.033* (0.020)	-0.003 (0.024)	-0.085*** (0.030)	0.024 (0.030)	-0.051 (0.038)	-0.043 (0.038)	-0.155*** (0.049)
Main occupation in wage job or self-employment	0.015 (0.011)	-0.001 (0.015)	0.037** (0.015)	-0.008 (0.020)	0.009 (0.021)	0.022 (0.020)	0.061*** (0.021)
Number of observations	2,855	2,704	2,572	2,604	2,521	2,508	2,485

Note: Base outcome is to list one of these four categories as main occupation: studies, household maintenance, unemployment, or disability. Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Table 3.19. Migration and the probability to stay engaged in work; NBS definition of “rural”

	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a town	Indicator for migration to a city
Difference in means between non-migrants and migrants	-0.006 (0.010)	-0.076*** (0.016)	-0.006 (0.011)	-0.006 (0.016)	-0.068*** (0.018)	-0.096*** (0.026)
<i>Logistic regression</i>						
Without controls	0.004 (0.013)	-0.032** (0.013)	0.007 (0.016)	-0.002 (0.020)	-0.026* (0.015)	-0.037** (0.018)
With a small set of controls	0.004 (0.013)	-0.014 (0.012)	0.004 (0.016)	0.005 (0.020)	-0.012 (0.015)	-0.018 (0.017)
With a large set of controls	-0.001 (0.012)	-0.019 (0.012)	-0.005 (0.015)	0.002 (0.019)	-0.018 (0.015)	-0.025 (0.017)
<i>Propensity score matching</i>						
With a small set of variables	-0.005 (0.014)	-0.044 (0.036)	-0.014 (0.014)	-0.014 (0.026)	-0.017 (0.037)	0.065 (0.092)
With a large set of variables	-0.014 (0.013)	-0.022 (0.041)	-0.007 (0.017)	-0.014 (0.022)	-0.051** (0.022)	-0.065 (0.057)
<i>Nearest neighbor matching</i>						
With a small set of variables	0.004 (0.017)	-0.047 (0.035)	0.012 (0.021)	-0.015 (0.025)	-0.027 (0.045)	-0.097* (0.053)
With a large set of variables	-0.014 (0.013)	-0.033 (0.034)	-0.021* (0.012)	0.012 (0.033)	-0.031 (0.036)	-0.075 (0.082)

Note: “Engaged in work” is defined as having main occupation in agriculture, non-agricultural wage job, non-agricultural self-employment, and household maintenance. Students, unemployed people, and disabled people are considered to not be engaged in work. For people engaged in work in the first survey wave, I estimate the probability to stay engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied.

Table 3.20. Migration and the probability to stay engaged in work excluding household maintenance; NBS definition of “rural”

	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a town	Indicator for migration to a city
Difference in means between non-migrants and migrants	-0.068*** (0.015)	-0.247*** (0.024)	-0.055*** (0.018)	-0.096*** (0.025)	-0.190*** (0.028)	-0.376*** (0.041)
<i>Logistic regression</i>						
Without controls	-0.037** (0.015)	-0.105*** (0.018)	-0.022 (0.019)	-0.059*** (0.022)	-0.081*** (0.022)	-0.129*** (0.024)
With a small set of controls	-0.030* (0.015)	-0.082*** (0.017)	-0.015 (0.019)	-0.051** (0.021)	-0.066*** (0.021)	-0.096*** (0.024)
With a large set of controls	-0.039*** (0.015)	-0.086*** (0.017)	-0.032* (0.019)	-0.050** (0.021)	-0.073*** (0.021)	-0.100*** (0.024)
<i>Propensity score matching</i>						
With a small set of variables	-0.052** (0.025)	-0.130*** (0.046)	-0.038 (0.031)	-0.131*** (0.046)	-0.120*** (0.047)	-0.333*** (0.082)
With a large set of variables	-0.052* (0.029)	-0.208*** (0.054)	-0.045 (0.030)	-0.066* (0.035)	-0.120*** (0.018)	-0.259** (0.108)
<i>Nearest neighbor matching</i>						
With a small set of variables	-0.042 (0.030)	-0.153** (0.063)	-0.017 (0.035)	-0.097* (0.052)	-0.079 (0.077)	-0.312*** (0.090)
With a large set of variables	-0.070*** (0.027)	-0.121** (0.054)	-0.053* (0.030)	-0.087* (0.049)	0.011 (0.070)	-0.362*** (0.098)

Note: “Engaged in work excluding household maintenance” is defined as having main occupation in agriculture, non-agricultural wage job, and non-agricultural self-employment. People with main occupation in household maintenance, students, unemployed people, and disabled people are considered to not be engaged in work. For people engaged in work in the first survey wave, I estimate the probability to stay engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied.

Table 3.21. Migration and the probability to become engaged in work; NBS definition of “rural”

	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a town	Indicator for migration to a city
Difference in means between non-migrants and migrants	0.232*** (0.057)	0.239*** (0.059)	0.219*** (0.074)	0.250*** (0.085)	0.329*** (0.081)	0.145* (0.084)
<i>Logistic regression</i>						
Without controls	0.363*** (0.082)	0.359*** (0.089)	0.350*** (0.109)	0.411*** (0.139)	0.651*** (0.231)	0.246** (0.106)
With a small set of controls	0.315*** (0.080)	0.375*** (0.087)	0.307*** (0.105)	0.347*** (0.133)	0.668*** (0.219)	0.254** (0.104)
With a large set of controls	0.314*** (0.080)	0.370*** (0.086)	0.315*** (0.105)	0.332** (0.131)	0.641*** (0.214)	0.268*** (0.103)
<i>Propensity score matching</i>						
With a small set of variables	0.278*** (0.072)	0.410*** (0.080)	0.305*** (0.104)	0.290*** (0.101)	0.464*** (0.108)	0.242** (0.113)
With a large set of variables	0.333*** (0.066)	0.361*** (0.075)	0.268*** (0.091)	0.097 (0.097)	0.429*** (0.088)	0.333*** (0.091)
<i>Nearest neighbor matching</i>						
With a small set of variables	0.385*** (0.081)	0.298*** (0.093)	0.419*** (0.102)	0.322*** (0.114)	0.260* (0.142)	0.341*** (0.127)
With a large set of variables	0.200*** (0.064)	0.377*** (0.084)	0.018 (0.092)	0.342*** (0.104)	-0.105 (0.348)	0.285*** (0.106)

Note: “Engaged in work” is defined as having main occupation in agriculture, non-agricultural wage job, non-agricultural self-employment, and household maintenance. Students, unemployed people, and disabled people are considered to not be engaged in work. For people not engaged in work in the first survey wave, I estimate the probability to become engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. For the propensity score matching, marital status is excluded due to low number of observations.

Table 3.22. Migration and the probability to become engaged in work excluding household maintenance; NBS definition of “rural”

	Indicator for migration to a rural area	Indicator for migration to an urban area	Indicator for migration to a low-density rural area	Indicator for migration to a high-density rural area	Indicator for migration to a town	Indicator for migration to a city
Difference in means between non-migrants and migrants	0.233*** (0.053)	-0.027 (0.058)	0.238*** (0.068)	0.225*** (0.080)	0.004 (0.077)	-0.064 (0.082)
<i>Logistic regression</i>						
Without controls	0.271*** (0.060)	0.010 (0.061)	0.357*** (0.088)	0.184** (0.087)	0.051 (0.085)	-0.031 (0.084)
With a small set of controls	0.259*** (0.060)	0.027 (0.061)	0.336*** (0.086)	0.176** (0.086)	0.074 (0.084)	-0.011 (0.083)
With a large set of controls	0.247*** (0.059)	0.017 (0.060)	0.316*** (0.084)	0.184** (0.084)	0.029 (0.083)	0.015 (0.083)
<i>Propensity score matching</i>						
With a small set of variables	0.294*** (0.067)	0.027 (0.082)	0.451*** (0.098)	0.218* (0.112)	-0.027 (0.135)	0.162 (0.104)
With a large set of variables	0.311*** (0.075)	0.135** (0.067)	0.373*** (0.107)	0.231** (0.112)	0.135*** (0.045)	-0.081 (0.096)
<i>Nearest neighbor matching</i>						
With a small set of variables	0.223*** (0.074)	0.039 (0.088)	0.314*** (0.097)	0.122 (0.115)	0.131 (0.113)	-0.013 (0.125)
With a large set of variables	0.203*** (0.075)	-0.006 (0.083)	0.177* (0.098)	0.225** (0.113)	0.033 (0.135)	0.009 (0.130)

Note: “Engaged in work excluding household maintenance” is defined as having main occupation in agriculture, non-agricultural wage job, and non-agricultural self-employment. People with main occupation in household maintenance, students, unemployed people, and disabled people are considered to not be engaged in work. For people not engaged in work in the first survey wave, I estimate the probability to become engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. For the propensity score matching, marital status and the indicator of being the head of the household are excluded due to low number of observations.

Table 3.23. Migration and the probability to have main occupation in agriculture in the last survey wave

	I(migr. to a rural area)	I(migr. to an urban area)	I(migr. to a low-density rural area)	I(migr. to a high-density rural area)	I(migr. to a peri-urban area)	I(migr. to a town)	I(migr. to a city)
Difference in means between non-migrants and migrants in the last survey wave	-0.060** (0.026)	-0.669*** (0.036)	-0.037 (0.032)	-0.101** (0.042)	-0.597*** (0.060)	-0.707*** (0.067)	-0.710*** (0.060)
Difference in means between non-migrants' and migrants' differences between the last and the first survey waves	-0.008 (0.039)	-0.346*** (0.054)	0.000 (0.048)	-0.021 (0.064)	-0.373*** (0.091)	-0.408*** (0.100)	-0.233** (0.091)
<i>Logistic regression</i>							
Without controls	-0.030 (0.028)	-0.653*** (0.057)	0.017 (0.035)	-0.112*** (0.043)	-0.513*** (0.078)	-0.784*** (0.147)	-0.843*** (0.145)
With I(ag.)	-0.019 (0.024)	-0.496*** (0.045)	0.021 (0.030)	-0.086** (0.038)	-0.401*** (0.064)	-0.615*** (0.112)	-0.602*** (0.111)
With a small set of controls	-0.068*** (0.025)	-0.577*** (0.054)	-0.035 (0.031)	-0.123*** (0.039)	-0.460*** (0.071)	-0.629*** (0.121)	-0.729*** (0.133)
With a small set of controls and I(ag.)	-0.061** (0.024)	-0.507*** (0.047)	-0.033 (0.029)	-0.111*** (0.037)	-0.397*** (0.063)	-0.585*** (0.106)	-0.573*** (0.109)
With a large set of controls	-0.062** (0.025)	-0.541*** (0.050)	-0.048 (0.030)	-0.086** (0.038)	-0.421*** (0.067)	-0.601*** (0.113)	-0.663*** (0.119)
With a large set of controls and I(ag.)	-0.058** (0.023)	-0.486*** (0.045)	-0.043 (0.028)	-0.088** (0.036)	-0.375*** (0.060)	-0.565*** (0.102)	-0.545*** (0.103)
<i>Propensity score matching</i>							
With a small set of variables	-0.076* (0.039)	-0.550*** (0.051)	-0.031 (0.042)	-0.102 (0.075)	-0.455*** (0.079)	-0.660*** (0.078)	-0.575*** (0.074)
With a small set of variables and I(ag.)	-0.102*** (0.036)	-0.494*** (0.049)	-0.030 (0.044)	-0.213*** (0.065)	-0.459*** (0.082)	-0.565*** (0.095)	-0.475*** (0.074)
With a large set of variables	-0.032 (0.041)	-0.560*** (0.056)	-0.043 (0.044)	-0.124** (0.059)	-0.465*** (0.087)	-0.583*** (0.089)	-0.600*** (0.077)
With a large set of variables and I(ag.)	-0.078** (0.038)	-0.532*** (0.051)	0.027 (0.049)	-0.062 (0.065)	-0.395*** (0.054)	-0.417*** (0.104)	-0.533*** (0.082)
<i>Nearest neighbor matching</i>							
With a small set of variables	-0.103*** (0.036)	-0.591*** (0.047)	-0.069* (0.040)	-0.170** (0.069)	-0.539*** (0.079)	-0.670*** (0.082)	-0.572*** (0.072)
With a small set of variables and I(ag.)	-0.052 (0.035)	-0.493*** (0.047)	-0.006 (0.040)	-0.141** (0.058)	-0.424*** (0.081)	-0.588*** (0.084)	-0.477*** (0.070)
With a large set of variables	-0.109*** (0.036)	-0.483*** (0.054)	-0.092** (0.040)	-0.158** (0.064)	-0.425*** (0.104)	-0.499*** (0.077)	-0.494*** (0.082)
With a large set of variables and I(ag.)	-0.094*** (0.035)	-0.405*** (0.050)	-0.067* (0.040)	-0.141** (0.061)	-0.320*** (0.095)	-0.522*** (0.082)	-0.358*** (0.082)

Note: "I(migr.)" is an indicator for migration. I(ag.) is an indicator for having main occupation in agriculture at baseline. In the small set of variables, indicator of being a household head and indicator of being married are excluded. In the large set of variables, land area under cultivation and asset index are replaced with indicators of living in a household that is above median in the respective variable. They are omitted in the logistic regressions with the indicator of migration to a city. For the computation of the differences in means, sample weights from the 2008/2009 survey wave were applied. Constructed definition of "rural" is used.

Table 3.24. Migration and the probability to have main occupation in non-agricultural wage job or self-employment in the last survey wave; NBS definition of “rural”

	I. for migration to a rural area	I. for migration to an urban area	I. for migration to a low- density rural area	I. for migration to a high- density rural area	I. for migration to a town	I. for migration to a city
Difference in means between non-migrants and migrants in 2012/13	0.082*** (0.019)	0.416*** (0.026)	0.055** (0.023)	0.131*** (0.031)	0.421*** (0.033)	0.407*** (0.041)
Difference in means between non-migrants' and migrants' differences between 2008/09 and 2012/13	0.064*** (0.024)	0.372*** (0.032)	0.042 (0.030)	0.094** (0.039)	0.388*** (0.041)	0.320*** (0.052)
<i>Logistic regression</i>						
Without controls	0.059*** (0.018)	0.227*** (0.018)	0.034 (0.022)	0.093*** (0.025)	0.225*** (0.023)	0.209*** (0.027)
With I(NA)	0.067*** (0.016)	0.211*** (0.017)	0.047** (0.020)	0.094*** (0.023)	0.214*** (0.021)	0.184*** (0.025)
With a small set of controls	0.079*** (0.018)	0.227*** (0.019)	0.058*** (0.022)	0.103*** (0.025)	0.222*** (0.023)	0.214*** (0.027)
With a small set of controls and I(NA)	0.077*** (0.017)	0.206*** (0.017)	0.059*** (0.020)	0.098*** (0.023)	0.206*** (0.021)	0.182*** (0.026)
With a large set of controls	0.088*** (0.017)	0.206*** (0.018)	0.076*** (0.021)	0.099*** (0.024)	0.206*** (0.022)	0.187*** (0.027)
With a large set of controls and I(NA)	0.085*** (0.016)	0.191*** (0.017)	0.074*** (0.020)	0.095*** (0.023)	0.196*** (0.021)	0.162*** (0.025)
<i>Propensity score matching</i>						
With a small set of variables	0.085*** (0.029)	0.401*** (0.043)	0.022 (0.037)	0.160*** (0.050)	0.425*** (0.060)	0.359*** (0.058)
With a small set of variables and I(NA)	0.081*** (0.031)	0.361*** (0.044)	0.060* (0.035)	0.140*** (0.050)	0.356*** (0.066)	0.344*** (0.041)
With a large set of variables	0.087*** (0.028)	0.371*** (0.050)	0.055* (0.031)	0.140*** (0.049)	0.368*** (0.058)	0.344*** (0.058)
With a large set of variables and I(NA)	0.099*** (0.028)	0.377*** (0.052)	0.071** (0.034)	0.130*** (0.049)	0.368*** (0.058)	0.297*** (0.072)
<i>Nearest neighbor matching</i>						
With a small set of variables	0.082*** (0.030)	0.356*** (0.050)	0.058 (0.035)	0.129** (0.055)	0.396*** (0.065)	0.309*** (0.077)
With a small set of variables and I(NA)	0.071** (0.030)	0.323*** (0.051)	0.046 (0.035)	0.120** (0.053)	0.357*** (0.067)	0.256*** (0.076)
With a large set of variables	0.060* (0.031)	0.396*** (0.049)	0.063* (0.035)	0.072 (0.059)	0.397*** (0.062)	0.350*** (0.074)
With a large set of variables and I(NA)	0.056* (0.029)	0.358*** (0.050)	0.078** (0.034)	0.071 (0.056)	0.351*** (0.062)	0.329*** (0.077)

Note: “I.” stands for “indicator”. I(NA) is an indicator for having main occupation in non-agricultural wage job or self-employment at baseline. For the computation of the differences in means, sample weights from the 2008/2009 survey wave were applied.

Table 3.25. Migration and the probability to stay engaged in work including studies

	Migration to a rural area	Migration to an urban area	Migration to a low- density rural area	Migration to a high- density rural area	Migration to a peri- urban area	Migration to a town	Migration to a city
Difference in means between non-migrants and migrants	-0.044*** (0.015)	-0.231*** (0.023)	-0.031* (0.018)	-0.067*** (0.024)	-0.295*** (0.036)	-0.177*** (0.041)	-0.208*** (0.036)
<i>Logistic regression</i>							
Without controls	-0.022 (0.016)	-0.134*** (0.017)	0.003 (0.021)	-0.057** (0.023)	-0.154*** (0.026)	-0.114*** (0.031)	-0.109*** (0.026)
With a small set of controls	-0.015 (0.016)	-0.104*** (0.017)	0.006 (0.021)	-0.049** (0.022)	-0.135*** (0.026)	-0.090*** (0.031)	-0.072*** (0.025)
With a large set of controls	-0.018 (0.016)	-0.113*** (0.017)	0.000 (0.021)	-0.046** (0.022)	-0.140*** (0.026)	-0.103*** (0.031)	-0.079*** (0.025)
<i>Propensity score matching</i>							
With a small set of variables	-0.056** (0.023)	-0.205*** (0.047)	0.011 (0.029)	-0.062 (0.049)	-0.317*** (0.066)	-0.062 (0.102)	-0.143** (0.068)
With a large set of variables	-0.029 (0.025)	-0.205*** (0.047)	0.011 (0.030)	-0.083** (0.038)	-0.317*** (0.078)	-0.234** (0.091)	-0.163** (0.067)
<i>Nearest neighbor matching</i>							
With a small set of variables	-0.033 (0.025)	-0.139*** (0.052)	-0.034 (0.028)	-0.022 (0.049)	-0.200** (0.088)	-0.080 (0.091)	-0.117 (0.083)
With a large set of variables	-0.032 (0.026)	-0.195*** (0.048)	-0.000 (0.033)	-0.105** (0.044)	-0.307*** (0.086)	-0.109 (0.105)	-0.161** (0.079)

Note: “Engaged in work including studies” is defined as having main occupation in agriculture, non-agricultural wage job, non-agricultural self-employment, or studies. People with main occupation in household maintenance, unemployed people, and disabled people are considered to not be engaged in work. For people engaged in work in the first survey wave, I estimate the probability to stay engaged in work by the last survey wave. The outside option is to not be engaged in work during the last survey wave. For the computation of the differences in means, sample weights from the 2008/2009 survey wave are applied. Constructed definition of “rural” is used.

Table 3.26. Contribution of migration to various destinations to the total shift in and out of engagement in work

	Non-migrants	Migrants to low-density rural areas	Migrants to high-density rural areas	Migrants to peri-urban areas	Migrants to towns	Migrants to cities
A = Share of population that this group represents	83.76%	7.15%	3.91%	1.85%	1.51%	1.83%
F = Among people in this group, share of those who were engaged in work at baseline	73.92%	71.96%	72.52%	55.29%	46.48%	33.38%
G = Among people in this group engaged in work at baseline, share of those who are not engaged during the last survey wave	4.22%	9.94%	11.15%	25.78%	18.13%	40.02%
A * F * G = Contribution of this group to the total shift from being engaged in work to not being engaged in work (4.08%)	2.61%	0.51%	0.32%	0.26%	0.13%	0.24%
H = 100% - D = Among people in this group, share of those who were not engaged in work at baseline at baseline	26.08%	28.04%	27.48%	44.71%	53.52%	66.62%
I = Among people in this group not engaged in work at baseline, share of those who are engaged during the last survey wave	55.90%	80.34%	75.34%	53.16%	64.39%	52.49%
A * H * I = Contribution of this group to the total shift from not being engaged in work to being engaged in work (16.23%)	12.21%	1.61%	0.81%	0.44%	0.52%	0.64%

Note: A continuation of Table 3.13. “Engaged in work” refers to main occupation in agriculture or non-agricultural wage job or self-employment. “Not engaged in work” refers to main occupation in studies, household maintenance, unemployment, or disability. Sample weights from 2008/2009 are applied.

Table 3.27. Contribution of migration to various destinations to the total shift from being a student into engaging in certain types of work

	Non-migrants	Migrants to low-density rural areas	Migrants to high-density rural areas	Migrants to peri-urban areas	Migrants to towns	Migrants to cities
A = Share of population that this group represents	83.76%	7.15%	3.91%	1.85%	1.51%	1.83%
J = Among people in this group, share of those with main occupation in studies at baseline	20.73%	20.57%	22.66%	30.67%	32.27%	53.82%
K = Among people in this group with main occupation in studies at baseline, share of those who are engaged in work during the last survey wave	54.84%	78.88%	76.73%	47.93%	61.09%	55.15%
A * J * K = Contribution of this group to the total shift from being a student to being engaged in work (12.47%)	9.52%	1.16%	0.68%	0.27%	0.30%	0.54%
L = Among people in this group with main occupation in studies at baseline, share of those who have main occupation in non-agricultural wage job or self-employment during the last survey wave	9.72%	23.40%	30.17%	40.71%	61.09%	55.15%
A * J * L = Contribution of this group to the total shift from being a student to having main occupation in non-farm wage job or self-employment (3.37%)	1.69%	0.34%	0.27%	0.23%	0.30%	0.54%

Note: A continuation of Table 3.26. “Engaged in work” refers to main occupation in agriculture or non-agricultural wage job or self-employment. Sample weights from 2008/2009 are applied.

REFERENCES

REFERENCES

- Beauchemin, C., and P. Bocquier. 2004. Migration and urbanisation in Francophone West Africa: An overview of the recent empirical evidence. *Urban studies* 41(11): 2245-2272.
- Blekking, J., K. B. Waldman, C. Tuholske, and T. Evans. 2020. Formal/informal employment and urban food security in Sub-Saharan Africa. *Applied Geography* 114: 102131.
- Blekking, J., K. B. Waldman, S. Lopus, and S. Giroux. 2020. Migration and urban food accessibility in Mumbwa, a tertiary city in Zambia. *Migration and Development*: 1-17.
- Bridges, S., L. Fox, A. Gaggero, and T. Owens. 2017. Youth unemployment and earnings in Africa: Evidence from Tanzanian retrospective data. *Journal of African Economies* 26(2): 119-139.
- Christiaensen, L., and R. Kanbur. 2017. Secondary towns and poverty reduction: Refocusing the urbanization agenda. *Annual Review of Resource Economics* 9: 405-419.
- Cockx, L., L. Colen, and J. De Weerd. 2018. From corn to popcorn? Urbanization and dietary change: Evidence from rural-urban migrants in Tanzania. *World Development* 110: 140-159.
- Corbane, C., A. Florczyk, M. Pesaresi, P. Politis, and V. Syrris. 2018. GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A. *European Commission, Joint Research Centre (JRC)* doi: 10.2905/jrc-ghsl-10007 PID: <http://data.europa.eu/89h/jrc-ghsl-10007>
- Crush, J. 2013. Linking food security, migration and development. *International Migration* 51(5): 61-75.
- Davis, B., S. Di Giuseppe, and A. Zezza. 2017. Are African household (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy* 67: 153-174.
- De Brauw, A., V. Mueller, and H. L. Lee. 2014. The role of rural-urban migration in the structural transformation of Sub-Saharan Africa. *World Development* 63: 33-42.
- Diao, X., E. Magalhaes, and M. McMillan. 2018. Understanding the role of rural non-farm enterprises in Africa's economic Transformation: Evidence from Tanzania. *The Journal of Development Studies* 54(5): 833-855.
- Emran, M. S., and F. Shilpi. 2018. Beyond dualism: Agricultural productivity, small towns, and structural change in Bangladesh. *World Development* 107: 264-276.
- Harris, J. R., and M. P. Todaro. 1970. Migration, unemployment and development: A two-sector analysis. *American Economic Review* 60(1): 126-142.
- Herrendorf, B., R. Rogerson, and A. Valentinyi. 2014. Growth and structural transformation. *Handbook of Economic Growth* 2, 855-941.

- Ingelaere, B., L. Christiaensen, J. De Weerd, and R. Kanbur. 2018. Why secondary towns can be important for poverty reduction – A migrant perspective. *World Development* 105: 273-282.
- Lewis, W. A. 1954. Economic development with unlimited supply of labour. *The Manchester School* 22(2): 139-191.
- Lucas, R. 2016. Internal migration in developing economies: An overview of recent evidence. *Geopolitics, History, and International Relations* 8(2): 159-191.
- Losch, B. 2017. A lastly booming rural population and the youth employment challenge. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in Motion. Dynamics and Drivers of Migration South of the Sahara*. Rome: FAO and CIRAD, pp. 20-21.
- McKenzie, D., S. Stillman, and J. Gibson. 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.
- Mercandalli, S., B. Losch, C. Rapone, R. Bourgeois, and C. A. Khalil. 2017. Rural migration and the new dynamics of structural transformation in Sub-Saharan Africa. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in Motion. Dynamics and Drivers of Migration South of the Sahara*. Rome: FAO and CIRAD, pp. 14-17.
- Mueller, V., E. Schmidt, N. Lozano, and S. Murray. 2019. Implications of migration on employment and occupational transitions in Tanzania. *International Regional Science Review* 42(2): 181-206.
- Tatem, A. J. 2017. WorldPop: Open data for spatial demography. *Scientific Data*, 4:170004.
- World Bank. 2017. *Living Standards Measurement Study – Integrated Surveys on Agriculture, Tanzania*.
 Wave 1 (2008-2009) retrieved from <http://microdata.worldbank.org/index.php/catalog/76>.
 Wave 2 (2010-2011) retrieved from <http://microdata.worldbank.org/index.php/catalog/1050>.
 Wave 3 (2012-2013) retrieved from <http://microdata.worldbank.org/index.php/catalog/2252>.

4. IMPACTS OF YOUTH OUTMIGRATION ON THE LIVELIHOOD OF HOUSEHOLDS LEFT BEHIND: EVIDENCE FROM TANZANIA

Abstract

Labor supply to the household farm can decrease after young adults move away from their original households, which they often do as migration of youth is a prominent phenomenon in Sub-Saharan Africa and in Tanzania in particular. The households then can undertake certain actions to maintain their livelihood, and this paper in particular focuses on the reallocation of labor. Using the data from the Living Standards Measurement Study in Tanzania from 2008/2009 and 2012/2013, I look at the impacts of youth outmigration on the time that the household members who stayed in the origin spend doing various agricultural tasks, on the adjustments to hired labor and land area under cultivation, and on the attraction of new household members. I investigate whether the observed effects differ by migrant's age, gender, and destination. Also, I look at the changes to labor patterns of household members of different age and gender. I apply difference-in-differences strategy with matching methods to account for selection into migrant-sending households. I find that women of age 34-65 living in households that experienced outmigration of youth significantly increase their labor inputs to the household farm, compared to women in households that did not experience outmigration of youth. At the same time, migrant-sending households tend to attract new household members through marriage, return migration, or extended family ties. Although I observe migrants' labor input to the household farm prior to migration to be lower than that of non-migrant youth with similar characteristics, the remaining household members still adjust their time spent on the farm in response to the loss of this labor.

4.1. Introduction

Internal migration of rural youth is common in Sub-Saharan Africa (Dinbabo, Mensah, and Belebema, 2017), but the question of how it affects the livelihood of migrants' families staying in the origin is still understudied (Mueller, Doss, and Quisumbing, 2018). In the contexts of Asia and South America, an extensive literature already exists on the effects of adult's and young adults' migration on the livelihood of their elderly parents and children left behind (e.g., studies by Ye et al., 2013; Qin and Liao, 2016; Antman, 2012b). Still, most studies done in Sub-Saharan African countries focus on the impacts of male labor migration on health, time use, and other outcomes of the spouses and children left behind (e.g., Agadjanian, Arnaldo, and Cau, 2011; Bennett et al., 2015; Agadjanian and Hayford, 2018). But if the majority of migrants on the continent are rural youth, then the majority of the population staying in the origin and affected by outmigration consists of migrant's parents of working age and siblings (usually younger siblings). With agricultural productivity among smallholder farmers still being one of the main targets for policies in many Sub-Saharan African countries (Collier and Dercon, 2014) and family farms being integral for poverty reduction and food security in rural areas (Graeub et al., 2016), one should be wary of any negative changes to the productivity of such farms. Since most smallholder farms rely heavily on family labor (Graeub et al., 2016), outmigration might be detrimental to their productivity. The goal of this study is to determine what actions related to the household farm does the family undertake after the outmigration of a young relative in rural Tanzania.

There are a few reasons why researchers could deem the outmigration of rural youth to be of a lesser importance than outmigration of adults. I state some of these reasons below along with suggesting why the study of outmigration of youth can still provide an important insight on

the productivity of small farms. There is a knowledge gap on the impact of youth outmigration on the livelihood of non-migrant household members in Sub-Saharan Africa, although recently there is an increased interest to the topic (Mueller, Doss, and Quisumbing, 2018). Hence, the impact of youth outmigration compared to adult outmigration is still an important empirical question for future studies.

Classical dual-sector models (Lewis, 1954; Ranis and Fei, 1961) assume a negligible marginal productivity of agricultural labor (regardless of the age group), which results in a labor surplus in the agricultural sector of rural areas. A wage gap then leads to shifts of labor from the rural agricultural sector to the urban manufacturing sector. After some people migrate, those staying in place are able to maintain the same level of output with no (or limited) increase in their own labor supply. At the same time, a decrease in consumption needs due to outmigration inevitably leads to agricultural output being excessive (Lewis, 1954; Ranis and Fei, 1961). Later developments of the two-sector models update the assumption of the insignificance of marginal productivity of labor suggesting it to be positive (Harris and Todaro, 1970; Gollin, 2014). It means that migrants contribute productive labor to the household farm prior to their move, so their outmigration is related to a decrease in both consumption and production.

There is a caveat to the updated theoretical model with significant marginal productivity of agricultural labor, which is specific to youth outmigration. Youth might have lower productivity of labor than adults do, both in farming and off-farm activities. Young adults can go to school, participate in low-paid or unpaid apprenticeship, or have a higher burden of household chores. Consequently, they would have less time to spend on the farm and paid off-farm jobs or self-employment. Also, their labor productivity can be lower because of the lack of knowledge or skills. Hence, youth outmigration might have a smaller effect on the household's livelihood.

Abay et al. (2021) show a slightly smaller average participation rate in agriculture among youth than among adults in Tanzania, although the participation rate is much smaller for younger people who are also more likely to be at school or report no activity. In this study, I find that migrant youth spend significantly less time on the household farm than non-migrant youth and adults. At the same time, their outmigration is associated with an increase in labor contribution of other household members. It suggests that the decrease in consumption needs is not enough to offset the decrease in labor supply, even though labor supply of migrant youth is indeed smaller.

Another reason why conventional wisdom might view outmigration as an unimportant or even a positive event for the household's livelihood is the view of migration as a strategy to diversify risks across space and increase the expected income. This theory holds among the households that receive remittances after sending out migrants to urban areas and other countries for employment purposes (Wouterse and Taylor, 2008). Migration of youth is different in the following aspects. First, most destinations are rural – as it was shown in the first essay. Migration rates to low-density rural areas are comparable to migration rates to all other destination types combined. Hence, the expected earnings for most migrants do not exceed drastically their earnings prospects at the origin. Second, youth might have troubles finding a job in an urban area, as it was shown in the second essay, opting for working mainly in household maintenance or being unemployed and searching for a job. Finally, youth can move for reasons other than employment. They can view migration as the way to transition into adulthood and start their own household. For these reasons, I do not expect to see a significant inflow of remittances in the migrant-sending households (and, unfortunately, there is no way to check for this with the data I use). It leaves the positive effect from a possible decrease in consumption

needs and a negative effect from a decrease in labor supply to the household farm as results of outmigration.⁶⁸

The objective of this paper is to assess the impacts of youth outmigration on the household's livelihood among rural and agricultural households in Tanzania. I look at the following outcomes: changes to the labor supplied by non-migrant household members to the household farm, off-farm activities, and household chores; attraction of new household members; changes to the hired labor; and adjustments to the size of the family farm. Following other studies on the changes to time use among the left-behind population (e.g., Mueller, Doss, and Quisumbing, 2018; Chang, Dong, and MacPhail, 2011; Xu, 2017; and Antman, 2011b), I look at the variations of impact to population groups divided by gender and age. Changes to the hired labor as a response to outmigration have been studied previously, for example, by Mueller, Doss, and Quisumbing (2018), Davis and Lopez-Carr (2014), and Radel, Schmook, and McCandless (2010). The study of the impact of migration on the household formation and dissolution, including the attraction of new household members, has been advanced by Bertoli and Murard (2020). Adjustments to land area under cultivation as a response to outmigration has been studied previously, for example, by Gray and Bilsborrow (2014), Davis and Lopez-Carr (2014), and Chen et al. (2014).

Selection into migration, reverse causality, and the simultaneity of the labor supply decision and the decision to send a young household member into migration complicate the research on the impacts of migration (Adams, 2011). Following other studies, I employ the difference-in-differences strategy (Murard, 2016; Dinkelman and Mariotti, 2016; Antman, 2011a) along with matching techniques (Adams, 2011; Mueller, Doss, and Quisumbing, 2018;

⁶⁸ A household can keep supporting the migrant after outmigration, hence there might be no decrease in consumption needs immediately after outmigration (Lucas, 2016).

Kuhn, Everett, and Silvey, 2011; Démurger and Wang, 2016). I use the 2008/2009 and the 2012/2013 waves of the Living Standards Measurement Study (LSMS) dataset for Tanzania (World Bank, 2017). My paper is one of the few focusing on the impacts of the outmigration of youth in Sub-Saharan Africa, following the work done by Mueller, Doss, and Quisumbing (2018), who look at the outmigration of the household head's children in Ethiopia and Malawi. I confirm that the response to a reduction in labor supply created by outmigration is not uniform across the remaining household members, it differs by gender and age. I complement this analysis by distinguishing new household members' contribution and find that it differs by gender and age too. I also test whether the impact of migration differs by migrant's gender, age, and destination.

I find that the outmigration of youth in rural Tanzania results in a reduction of labor supplied to the household farm. Certain groups of people, namely men of age 15-34 and women of age 35-64, who are likely to be a brother and a mother of the migrant, increase their labor supply in households that experience outmigration. Elderly people in these households are likely to delay exiting agriculture, while female children are more likely to enter agriculture. These households are more likely to attract new household members than households that did not send out a migrant. Interestingly, new members in households that experienced outmigration, on average, spend less time on agricultural activities and off-farm employment; except for women of working age who spend more time on agriculture than new female members in households without young migrants. Households that sent out an older migrant are more likely to increase their use of hired labor and the amount of cultivated land, which could be related to an inflow of remittances.

4.2. Literature Review

4.2.1. The effects of labor withdrawal and remittances on labor outcomes

Studies on the impact of outmigration on the livelihood of household members left behind have conflicting results: some researchers find positive effects while others find negative effects (Ye et al., 2013; Antman 2012b; Murard, 2016). The contexts of the studies matter, but even in a similar setting the observed patterns may differ (Qin and Liao, 2016). An ambiguity in expectations based on the economic theory of migration can explain this discrepancy. On the one hand, outmigration has a negative impact on the household of origin through a decrease in family labor available. The pre-migration labor migrant spent on farming and non-farm activities that brought income to the household and the labor migrant spent on household chores and care for the children and the elderly are not available to the household after outmigration. The effect of outmigration on time use of the family members who stay in the origin is studied, among others, by Murard (2016), Antman (2012a), Ao, Jiang, and Zhao (2016), and Xu (2017). On the other hand, consumption needs may decrease with outmigration while the loss of income can be covered by remittances that the migrant starts to send. Remittances are the main channel of the positive impact of outmigration on the agricultural production (Abebaw et al., 2019). In this subsection, I discuss the impact of labor loss and remittances on the labor allocation of non-migrant household members.

The negative effect of the withdrawal of labor associated with outmigration depends on the productivity of migrants' labor. As discussed earlier, classical models of rural-urban migration and the associated shift from agricultural to a non-agricultural sector assume marginal productivity of agricultural labor to be insignificant and very close to zero (Lewis, 1954; Gollin, 2014). As more people leave rural areas, the productivity of labor in the agricultural sector

increases, and eventually the wage differential between the rural and the urban areas become small enough to significantly limit or stop the migration flow. From the household's perspective, this would imply that outmigration should cause no serious decline in the household's agricultural productivity, and, on the contrary, may even cause an increase in agricultural surplus as the consumption requirements become lower. Todaro (1980) shows how the perception of internal migration as the way to shift labor from a less productive rural agricultural sector to a more productive urban manufacturing sector changed over time and discusses how disruptive labor migration could be for the productivity of labor in rural areas and rural incomes.

Withdrawal of productive labor could force the remaining household members to spend more time working to replace the lost labor and uphold the same level of income or to prevent income from falling significantly. The loss of labor can be associated with a decrease in consumption needs, but, on the other hand, it is possible that the household needed to save for some time in order to be able to send out a migrant (Dustmann and Okatenko, 2014) or that the household needed to provide remittances until the migrant finds a job at destination (Lucas, 2016). Then, the household may work more to return to the previous consumption and savings levels. Overall, holding other things fixed, the withdrawal of labor is usually found to lead to an increase in labor supplied by non-migrant household members (Chang, Dong, and MacPhail, 2011). With the data from Malawi and Ethiopia, Mueller, Doss, and Quisumbing (2018) find that outmigration of a child of household head increases labor supply of the migrant's mother and siblings; the authors also consider changes in hired labor and find a significant increase in Malawi. Murard (2016) also observes an increase in farm labor supply by the members of the household left behind in Mexico, but suggested reason behind the observed effects is labor reallocation rather than an increase in total labor supply. He and Ye (2014) find that left-behind

elderly parents continue to work in agriculture even after reaching physical limits and/or at high age. They also observe that, in many cases, households of origin cannot rely on unpaid help from family networks at the location and must hire labor.

Adams (2011) describes two channels for the effect of remittances on labor supply and labor market participation commonly found in the literature: through an increase in consumption and through an increase in investment. A growth in consumption often follows an inflow of remittance in the households left behind (Kangmennaang, Bezner-Kerr, and Luginaah, 2018; Démurger and Wang, 2016), which can lead to a reduction in labor supply by the household members. Alternatively, some studies suggest that the household raises consumption prior to decreasing labor supply (for example, due to a longer experience of receiving remittances: Justino and Shemyakina, 2012). In addition to increasing consumption, an inflow of remittances allows the household to invest into the means of production (agricultural or non-agricultural) enabling higher productivity of labor, so that now an increase in labor supply would bring more benefits than before (Taylor, 1999; de Brauw and Giles, 2018).

There could be reasons for youth to not be sending remittances for some time after their move. While in my sample I observe more people moving for non-monetary reasons, even those who move for work may not be able to find a good job right away (Filmer and Fox, 2014; Tanle, 2018). Sometimes, youth want to avoid employment paths of their parents by migrating to more urbanized communities striving to exit agriculture and find an off-farm job (Fox and Thomas, 2016). In these destination areas, the fact of getting preferable employment (e.g., formal employment, wage job) as a first job may have a significant effect on future earnings (Bridges et al., 2016). Hence, youth might spend more time trying to find a more desirable job. Moreover, as I look at short-term consequences (at most four years after outmigration), some migrants could

not have settled yet at their destinations. For some people it takes time to find a job at the new location, especially for those moving to places with higher unemployment levels (Roubaud and Torelli, 2013), for example, to urban areas; or for people lacking behavioral skills (Fox, Senbet, and Simbanegavi, 2016). Many young people in Africa find themselves underemployed which hinders their earning potential (Filmer and Fox, 2014). So, even if employment opportunities and expected income are higher at destination (Beegle, De Weerd, and Dercon, 2011), they may not be realized soon after the migrant arrives, which delays migrant's contributions to the income of the household at the origin.

4.2.2. Other outcomes of interest

Attraction of a new household member of working age could be another way for the household to adjust family labor to the loss of migrant's labor. In Mexico, Bertoli and Murard (2020) observe households sending migrants to change structure: many households either attract new members or join a different household at the origin soon after migrant moves. Klasen and Woolard (2008) suggest that changes to the household structure could become a strategy to cope with unemployment. In their analysis of households in South Africa, Klasen and Woolard (2008) find that employment is correlated with the establishment of one's own household (and becoming the household head or the head's spouse). Hence, people who have troubles with sustaining their livelihood are more likely to stay with their parents or join a household of their relatives seeing extended family as a safety net. On the other hand, a household that recently had an adult child move elsewhere may decide to welcome a relative to help with the farm or join the family business.

Another strategy I look at is the change to land under agriculture through land markets. Chen et al. (2014) discuss the impact of rural outmigration on land use and its transformation in

China. They describe a decrease in land use and a positive effect on land conservation in the households left behind. The proposed mechanisms behind these changes are a decrease in the demand for food and fuelwood and an increase in the investment in energy-saving technologies with funds available through remittances. Taylor et al. (2016) study the effects of international migration on the land use at the communities of origin in Guatemala. They show that money, knowledge, and technology transmission from migrants to their families provide opportunity for the transition from agriculture to non-agricultural activities at the origin, which is associated with an increase in conversion of fields into forests.

In their overview study, Ye et al. (2013) look at the literature evaluating the impact of outmigration in China. They find most of the reviewed studies to focus on the impacts on migrants' children, followed by female spouses and elderly parent. Among the outcomes of interest, researchers look at children's educational attainment, physical and mental health of the left-behind family members, and gendered patterns and roles in the household. Antman (2012b) reviews the literature on the impacts of outmigration on migrants' elderly parents, children, and spouse, mostly looking at the effect of international migration and the role of remittances. Most of the review studies on the welfare of migrants' parents look at the time use patterns and health. Antman is also interested in the old-age care that elderly parents receive from all their children and infer that migration decreases the number of hours of care that parents receive. The main channel for such decrease is the disappearance of migrant's time contribution, which suggests that migrants' siblings do not replace in full the time migrants spent attending to the needs of the elderly parents prior to migration.

Other characteristics of the household's livelihood could also change with outmigration, and usually the patterns of change differ by gender and age of the non-migrant household

members. Chang, Dong, and MacPhail (2011) found that internal migration of adult children in China leads to increased hours of farm and domestic labor among elderly parents and children of migrants, especially among women, and doesn't increase off-farm labor. Luis et al. (2015) show how diverse the impacts of rural outmigration on the technology of rice production by the household of origin in the Philippines could be, depending on the type of migration and migrant's gender. Mueller, Doss, and Quisumbing (2018) find that outmigration of adult children from agricultural households leads to increase in labor supplied to the farm by their parents and siblings in Malawi and Ethiopia and to increase in hired labor in Malawi. They also stress that some of the patterns in the observed coping mechanisms are gendered.

Incentives for future migration could also play a very important role in labor and educational choices that the household makes. With the outmigration of youth, migrants' siblings gain additional knowledge about such opportunity, hence they could adjust their educational attainment or work experience to become more suitable for employment in a different area once they reach the age of migration. Shrestha and Palaniswamy (2017) show how a biased demand towards male migrants could negatively affect educational attainment of migrants' female siblings and positively affect education attainment of migrants' male siblings staying at home in Nepal. Dinkelman and Mariotti (2016) demonstrate positive long-run effect of circular international labor migration on the education at the community of origin in Malawi.

4.3. Data and definitions

I use the first wave of the Living Standards Measurement Study for Tanzania (World Bank, 2017), conducted in 2008/2009, to get the baseline characteristics for individuals and households. Then, I use the third wave, conducted in 2012/2013, to define migrant-sending households and compute changes in the outcomes of interest. The second wave of survey

conducted in 2010/2011 is used to check for pre-migration trends. I compare the changes in characteristics from 2008/2009 to 2010/2011 in households without migrants to households that sent out a migrant between 2010/2011 and 2012/2013 (which is a subsample of households that sent out a migrant between 2008/2009 and 2012/2013). No information about the pre-migration trends before 2008/2009 is available. I include into my analysis every household that had at least one member of age 15 to 34 in the first survey wave and at least one member (of any age) staying at the baseline location in the last survey wave. The households are then differentiated based on the outmigration of young adults (people of age 15-34). I define a migrant as an individual for whom the distance between the locations in the first and the last survey wave is at least 5km⁶⁹.

There are 2,258 non-migrant households⁷⁰ which had at least one household member of age 15-34 present in the first survey wave. Out of them, 1,458 households lived in areas defined as “rural” by the National Bureau of Statistics of Tanzania (NBS) – I will call these households rural; and 1,683 households reported spending time on agricultural activities or using land for agricultural activities – I will call these households agricultural.⁷¹ Among rural households, I

⁶⁹ In most cases, I am able to use the distance provided in the dataset. If the distance travelled was below 5 km, it was recorded as zero by the survey team. When the distance is missing, I compute it using the coordinates provided in the dataset and apply the same threshold of 5 km. These coordinates are unique to each enumeration area: households’ coordinates were averaged, and a random offset was applied. For households identified as rural in the sample by the National Bureau of Statistics, the offset ranges from 0 to 5 km; and for 1% of rural households an offset of 0-10 km is applied.

⁷⁰ At least one member of the original households must be present in the origin during the last survey wave for the household to be included into the sample. There are split-off members: members of the original household who are listed as a part of a new household during the last survey wave but did not travel more than five km to consider them migrants. Split-off members are not considered present in the origin, and households consisting only of new members are excluded from the sample. For example, a household that attracted new members during the second survey wave, experienced outmigration of youth, and split-off by the last survey wave leaving only those who joined between the first and the second survey wave in the origin would be excluded from the sample.

⁷¹ Among 2,258 households with youth in the first survey wave and non-migrant household members remaining in place by the last survey waves, 1,416 household were defined as “rural” by the NBS and were involved in agricultural activities; 42 households were defined as “rural” and were not involved in agricultural activities; 267 households were defined as “urban” and were involved in agricultural activities; 533 were defined as “urban” and were not involved in agricultural activities.

observe 305 young adults moving from 255 households. Among agricultural households, I observe 360 young adults moving from 294 households.

Summary statistics for household-level and individual-level outcomes in agricultural households are presented in Table 4.1. Average share of income coming from farming⁷² falls over time, from 61% in 2008/2009 to 52% in 2012/2013. The number of households specializing on agriculture (those with the share of agricultural income over 75%) also falls, from 48% of the households to 37%. Nevertheless, more people shift into farming, and the time household members spend on farming increases. The average number of household members participating in agriculture raises from 2.3 to 2.7 while the total number of days household members spend working raises from 146 to 191. Hence, the average number of working days per person increases by 6, which is also reflected in the individual-level characteristics.

⁷² The data come from the Rural Income Generating Activities database (RIGA), <https://www.fao.org/economic/riga>

Table 4.1. Summary statistics for agricultural households with youth at baseline and non-migrant members during the last survey wave

	Mean	Std. dev.	25 th percentile	Median	75 th percentile
<i>Household-level variables (1,683 observations)</i>					
Number of household members participating in ag., 2008/2009	2.3	1.8			
Number of household members participating in ag., 2012/2013	2.7	1.7			
Days spent on ag. activities by household members, 2008/2009	145.7	180.5	11	91	213
Days spent on ag. activities by household members, 2012/2013	190.6	213.7	53	126	259
1 = Household uses hired labor, 2008/2009	0.42	0.49			
1 = Household uses hired labor, 2012/2013	0.39	0.49			
Hired labor, days; 2008/2009	13.0	30.7	0	0	12
Hired labor, days; 2012/2013	16.3	52.6	0	0	12
Land area under cultivation, acres; 2008/2009	4.7	18.9	1	2.5	5
Land area under cultivation, acres; 2012/2013	4.9	11.4	1	2.5	6
Land area owned, 2008/2009	4.9	19.7	0.8	2.5	5
Land area owned, 2012/2013	5.9	15.4	0.8	3	6.5
Share of income coming from ag. activities (1 = 100%), 2008/2009	0.61	0.39			
Share of income coming from ag. activities (1 = 100%), 2012/2013	0.52	0.39			
1 = Household specializes on agricultural activities, 2008/2009	0.48	0.50			
1 = Household specializes on agricultural activities, 2012/2013	0.37	0.48			
Number of new household members, 2012/2013	1.4	1.7	0	1	2
<i>Individual-level variables for people present in both the first and the last survey waves (8,102 observations)</i>					
Age, 2008/2009	21.5	18.6	7	15	33
1 = Male	0.50	0.50			
1 = Completed primary school, 2008/2009	0.30	0.46			
1 = Married, 2008/2009	0.29	0.46			
1 = Head of the household, 2008/2009	0.19	0.39			
1 = Child of household head, 2008/2009	0.51	0.50			
1 = Spent any time working on the household farm, 2008/2009	0.47	0.50			
1 = Spent any time working on the household farm, 2012/2013	0.51	0.50			
Days spent on ag. activities, 2008/2009	30.4	53.3	0	0	43
Days spent on ag. activities, 2012/2013	36.0	56.9	0	4	55
1 = Main occupation in farming or fishing, 2008/2009	0.37	0.48			
1 = Main occupation in farming or fishing, 2012/2013	0.41	0.49			
1 = Main occupation in non-ag. sector, 2008/2009	0.05	0.22			
1 = Main occupation in non-ag. sector, 2012/2013	0.07	0.26			

Note: A household is considered to be agricultural if any of its members participated in agricultural activities or if the household cultivated any land at baseline. Share of income coming from agriculture and an indicator for specializing on agricultural activities come from the RIGA dataset.

4.4. Empirical strategy

I use the methodology that Mueller, Doss, and Quisumbing (2018) used for Ethiopia and Malawi, adding certain outcome variables, controls, and robustness checks. With the LSMS data for Tanzania, I observe households and their members in two moments in time: the first survey wave was conducted in 2008/2009 and the last survey wave was conducted in 2012/2013. I want to separate the outcomes for two types of households: with and without migrants. To determine if there is any response to outmigration, I apply difference-in-differences technique for the outcomes of interest. In this setting, the time dimension from the classical difference-in-differences setup is the survey wave. The binary variable that distinguishes treatment and control observations is an indicator of having a household member of age 15 to 34 at baseline move away between the survey waves.

The key assumption for the difference-in-differences method, parallel trends of the outcomes in the absence of treatment (migration), could be violated due to selection into migration. If this is the case, households with migrant youth would have had a different change in outcome if no youth had moved compared to the change in outcome that households with no migrant youth experience; and households with no migrant youth would have had a different change in outcome if any youth had moved compared to the change in outcome that the households with migrant youth experience. To account for non-random selection into migration, I apply the difference-in-differences matching approach (with bias-adjusted nearest neighbor matching and propensity score matching) and match household with migrant youth to households with no migrant youth based on their observed characteristics.⁷³ Then, the effect of unobservable

⁷³ I match households based on their size, age and gender of the household head, land area under cultivation, amount of livestock owned by the household, asset index that compares the household's assets (excluding land and livestock) to the assets of other rural (or agricultural) households, indicators for experiencing a negative agricultural and non-agricultural shocks in the past year, population density, distance to the nearest road,

time-invariant characteristics that could shift the trend for households with migrants and that cannot be accounted for in the matching procedure is eliminated by the first differencing.

I estimate the impacts of young adult's out-migration on various household-level and individual-level outcomes. At the household level, I am interested in total household labor supply to the household farm, area under cultivation, and number of household members attracted between the survey waves. Then, the classical difference-in-differences setup could be written as follows:

$$Y_{ht} = \alpha_0 + \alpha_T t + \alpha_M M_h + \alpha_{MT} M_{ht} + \sum_n \alpha_n H_{htn} + \varepsilon_{ht} \quad (3.1)$$

In this equation, Y_{ht} is the outcome variable for household h , where t denotes the wave of survey (0 for the first and 1 for the last wave of survey). M_h is an indicator for having any household member of age 15 to 34 at baseline migrate from the household h by the last survey wave:

$$M_h = \begin{cases} 1, & \text{any youth out – migrated from the household } h \text{ between } t = 0 \text{ and } t = 1, \\ 0, & \text{no youth out – migrated from the household } h \text{ between } t = 0 \text{ and } t = 1. \end{cases}$$

Household-level control variables are denoted H_{htn} (n indicates a set of household-level control variables). In this model, α_T captures the time trend in outcome for households with no migrant youth and α_M captures the difference between households with and without migrant youth at baseline, before migration happened. Then, α_{MT} is the coefficient of interest, the difference-in-differences estimator. It shows the impact of having migrant youth in the household on the outcome of interest between the waves.

and distance for the nearest town with population of at least 50,000 people. For the individual-level specifications, I additionally match non-migrant individuals living in households that experienced outmigration to non-migrant individuals living in households that did not experience outmigration. This matching is based on age, gender, marital status, and education (indicator for the completion of primary school) as well on the household-level characteristics mentioned above.

When I calculate the difference in the outcome variable across time, I get the following equation:

$$\Delta Y_h = Y_{h1} - Y_{h0} = \alpha_T + \alpha_{MT}M_h + \sum_n \alpha_n(H_{h1n} - H_{h0n}) + (\varepsilon_{h1} - \varepsilon_{h0}) \quad (3.2)$$

All time-invariant variables then will be excluded from the regression, since for them $H_{h1*} = H_{h0*}$, and for all variables with linear time trend the coefficient, α_* , would be captured by the constant term in the regression since $H_{h1*} - H_{h0*} = H_{h'1*} - H_{h'0*}$.

To increase the precision of my estimates, I can additionally control for these baseline characteristics that were excluded from the regression due to first differencing. Then, the final version of my model is:

$$\begin{aligned} \Delta Y_h = Y_{h1} - Y_{h0} = \\ = \alpha'_T + \alpha_{MT}M_h + \sum_n \alpha_n(H_{h1n} - H_{h0n}) + \sum_q \alpha_q H_{h0q} + \\ + (\varepsilon_{h1} - \varepsilon_{h0}) \end{aligned} \quad (3.3)$$

Depending on the outcome variable, I include the following controls: gendered composition of the household, area under cultivation, number of livestock owned, asset index built based on the household's assets except for land and livestock.

For the individual-level outcomes, I look at labor allocation towards farm and non-farm activities. Hence, the difference-in-differences model could be rewritten as:

$$Y_{iht} = \beta_0 + \beta_T t + \beta_M M_h + \beta_{MT} M_h t + \sum_m \beta_m X_{ihm} + \sum_n \beta_n H_{htn} + \delta_{iht} \quad (3.4)$$

In contrast to the household-level model, I replace the outcome with an individual-level variable for individual i from household h , Y_{iht} , and add individual-level control variables, X_{ihm}

(m indicates a set of individual-level control variables). In this specification, the coefficient of interest is β_{MT} .

Again, I calculate the difference in the outcome variable across time and bring back control variables measured at baseline that were excluded by first differencing:

$$\begin{aligned}
 \Delta Y_{ih} &= Y_{ih1} - Y_{ih0} = \\
 &= \beta'_T + \beta_{MT} M_h + \sum_m \beta_m (X_{ih1m} - X_{ih0m}) + \\
 &+ \sum_n \beta_n (H_{h1n} - H_{h0n}) + \sum_p \beta_p X_{ih0p} + \\
 &+ \sum_q \beta_q H_{h0q} + (\delta_{ih1} - \delta_{ih0})
 \end{aligned} \tag{3.5}$$

In this specification, I control for the following characteristics captured at baseline: age, gender, indicator for the completion of primary school, marital status, relationship to the household head (indicators for being a household head, a spouse, and a child of household head). For both household-level and individual-level outcomes, the matching procedure I use is based on household-level characteristics. I match every household where any young household member migrated between the survey waves to a household where no youth migrated. I compare models' fit by information criterion as a robustness check (Cattaneo et al., 2013). I perform bias adjustment to correct bias coming from matching on more than one continuous variable (Abadie et al., 2004).

4.5. Results

4.5.1. Migrant youth at baseline

I start with the analysis of youth' characteristics and activities at baseline aiming to understand the contribution migrants made to their households' livelihood prior to migration. Table 4.12 in Appendix 2 contains basic individual and household characteristics of youth and compares migrants to non-migrants. Migrants are on average younger and less likely to be married, there are more women among them. The share of people who were away from the household for at least a month is 7% higher among migrant youth. There are more children and grandchildren of the household head among migrants and fewer household heads or spouses of the household head. Hence, the sample of the household members left behind after the outmigration of youth mostly consists of migrants' parents and siblings. Migrant-sending households are wealthier, living in more densely populated areas, and larger, with more children of the household head living in the household.

The activities of migrant and non-migrant youth are summarized in Table 4.2. The share of people with self-reported main occupation in farming or fishing is 15% lower among migrants. Moreover, the share of people participating in any agricultural activity (based on the time spent) is lower among migrants. Among people who report spending some time on agriculture in the past year, migrants on average spend 10.8 days less which equals to a 17.6% difference from the time non-migrant youth spend on agriculture. In addition to farming, migrants are less likely to have participated in any non-agricultural wage work or self-employment. At the same time, the share of people with main occupation in studies is 11% higher among migrants, and they are more likely to be at school during the time of the survey or the year prior.

Table 4.2. Activities of people of age 15 to 34 who lived in rural areas in 2008/2009 (NBS definition)

	Non-migrants	Migrants	Migrants - Non-migrants	Std. error
1 = Main occupation is farming or fishing	0.70	0.55	-0.15***	(0.03)
1 = Main occupation is wage job	0.01	0.02	0.00	(0.01)
1 = Main occupation is self-employment	0.03	0.03	0.00	(0.01)
1 = Main occupation is studies	0.21	0.32	0.11***	(0.02)
1 = Main occupation is household maintenance	0.03	0.07	0.03***	(0.01)
1 = Main occupation is unemployment or disability	0.01	0.01	-0.00	(0.01)
1 = Spent any days performing agricultural activities in the past year	0.80	0.67	-0.13***	(0.02)
Days spent on land preparation and planting, past year	19.09	13.33	-5.76***	(1.39)
Days spent on weeding, past year	17.18	11.58	-5.61***	(1.22)
Days spent on harvesting, past year	12.75	8.92	-3.83***	(1.15)
Total number of days spent on agricultural activities, past year	49.02	33.82	-15.20***	(3.27)
Total number of days spent on agricultural activities, past year; among those who spent any	61.20	50.42	-10.78***	(4.14)
1 = Spent any time on agriculture in the past week	0.63	0.55	-0.08***	(0.03)
Hours spent on household agricultural activities, past week	17.15	13.41	-3.74***	(1.11)
Hours spent as an unpaid family worker on a non-farm household business, past week	17.73	16.61	-1.13	(1.02)
Hours spent collecting firewood or water, yesterday	0.76	0.69	-0.07	(0.09)
1 = Currently attending school	0.20	0.29	0.10***	(0.02)
1 = Was in school last year (if not attending currently)	0.05	0.10	0.05***	(0.01)
1 = Did any work for pay, profit, barter, or home use in the past week	0.59	0.47	-0.12***	(0.03)
1 = Have work or own farm or enterprise to return to (if didn't work in the past week)	0.16	0.17	0.01	(0.02)
1 = Did any wage work, past week	0.11	0.09	-0.02	(0.02)
1 = Did any wage work, past year	0.16	0.12	-0.04*	(0.02)
Hours worked at wage job, past week	3.62	3.17	-0.45	(0.77)
1 = Did non-agricultural self-employed activity, past week	0.11	0.05	-0.06***	(0.02)
1 = Did non-agricultural self-employed activity, past year	0.14	0.08	-0.06***	(0.02)
Months operating a business, past year	0.88	0.46	-0.41***	(0.16)

Note: Sample weights from 2008/2009 are applied. t-test for difference in means: *** 0.01; ** 0.05; * 0.1.

Disparities between migrants and non-migrant with the same type of main occupation, for three most frequent types, are presented in Table 4.13. Migrants with main occupation in farming on average spend less time farming and are more likely to be at school, while migrants with main occupation in household maintenance are more likely to have a wage job. At the same time, there are almost no significant differences in activities of migrant and non-migrant students. In Table 4.14, I differentiate migrant and non-migrant youth by gender and age, and additionally split migrants into groups by their destination type.⁷⁴ Older migrants are less likely to have main occupation in farming, on average spend less time on the household farm, and are more likely to attend school than non-migrants of the same age, while younger migrants spend more time on wage work. Although urban-destined migrants are less likely to have main occupation in farming and more likely to be students than rural-destined migrants, the time spent on the household farm is somewhat comparable between migrants to different destinations.

4.5.2. Descriptive results

In Table 4.3, I compare means of household-level outcomes for households that experienced and did not experience outmigration of youth. At baseline (2008/2009), households that would send out a migrant are on average larger, with more people working on the family farm, and larger farms (both in terms of cultivated land and owned land). Consequently, these households have higher average total family labor spent on agriculture. But the average number

⁷⁴ In this chapter, I distinguish rural and urban destinations and use the NBS definition of “rural”. If I was to use a more elaborate categorization of destination types employed in the previous two chapters, the number of observations in smaller groups would not be sufficient for estimation. For a household to be included into the sample for this chapter, there must be at least one non-migrant member present at the location of the origin during the last wave of survey. This leads to a decrease in the number of observations of migrants compared to the previous two chapters.

As shown by Bertoli and Murard (2020), outmigration can be associated with household dissolution, when the remaining household members join another household, as well as with an invitation of new household members to the original household. In this study, I focus on the non-migrant household members in the household of origin and the new members who join this household, while household dissolution is beyond the scope of this work.

of days spent on agriculture per worker in these households is actually lower. Interestingly, if I exclude migrants of any age from the analysis of total family agricultural labor, I see little difference between households with and without migrant youth. In particular, in both types of households, the average number of people participating in agriculture is 2.6, and the average total number of days spent on the family farm is 170.4 for households that will not send out a young migrant and 161.0 for households that will send out a young migrant.

Table 4.3. Descriptive results: household-level outcomes in rural households (according to the NBS definition)

	Households that did not experience outmigration of youth		Households that experienced outmigration of youth	
	2008/ 2009	2012/ 2013	2008/ 2009	2012/ 2013
Number of household members	5.66	6.17	7.37	6.49
Number of household members performing agricultural tasks	2.64	2.83	3.53	3.00
Total family labor spent on agriculture, days	181.20	200.16	214.55	215.54
1 = Use hired labor	0.44	0.40	0.46	0.44
Hired agricultural labor, days	15.36	15.68	13.99	21.95
Land under cultivation, acres	5.00	4.44	7.96	7.16
Owned land, acres	5.29	5.44	8.14	7.67
Share of income coming from agriculture	67.86%	58.25%	67.17%	55.75%
1 = Specialize on farming	0.55	0.43	0.51	0.39
1 = Have a new member		0.67		0.61
Number of new household members		1.37		1.82
Number of new household members of age below 5		0.83		0.83
Number of new household members of age above 5		0.55		0.99
Number of new household members of age 15-64		0.34		0.61
Number of new household members working on the family farm		0.24		0.42
Total labor spent on agriculture by new members, days		10.16		20.92
Number of observations	1,203		255	

Note: Sample weights are applied - except when count the number of household members.

Over time, the gap in the household size and the number of household members working on the family farm between households that did and did not experience outmigration of youth becomes lower, as well as the consequential gap in the total family agricultural labor. But while the average number of days the household members spent on the family farm increase by 3% in households that did not send out a migrant, in households with migrants this number increased by 18%. By 2012/2013, a gap appeared in the use of hired labor: households that experienced outmigration of youth start using more hired labor than households that did not experience outmigration, both in terms of frequency (number of households using it) and quantity (number of workdays). The gap in the land cultivated by the household increased slightly.⁷⁵ At the same time, the gap in the owned land decreased as households without migrants acquired some land between the surveys. These differences are mainly driven by households with larger farms: the scale of changes is much smaller among households cultivating less than 10 acres of land. Interestingly, these changes to the agricultural inputs, labor and land, are not reflected in the structure of household income. I observe a similar decrease in the share of income coming from agriculture and the share of households specializing on agriculture regardless of migration status of the households' youth.

Over two thirds of households in the sample attracted new household members by the last survey wave. Although households that experienced outmigration of youth are on average slightly less likely to invite any new members to the household, the average number of new members in these households is higher. This difference does not come from the number of

⁷⁵ This explains the rising gap in the total family labor spent on the household farm per acre of cultivated land. In households that did not experience outmigration of youth, labor per acre increased from 64.7 days to 75.6 days, whereas in households that experienced outmigration, labor per acre declined from 64.9 days to 57.6 days. Among households using hired labor, the quantity of hired labor per acre of cultivated land in households without migrants decreased from 9.2 days to 8.9 days, and in households with migrants it increased from 6.7 to 7.8 days.

newborns in the household, which is the same regardless of outmigration of youth, but from the number of new members of age 15-64. The number of new members working on the family farm is almost twice as high for households that experienced outmigration of youth, same as the total amount of labor days supplied by the new household members. It suggests that the average time spent by an individual new member is similar between the two types of households.

In Table 4.4, I present an overview of individual labor supply to the family farm, categorized by gender, age, and presence in the household at baseline. I distinguish four age groups: children under the age of 15, people from age 15 to 34, from age 35 to 64, and over the age of 64.⁷⁶ At baseline, men in the households that will send out a migrant on average participate in agriculture less than in households that will not experience youth outmigration, while women participate more – except for women of age 15-34. The average time spent on the household farm at baseline is lower in migrant-sending households for all groups except for girl below the age of 15 and men of age 15-34.

⁷⁶ The age I look at for people who were present in the household at baseline was reported in 2008/2009, so for comparison I revert the age of new members, which they reported in 2012/2013, to their age in 2008/2009. Hence, for both the old and the new household members, their age at the time of the survey in 2012/2013 was 4 years higher than the brackets I report for 2008/2009. Therefore, time spent on agriculture in 2012/2013 is summarized for age group of 4-18, 19-38, 39-68, and above 68. This distinction is important for younger people: average time spent on agriculture increases with age up to a certain age, peaks at 49-53 years of age, and then declines as age increases (see Figure 4.3).

Table 4.4. Descriptive results: individual-level outcomes, by gender, age, and presence in the household; rural households (NBS definition)

	Present in both 2008/2009 and 2012/2013					Present only in 2012/2013		
	Frequency	Share of people who spent any time on agriculture, 2008/2009	Days spent on agricultural activities - among those who participated, 2008/2009	Share of people who spent any time on agriculture, 2012/2013	Days spent on agricultural activities - among those who participated, 2012/2013	Frequency	Share of people who spent any time on agriculture, 2012/2013	Days spent on agricultural activities - among those who participated, 2012/2013
<i>A. Households that did not experience outmigration of youth</i>								
Girls of age below 15	25.5%	12.0%	25.6	27.8%	50.2	31.1%	21.8%	33.2
Women of age 15-34	15.5%	84.5%	64.0	85.0%	72.9	18.7%	66.9%	90.4
Women of age 35-64	8.4%	92.5%	87.9	89.4%	90.5	3.0%	73.2%	70.2
Women of age 65 and above	1.4%	68.9%	64.5	63.2%	61.9	1.4%	20.1%	110.7
Boys of age below 15	24.9%	15.4%	29.4	32.2%	44.6	28.9%	26.8%	49.4
Men of age 15-34	14.8%	79.8%	59.4	81.9%	70.1	11.3%	59.4%	61.6
Men of age 35-64	8.2%	91.5%	81.9	92.7%	83.3	4.1%	80.3%	59.2
Men of age 65 and above	1.3%	80.6%	91.8	70.5%	78.6	1.3%	79.7%	107.7
<i>B. Households that experienced outmigration of youth</i>								
Girls of age below 15	24.3%	16.8%	32.6	33.3%	43.9	32.3%	26.6%	22.1
Women of age 15-34	8.1%	72.7%	58.8	74.8%	88.5	25.0%	65.0%	78.9
Women of age 35-64	15.2%	94.5%	73.9	91.6%	93.9	3.0%	90.3%	124.0
Women of age 65 and above	1.7%	57.7%	95.3	70.0%	103.0	1.8%	18.3%	144.0
Boys of age below 15	23.7%	14.7%	24.3	37.2%	52.6	27.9%	21.6%	42.0
Men of age 15-34	12.2%	74.9%	59.2	76.8%	63.9	7.2%	57.7%	72.4
Men of age 35-64	12.8%	89.8%	76.7	83.7%	79.1	2.6%	87.2%	90.7
Men of age 65 and above	2.1%	92.1%	59.2	72.8%	74.5	0.3%	0.0%	-

Note: Age groups are based on individuals' age in 2008/2009, hence newborns are excluded from the sample of people present only in 2012/2013. Sample weights are applied.

By the last survey wave, women in households that experienced outmigration drastically increase the time they spend on agricultural activities: the average number of days for women of age 15-34 increases by 30 (by 9 in non-migrant-sending households), and for women of age 35-64 it increases by 20 (by 3 in non-migrant-sending households). Among children, I see a dramatic increase in the participation in agriculture and the time spent on it, especially for boys. A 28-days increase is observed for boys in households that experienced outmigration, while in household that did not experience outmigration there is a 15-days increase. Girls in households that experienced outmigration spent more time on agriculture at baseline, but over time an increase in their labor supply is smaller than in households that did not send out a migrant.

Households that experienced outmigration of youth attract more women of age 15-34 and less men, compared to households that did not experience outmigration of youth. These women are less likely to supply any labor to the household farm, and they supply less labor when they do, than women of the same age who were present in the household in the first survey wave and women in households without migrants. Conversely, women of age 35-64 and men of age 15-64 who join the households that experienced outmigration supply more labor to the household farm than those who join the households which did not experience outmigration. Children who recently joined the households with migrants on average spend less time on agricultural activities, and boys are also less likely to participate in agriculture.

I also must note that the average age within certain age groups between people in households that did and did not experience outmigration is not balanced. As I described before, the average time spent on agriculture increases with age for people younger than 54, and the bulk of the growth happens during youth (see Figure 4.3). The average time spent on agricultural activities among people of age 15 is 21 days, while people of age 35 on average spend 69 days

farming.⁷⁷ Hence, an additional year of age increases the average time spent on agriculture by 2.3 days for people of age from 15 to 35. In households with migrant youth, both men and women in the age range from 15 to 34 years old in 2008/2009 and men and women of this age joining the household by the last survey wave are on average 0.9-1.9 years younger than people from this age group in household without migrant youth. Therefore, I can conclude that the observed positive correlation of youth outmigration with the time spent on agriculture will be even higher among young people once age is controlled for.

Additionally, I analyze changes in the self-reported main occupation categorized into four groups: farming and fishing, wage job and self-employment, studies, and other occupations which include household maintenance, unemployment, and disability. The results are presented in Table 4.15 in Appendix 2. At baseline, there are some gaps – mostly among women – between households that will and will not send out a migrant. In households that will experience outmigration of youth, the share of children with main occupation in farming is lower: boys are more likely to have main occupation in farming, while girls are more likely to have main occupation in the other category. Women of age 15-34 are less likely to have main occupation in farming but are more likely to be in school or have main occupation in the other category. Women of age 35-64 are more likely to have main occupation in non-agricultural wage job or self-employment. Women of age 65 and older are less likely to have main occupation in farming and are more likely to have main occupation in the other category.

Over time, the described gaps become narrower. In particular, in households that experienced youth outmigration, children are more likely to stay in school, while in other households children are more actively shifting into the other category. Older women in

⁷⁷ I do not observe a similar pattern for participation in agriculture: 61% of people of age 15 participate in some agricultural activity, and the bulk of the growth of participation happens among children.

households with migrants are less likely to exit agriculture. In households that experienced outmigration of youth, women of age 15-34 are more likely to shift into farming than in other households, but the share of people with main occupation in farming among them is still lower than among women from households without migrants. A new gap forms among men of age 15-64: in households with migrants, they are more likely to exit farming and shift into non-agricultural wage job or self-employment.

There are some interesting patterns in the main occupation of new household members. Children who join households that experienced youth outmigration are less likely to have main occupation in farming: girls are more likely to be at school, while boys are more likely to have main occupation in the other category. Young women are less likely to have main occupation in farming, wage job, or self-employment but more likely to be students or have main occupation in the other category. Young men are also less likely to have main occupation in farming, but they are more likely to have main occupation in non-agricultural wage job or self-employment. On the contrary, women of age 35-64 are more likely to have main occupation in farming and are less likely to have other type of main occupation.

4.5.3. Main results

I begin with analyzing the impact of youth outmigration on the household-level outcomes and focus on the results for agricultural households. The results of a simple ordinary least squares (OLS) regressions and the nearest neighbor matching (NNM) estimation are presented in Table 4.5 and Table 4.6 respectively. For comparison, the results of OLS, NNM, and propensity score matching (PSM) estimation for the sample of rural households are presented in Table 4.16, Table 4.17, and Table 4.18 respectively. I find youth outmigration to have a negative effect on the change to the family agricultural labor, some positive effect on the change to the amount of

cultivated land, and a positive effect on the number of new household members working on the household farm. For agricultural outcomes, the impact of the outmigration to rural areas and the outmigration of older young adults stands out. I confirm the observation made with simple differences that there is no significant difference in the share of income coming from agriculture.

The outcome of interest is the change in a certain household characteristic, a difference between the level observed during the last survey wave and the baseline. The first characteristic I look at is the change to the number of household members who supply any amount of labor to the household farm. Over time, household composition changes: people can form a split-off household nearby, migrate⁷⁸, stop or start participating in agriculture, or die. Without these processes, the number of people participating in farming would not change in households that did not experience outmigration, hence the change in that number would be zero. In households that did experience outmigration, on the other hand, the change would be negative one if the migrant participated in farming prior to migration. With controls and matching, I try to separate the effect of outmigration from the impacts of the processes described above. Hence, the impact of outmigration should be negative one assuming one migrant moved out and the migrant participated in farming. As shown before, less than 70% of migrant youth supplied some labor to the farm at baseline. On the other hand, among 294 agricultural households that experienced outmigration of youth, 240 had only one migrant. A simple calculation⁷⁹ shows a smaller impact of outmigration.

⁷⁸ Split-off households are differentiated from migrants in the sample. If a split-off household's new location is within five kilometers of the origin, then the members of this household are not considered migrants. If a split-off household's new location is further than five kilometers away, then the members of this household are considered migrants.

⁷⁹ For agricultural households, the number of migrant youth times the share of migrants participating in agriculture divided by the number of households that experienced outmigration equals 0.79 (0.81 for rural households). So, on average, a household that experience outmigration has an 80% chance to lose an agricultural worker to outmigration.

Table 4.5. OLS regressions, household-level outcomes in agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
Change in the number of household members who supply labor to the household's farm	-0.67*** (0.11)	-0.65*** (0.13)	-0.81*** (0.16)	-0.58*** (0.14)	-0.67*** (0.17)	-0.55*** (0.14)	-0.79*** (0.16)
Change in the total family labor supplied to the household's farm, days	-22.72 (13.88)	-15.78 (16.02)	-36.39* (19.98)	-3.90 (16.76)	-45.54** (20.32)	-23.60 (16.51)	-6.51 (19.90)
Change in the indicator of using any hired labor	0.01 (0.04)	-0.00 (0.04)	0.03 (0.06)	-0.05 (0.05)	0.09 (0.06)	0.02 (0.05)	-0.00 (0.06)
Change in the hired labor, days	1.68 (3.63)	3.72 (4.19)	-1.97 (5.23)	2.84 (4.38)	-4.27 (5.32)	0.90 (4.32)	8.25 (5.20)
Change in the land under cultivation, acres	1.24* (0.68)	1.58** (0.78)	1.73* (0.97)	1.86** (0.82)	0.02 (0.99)	0.37 (0.81)	2.39** (0.97)
Change in the land owned, acres	0.85 (0.66)	0.67 (0.77)	1.41 (0.95)	0.97 (0.80)	0.69 (0.97)	-0.25 (0.79)	1.60* (0.95)
Change in the share of income coming from agriculture, decimal	-0.01 (0.03)	0.01 (0.03)	-0.04 (0.04)	-0.02 (0.03)	0.00 (0.04)	-0.02 (0.03)	0.01 (0.04)
Change in the indicator of specializing on farming	0.02 (0.04)	0.00 (0.05)	0.04 (0.06)	-0.01 (0.05)	0.06 (0.06)	0.02 (0.05)	0.02 (0.06)
1 = Have any new household members	-0.02 (0.03)	-0.03 (0.04)	-0.01 (0.04)	0.01 (0.04)	-0.06 (0.04)	-0.05 (0.04)	0.03 (0.04)
Number of newborns	-0.20*** (0.06)	-0.19*** (0.07)	-0.34*** (0.08)	-0.10 (0.07)	-0.31*** (0.08)	-0.16** (0.07)	-0.29*** (0.08)
Number of new household members, excluding newborns	0.21*** (0.06)	0.22*** (0.07)	0.14 (0.09)	0.19** (0.08)	0.24** (0.09)	0.06 (0.08)	0.41*** (0.09)
Number of new household members who supply labor to the household's farm	0.10*** (0.04)	0.09** (0.04)	0.08 (0.05)	0.09** (0.04)	0.11** (0.05)	0.06 (0.04)	0.13** (0.05)
Total amount of labor supplied to the household's farm by new members, days	5.40* (2.88)	6.95** (3.32)	4.89 (4.17)	9.54*** (3.48)	0.30 (4.21)	2.59 (3.40)	10.73** (4.19)

Note: Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Table 4.6. Nearest neighbor matching, household-level outcomes in agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
Change in the number of household members who supply labor to the household's farm	-0.79*** (0.18)	-0.85*** (0.20)	-0.93*** (0.26)	-0.77*** (0.20)	-0.77*** (0.27)	-0.57*** (0.21)	-1.24*** (0.25)
Change in the total family labor supplied to the household's farm, days	-44.90** (20.05)	-46.28** (22.19)	-55.15** (25.09)	-25.90 (22.90)	-54.88* (28.20)	-58.84** (25.34)	-16.81 (26.38)
Change in the indicator of using any hired labor	0.01 (0.05)	-0.00 (0.06)	-0.00 (0.07)	-0.03 (0.06)	0.08 (0.07)	-0.01 (0.06)	0.03 (0.07)
Change in the hired labor, days	-1.56 (6.54)	-3.22 (8.64)	1.97 (6.51)	6.49 (6.27)	-20.07 (13.22)	-7.84 (8.94)	6.86 (10.87)
Change in the land under cultivation, acres	0.71 (1.31)	0.82 (1.79)	-0.51 (1.96)	1.20 (1.77)	-0.33 (0.91)	-1.22 (2.05)	-0.20 (2.04)
Change in the land owned, acres	0.84 (1.27)	0.55 (1.72)	0.72 (1.94)	0.89 (1.68)	0.62 (1.04)	-1.08 (1.97)	-0.12 (1.97)
Change in the share of income coming from agriculture, decimal	0.02 (0.03)	0.02 (0.04)	-0.02 (0.05)	-0.01 (0.04)	0.06 (0.05)	-0.01 (0.04)	0.06 (0.05)
Change in the indicator of specializing on farming	0.05 (0.05)	0.03 (0.06)	0.07 (0.08)	0.01 (0.07)	0.10 (0.07)	0.01 (0.06)	0.08 (0.08)
1 = Have any new household members	-0.02 (0.04)	-0.02 (0.05)	-0.02 (0.06)	0.05 (0.05)	-0.10 (0.07)	-0.02 (0.05)	0.00 (0.05)
Number of newborns	-0.07 (0.09)	-0.08 (0.11)	-0.16 (0.11)	0.09 (0.12)	-0.31*** (0.11)	0.02 (0.11)	-0.26** (0.12)
Number of new household members, excluding newborns	0.37*** (0.10)	0.41*** (0.13)	0.18 (0.14)	0.44*** (0.13)	0.33** (0.14)	0.13 (0.12)	0.73*** (0.18)
Number of new household members who supply labor to the household's farm	0.13** (0.06)	0.10 (0.07)	0.05 (0.09)	0.16** (0.07)	0.12 (0.08)	0.06 (0.06)	0.20* (0.10)
Total amount of labor supplied to the household's farm by new members, days	9.98** (4.74)	8.51 (6.81)	2.71 (7.25)	13.98** (6.48)	3.53 (5.07)	0.93 (5.34)	19.25* (10.02)

Note: Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Indeed, all estimations point to the absolute value of the outmigration impact to be at most 0.79, not one. For households that experienced outmigration of youth, the change in the number of people supplying labor to the household farm is smaller by 0.79 than the change in households without migrants. At the same time, outmigration significantly increases the chances for the household to invite new members to work on the farm: the estimates for the number range from 0.09 to 0.16 and are significant. This gap indicates changes to the labor supply of the members of the original household not captured by the household-level outcomes. It will be explored further with the individual-level outcomes.

Next, I look at the impact on the change to the total amount of labor supplied by household members to the farm.⁸⁰ NNM estimates this change to be 45 days smaller in households that experienced outmigration, which is roughly 28% of the baseline labor supply. The estimated loss is larger in households that experienced outmigration of men, those who moved to an urban area, and migrants of age 15-20. PSM estimates show the largest effect: 60-100 days less in households with migrants. OLS estimates are only significant for the outmigration of men and younger people. One of the mechanisms behind such a big decline in labor dedicated towards farming can be the need to support the migrant after migration. Almost half of young women move for marriage, and the household can expect the other party to partly contribute to the migrant's welfare. People moving to urban areas and younger migrants might need extra funds right after their move as a safety net in case of unemployment. This can make the household members shift from farming to off-farm activities. Since the change in the number

⁸⁰ I tried various specifications considering the amount of land the household cultivated at baseline, as the results could depend on farm size. In the estimations with the changes in family labor and hired labor, I replaced the dependent variables with the changes in the amount of family and hired labor per acre of cultivated land. In the estimations with the changes in the amount of cultivated land, owned land, and hired labor, I ran additional regressions for the subsample of households with under 10 acres of cultivated land at baseline (around 90% of the sample). As discussed in section 4.5.2, some of the observed results might be driven by larger farms.

of people participating in agriculture is smaller in households with migrants, while the total amount of labor doesn't change in some types of households, in these households those who did not stop farming started to work more.

The results for the impact on the hired labor are not significant, neither for the fact of using hired labor nor for the amount of labor used. I additionally estimate the impact on the amount of hired labor used per acre of cultivated land, with two specifications: for all farms and for farms under 10 acres (not presented in the tables). OLS results are not significant, while PSM and NNM results show a positive impact of outmigration on the change in hired labor. On smaller farms, the change in hired labor used per acre is 5 days higher in households that experienced outmigration of youth (31% of the baseline level). When I differentiate by migrant's gender, destination, and age, this result holds only for households with older migrants. A similar pattern holds for the land under cultivation and owned land: although the results are not consistent across models and not always significant, they are often significant for households that experienced outmigration of older youth. The estimates range from 0.9 to 1.6 acres for the change in land cultivated or owned (16% of the baseline level). One of the possible mechanisms behind this observation is the opportunity for the household to invest into farm inputs like land and hired labor when a migrant sends remittances.⁸¹

Although youth outmigration is shown to affect household composition, labor supply to the household farm, and – in some cases – the size of the farm, I see no significant impact to the change in the share of income coming from agriculture and the probability to specialize in farming. Even in households that sent out a male migrant, a younger migrant, a migrant moving to an urban area, which were shown to decrease the total family labor supply to the farm, the

⁸¹ Older migrants are more likely to send back remittances (see Appendix 1).

structure of income follows the same pattern as in households that did not send out any migrants. Since there is no disturbance to their livelihood, household members were able to offset a decrease in labor supply associated with outmigration.

The results for the number of new household members confirm the observations made using descriptive statistics: outmigration of youth does not have any significant effect on the probability to attract new household members, but it positively affects the number of new members attracted. There is no difference in the average number of newborns, yet, after I account for selection into migrant-sending households, outmigration is shown to have a negative impact. It can be explained by natural reasons as migrants are of child-bearing age. There is a significant positive impact of outmigration on the number of new household members of working age. This effect remains significant for the outmigration of women and migrants of age above 20. The number of new members who supply labor to the household farm is also positively affected by outmigration, as well as the total amount of labor supplied solely by new members. In households that experienced the outmigration of youth, new members work on the farm for 5-10 more days, which is equivalent to 54-77 more days of work per an active worker. The estimate is higher for households with rural-destined and older migrants, which can reflect higher needs in offsetting the loss of labor, occurring either due to specialization on agriculture or higher initial labor supply by the migrant.

For the individual-level outcomes, I look at the probability to stop working on the household farm, probability to start working on the household farm, and the number of hours spent working on the farm. In Table 4.7, I present the results for the impact of outmigration on the indicator to have worked any number of days in the first survey wave and no days in the last survey wave. Outmigration of youth to urban areas has a positive impact on the probability to

stop working in agriculture among women of age 15-64 and male children staying in the origin. Outmigration of people of age 15-20 has a significant negative effect on the probability to stop working on the household farm for women of age 65 and above. In Table 4.19 in Appendix 2, I present the results for the impact on the indicator to not have worked on the farm in the first survey wave and to have worked for any number of days during the last survey wave. Outmigration of youth negatively affects the probability to begin working in agriculture among women of age 15-34 and men of age 35-64.

The results for the impact on the change to the total number of days the household member spent on the farm are presented in Table 4.20 for the full sample and in Table 4.8 for the subsample of people who participated in agriculture in both survey waves. I find that people in households that experienced outmigration of women, youth who moved to rural areas, and youth of age 21-34 have a higher change to the number of days they spend on the household farm by 10-12 days (the NNM estimate for the outmigration of older young adults is 27 days). This result is consistent with the theory of migration-induced labor shortage, as youth from these groups contributed more labor to the household farm prior to migration. After disaggregating by age group and gender, I conclude that this result comes from the increase in working days of women of age 35-64 and men of age 65 and above. The comparison of the full sample and the sample of farmers, in addition to the results from Table 4.7, shows that women of age 15-34 exited farming while those who stayed employed in agriculture did not change significantly the number of working days compared to women in households that did not experience outmigration of youth.

Table 4.7. Dependent variable: indicator for working any number of days on the household farm in the first survey wave and not working in the last survey wave (probability to stop working in agriculture); agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
<i>A. Logistic regressions</i>							
All household members	0.01 (0.02)	0.01 (0.02)	0.03 (0.02)	-0.02 (0.02)	0.05** (0.02)	0.01 (0.02)	0.00 (0.03)
Girls of age below 15	-0.05 (0.10)	-0.04 (0.11)	0.07 (0.20)	0.06 (0.12)	-0.48 (0.30)	-0.04 (0.12)	-0.04 (0.14)
Women of age 15-34	0.04 (0.04)	0.04 (0.05)	0.09 (0.06)	-0.05 (0.05)	0.19*** (0.06)	0.07 (0.05)	-0.01 (0.07)
Women of age 35-64	0.02 (0.03)	0.02 (0.03)	0.04 (0.04)	-0.02 (0.04)	0.08** (0.03)	0.01 (0.03)	0.04 (0.04)
Women of age 65 and above	-0.07 (0.10)	-0.00 (0.11)	-0.31 (0.19)	-0.00 (0.11)	- (-)	-0.36** (0.16)	0.14 (0.14)
Boys of age below 15	0.24 (0.19)	0.39 (0.24)	0.23 (0.30)	0.26 (0.25)	0.46* (0.25)	0.17 (0.21)	- (-)
Men of age 15-34	-0.06 (0.04)	-0.07 (0.05)	-0.03 (0.06)	-0.07 (0.05)	-0.04 (0.06)	-0.02 (0.04)	-0.17** (0.07)
Men of age 35-64	0.01 (0.03)	0.01 (0.03)	0.06 (0.04)	0.00 (0.04)	0.04 (0.04)	0.01 (0.03)	-0.01 (0.05)
Men of age 65 and above	-0.02 (0.08)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)
<i>B. Nearest neighbor matching</i>							
All household members	0.02 (0.02)	0.04 (0.03)	0.04 (0.04)	0.00 (0.03)	0.08** (0.04)	0.02 (0.03)	0.05 (0.04)
Girls of age below 15	0.06 (0.17)	0.06 (0.16)	0.40 (0.36)	0.08 (0.13)	-0.40 (0.36)	0.12 (0.17)	0.14 (0.24)
Women of age 15-34	0.11 (0.07)	0.14* (0.08)	0.14 (0.12)	-0.07 (0.08)	0.42*** (0.11)	0.16* (0.08)	0.06 (0.12)
Women of age 35-64	0.08*** (0.03)	0.08** (0.03)	0.11* (0.06)	0.04 (0.04)	0.15*** (0.05)	0.08*** (0.03)	0.09 (0.06)
Women of age 65 and above	-0.15 (0.15)	-0.06 (0.19)	-0.25 (0.18)	-0.05 (0.18)	-0.44*** (0.17)	-0.41** (0.16)	0.08 (0.23)
Boys of age below 15	0.33 (0.20)	0.50*** (0.18)	0.30 (0.27)	0.22 (0.25)	0.67** (0.27)	0.30 (0.23)	0.60*** (0.22)
Men of age 15-34	-0.06 (0.05)	-0.12* (0.06)	0.04 (0.08)	-0.10 (0.07)	-0.02 (0.08)	-0.06 (0.07)	-0.08 (0.07)
Men of age 35-64	0.02 (0.04)	0.03 (0.05)	0.00 (0.07)	0.01 (0.05)	-0.02 (0.07)	0.03 (0.05)	0.00 (0.06)
Men of age 65 and above	-0.03 (0.10)	-0.16 (0.13)	0.00 (0.12)	-0.09 (0.11)	0.00 (0.22)	-0.12 (0.14)	-0.05 (0.12)

Table 4.8. Dependent variable: change in the number of days of working on the household farm; subsample of people who participated in agriculture in both the first and the last survey wave; agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
<i>A. OLS</i>							
All household members	5.26 (3.69)	9.56** (4.33)	4.54 (5.62)	9.84** (4.41)	-0.06 (5.69)	5.20 (4.29)	12.20** (5.80)
Girls of age below 15	14.29 (36.02)	18.02 (45.88)	44.37 (118.58)	-6.05 (71.24)	35.66 (57.22)	13.30 (42.40)	-52.50 (68.18)
Women of age 15-34	6.55 (9.89)	13.62 (11.91)	-5.32 (15.86)	7.82 (10.93)	5.40 (19.46)	15.78 (11.77)	-4.58 (15.89)
Women of age 35-64	11.32 (7.05)	15.07* (8.25)	11.29 (10.29)	14.34* (8.58)	11.23 (10.17)	10.38 (7.81)	22.37* (12.28)
Women of age 65 and above	-7.17 (20.70)	-9.74 (26.46)	-4.30 (26.23)	3.75 (24.98)	-20.46 (29.38)	-18.65 (23.53)	19.81 (34.37)
Boys of age below 15	-28.16 (23.04)	-15.96 (29.66)	-46.96 (39.08)	-43.57 (29.35)	-4.88 (36.54)	-28.16 (23.04)	- -
Men of age 15-34	-9.66 (7.89)	-8.66 (9.12)	-1.22 (12.82)	-2.13 (9.40)	-23.55* (12.82)	-11.62 (9.92)	-0.87 (11.12)
Men of age 35-64	2.03 (7.57)	7.96 (8.71)	-5.78 (11.91)	9.12 (9.19)	-6.11 (11.37)	-1.97 (8.64)	15.19 (12.38)
Men of age 65 and above	42.44** (19.07)	71.31*** (26.15)	34.51 (24.42)	47.36** (23.33)	27.67 (31.68)	69.39*** (25.73)	34.17 (23.83)
<i>B. Nearest neighbor matching</i>							
All household members	7.51 (4.88)	10.40* (5.74)	12.80 (7.82)	10.50* (6.04)	10.56 (7.68)	2.51 (5.74)	27.23*** (7.61)
Girls of age below 15	-15.17 (47.26)	- -	- -	- -	- -	- -	- -
Women of age 15-34	7.69 (8.36)	4.01 (10.38)	3.38 (11.83)	-0.30 (9.01)	22.40 (16.48)	11.91 (10.36)	16.76 (21.40)
Women of age 35-64	13.78 (9.87)	18.53 (12.42)	19.77 (14.16)	19.85* (11.58)	15.64 (13.81)	9.48 (10.89)	34.98** (17.03)
Women of age 65 and above	-14.19 (27.96)	-3.92 (29.99)	-17.91 (30.33)	11.46 (37.61)	-32.78 (33.83)	-32.00 (27.69)	54.86*** (20.38)
Boys of age below 15	-16.00 (32.38)	- -	- -	- -	- -	- -	- -
Men of age 15-34	-4.05 (11.45)	1.75 (12.64)	5.81 (23.20)	6.27 (12.91)	-21.76 (23.57)	-6.92 (13.96)	10.70 (19.22)
Men of age 35-64	7.98 (10.50)	12.65 (11.58)	0.93 (17.82)	18.77 (12.95)	-0.32 (15.16)	1.78 (11.00)	21.10 (20.07)
Men of age 65 and above	45.00** (17.73)	63.31** (25.73)	65.43* (34.05)	48.26** (23.07)	75.00** (34.67)	71.67** (29.46)	46.00* (24.08)

Table 4.9. Dependent variable: the number of days of working on the household farm, for new household members; agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
<i>A. OLS</i>							
All household members	1.08 (1.66)	0.54 (1.84)	2.09 (2.53)	3.42* (1.91)	-4.57* (2.62)	1.54 (2.00)	0.59 (2.28)
Girls of age below 15	-0.69 (0.66)	-0.87 (0.75)	-0.47 (1.09)	-0.55 (0.78)	-0.92 (1.12)	-0.07 (0.81)	-1.87* (1.01)
Women of age 15-34	-2.20 (6.63)	-2.81 (7.31)	-1.94 (10.25)	0.89 (7.45)	-14.29 (11.48)	1.78 (8.46)	-4.45 (8.59)
Women of age 35-64	26.20 (26.15)	20.96 (28.09)	92.03** (38.48)	46.51 (29.60)	-3.70 (37.23)	22.08 (31.42)	55.53 (34.41)
Women of age 65 and above	10.80 (23.19)	13.42 (25.14)	-12.45 (26.65)	14.59 (25.18)	-13.79 (31.98)	-12.45 (26.65)	19.41 (27.80)
Boys of age below 15	-1.12 (1.45)	-0.74 (1.67)	-2.28 (2.29)	-0.03 (1.66)	-4.50* (2.58)	-1.09 (1.67)	-1.61 (2.23)
Men of age 15-34	5.66 (6.94)	0.77 (7.63)	7.63 (9.72)	9.43 (7.82)	1.06 (10.53)	3.85 (8.99)	2.94 (8.28)
Men of age 35-64	7.55 (18.05)	17.90 (23.98)	6.78 (19.26)	14.60 (26.73)	2.57 (23.21)	5.98 (26.01)	4.07 (20.57)
Men of age 65 and above	-8.54 (44.25)	-37.23 (48.88)	27.73 (62.53)	7.93 (47.85)	-60.03 (73.30)	107.18 (70.37)	-45.17 (40.95)
<i>B. Nearest neighbor matching</i>							
All household members	0.72 (2.00)	-0.15 (2.17)	3.12 (3.44)	3.02 (2.43)	-4.49** (2.17)	1.59 (2.37)	-0.13 (2.99)
Girls of age below 15	-0.71 (0.70)	-0.98 (0.83)	-0.06 (0.92)	-0.15 (0.54)	-1.68 (1.46)	0.09 (0.63)	-2.12 (1.35)
Women of age 15-34	-6.90 (7.68)	-7.86 (7.99)	-5.43 (12.41)	-2.17 (8.43)	-23.93** (9.55)	-1.30 (9.24)	-10.95 (10.03)
Women of age 35-64	30.59 (34.56)	18.21 (38.26)	117.75*** (36.66)	53.34 (41.53)	-6.00 (35.91)	28.83 (43.01)	60.38 (53.52)
Women of age 65 and above	24.00 (29.45)	- -	- -	- -	- -	- -	- -
Boys of age below 15	-1.23 (1.59)	-0.71 (1.88)	-2.67** (1.33)	0.40 (1.56)	-5.61** (2.67)	-0.83 (1.07)	-2.50 (3.27)
Men of age 15-34	5.00 (7.71)	-1.11 (8.36)	10.01 (10.71)	8.06 (9.06)	0.45 (10.05)	3.89 (11.10)	3.14 (7.54)
Men of age 35-64	2.17 (18.68)	2.43 (32.44)	1.25 (20.99)	6.04 (37.20)	-0.93 (9.17)	-24.71 (26.89)	2.19 (22.60)
Men of age 65 and above	48.75 (38.66)	- -	- -	- -	- -	- -	- -

Finally, I look at the labor supply of the new household members (Table 4.9). I find that women of age 15-34 in households that experienced outmigration of youth to urban areas spend 24 days less on the household farm than women in households that did not experience outmigration. In these households with migrants to urban areas, male children spend 4-6 less days on the farm. At the same time, women of age 35-64 who join households with male migrants spend significantly more time on the farm than women who join households without migrants. Thus, the gap in the amount of labor supplied by new household members observed in the descriptive results occurs mainly due to higher labor supply by women of age 35-64.

4.6. Discussion

The discussion about the effects of outmigration on the agricultural inputs is based on the assumption that migrants contributed a significant amount of labor to the household farm prior to their migration. Mueller, Doss, and Quisumbing (2018) show a significant difference in labor inputs at baseline between future migrant and non-migrant children of the household head in Ethiopia and Malawi. In Ethiopia, the average time spent on planting at baseline by women (men) who will move is 57% (34%) higher than for women (men) who will stay in the origin. In Tanzania, I find the opposite: women (men) who will move spend 19% (16%) less time on the household farm at baseline. In Malawi, Mueller, Doss, and Quisumbing (2018) find no significant differences in the amount of farm labor, but migrant men are less likely to work on the farm and spend almost six times more hours on wage labor at baseline than non-migrant men. In Tanzania, I observe no significant differences in wage job or self-employment, but migrants are more likely than non-migrants to be at school.

These fundamental differences in migrants' activities prior to migration can explain the differences in the observed results: Mueller, Doss, and Quisumbing (2018) show a strong impact

of outmigration of children of the household head on the remaining household members. In Ethiopia and Malawi, migrants' mothers and brothers respectively are the most affected by outmigration. In Tanzania, I find a significant effect on women of age 35-64, who are most likely to be migrants' mothers, and people of age 65 and above, who are most likely to be migrants' grandparents. The impact on men of age 15-34, who are most likely to be migrants' brothers, is observed only in descriptive results and disappears after I account for selection into migrant-sending households. Whenever my estimates are significant, they are comparable to those of Mueller, Doss, and Quisumbing (2018) and show a 10-20% increase in labor supplied to the household farm by certain groups of household members.

In Ethiopia and Malawi, Mueller, Doss, and Quisumbing (2018) find a significant increase in the farm labor supplied by brothers (in Ethiopia) and sisters (in Malawi) of migrants to urban areas. In Tanzania, I observe the opposite: youth in households with urban-destined migrants are more likely to stop working in agriculture, and the number of days spent on the farm decreases among those who keep working. Also, I find significant effects on children in households that sent out migrants to urban areas: they spend less time on farming and are more likely to be students.

Mueller, Doss, and Quisumbing (2018) find a positive impact of outmigration on the probability to use hired labor in Malawi. I find no significant impact on hired labor in Tanzania after accounting for selection. Descriptive results suggest no difference in the probability to use hired labor, but the change to the amount of hired labor used among those who already use it is higher in households that experienced outmigration. A similar pattern holds for the results on the attraction of new household members. Bertoli and Murard (2020) find that the probability to receive any new members is three times higher in households in Mexico that sent an

international migrant. For internal migrants in Tanzania, I find no significant impact on the probability to attract a migrant, but I see significantly higher number of new members in households that experienced the outmigration of youth and attracted any new members.

The main drawback of this study is the inability to confirm pre-migration trends. Households that sent out a migrant differ significantly from households without migrants, which suggests that there could be factors affecting the decisions of these two types of households in a different way. For example, all household members can increase their labor supply to the farm in a few years prior to migration because they expect to send out a migrant. This short-term increase in the labor supply can be necessary, for example, to increase income and generate savings that would later be used to finance outmigration. I would then observe unusually high family labor supply at baseline in households that will send out a migrant in subsequent years.

4.7. Conclusion

Outmigration of youth can be associated with a reduction in labor supplied to the household farm in agricultural households remaining in place. Non-migrant household members then can increase their own labor supply, invite new household members, hire workers, or decrease the size of cultivated land. I compare these outcomes in households with and without migrant youth and estimate the impact of outmigration. The findings suggest that households indeed experience a reduction in labor supply of up to 20%, although it does not affect its structure of income. With the outmigration of youth, household has 80% chance to lose a member who supplied labor to the farm. On the other hand, the probability to invite a new agricultural worker to the household that sent out a migrant increases by 10-13%.

My results for individual-level outcomes are similar to the results of the study conducted by Mueller, Doss, and Quisumbing (2018) for Ethiopia and Malawi, which show that

outmigration of children of the household head is associated with a 10-20% increase in the time spent on agriculture by certain groups of non-migrant household members differentiated by gender and relation to the migrant. For Tanzania, I find that women of age 35-64 are the ones who significantly increase their labor supply to the farm, both if they have been present in the household before the outmigration and if they joined a household that recently sent out a migrant. Also, I see a negative impact of youth outmigration on the welfare of the elderly who are less likely to stop working in agriculture and do not decrease their supply of labor as much as elderly in households without migrants. On the other hand, children who live in households that experienced outmigration to urban areas or outmigration of people of age 21-34 are more likely to stop working in agriculture and be students.

There are two ways of gendered impact in this study. First, I find outmigration of a young woman to be associated with higher increase in the labor of non-migrant household members than outmigration of a young man. This could happen because migrant women spend more time on the household farm at baseline than migrant men do, hence their outmigration leads to a larger decrease in labor supply. Second, I find the increase in the labor supply of non-migrant women in households that experienced outmigration to be more significant than the increase in the labor supply of non-migrant men in these households. This pattern holds among new household members as well: women of age 35-64 who join households that experienced outmigration, on average, spend more time on the household farm than men of age 35-64. On the other hand, men of age 15-34 who join households with migrants are more likely to have main occupation in non-agricultural wage job or self-employment while women of age 15-34 are more likely to be at school.

Unlike Mueller, Doss, and Quisumbing (2018), I do not distinguish people by their role in the household. Instead, I make groups based on gender and age. Hence, children present in the household can be migrant's siblings, children, or other relatives; people of working age can be migrant's parents, siblings, spouses, or other relatives; and elderly can be migrant's parents, grandparents, or other relatives. Given the non-linear relationship between age and the amount of time spent on the household farm, this approach allows to separate more vulnerable groups (children and elderly) regardless of their tie to the migrant. Like the studies on internal migration in China (e.g., Chang, Dong, and MacPhail, 2011), I find that the welfare of people in these groups in Tanzania can be negatively affected by outmigration.

A further investigation of the changes to the time use among non-migrant household members in African households is needed. This study describes some of the patterns of transition in and out of agriculture, an increase of the labor supply to the household farm, and the role of new household members. Still, more work needs to be done to uncover the mechanisms behind the observed changes. In this study, I suggest the main channel to be the reduction in labor supply, although I observe migrant youth to spend significantly less time on the household farm prior to their migration than non-migrant youth. Still, the increase in the total labor supply in households that sent out a migrant over time is much smaller than in households that did not. It suggests that an increase in time spent by non-migrant household members was not enough to offset the loss of labor. The characteristics of a migrant matter as well. In households where the migrant is female, older, or moved to a rural area, the response is more drastic than in other households. It can be caused by a more severe decrease in the supply of labor, as migrants in these groups were more likely to supply labor to the household farm and to supply more labor prior to migration. On the other hand, households that sent out an older migrant are more likely

to increase the amount of cultivated land and hire more labor than before. Older migrants are more likely to send remittances and send more remittances when they do, which can allow the household of origin to invest more into the farm.

I find that the structure of the household is more likely to change with youth outmigration. Moreover, households that sent out a migrant do not necessarily attract workers among the new members. Yet women of age 35-64 who join the households with migrants do indeed work more on the farm, children and youth who join these households are more likely to be students and supply less labor to the farm than new members in households without migrants. At baseline, households that will experience outmigration are larger and wealthier. Future migrants in these households tend to spend less time on agriculture. It suggests that these households are able to withstand the reduction in labor supply and sustain new members who, at the time of joining, contribute less. On the other hand, households with less resources are much more vulnerable to the reduction in labor supply associated with outmigration when they cannot invite new members.

APPENDICES

APPENDIX 1. Remittances

In this subsection, I summarize the information on remittances available from the LSMS dataset for Tanzania. The focus is on the 2008/2009 and the 2012/2013 survey waves. For comparison and to verify the information from the latest wave used in the main analysis, I add the 2014/2015 wave which lists a new set of households. In Table 4.10, I present summary statistics for the remittances a child who moved within Tanzania sends back to the household of origin. Unfortunately, it is beyond the scope of this essay to link the certain remitted amount to the sender. Hence, in households with multiple migrant children, it is hard to verify whether an increase in remittances is due to recent outmigration or an increase in the amount sent by people who moved away earlier.

Table 4.10. Summary statistics for remittances received in the past 12 months from children living elsewhere in Tanzania, thousand Tanzanian Shilling (TSh); by child

Wave (years)	2008/2009		2012/2013		2014/2015	
Location of the recipient household	Rural	Urban	Rural	Urban	Rural	Urban
10 th percentile	5	10	15	30	20	30
25 th percentile	10	20	35	60	40	60
50 th percentile	20	50	95	180	100	150
75 th percentile	50	100	200	350	200	340
90 th percentile	85	300	400	700	450	700
Mean	41	107	213	309	202	285
Number of observations	1270	477	444	189	427	179

Note: For 2008/2009, the location is considered to be rural if its type is listed as rural. For 2014/2015, the location is considered to be rural if its cluster type is listed as rural.⁸²

⁸² These clarifications are made since for the 2008/09 and 2014/2015 waves there are additional variables (locality and ward type, respectively) that divide locations into three types: rural, urban, and mixture. If I require both available variables to be listed as rural for the location to be rural, then 291 observations in 2008/2009 and 62 observations in 2014/2015 shift to “urban”, and the mean and the median of remittances received by urban households decrease by 5-30%.

The distribution of remittances is similar between the 2012/2013 and the 2014/2015 survey waves. In 2008/2009, though, the amount of remittances was much smaller. The median amount of remittances received from one child in rural households is almost twice as low as in urban households. Remittances received from international migrants are at least twice as high as remittances from internal migrants (see Table 4.11). The number of children of the household head sending remittances exceeds the number of households receiving remittances, meaning that households usually have more than one source of remittances. Moreover, children of the household head are not the only ones sending remittances; and the average amount the households receive from migrant children is comparable to the amount they receive from other individuals.

Figure 4.1 and Figure 4.2 show the distribution of remittances by sender's age and time spent at destination.⁸³ Migrants of age 30-39 are more likely to send remittances, but migrants of age 45-54, on average, sent the largest amounts. People who moved recently are at least as likely to send back remittances as people who have already spent more time at destination, but the amount they send back is on average smaller. In the main analysis, I look at people of age 15-34 (in gray on the figures) who moved within the past four years. Among them, older young adults are more likely to remit and send a higher amount. These observations are consistent with the ones made in the main body of the essay: youth find it harder to remit for various reasons, from unemployment to focusing on the new household.

⁸³ Note that the information recorded is the number of years spent at that specific destination, not the number of years spent away from the household.

Table 4.11. Remittances received in the past 12 months, by information about the sender

Wave (years)	2008/2009		2010/2011	2012/2013		2014/2015	
Source of remittances	domestic	abroad	abroad (cash)	domestic	abroad	domestic	abroad
Number of households	3,265		3,924	5,010		3,352	
Number of households receiving remittances from children living elsewhere	759	16	4	412	16	377	12
Number of households receiving remittances from other individuals	-	-	31	688	23	740	29
Number of children sending remittances	1,747	21	4	633	17	606	13
Number of other individuals sending remittances	-	-		860	23	996	29
Average age of a child sending remittances	33.48	34.57	-	35.39	35.29	33.87	33.31
Average age of other individual sending remittances	-	-	-	41.42	45.00	42.19	47.10
Average value of remittances sent by children (cash and in-kind), by sender, TSh	58,898	198,857	525,000	241,260	486,000	226,177	1,353,846
Average value of remittances sent by someone else (cash and in-kind), by sender, TSh	-	-	1,036,903	199,989	1,036,870	230,814	1,430,517
Average value of remittances sent by children (cash and in-kind), by household, TSh	135,566	261,000	525,000	370,674	516,375	363,563	1,466,667
Average value of remittances sent by someone else (cash and in-kind), by household, TSh	-	-	1,036,903	249,986	1,036,870	310,663	1,430,517

Note: In the 2008/2009 wave the questions on remittances were asked only for children living outside of the household whose mothers live in the household. In the 2010/2011 wave the questions about the sender were asked only about remittances in a form of cash sent from abroad; the question about the relation of the sender asks about the person interviewed (could be either the head of the household or the spouse of household head). In 2012/2013 and 2014/2015 waves the questions about the sender were asked about both remittances from abroad and domestic remittances; the question about the relation to the sender asked about the household head. New households were sampled for the 2014/2015 wave, this wave is not included in the panel.

Figure 4.1. Remittances received from children of the household head, by sender's age group

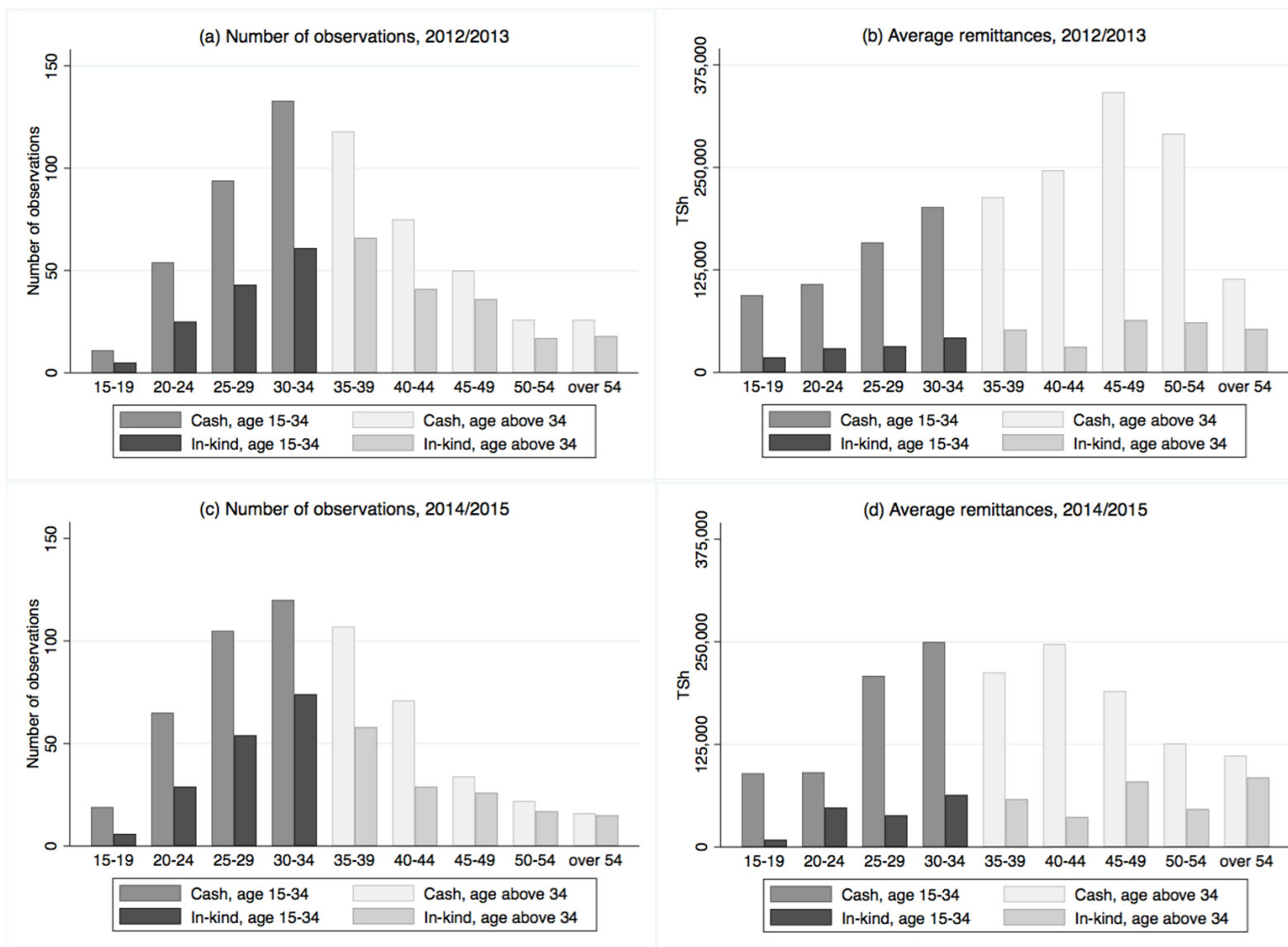
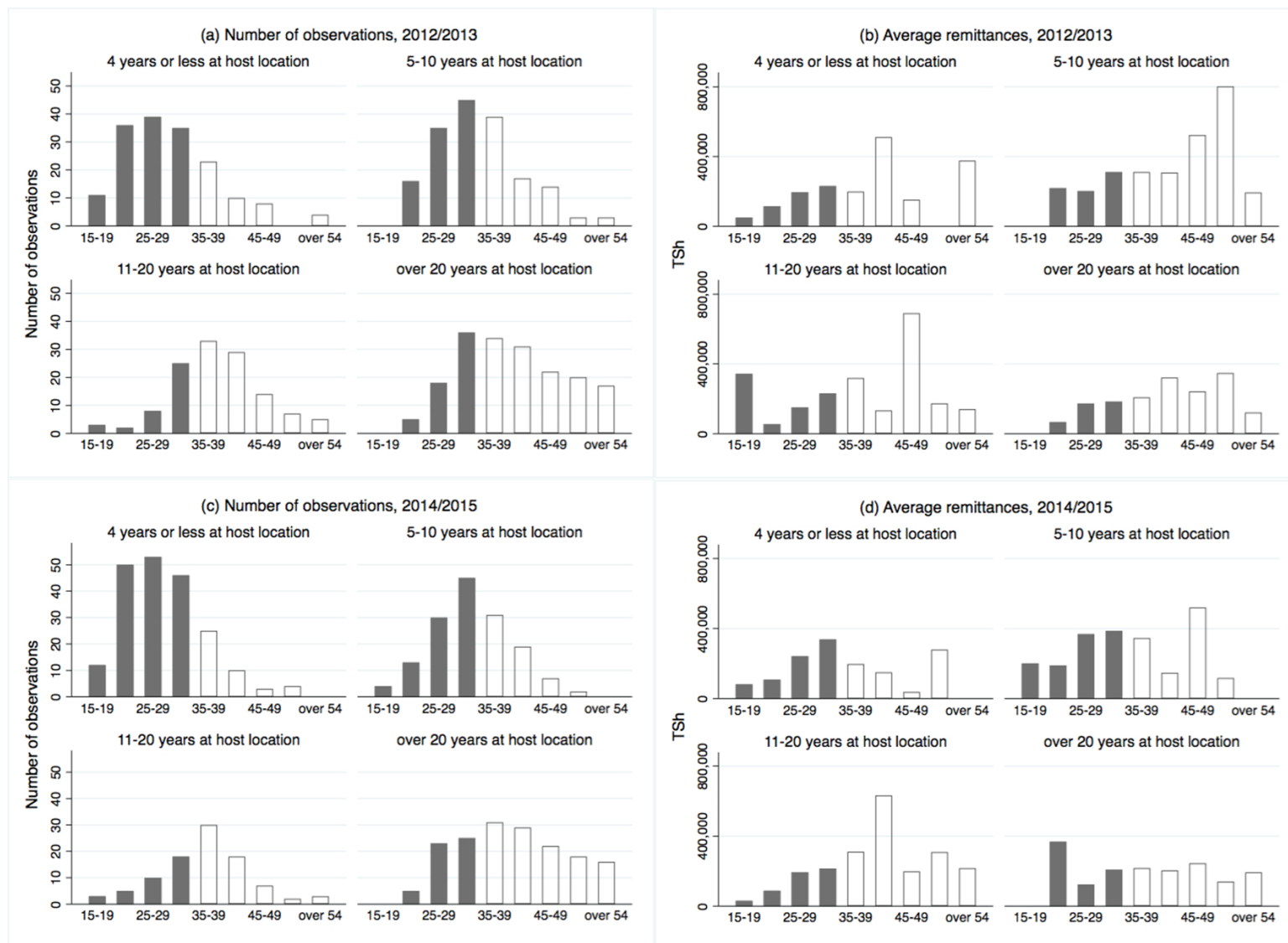


Figure 4.2. Remittances received from children of the household head, by years lived at the host location and age group (colored gray for people of age 15-34)



APPENDIX 2. Additional tables and figures

Table 4.12. Characteristics of people of age 15 to 34 who lived in rural areas in 2008/2009

	Non-migrants	Migrants	Migrants - Non-migrants	Std. error
Age	23.23	20.44	-2.78***	(0.35)
1 = Male	0.51	0.35	-0.16***	(0.03)
1 = Married	0.43	0.23	-0.20***	(0.03)
1 = Completed primary school	0.58	0.62	0.04	(0.03)
1 = Born in his village	0.82	0.76	-0.06***	(0.02)
1 = Was away from the household for at least 1 month in the past year	0.09	0.16	0.07***	(0.02)
1 = Household head or a spouse of household head	0.43	0.11	-0.32***	(0.03)
1 = Child of the household head	0.43	0.56	0.14***	(0.03)
Asset index	0.62	1.35	0.73***	(0.17)
Land area cultivated by the household, acres	7.57	9.80	2.23	(1.91)
Livestock units (TLU)	3.41	5.01	1.60*	(0.83)
Age of the household head	44.58	50.38	5.79***	(0.90)
1 = Household head is male	0.82	0.80	-0.02	(0.02)
Size of the household	6.87	7.86	1.00***	(0.26)
Number of children of the household head living in the household	3.42	3.80	0.38**	(0.15)
1 = Household experienced a negative agricultural shock in the past year	0.29	0.28	-0.01	(0.03)
1 = Household experienced a negative non-agricultural shock in the past year	0.29	0.33	0.04	(0.03)
Population density, people per sq. km	142.16	171.15	28.99**	(14.54)
Distance to the nearest road, km	20.39	19.86	-0.52	(1.21)
Distance to the nearest town with population of at least 50,000 people, km	60.85	60.28	-0.57	(2.40)
Number of observations	2,377	305		

Note: Sample weights from 2008/2009 are applied. t-test for difference in means: *** 0.01; ** 0.05; * 0.1.

Table 4.13. Difference in means between migrant and non-migrant youth from rural areas according to the NBS definition, for people with certain types of main occupation

	Farming	Studies	Household maintenance
1 = Spent any days performing agricultural activities in the past year	-0.11***	-0.09*	0.11
Days spent on land preparation and planting, past year	-6.20***	0.16	2.57
Days spent on weeding, past year	-5.93***	-0.18	1.70
Days spent on harvesting, past year	-5.50***	0.97	4.71
Total number of days spent on agricultural activities, past year	-17.62***	0.95	8.98
Total number of days spent on agricultural activities, past year; among those who spent any	-12.70**	8.18	9.07
1 = Spent any time on agriculture in the past week	-0.06	-0.00	0.01
Hours spent on household agricultural activities, past week	-3.08**	0.01	-3.35
Hours spent as an unpaid family worker on a non-farm household business, past week	0.47	-0.62	-3.80
Hours spent collecting firewood or water, yesterday	-0.00	-0.06	-0.30
1 = Currently attending school	0.02*	-0.03	-0.02
1 = Was in school last year (if not attending currently)	0.05***	0.04	-0.06
1 = Did any work for pay, profit, barter, or home use in the past week	-0.13***	0.01	0.22**
1 = Have work or own farm or enterprise to return to (if didn't work in the past week)	0.08**	-0.03	-0.04
1 = Did any wage work, past week	-0.01	0.01	0.18***
1 = Did any wage work, past year	-0.05*	0.01	0.22***
Hours worked at wage job, past week	-0.79	0.09	12.75***
1 = Did non-agricultural self-employed activity, past week	-0.08***	-0.00	0.03
1 = Did non-agricultural self-employed activity, past year	-0.07**	-0.01	0.03
Months operating a business, past year	-0.42*	-0.02	0.34

Note: Sample weights from 2008/2009 are applied. t-test for difference in means: *** 0.01; ** 0.05; * 0.1.

Table 4.14. Activities of youth from rural areas according to the NBS definition – by gender, age, and migration destination

	Non-migrants		Migrants		Non-migrants		Migrants		Migrants	
	Men	Women	Men	Women	Of age 15-20	Of age 21-34	Of age 15-20	Of age 21-34	To rural	To urban
1 = Main occupation is farming or fishing	0.65	0.75	0.47	0.60	0.44	0.88	0.45	0.73	0.66	0.36
1 = Main occupation is wage job	0.02	0.01	0.03	0.01	0.00	0.02	0.01	0.02	0.01	0.03
1 = Main occupation is self-employment	0.03	0.02	0.05	0.02	0.01	0.04	0.01	0.08	0.02	0.05
1 = Main occupation is studies	0.25	0.17	0.35	0.30	0.48	0.03	0.44	0.10	0.25	0.45
1 = Main occupation is household maintenance	0.03	0.04	0.08	0.06	0.05	0.02	0.07	0.05	0.05	0.10
1 = Main occupation is unemployment or disability	0.01	0.02	0.02	0.00	0.02	0.01	0.01	0.01	0.01	0.00
1 = Spent any days performing agricultural activities in the past year	0.79	0.81	0.63	0.69	0.69	0.88	0.64	0.72	0.67	0.67
Days spent on land preparation and planting, past year	18.65	19.55	13.06	13.47	12.10	23.89	12.20	15.29	13.65	12.72
Days spent on weeding, past year	16.23	18.17	10.26	12.28	11.82	20.86	10.94	12.68	12.11	10.56
Days spent on harvesting, past year	11.96	13.56	7.52	9.67	8.46	15.69	8.00	10.51	8.88	8.99
Total number of days spent on agricultural activities, past year	46.84	51.29	30.84	35.42	32.38	60.44	31.15	38.48	34.64	32.26
Total number of days spent on agricultural activities, past year; among those who spent any	59.20	63.23	48.63	51.30	47.24	68.66	48.31	53.73	51.50	48.33
1 = Spent any time on agriculture in the past week	0.63	0.62	0.53	0.56	0.55	0.68	0.56	0.52	0.58	0.48
Hours spent on household agricultural activities, past week	18.26	16.00	15.33	12.38	13.71	19.52	13.04	14.05	14.59	11.17
Hours spent as an unpaid family worker on a non-farm household business, past week	9.67	26.13	8.49	20.97	15.11	19.53	15.58	18.39	18.88	12.27
Hours spent collecting firewood or water, yesterday	0.39	1.15	0.30	0.90	0.67	0.82	0.71	0.66	0.83	0.42
1 = Currently attending school	0.24	0.15	0.32	0.28	0.45	0.02	0.39	0.12	0.21	0.44
1 = Was in school last year (if not attending currently)	0.05	0.04	0.11	0.09	0.10	0.01	0.14	0.01	0.10	0.08
1 = Did any work for pay, profit, barter, or home use in the past week	0.61	0.57	0.47	0.47	0.43	0.71	0.42	0.57	0.52	0.39
1 = Have work or own farm or enterprise to return to (if didn't work in the past week)	0.14	0.19	0.17	0.18	0.11	0.20	0.16	0.21	0.21	0.11
1 = Did any wage work, past week	0.15	0.07	0.13	0.07	0.04	0.15	0.07	0.13	0.11	0.06
1 = Did any wage work, past year	0.21	0.10	0.18	0.08	0.07	0.22	0.10	0.15	0.12	0.11
Hours worked at wage job, past week	5.42	1.74	6.11	1.60	1.20	5.28	3.42	2.75	3.04	3.42
1 = Did non-agricultural self-employed activity, past week	0.11	0.10	0.02	0.06	0.04	0.15	0.02	0.09	0.04	0.05
1 = Did non-agricultural self-employed activity, past year	0.14	0.13	0.06	0.09	0.05	0.20	0.04	0.15	0.07	0.08
Months operating a business, past year	0.93	0.82	0.37	0.51	0.22	1.33	0.17	0.96	0.40	0.58
Number of observations	1,189	1,188	106	199	995	1,382	195	110	197	108

Note: Sample weights from 2008/2009 are applied.

Figure 4.3. Average number of days spent on agricultural activities, 2008/2009, by age and outmigration experience: for non-migrant household members living in rural areas (according to the NBS definition) in households with youth at baseline

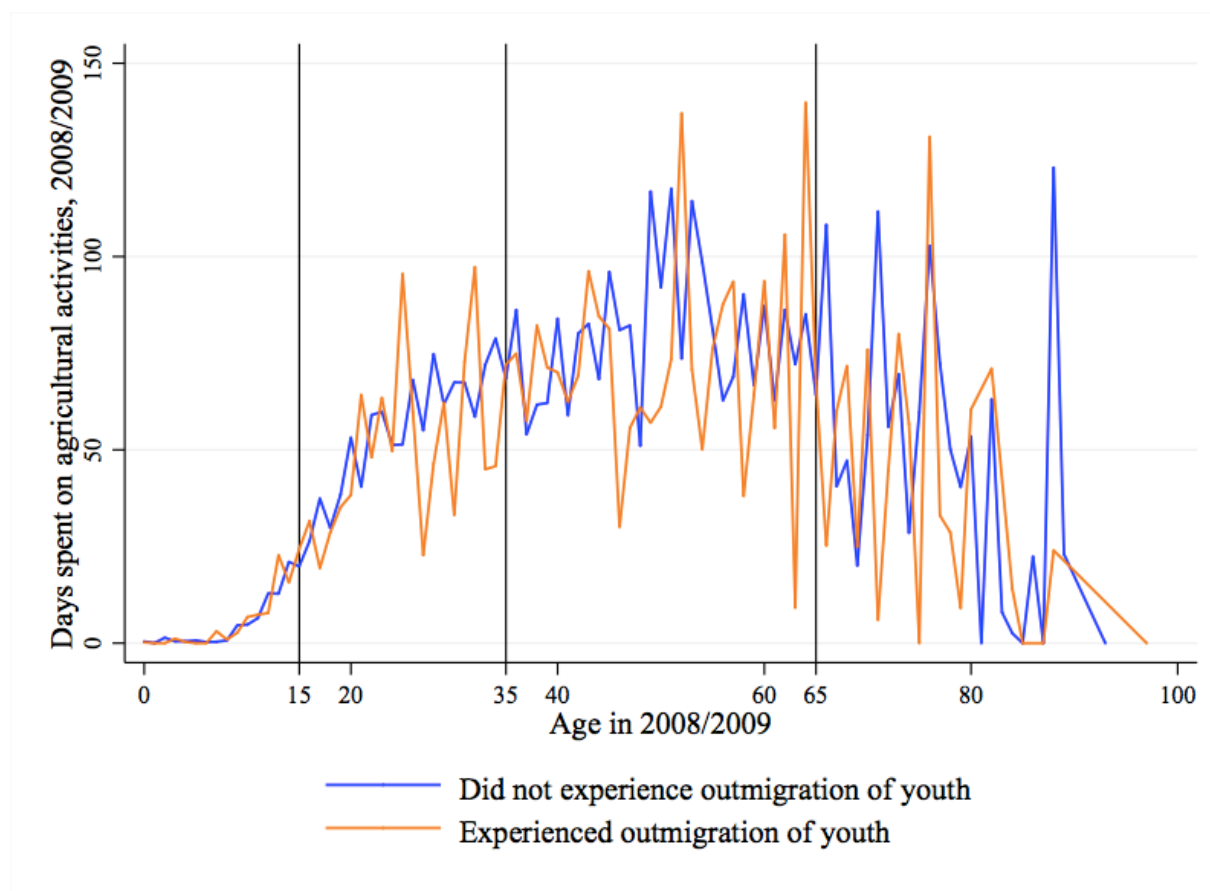


Table 4.15. Main occupation by gender, age, and presence in the household; rural areas (according to the NBS definition)

	Present in both 2008/2009 and 2012/2013							
	2008/2009: Share of people with main occupation in:				2012/2013: Share of people with main occupation in:			
	Farming	Wage job / self-employment	Studies	Other	Farming	Wage job / self-employment	Studies	Other
<i>A. Households did not experience outmigration of youth</i>								
Girls of age below 15	4.2%	0.2%	88.1%	7.6%	12.8%	0.4%	72.3%	14.6%
Women of age 15-34	79.6%	3.1%	13.2%	4.1%	81.2%	6.1%	5.6%	7.2%
Women of age 35-64	93.3%	3.9%	0.0%	2.8%	91.4%	4.7%	0.0%	4.0%
Women of age 65 and above	73.8%	1.5%	0.0%	24.7%	59.9%	1.6%	0.0%	38.6%
Boys of age below 15	4.9%	0.4%	86.5%	8.2%	16.1%	0.2%	67.7%	15.9%
Men of age 15-34	64.7%	5.2%	26.1%	4.1%	71.1%	12.0%	10.9%	6.0%
Men of age 35-64	87.7%	11.0%	0.0%	1.2%	87.5%	11.3%	0.0%	1.2%
Men of age 65 and above	87.7%	1.5%	0.0%	10.7%	79.5%	2.0%	0.0%	18.5%
<i>B. Households experienced outmigration of youth</i>								
Girls of age below 15	5.5%	0.0%	80.0%	14.6%	14.9%	0.0%	69.6%	15.5%
Women of age 15-34	61.0%	3.2%	24.2%	11.6%	69.8%	4.3%	13.9%	12.1%
Women of age 35-64	92.2%	6.9%	0.0%	1.0%	90.4%	6.2%	0.0%	3.4%
Women of age 65 and above	64.1%	1.1%	0.0%	34.8%	62.6%	0.0%	0.0%	37.4%
Boys of age below 15	9.6%	0.0%	80.0%	10.4%	16.8%	0.0%	69.2%	14.0%
Men of age 15-34	62.0%	6.8%	28.1%	3.0%	66.8%	17.4%	11.2%	4.5%
Men of age 35-64	88.9%	10.6%	0.0%	0.5%	81.1%	16.9%	0.0%	2.1%
Men of age 65 and above	88.0%	3.9%	0.0%	8.0%	79.7%	0.0%	0.0%	20.3%

Note: "Farming" includes farming and fishing. "Wage job / self-employment" includes non-agricultural wage job or self-employment. "Other" category includes household maintenance, unemployment, and disability. Age groups are based on individuals' age in 2008/2009, hence newborns are excluded from the sample of people present only in 2012/2013. Sample weights are applied.

Table 4.15 (cont'd)

	Present only in 2012/2013 2012/2013: Share of people with main occupation in:			
	Farming	Wage job / self-employment	Studies	Other
<i>A. Households did not experience outmigration of youth</i>				
Girls of age below 15	24.0%	0.5%	57.9%	17.6%
Women of age 15-34	78.5%	7.9%	3.6%	10.0%
Women of age 35-64	78.1%	2.1%	4.8%	15.0%
Women of age 65 and above	28.8%	4.1%	0.0%	67.2%
Boys of age below 15	19.3%	2.5%	65.7%	12.5%
Men of age 15-34	60.6%	16.6%	10.9%	12.0%
Men of age 35-64	76.4%	16.0%	0.0%	7.6%
Men of age 65 and above	94.4%	0.0%	0.0%	5.6%
<i>B. Households experienced outmigration of youth</i>				
Girls of age below 15	12.8%	0.0%	68.6%	18.7%
Women of age 15-34	73.8%	3.8%	7.2%	15.2%
Women of age 35-64	94.7%	5.3%	0.0%	0.0%
Women of age 65 and above	56.4%	0.0%	0.0%	43.6%
Boys of age below 15	9.7%	2.6%	64.8%	22.9%
Men of age 15-34	49.9%	28.6%	9.5%	12.0%
Men of age 35-64	100.0%	0.0%	0.0%	0.0%
Men of age 65 and above	0.0%	0.0%	0.0%	100.0%

Note: "Farming" includes farming and fishing. "Wage job / self-employment" includes non-agricultural wage job or self-employment. "Other" category includes household maintenance, unemployment, and disability. Age groups are based on individuals' age in 2008/2009, hence newborns are excluded from the sample of people present only in 2012/2013. Sample weights are applied.

Table 4.16. OLS regressions for household-level outcomes in rural households (according to the NBS definition)

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
Change in the number of household members who supply labor to the household's farm	-0.66*** (0.12)	-0.64*** (0.13)	-0.75*** (0.18)	-0.61*** (0.14)	-0.61*** (0.18)	-0.58*** (0.14)	-0.82*** (0.17)
Change in the total family labor supplied to the household's farm, days	-26.89* (15.40)	-17.08 (17.34)	-44.44* (23.25)	-2.31 (18.16)	-60.43** (23.68)	-25.84 (18.03)	-20.30 (22.57)
Change in the indicator of using any hired labor	0.01 (0.04)	-0.02 (0.05)	0.03 (0.06)	-0.04 (0.05)	0.08 (0.06)	0.00 (0.05)	0.00 (0.06)
Change in the hired labor, days	2.33 (3.60)	4.85 (4.04)	1.55 (5.43)	4.10 (4.23)	-1.91 (5.54)	0.90 (4.21)	9.17* (5.26)
Change in the land under cultivation, acres	1.41* (0.76)	1.69** (0.85)	2.40** (1.14)	2.00** (0.89)	0.11 (1.17)	0.32 (0.89)	2.66** (1.11)
Change in the land owned, acres	0.76 (0.69)	0.96 (0.78)	1.07 (1.05)	1.17 (0.82)	0.21 (1.07)	-0.12 (0.81)	1.01 (1.02)
Change in the share of income coming from agriculture, decimal	-0.00 (0.03)	0.01 (0.03)	-0.03 (0.04)	-0.01 (0.03)	0.02 (0.04)	-0.00 (0.03)	0.01 (0.04)
Change in the indicator of specializing on farming	0.03 (0.04)	0.02 (0.05)	0.04 (0.07)	0.01 (0.05)	0.05 (0.07)	0.04 (0.05)	0.01 (0.06)
1 = Have any new household members	-0.04 (0.03)	-0.03 (0.04)	-0.05 (0.05)	0.01 (0.04)	-0.11** (0.05)	-0.06 (0.04)	0.02 (0.05)
Number of newborns	-0.19*** (0.06)	-0.19*** (0.07)	-0.38*** (0.10)	-0.12 (0.07)	-0.28*** (0.10)	-0.17** (0.07)	-0.26*** a(0.09)
Number of new household members, excluding newborns	0.16** (0.07)	0.20** (0.08)	0.02 (0.10)	0.17** (0.08)	0.12 (0.11)	0.03 (0.08)	0.35*** (0.10)
Number of new household members who supply labor to the household's farm	0.09** (0.04)	0.07* (0.04)	0.09 (0.06)	0.09* (0.05)	0.08 (0.06)	0.05 (0.05)	0.10* (0.06)
Total amount of labor supplied to the household's farm by new members, days	5.09 (3.26)	5.84 (3.67)	5.94 (4.95)	9.14** (3.85)	-1.57 (5.00)	2.10 (3.79)	10.20** (4.86)

Note: Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Table 4.17. Nearest neighbor matching, household-level outcomes in rural households (according to the NBS definition)

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
Change in the number of household members who supply labor to the household's farm	-0.71*** (0.18)	-0.77*** (0.20)	-0.96*** (0.27)	-0.72*** (0.20)	-0.67*** (0.25)	-0.63*** (0.20)	-1.14*** (0.25)
Change in the total family labor supplied to the household's farm, days	-33.13 (23.66)	-22.30 (25.37)	-49.54 (32.84)	-12.20 (27.40)	-54.55 (33.50)	-48.11* (28.76)	-16.77 (31.61)
Change in the indicator of using any hired labor	0.03 (0.05)	0.00 (0.06)	0.03 (0.07)	0.02 (0.06)	0.05 (0.09)	0.02 (0.06)	0.03 (0.07)
Change in the hired labor, days	2.64 (5.93)	8.42 (5.52)	-4.78 (12.39)	5.71 (8.18)	-3.20 (7.03)	-1.56 (7.81)	12.29 (7.75)
Change in the land under cultivation, acres	0.71 (1.47)	1.12 (1.94)	-1.54 (2.42)	0.82 (1.95)	0.09 (0.57)	-1.59 (2.27)	-0.44 (2.32)
Change in the land owned, acres	0.70 (1.41)	1.14 (1.85)	-1.28 (2.30)	0.80 (1.84)	-0.07 (0.64)	-1.19 (2.15)	-0.95 (2.25)
Change in the share of income coming from agriculture, decimal	-0.01 (0.03)	-0.01 (0.04)	-0.03 (0.05)	-0.06 (0.04)	0.06 (0.05)	-0.00 (0.04)	-0.01 (0.06)
Change in the indicator of specializing on farming	-0.00 (0.06)	-0.03 (0.07)	0.01 (0.09)	-0.07 (0.07)	0.11 (0.08)	-0.01 (0.06)	0.00 (0.09)
1 = Have any new household members	-0.04 (0.04)	-0.05 (0.05)	-0.04 (0.07)	0.01 (0.05)	-0.11 (0.07)	-0.04 (0.05)	-0.04 (0.06)
Number of newborns	-0.06 (0.10)	-0.12 (0.13)	-0.11 (0.12)	0.06 (0.14)	-0.22* (0.11)	0.04 (0.12)	-0.26** (0.13)
Number of new household members, excluding newborns	0.37*** (0.12)	0.36** (0.14)	0.32* (0.18)	0.44*** (0.15)	0.22 (0.16)	0.18 (0.13)	0.72*** (0.19)
Number of new household members who supply labor to the household's farm	0.16** (0.07)	0.10 (0.08)	0.16 (0.11)	0.20** (0.08)	0.12 (0.10)	0.09 (0.08)	0.18 (0.12)
Total amount of labor supplied to the household's farm by new members, days	8.79 (7.75)	1.21 (9.48)	17.42 (12.36)	19.68** (8.54)	-7.26 (13.02)	1.92 (9.90)	6.13 (15.10)

Note: Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Table 4.18. Propensity score matching, household-level outcomes in rural households (according to the NBS definition)

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
Change in the number of household members who supply labor to the household's farm	-0.56*** (0.18)	-0.66*** (0.21)	-0.88*** (0.28)	-0.51** (0.21)	-0.73*** (0.27)	-0.88*** (0.23)	-0.69*** (0.26)
Change in the total family labor supplied to the household's farm, days	-13.43 (29.26)	-59.21* (34.33)	-103.29*** (25.55)	-34.72 (31.05)	-100.42** (41.21)	-62.63** (30.02)	-45.70 (51.07)
Change in the indicator of using any hired labor	-0.04 (0.06)	-0.07 (0.06)	0.09 (0.07)	-0.07 (0.07)	0.04 (0.09)	-0.01 (0.06)	-0.03 (0.08)
Change in the hired labor, days	3.63 (5.05)	-4.78 (10.59)	8.17 (7.00)	3.46 (6.12)	3.32 (6.19)	-0.21 (5.82)	3.07 (9.29)
Change in the land under cultivation, acres	0.81 (1.63)	-0.35 (2.11)	-2.04 (2.58)	0.13 (2.15)	-0.20 (0.83)	-0.19 (1.38)	-2.42 (3.72)
Change in the land owned, acres	0.69 (1.43)	-1.33 (1.76)	-1.19 (2.48)	-0.09 (1.90)	-0.34 (0.88)	0.83 (1.41)	-2.75 (2.34)
Change in the share of income coming from agriculture, decimal	0.00 (0.03)	-0.01 (0.04)	-0.07 (0.06)	0.03 (0.04)	-0.02 (0.05)	0.00 (0.04)	-0.02 (0.05)
Change in the indicator of specializing on farming	0.05 (0.06)	0.03 (0.06)	0.15 (0.09)	0.05 (0.07)	0.12 (0.10)	-0.01 (0.07)	0.04 (0.09)
1 = Have any new household members	-0.01 (0.05)	-0.06 (0.05)	-0.13** (0.05)	0.06 (0.05)	-0.10 (0.07)	-0.05 (0.05)	-0.09 (0.06)
Number of newborns	-0.03 (0.11)	-0.23* (0.13)	-0.34** (0.15)	0.09 (0.15)	-0.18 (0.14)	0.02 (0.14)	-0.46** (0.23)
Number of new household members, excluding newborns	0.30** (0.12)	0.17 (0.16)	0.03 (0.15)	0.49*** (0.14)	0.11 (0.16)	0.05 (0.15)	0.30 (0.25)
Number of new household members who supply labor to the household's farm	0.12 (0.07)	0.02 (0.08)	0.13 (0.12)	0.21** (0.09)	0.13 (0.09)	-0.06 (0.10)	0.16 (0.17)
Total amount of labor supplied to the household's farm by new members, days	8.72 (6.53)	-1.00 (7.25)	19.32** (9.77)	20.73** (8.31)	3.70 (5.02)	-6.37 (10.50)	15.44 (22.85)

Note: Standard errors are in parentheses. *** 0.01; ** 0.05; * 0.1.

Table 4.19. Dependent variable: indicator for not working on the household farm in the first survey wave and working any number of days in the last survey wave (probability to start working in agriculture); agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
<i>A. Logistic regressions</i>							
All household members	-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.03)	-0.02 (0.02)	-0.02 (0.03)	0.04* (0.02)	-0.16*** (0.03)
Girls of age below 15	-0.00 (0.03)	0.01 (0.03)	-0.03 (0.04)	0.00 (0.03)	-0.00 (0.04)	0.04 (0.03)	-0.08* (0.04)
Women of age 15-34	-0.03 (0.05)	-0.05 (0.06)	-0.08 (0.08)	-0.03 (0.06)	-0.09 (0.08)	0.01 (0.06)	-0.22** (0.10)
Women of age 35-64	-0.10 (0.12)	-0.15 (0.14)	-0.11 (0.14)	-0.21 (0.17)	-0.04 (0.15)	-0.07 (0.15)	-0.12 (0.14)
Women of age 65 and above	0.16 (0.16)	0.12 (0.18)	0.32 (0.27)	0.24 (0.21)	0.13 (0.24)	0.18 (0.19)	0.05 (0.31)
Boys of age below 15	0.00 (0.02)	-0.01 (0.03)	-0.00 (0.04)	-0.04 (0.03)	0.05 (0.04)	0.02 (0.03)	-0.09* (0.05)
Men of age 15-34	-0.05 (0.05)	-0.06 (0.06)	-0.08 (0.08)	-0.05 (0.06)	-0.11 (0.08)	-0.02 (0.05)	-0.23** (0.10)
Men of age 35-64	-0.17 (0.11)	-0.16 (0.12)	-0.24* (0.14)	-0.13 (0.13)	-0.25* (0.14)	-0.14 (0.13)	-0.24* (0.15)
Men of age 65 and above	0.11 (0.32)	- -	- -	- -	- -	- -	- -
<i>B. Nearest neighbor matching</i>							
All household members	-0.04 (0.03)	-0.05 (0.03)	-0.02 (0.04)	-0.05* (0.03)	-0.04 (0.04)	-0.01 (0.03)	-0.10** (0.04)
Girls of age below 15	0.02 (0.04)	0.06 (0.04)	0.01 (0.06)	0.01 (0.05)	0.01 (0.05)	0.07 (0.04)	-0.04 (0.07)
Women of age 15-34	-0.17** (0.07)	-0.15* (0.08)	-0.21* (0.11)	-0.16* (0.09)	-0.28** (0.11)	-0.10 (0.09)	-0.27** (0.12)
Women of age 35-64	-0.18 (0.18)	-0.31 (0.21)	0.00 (0.22)	-0.44 (0.28)	-0.08 (0.23)	-0.07 (0.22)	-0.08 (0.23)
Women of age 65 and above	0.07 (0.13)	0.08 (0.17)	0.33 (0.24)	0.11 (0.19)	0.25 (0.23)	0.18 (0.20)	0.00 (0.24)
Boys of age below 15	-0.03 (0.04)	-0.04 (0.05)	0.00 (0.05)	-0.07 (0.05)	-0.01 (0.07)	-0.03 (0.05)	-0.11* (0.06)
Men of age 15-34	-0.08 (0.08)	-0.10 (0.08)	-0.05 (0.11)	-0.07 (0.09)	-0.19* (0.11)	-0.03 (0.09)	-0.22* (0.13)
Men of age 35-64	-0.26* (0.15)	-0.31* (0.18)	-0.17 (0.16)	-0.15 (0.15)	-0.31* (0.18)	-0.27 (0.18)	-0.36* (0.19)
Men of age 65 and above	0.00 (0.40)	- -	- -	- -	- -	- -	- -

Table 4.20. Dependent variable: change in the number of days of working on the household farm; agricultural households

	1 = Sent a migrant	1 = Sent a female migrant	1 = Sent a male migrant	1 = Sent a migrant to a rural area	1 = Sent a migrant to an urban area	1 = Sent a migrant of age 15-20	1 = Sent a migrant of age 21-34
<i>A. OLS</i>							
All household members	1.28 (1.73)	3.48* (2.01)	-1.59 (2.56)	4.27** (2.11)	-3.49 (2.53)	1.50 (2.00)	2.84 (2.65)
Girls of age below 15	0.37 (2.03)	2.10 (2.31)	-4.89 (3.11)	0.07 (2.49)	1.19 (3.01)	0.09 (2.27)	0.26 (3.13)
Women of age 15-34	-5.91 (5.12)	-6.27 (6.03)	-6.46 (7.43)	2.74 (6.16)	-20.13*** (7.53)	-5.61 (5.96)	-8.45 (8.11)
Women of age 35-64	9.50 (6.04)	10.50 (7.11)	12.19 (8.50)	14.27* (7.64)	4.93 (8.20)	9.13 (6.80)	17.45* (9.79)
Women of age 65 and above	-0.50 (13.24)	4.00 (15.71)	14.55 (19.13)	7.31 (15.76)	7.40 (19.88)	-0.48 (16.07)	11.41 (18.98)
Boys of age below 15	-1.01 (1.64)	-0.16 (1.93)	-3.69 (2.53)	-2.25 (1.98)	0.24 (2.54)	-0.86 (1.93)	-3.41 (2.57)
Men of age 15-34	-1.65 (4.42)	2.27 (5.19)	-8.11 (6.63)	3.10 (5.41)	-10.60 (6.56)	-1.89 (5.17)	2.22 (6.91)
Men of age 35-64	-0.49 (6.46)	3.43 (7.44)	-9.96 (9.63)	4.47 (7.95)	-10.06 (9.23)	-5.43 (7.48)	14.72 (10.09)
Men of age 65 and above	32.70** (15.67)	57.09*** (19.40)	21.44 (21.45)	38.70* (19.85)	22.73 (22.98)	48.97** (21.07)	28.59 (19.73)
<i>B. Nearest neighbor matching</i>							
All household members	2.01 (2.37)	4.92* (2.69)	-0.12 (3.66)	3.86 (2.83)	-0.22 (3.60)	-1.54 (2.78)	7.36* (3.80)
Girls of age below 15	3.19 (2.99)	4.77 (3.31)	-0.96 (4.51)	-0.00 (3.91)	7.47** (3.68)	1.85 (3.69)	3.31 (4.69)
Women of age 15-34	-17.66** (7.06)	-18.36** (8.44)	-6.03 (9.65)	-11.59 (8.77)	-22.41** (9.48)	-19.45** (8.41)	-18.20 (11.69)
Women of age 35-64	3.67 (8.32)	3.91 (10.16)	9.90 (10.30)	7.73 (10.62)	2.77 (11.28)	-1.28 (9.45)	29.25** (12.36)
Women of age 65 and above	-2.14 (16.82)	0.07 (19.63)	-1.89 (28.83)	-1.93 (20.99)	-0.12 (27.93)	-6.46 (21.00)	-1.83 (22.99)
Boys of age below 15	0.39 (2.24)	0.05 (2.53)	0.26 (3.79)	-2.48 (2.56)	3.20 (4.14)	-1.83 (2.64)	-0.98 (3.16)
Men of age 15-34	4.27 (5.70)	9.78 (6.72)	-3.62 (7.59)	10.73 (6.74)	-2.45 (7.96)	3.17 (6.57)	14.95 (9.69)
Men of age 35-64	0.69 (8.77)	3.30 (9.79)	-12.18 (13.19)	8.82 (11.50)	-7.62 (12.11)	-11.91 (9.67)	32.14** (15.34)
Men of age 65 and above	29.50 (19.59)	43.79* (23.22)	24.53 (30.64)	41.65 (27.41)	22.31 (29.07)	45.10* (26.66)	48.14* (26.42)

REFERENCES

REFERENCES

- Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens. 2004. Implementing matching estimators for average treatment effects in Stata. *The Stata Journal* 4(3): 290-311.
- Abay, K. A., W. Asnake, H. Ayalew, J. Chamberlin, and J. Sumberg. 2021. Landscapes of opportunity: Patterns of young people's engagement with the rural economy in Sub-Saharan Africa. *The Journal of Development Studies* 57(4): 594-613.
- Abebaw, D., A. Admassie, H. Kassa, and C. Padoch. 2019. Does rural outmigration affect investment in agriculture? Evidence from Ethiopia. *Migration and Development* 10(1): 144-168.
- Adams, R. H. 2011. Evaluating the economic impact of international remittances on developing countries using household surveys: A literature review. *Journal of Development Studies* 47(6): 809-828.
- Agadjanian, V., C. Arnaldo, and B. Cau. 2011. Health costs of wealth gains: Labor migration and perceptions of HIV/AIDS risks in Mozambique. *Social Forces* 89(4): 1097-1118.
- Agadjanian, V., and S. R. Hayford. 2018. Men's migration, women's autonomy, and union dissolution in rural Mozambique. *Journal of Family Issues* 39(5): 1236-1257.
- Antman, F. M. 2011. International migration and gender discrimination among children left behind. *The American Economic Review: Papers & Proceedings* 101(3): 645-649.
- Antman, F. M. 2011. The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations? *Journal of Development Economics* 96: 200-208.
- Antman, F. M. 2012. Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics* 25: 1187-1214.
- Antman, F. M. 2012. The impact of migration on family left behind. In A. F. Constant and K. F. Zimmerman (eds.) *International Handbook on the Economics of Migration*. Cheltenham, UK, and Northampton, MA, USA: Edward Elgar, pp. 293-308.
- Ao, X., D. Jiang, and Z. Zhao. 2016. The impact of rural-urban migration on the health of the left-behind parents. *China Economic Review* 37: 126-139.
- Beegle, K., J. De Weerdt, and S. Dercon. 2011. Migration and economic mobility in Tanzania: Evidence from a tracking survey. *The Review of Economics and Statistics* 93(3): 1010-1033.
- Bennett, R., V. Hosegood, M. Newell, and N. McGrath. 2015. An approach to measuring dispersed families with a particular focus on children 'left behind' by migrant parents: Findings from rural South Africa. *Population, Space and Place* 21: 322-334.

- Bertoli, S., and E. Murard. 2020. Migration and co-residence choices: Evidence from Mexico. *Journal of Development Economics* 142: 102330.
- Bridges, S., L. Fox, A. Gaggero, and T. Owens. 2016. Youth unemployment and earnings in Africa: Evidence from Tanzanian retrospective data. *Journal of African Economies* 26(2): 119-139.
- Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *The Stata Journal* 13(3): 407-450.
- Chang, H., X. Dong, and F. MacPhail. 2011. Labor migration and time use patterns of the left-behind children and elderly in rural China. *World Development* 39(12): 2199-2210.
- Chen, R., C. Ye, Y. Cai, X. Xing, and Q. Chen. 2014. The impact of rural out-migration on land use transition in China: Past, present, and trend. *Land Use Policy* 40: 101-110.
- Collier, P., and S. Dercon. 2014. African agriculture in 50 years: Smallholders in a rapidly changing world? *World Development* 63: 92-101.
- Davis, J., and D. Lopez-Carr. 2014. Migration, remittances and smallholder decision-making: Implications for land use and livelihood change in Central America. *Land Use Policy* 36: 319-329.
- de Brauw, A., and J. Giles. 2018. Migrant labor markets and the welfare of rural households in the developing world: Evidence from China. *The World Bank Economic Review* 32(1): 1-18.
- Démurger, S., and X. Wang. 2016. Remittances and expenditure patterns of the left behinds in rural China. *China Economic Review* 37: 177-190.
- Dinbabo, M. F., C. Mensah, and M. N. Belebema. 2017. Diversity of rural migrants' profiles. In Mercandalli, S., and B. Losch (eds.) *Rural Africa in Motion. Dynamics and Drivers of Migration South of the Sahara*. Rome: FAO and CIRAD, pp. 24-25.
- Dinkelman, T., and M. Mariotti. 2016. The long-run effects of labor migration on human Capital formation in communities of origin. *American Economic Journal: Applied Economics* 8(4): 1-35.
- Dustmann, C., and A. Okatenko. 2014. Out-migration, wealth constraints, and the quality of local amenities. *Journal of Development Economics* 110: 52-63.
- Filmer, D., and L. Fox. 2014. *Youth Employment in Sub-Saharan Africa*. Africa Development Series. Washington, D. C.: World Bank.
- Fox, L., L. W. Senbet, and W. Simbanegavi. 2016. Youth employment in Sub-Saharan Africa: Challenges, constraints, and opportunities. *Journal of African Economies* 25(AERC supplement 1): i3-i15.

- Fox, L., and A. Thomas. 2016. Africa's got work to do: A diagnostic of youth employment challenges in Sub-Saharan Africa. *Journal of African Economies* 25(AERC supplement 1): i16-i36.
- Gollin, D. 2014. The Lewis model: A 60-year retrospective. *Journal of Economic Perspectives* 28(3): 71-88.
- Graeub, B. E., M. J. Chappell, H. Wittman, S. Ledermann, R. B. Kerr, and B. Gemmill-Herren. 2016. The state of family farms in the world. *World Development* 87: 1-15.
- Gray, C. L., and R. E. Bilsborrow. 2014. Consequences of out-migration for land use in rural Ecuador. *Land Use Policy* 36: 182-191.
- Harris, J. R., and M. P. Todaro. 1970. Migration, unemployment and development: A two-sector analysis. *The American Economic Review* 60: 126-142.
- He, C., and J. Ye. 2014. Lonely sunsets: Impacts of rural-urban migration on the left-behind elderly in rural China. *Population, Space, and Place* 20: 352-369.
- Justino, P., and O. Shemyakina. 2012. Remittances and labor supply in post-conflict Tajikistan. *IZA Journal of Labor & Development* 1(1): 1-28.
- Kangmennaang, J., R. Bezner-Kerr, and I. Luginaah. 2018. Impact of migration and remittances on household welfare among rural households in Northern and Central Malawi. *Migration and Development* 7(1): 55-71.
- Klasen, S., and I. Woolard. 2008. Surviving unemployment without state support: unemployment and household formation in South Africa. *Journal of African Economies* 18(1): 1-51.
- Kuhn, R., B. Everett, and R. Silvey. 2011. The effects of children's migration on elderly kin's health: A counterfactual approach. *Demography* 48: 183-209.
- Lewis, W. A. 1954. Economic development with unlimited supplies of labour. *The Manchester School* 22(2): 139-191.
- Lucas, R. E. B. 2016. Internal migration in developing economies: An overview of recent evidence. *Geopolitics, History, and International Relations*, 8(2), 159-191.
- Luis, J. S., M. F. Rola-Rubzen, T. R. Paris, and V. O. Pede. 2015. Rural labor outmigration and gender dimension in an assessment of farm technical efficiency: A case study in selected rice villages in the Philippines. *Asian Journal of Agriculture and Development* 12(1): 53-65.
- Mueller, V., C. Doss, and A. Quisumbing. 2018. Youth migration and labour constraints in African agrarian households. *The Journal of Development Studies* 54(5): 875-894.
- Murard, E. 2016. Consumption and leisure: The welfare impact of migration on family left behind. *IZA Discussion Paper* No. 10305.

- Qin, H., and T. F. Liao. 2016. Labor out-migration and agricultural change in rural China: A systematic review and meta-analysis. *Journal of Rural Studies* 47: 533-541.
- Radel, C., B. Schmook, and S. McCandless. 2010. Environment, transnational labor migration, and gender: Case studies from Southern Yucatán, Mexico and Vermont, USA. *Population and Environment* 32(2-3): 177-197.
- Ranis, G., and J. C. H. Fei. 1961. A theory of economic development. *The American Economic Review* 51(4): 533-565.
- Roubaud, F., and C. Torelli. 2013. Employment, unemployment and working conditions in urban labor markets of Sub-Saharan Africa: Main stylized facts. In P. de Vreyer and F. Roubaud, eds. *Urban Labor Markets in Sub-Saharan Africa*. Washington, D. C.: World Bank, pp. 37-79.
- Shrestha, S. A., and N. Palaniswamy. 2017. Sibling rivalry and gender gap: Intrahousehold substitution of male and female educational investments from male migration prospects. *Journal of Population Economics* 30(4): 1355-1380.
- Tanle, A. 2018. Concerns and intentions among young migrants in the shoe-shine business in the Cape Coast Metropolis, Ghana. *Ghana Journal of Development Studies* 15(1): 37-54.
- Taylor, J. E. 1999. The new economics of labour migration and the role of remittances in the migration process. *International Migration* 37(1): 63-88.
- Taylor, M. J., M. Aguelar-Støen, E. Castellanos, M. J. Moran-Taylor, and K. Gerkin. 2016. International migration, land use change and the environment in Ixcán, Guatemala. *Land Use Policy* 54: 290-301.
- Todaro, M. P. 1980. Internal migration in developing countries: A survey. In R. A. Easterlin, ed. *Population and Economic Change in Developing Countries*. University of Chicago Press, pp. 361-402.
- World Bank. 2017. *Living Standards Measurement Study – Integrated Surveys on Agriculture, Tanzania*.
Wave 1 (2008-2009) retrieved from <http://microdata.worldbank.org/index.php/catalog/76>.
Wave 2 (2010-2011) retrieved from <http://microdata.worldbank.org/index.php/catalog/1050>.
Wave 3 (2012-2013) retrieved from <http://microdata.worldbank.org/index.php/catalog/2252>.
Wave 4 (2014-2015) retrieved from <http://microdata.worldbank.org/index.php/catalog/2862>.
- Wouterse, F., and J. E. Taylor. 2008. Migration and income diversification: Evidence from Burkina Faso. *World Development* 36(4): 625-640.
- Xu, H. 2017. The time use pattern and labour supply of the left behind spouse and children in rural China. *China Economic Review* 46: S77-S101.
- Ye, J., C. Wang, H. Wu, C. He, and J. Liu. 2013. Internal migration and left-behind populations in China. *The Journal of Peasant Studies* 40(6): 1119-1146.

5. CONCLUSION

This study revolves around the issues of internal migration of people of age 15-34 from rural Tanzania, a phenomenon that is frequent yet understudied. I introduce a new categorization of location types on the rural-urban spectrum and test whether using this wider set of destinations is beneficial to our understanding of migration patterns. I document the direction of migration flows, disaggregate them by age and gender, and shed light on the dominance of low-density rural areas as the main migration destination. I look at four to six destination types and find several distinct migration flows: to remote low-density rural areas, to more densely populated rural and peri-urban areas, to secondary towns, and to cities.

I find that certain factors associated with migration decision are specific to destination choice – a fact hidden in a generalized model of migration with a simple rural/urban categorization. For example, the observed negative impact of agricultural shocks on the probability to migrate is only important for the probability to move to a high-density rural area when a larger set of destinations is considered. At the same time, having a history of temporary migration during a year prior to the survey positively affects the probability to move in general, with no specific impact on destination choice. Other factors, like lack of education and remoteness of the origin location positively impact the probability to move to a low-density rural area and negatively affect the probability to move to a town, while a simpler model shows no effect. The impact of some factors is so strong that it appears in the binary choice (to move or to stay in place): for example, having main occupation in agriculture shows as having a negative effect on the probability to migrate, while a model with a wider set of destinations finds no effect on the probability to move to any type of rural area, a strong negative correlation with the probability to move to a city, and a weaker association with migration to other urban destinations.

The same set of destinations is useful when looking at the impact of migration on occupational transitions. I compare several employment outcomes of migrants to those of non-migrants using matching based on individual, household, and community characteristics. I find migration to any destination, including low-density rural areas, to be associated with a shift towards non-agricultural wage job and self-employment. At the same time, the rates of shifting from main occupation in agriculture to non-agricultural employment observed among rural-to-urban migrants are underwhelming, especially among migrants to cities. There are two reasons for this phenomenon: (i) many migrants to cities did not have main occupation in farming at baseline (many of them were students⁸⁴), and (ii) among those who had main occupation in agriculture and moved to a city, many people chose occupation other than non-agricultural wage job or self-employment (instead, they chose household maintenance⁸⁵, unemployment, or staying in agriculture). Hence, migration to towns is the main driver of occupational shifts tied to structural transformation, although some farmers who move to peri-urban areas and towns maintain their main occupation in agriculture after migration. The rates of underemployment and unemployment among migrants to peri-urban areas and cities respectively are alarming. At the same time, migration to low-density rural areas is associated with a higher chance to become engaged in work for people who were not engaged in work at baseline.

At baseline, youth who will move in the next four years spend less time on the household farm than youth who will not move: migrants are, on average, younger and are more likely to be at school. Still, I find that outmigration is related to a reduction in labor supply that is not

⁸⁴ Average age of those who moved to a city is much smaller than of those who picked a different destination or did not move.

⁸⁵ Among migrants to cities, the share of men participating in household maintenance is the highest compared to migrants to any other destination and non-migrants, where the participation rates are almost zero among men.

covered by the non-migrant household members despite an increase in labor supply of some groups differentiated by gender and age. Women of age 35-64 are the ones who significantly increase their time spent on the household farm, and elderly household members are less likely to exit agriculture. On the other hand, households that experienced outmigration of youth are more likely to attract new household members, while new members in these households supply less labor to the farm. Children and youth, especially girls, both new and non-migrant, are more likely to be in school and spend less time on farming. The characteristics of the migrant also matter for the impact of outmigration on the household's livelihood. Female migrants, rural-destined migrants, and migrants in the age group of 25-34 on average contributed more time to the household farm, and their outmigration is associated with more drastic changes to the time use of the non-migrant household members. At the same time, outmigration of people of age 25-34 is correlated with an increase in the use of hired labor and an increase in farm size, which can be explained by an increase in remittances.

My work makes several contributions to the literature on migration in developing countries. It broadens the conceptualization of migration decision and studies both causes and consequences of migration considering various destinations on the rural-urban spectrum. Hence, it helps the stream of literature that shows how different aspects of migration flows appear once we step away from the binary approach and distinguish migration destinations (Lucas, 2016). It contributes to the ongoing discussion about the role of secondary towns in migration from rural areas (Christiaensen and Todo, 2014; Ingelaere et al., 2018). It adds to the growing interest in peri-urban areas (Mueller et al., 2018a; Chen and Zhao, 2017; Ward and Shackleton, 2016) and stresses the importance of rural destinations. I test different definitions for location types, based on population density, built-up area density, distance to the nearest town, and access to

amenities. The use of administrative definitions of “rural” and “urban” was shown to distort the observations on urbanization (Potts, 2017a; Potts, 2017b), and I show how it can affect our perception of migration flows. I also contribute to the growing literature on the impacts of internal migration of youth on the livelihood of households left behind in the countries of Sub-Saharan Africa (Mueller, Doss, and Quisumbing, 2018), adding important observations on the attraction of new household members associated with outmigration.

REFERENCES

REFERENCES

- Chen, C., and M. Zhao. 2017. The undermining of rural labor out-migration by household strategies in China's migrant-sending areas: The case of Nanyang, Henan province. *Cities* 60: 446-453.
- Christiaensen, L., and Y. Todo. 2014. Poverty reduction during the rural-urban transformation – The role of the missing middle. *World Development* 63: 43-58.
- Ingelaere, B., L. Christiaensen, J. De Weerdt, and R. Kanbur. 2018. Why secondary towns can be important for poverty reduction – A migrant perspective. *World Development* 105: 273-282.
- Lucas, R. 2016. Internal migration in developing economies: An overview of recent evidence. *Geopolitics, History, and International Relations* 8(2): 159-191.
- Mueller, V., C. Doss, and A. Quisumbing. 2018. Youth migration and labour constraints in African agrarian households. *The Journal of Development Studies* 54(5): 875-894.
- Mueller, V., E. Schmidt, N. Lozano, and S. Murray. 2018. Implications of migration on employment and occupational transitions in Tanzania. *International Regional Science Review*: 1-26.
- Potts, D. 2017. Conflict and collisions in Sub-Saharan African urban definitions: interpreting recent urbanization data from Kenya. *World Development* 97: 67-78.
- Potts, D. 2017. Urban data and definitions in Sub-Saharan Africa: Mismatches between the pace of urbanization and employment and livelihood change. *Urban Studies* 55(5): 965-986.
- Ward, C. D., and C. M. Shackleton. 2016. Natural resource use, incomes, and poverty along the rural-urban continuum of two medium-sized, South African towns. *World Development* 78: 80-93.