East Africa Research Papers in Economics and Finance

East Africa Collaborative Ph.D. Program in Economics and Management

Spatial Spill-overs and Households Involvement in the Non-Farm Sector: Evidence from Rural Rwanda

Pia NILSSON

East Africa Research Papers in Economics and Finance

EARP-EF No. 2016:05

Jönköping International Business School (JIBS),
Jönköping University, P.O. Box 1026,
SE-551 11 Jönköping, Sweden,
Web: http://www.ju.se/earp, E-mail: EARP@ju.se

Preface

East Africa Research Papers in Economics and Finance is a series linked to the collaborative PhD program in Economics and Management among East Africa national universities. The program was initiated and is coordinated by the Jönköping International Business School (JIBS) at Jönköping University, Sweden, with the objective of increasing local capacity in teaching, supervision, research and management of PhD programs at the participating universities. The program is financed by the Swedish International Development Cooperation Agency (SIDA).

East Africa Research Papers is intended to serve as an outlet for publishing theoretical, methodological and applied research covering various aspects of the East African economies, especially those related to regional economic integration, national and regional economic development and openness, movement of goods, capital and labor, as well as studies on industry, agriculture, services sector and governance and institutions. In particular, submission of studies analyzing state-of-the-art research in areas of labor, technology, education, health, well-being, transport, energy, resources extraction, population and its movements, tourism, as well as development infrastructure and related issues and discussion of their implications and possible alternative policies are welcome.

The objective is to increase research capacity and quality, to promote research and collaboration in research, to share gained insights into important policy issues and to acquire a balanced viewpoint of economics and financial policymaking which enables us to identify the economic problems accurately and to come up with optimal and effective guidelines for decision makers. Another important aim of the series is to facilitate communication with development cooperation agencies, external research institutes, individual researchers and policymakers in the East Africa region.

Research disseminated through this series may include views on economic policy and development, but the series will not take any institutional policy positions. Thus, any opinions expressed in this series will be those of the author(s) and not necessarily the Research Papers Series.

Editor: Almas Heshmati
Professor of Economics
Jönköping International Business School (JIBS),
Jönköping University, Room B5017,
P.O. Box 1026, SE-551 11 Jönköping, Sweden,
E-mail: Almas.Heshmati@ju.se

Assisting Editor: Olivier Habimana
Candidate for PhD in Economics
College of Business and Economics, University of Rwanda
E-mail: Olivier.Habimana@ju.se

Spatial Spill-overs and Households Involvement in the Non-Farm Sector: Evidence from Rural Rwanda

Pia NILSSON

Jönköping International Business School,
Jönköping University, Sweden.
E-mail: pia.nilsson@ju.se

Abstract

This paper studies the effects of external economies of scale on households' involvement in the non-farm sector. The focus of the paper is on the type of knowledge spill-overs that occur at the level of individuals; these relate to learning processes and non-market interactions. Nationally representative data on 8,100 households surveyed in 2006 and 2009 are used and unobserved heterogeneity and spatial dependencies are modelled by employing a multi-level model and instruments in the form of clustered cantered means. The findings show that in addition to other household characteristics related to education, asset endowments and credit availability, measures of agglomeration economies are positively associated with smallholders' degree of involvement in non-farm activities. The results indicate that there exist significant scale efficiencies associated with local markets and that an important part of the capacity to diversify lies outside households, residing instead in their locations.

Keywords: Rwanda; income diversification; non-farm income; agglomeration economies; three-level multi-level model.

JEL Classification Codes: R12; R20; O12

1. Introduction

Literature focuses on the role played by non-farm diversification as a means of improving sustainability and economic growth in rural Africa (Barrett et al., 2001; Bigsten, 1983; Ellis, 1998, 2000; Reardon, 1997; Smith et al., 2001; Weldegebriel et al., 2015). Diversification beyond agriculture is often pointed out as a significant driver of rural growth and studies argue that diversification into non-farm business and wage work is able to protect farmers' incomes (Haggblade et al., 2007). In literature on non-farm diversification in Africa a lot of attention is devoted to the determinants of diversification and the associations between diversification, income growth and productivity. Hence, there exists a fairly good understanding of the role played by factors such as education, access to markets, credit availability and external shocks in explaining the degree to which rural households and firms involve in the non-farm sector (Bigsten, 1996; Barrett et al., 2005; Rijkers and Soderbom, 2013). Significantly less attention has been devoted to the potential importance of external economies of scale in the context of non-farm growth in sub-Saharan Africa (Owoo and Naude, 2016; Rijkers et al., 2010). This is despite the large number of studies that have looked into similar matters from the perspective of advanced economies and which document significant spill-overs attached to agglomeration economies in both urban and rural regions (Artz et al., 2016; Duranton and Puga, 2004; Rosenthal and Strange, 2004).

Much of the rural economy in Rwanda is a subsistence economy and households typically have few alternative sources of income besides agriculture (Dabalen et al., 2004). Although non-farm incomes are shown to account for a substantial share of farmers' incomes across rural Africa (between 40 and 50 per cent), such income accounts for a relatively small share in Rwanda where households have few income sources besides agriculture (Davis et al., 2014).[1] The important challenges that need to be addressed are how to promote the creation of rural non-farm enterprises and how to enhance households' opportunities to diversify their incomes.

This study focuses on the role played by spatial spill-overs on the degree to which households in rural Rwanda involve in the non-farm sector, a perspective that has received little attention in literature. Although the influence of spatial spill-overs on growth in the non-farm sector has been addressed in previous studies on sub-Saharan Africa, these focus exclusively on firm and enterprise productivity. Rijkers et al., (2010) find that Ethiopian manufacturing firms in urban areas have larger productivity as compared to those in rural areas. Owoo and Naudé, (2016) find a positive association between non-farm enterprise productivity and co-location in Ethiopia and Nigeria. Ali and Peerlings, (2011) focus on the handloom industry in Ethiopia and find a positive association between clustering and productivity at the industry level. Together, these studies point at agglomeration spill-overs as a significant determinant of productivity at the enterprise and industry levels in the context of sub-Saharan Africa.

This study makes a contribution to literature on diversification and spatial spill-overs and builds on prior research in several ways. The focus is on measures of agglomeration

[1] The share of non-farm income in Rwanda has been estimated at around 20 per cent (Andre and Platteau, 1998).

economies that have been associated with both urban and rural growth (Artz et al., 2016), but rarely applied in a study of income diversification in the context of rural Africa. Several measures are used to indicate the type of density driven externalities that can be associated with non-market interactions. The measure in focus is calculated with respect to the number of workers in the non-farm sector in relation to the nation as a whole, and is interpreted as a measure of local clustering in non-farm activities. Spatial spill-overs are hypothesized to increase smallholders' capacity to diversify as they offer opportunities for learning, sharing and other non-market interactions to take place through the process of knowledge diffusion (Duranton and Puga, 2004; Hoover, 1937; Ohlin, 1933). They are also hypothesized to provide opportunities for market interactions through improved connectivity between customers and suppliers through thicker markets (Fujita et al., 2001).

The empirical analysis uses nationally representative household-level data obtained from two rounds of the Comprehensive Food Security and Vulnerability Analysis (CFSVA). A total of 8,100 households surveyed during 2006 and 2009 are included in the sample. Since agglomeration externalities tend to attenuate very quickly with distance (Andersson et al., 2014; Arzagi and Henderson, 2008), their effects are place-based and might be critical predictors within rather than between regions. The implication is that spatial spill-overs related to agglomeration may not be reflected accurately in aggregate (regional) economic indicators. Instead they require a spatially disaggregated scale of analysis. This paper applies a three-level multi-level model that allows a study of the between-level and within-level affects in several geographical units (for example, the local level and the more aggregated regional level). Spatial spill-overs related to agglomeration effects are thus analysed more precisely as their effects are allowed to vary in geography.

This paper is structured as follows. The second part reviews relevant literature on income diversification and argues for the importance of agglomeration effects in the context of sub-Saharan Africa. The third part describes the data and the empirical approach followed for testing the outlined hypotheses and summarizes some key facts observed in the data. The fourth part presents the results and discusses their relevance in relation to theory and prior literature. The fifth part gives a conclusion.

2. Background and theoretical framework

More than 75 per cent of the population in Rwanda depends on agriculture for its livelihood, combining small-scale food cropping with livestock rearing. The average population density in Rwanda is among the highest in Africa and households face constraints related to land scarcity, dependence on rain fed agriculture, adverse weather conditions (for example, droughts and erosions) and low technological developments (Barrett et al., 2005; Kim and Heshmati, 2015). The absence of functioning markets for insurance imply that Rwandan households are vulnerable to changes in their environment (Ali et al., 2014). Households which operate under such conditions have several motives to engage in livelihood diversification and one way for them to address uncertainty and risk is to allocate resources over several income generating activities (Barrett et al., 2001; Ellis, 2000). Regarded from a macro-perspective, agriculture as a share of aggregate output is typically seen to decline with overall growth in GDP per capita as countries undergo structural transformations, implying that the expansion of rural non-farm activities and income diversification are likely features of the process of economic development in rural Africa (Bigsten, 1983; Ellis, 1998, 2000). According to the International Monetary Fund's (IMF) Regional Economic Outlook (2015) most sub-Saharan countries will experience a demographic transition, implying both a rapid growing population and a rising share of the working age population. This will provide a transition that can offer opportunities if countries succeed in implementing supportive policies like creating jobs to absorb new entrants into the labour force. Hence, the non-farm sector has a key role to play in this transition as a source of employment generation and in furthering economic diversification (Davis et al., 2014).

2.1 Incentives for engaging in diversification

Theoretical studies argue that there are various channels through which income diversification may influence risks, uncertainties and incomes of rural households (Ellis, 1998; Reardon et al., 2000). Some diversification incentives are related to risk minimization and distress as households may be forced to engage in diversification to stabilize income and consumption flows as a way of hedging for or responding to changes in their environments. Incentives may also be driven by factors found in the surrounding geography and access to markets and other relevant opportunities (for example, knowledge, input markets and customers) may create improved possibilities for households to engage in non-farm activities to accumulate incomes (Ellis, 2000). In literature on diversification in sub-Saharan Africa, a lot of attention has been on the role of education and on access to markets and credit (Abdulai and CroleRees, 2001; Weldegebriel et al., 2015). Most of the studies have found that educational attainment and access to educational infrastructure have a positive and significant influence on non-agricultural earnings and lack of education is commonly highlighted as a key constraint that prevents households from diversifying (Ellis, 2000). Access to capital (through credit and remittances) and asset endowments are also commonly seen as key determinants of income diversification in the context of sub-Saharan Africa (Abdulai and CroleRees, 2001; Barrett et al., 2005; Bigsten, 1996; Bigsten and Tengstam, 2011; Isaksson, 2013; Smith et al., 2001).

Locational factors are important for the development of both agricultural and non-agricultural activities. Several people argue why location matters for the choice of economic activities that households engage in. There are also arguments for diversification including backward and forward linkages from agriculture (Collier and Dercon, 2014; Gollin et al., 2014) and arguments related to the new economic geography. Barrett et al., (2005) focus on rural households in Rwanda, Kenya and Cote d'Ivoire and find that locational factors play an important role in explaining households' capacity to diversify. Market access through roads or transportation facilities can lower transportation and transaction costs and provide access to market potential and non-agricultural jobs (Reardon et al., 2000). Being located closer to the market may also imply a greater possibility for the use of land as collateral as well as improved possibilities for its alternative uses (Capozza and Helsley, 1989; Plantinga et al., 2002). Altogether, a fairly good understanding exists on the various push and pull factors that influence households' degree of involvement in the non-farm sector.

Considerably less attention has been devoted to the potential role played by spatial spill-over effects, particularly those related to external economies of scale.

2.2 Agglomeration economies and income diversification

This framework draws attention to several perspectives set forth in research on agglomeration economies and argues for their importance in increasing rural households' capacity to diversify incomes. The concept of agglomeration economies dates back to Marshall, (1920) who argued that similar firms benefit from co-location as this reduces transport costs, provides access to workers with relevant skills and gives rise to knowledge spill-overs that increase growth in both the industry and the region as a whole. These benefits arise as a result of industrial specialization and occur among firms that are related in terms of input factors, technology and customers (Malmberg and Maskell, 2002). These types of agglomeration economies are often connected to the rise and growth of cities and specialized industrial clusters and they have been studied mostly in advanced economies.[2] As suggested by the term agglomeration economies and on a prima-facie basis, such agglomeration economies may seem relevant only for firms in urban regions.

A number of theoretical and empirical studies argue for a broader understanding of the concept of agglomeration economies and show that these are also valid for rural regions, albeit on a different scale (Artz et al., 2016; Naldi et al., 2015). In Porter's (2000) view, spatial spill-overs are not restricted to within industries, rather they occur in many types of industries, even in some very local ones. They are present in rural and urban areas, although those in urban areas tend to be more developed.

Studies highlight that many of the spill-overs that take place in regions occur as a result of non-market interactions in the form of knowledge interconnections that take place between individuals rather than between firms. The types of agglomeration economies that arise as a result of knowledge spill-overs are different compared to the Marshallian ones in that they refer to a pure externality that is bounded and attenuating in space (Andersson et al., 2014; Arzagi and Henderson, 2008).

Duranton and Puga, (2004) distinguish between three types of mechanisms behind agglomeration economies: sharing of (for example, fixed costs and risks), matching of workers with relevant skills and learning due to knowledge accumulation and spatial spill-overs. They also emphasize that the heterogeneity of workers is necessary for these spill-overs to take place and that regional diversities can give rise to agglomeration economies, which may stimulate innovation and growth. Hence, both the matching and the learning arguments emphasize individuals rather than firms, implying that knowledge and information may not spill over between firms per se, but between individuals who channel the knowledge to firms (Wixe and Andersson, 2015).

This view dates back to Ohlin, (1933), Hoover, (1937) and Jacobs, (1969), who argued that firms and individuals benefit from being located in diverse economic environments because these provide them access to a broad knowledge base and shared services and infrastructure

[2] See Duranton and Puga, (2004) and Rosenthal and Strange, (2004) for reviews.

through the scale of economic activity. This improves the potential for non-market interactions and cross-fertilization of ideas, which may spur innovation and growth. Diverse environments also provide a greater potential for market interactions through better connectivity between customers and suppliers and thicker and more diverse transport and communication links (including face-to-face contact) (Storper and Venables, 2004).

Face-to-face (F2F) contact is often cited as one of the most fundamental aspects of proximity as it facilitates learning and social skills, motivation and trust and provides an efficient way of communication (Charlot and Duranton, 2004). It is also argued that these interactions are particularly important in societies where information is imperfect, rapidly changing and difficult to diffuse (Porter, 2000), as in much of rural Africa. Storper and Venables, (2004) take this a step further and argue that F2F contact is necessary for positive externalities of agglomeration economies to materialize. What follows is that social interaction and physical contact may be a particularly important way of diffusing knowledge in societies characterized by undeveloped infrastructure for communication and information exchange (McCormick, 1999). It is thus likely that improved connectivity spurs social interactions and may assist households in gaining ideas, skills and information which increase their capacity to diversify their income sources. Moreover, given that there exist within-variations, there will be areas within rural regions with relatively more potential for matching, sharing and learning processes to take place. Individuals in rural regions may also have more to gain from an increase in agglomeration. In other words, when starting from a very small scale, the marginal effect from increased agglomeration may be larger in rural regions than it is for already urbanized cities (Naldi et al., 2015).

What follows from this discussion is that some arguments support the framing of agglomeration economies in terms of rural individuals or households; this is also the focus of this paper. Given that human capital plays an important role in the urbanization process (Henderson and Wang, 2005) these locational factors can be hypothesized to explain a significant part of households' capacities to engage in non-farm activities as they imply greater opportunities for both non-market and market interactions to take place.

2.3 Relevant empirical studies

Compared to the large number of studies that have been carried out in the United States and Europe, relatively few studies have attempted to estimate the effects of agglomeration economies with a focus on sub-Saharan economies using a disaggregated scale of analysis. Ali and Peerlings, (2011) focus on the handloom industry in Ethiopia and find a positive association between co-location and productivity, indicating localization economies within the industry. Siba et al., (2012) use panel data on manufacturing firms in Ethiopia and find that while new entries lead to higher competitive pressure in the local economy, agglomerations of similar firms is positively associated with productivity, indicating that clustering brings positive externalities. Owoo and Naude, (2016) focus on rural enterprises in Ethiopia and Nigeria and find evidence of productivity gains associated with the clustering of non-farm enterprises in both the countries. Rijkers et al., (2010) compare enterprise productivity differences between urban and rural Ethiopia. They find that enterprise productivity was higher in urban areas as compared to rural areas, indicating positive spatial spill-over effects related to urbanization economies. The findings in Dorosh and Thurlow, (2014) also point to the existence of urbanization economies in both Ethiopia and Uganda.

In relation to these studies which focus exclusively on enterprise productivity, this paper focuses on rural households and the potential importance of spatial spill-overs in explaining their degree of involvement in non-farm activities. The basic hypotheses are that market and non-market interactions that take place at the individual level are at least as important as those that take place between firms in providing a breeding ground for local knowledge spill-overs and cross-fertilization and in creating pre-conditions for households to diversify.

3. Data and empirical model

This paper uses data from two rounds of the Comprehensive Food Security and Vulnerability Analysis (CFSVA) in Rwanda for 2006 and 2009. The unit of analysis in all the empirical analyses is households and the focus is on characteristics that describe their external local conditions at the local and more aggregated regional levels. The terms local areas and regions are used to denote the sector and district levels, as illustrated in Appendix A.

CFSVA is a nationwide household survey that provides quantitative data on key perspectives of Rwandan households. Although there are components that are not consistently measured in the survey the information needed to study income diversification is consistent across the surveys. The components of particular interest in this study are those that hold information on location, income sources, households' ownership of assets and their access to credit and remittances; 8,100 households were included in the two surveys and the data used for estimations are structured as a repeated cross-sectional dataset where each cross-section includes a new sample of surveyed households. These data can be seen as comparative as they are drawn from a consistent set of higher-level units (local areas and regions), but they also have a longitudinal dimension as households are surveyed at different points in time. Although repeatedly surveyed households may be there in the sample, it is not possible to identify these, which rules out the use of a panel approach. This gives rise to the problem of unobserved heterogeneity, which is more challenging to mitigate when dealing with cross-sectional data. As an alternative, this paper employs a multi-level model in which households are nested in the two geographical units (see Appendix A) and in the two time points in which they are surveyed. Using this approach, a pseudo (geographical) panel is created in which the 8,100 households are nested into 702 local areas, 35 regions and two years.

Combining the surveys there are some aspects related to sampling that need to be addressed. CFSVA in 2009 covered both urban and rural households which was not the case in the previous survey. The 2006 survey was conducted only among rural households. To make these data comparable, households sampled in Kigali in 2009 were removed from the dataset (67 observations in districts Nyarugenge (11), Kicukiro (13) and Gasabo (43)).

The surveys are designed to collect data on a number of key attributes but they do not address external local and regional conditions that may influence households' capacity to diversify. The CFSVA dataset is therefore combined with data from other sources to obtain measures of market access and external local conditions. These data are obtained from the General Census of Population and Housing, Rwanda Meteorology Agency and the Rwanda Transport and Development Agency.

3.1 Estimated model

The empirical approach is to estimate a three-level multi-level model that allows for a simultaneous but separate analysis of spatial and temporal effects. Using this approach also enables one to distinguish among between-level and within-level effects at the two geographical levels, which can give an indication of the degree of attenuation.

A particular concern in this study is the presence of both level-1 and level-2 endogeneity with regard to income diversification. Level-1 endogeneity may arise from the inability to control for key idiosyncratic factors (for example, abilities) and level-2 endogeneity may occur if the random effects are correlated with a level-1 covariate, for example, when households' capacity to diversify and engage in non-farm income generating activities is influenced by factors that are common in the village or in the region. The topic of endogeneity in multi-level models has been discussed in several papers (cf. Skrondal and Rabe-Hesketh, 2004). Since multi-level models have at least one random intercept at each of the higher levels in the hierarchy there is a potential for endogeneity between these random intercepts and the covariate in focus.

A way to deal with level-2 endogeneity in the framework of multi-level modelling is by including instruments in the form of cantered cluster means of the endogenous covariates (Snijders and Berkhof, 2006). The rationale is that a purely within-variable, for example, a variable that varies only within clusters, is necessarily uncorrelated with any between-variable, constant within the cluster (Hausman and Taylor, 1982; Mundlak, 1978). The cantered clustered mean of a level-1 covariate is thus a potential instrumental variable that is both internal and uncorrelated with the error term. Following multi-level literature, a three-level model with endogenous covariates can be expressed as:

(1) $$y_{ijk} = \alpha\beta\boldsymbol{X}_{ijk} + \gamma CH_j + u_{kj} + u_k + \epsilon_{ijk}$$

$$\epsilon_{ijk}|\boldsymbol{X}_{ijk} \sim N\{0, \sigma^2\}$$

where households indexed *i* are nested within higher-level societal units *j* and districts *k* and where each j and k have random intercepts which are assumed independent (given the covariates) and normally distributed with zero mean and constant variance (Goldstein, 2003). The fixed part of the model $\boldsymbol{X}_{ijk}$ contains a vector of characteristics of households and their economic and natural environments hypothesized to influence their incomes; β is a vector of coefficients. Moreover, CH_j denotes the clustered means of the endogenous variables and γ their estimated coefficients. Hence, the fixed part of the model contains variables that can be either variable within *j* and *k* (*ij* or *ijk*) or invariant (*j* or *k*). The cluster mean of the endogenous variable and the deviation from the cluster mean cantered covariate are defined as:

(2) $$\overline{H}_j = \frac{1}{n_j}\sum H_{ij}$$

(3) $$CH_j = H_{ijk} - \overline{H}_j$$

The model in Eq 1 can also be expanded to allow for a simultaneous and separate analysis of cross-sectional and longitudinal effects. Skrondal and Rabe-Hesketh, (2008) suggest the inclusion of group mean cantered covariates as a method of distinguishing separate longitudinal and cross-sectional associations. Using this approach, households are clustered both in geography and in the time-points that they are surveyed in and spatial; temporal heterogeneity is modelled by allowing for serial correlation among the higher levels in the hierarchy (Browne and Goldstein, 2010; Snijders and Berkhof; 2006). A three-level multi-level model with endogenous covariates and separability between cross-sectional and longitudinal effects can be expressed as:

(4) $$y_{ijkt} = \alpha + \beta X_{ijkt}\gamma CH_j + \eta CH_t + u_{kjt} + u_{kt} + u_t + \epsilon_{ijkt}$$

$$\epsilon_{ijkt}|X_{ijkt} \sim N(0, \sigma^2)$$

where the clustered of $\overline{H}_j$ are identical to above and $\overline{H}_t = \frac{1}{n_j}\sum H_j$ for H_{jt}. The clustered mean centered covariate is defined as $CH_t = H_{ijkt} - \overline{H}_{jt}$. The group mean centered covariate is able to identify separate longitudinal and cross-sectional associations, calculated as the mean of the variable across the higher level units (*jk* or *kt*) for each **j** and *k*. Since the cross-sectional component of the district-level variable and the longitudinal component of the district year-level variable are orthogonal, their effects can be estimated separately. If γ or η are significant it indicates endogeneity across levels, which is absorbed by the instrumental variables and therefore does not bias the estimates (Snijders and Berkhof, 2008).

3.2 Variables and summary statistics

In the CFSVA surveys, households report their main income activities and the amount of total income generated from each activity. In the surveys it is possible to distinguish between 26 different activities. Based on the approach in Davis et al., (2014), incomes are grouped into seven categories: (A) income from agricultural production; (B) selling of agricultural products (C) livestock; (D) non-agricultural wages; (E) non-agricultural self-employment; (F) remittances and credit; (G) others.[3] Summary statistics for households' sources of income are presented in Table 1. These show that agricultural production accounts for the largest share (63 per cent) followed by non-agricultural wage work (17 per cent) and incomes from the sale of livestock (8.4 per cent).

[3] Others include transfers and other non-labor sources like aid and pension.

Table 1. Summary statistics of income shares

	Income share (mean)	Standard deviation
A. Agricultural production	0.631	0.345
B. Selling of agricultural products	0.023	0.125
C. Livestock	0.084	0.186
D. Non-agricultural wage work	0.170	0.298
E. Non-agricultural self-employment	0.063	0.195
F. Credit and remittances	0.002	0.036
G. Others	0.021	0.118

Following the approach in Davis et al., (2014), the income categories described earlier are used to create four dependent variables by aggregation into shares that reflect off-farm and non-farm incomes (Table 2).

Table 2. Summary statistics of dependent variables

Income categories	Income share (mean)	Standard deviation
Non-agricultural (D-G) off-farm	0.256	0.341
Non-agricultural (D,E) non-farm	0.234	0.330
Agricultural total (A-C)	0.743	0.341
Agricultural (A)	0.631	0.345

The other dependent variables are two measures of income diversification calculated with regard to the number of income activities of each household and the dispersion of those activities' shares in the total. These variables are calculated using the entropy approach (Shannon, 1948) in the following:

(5) $$H = -\sum a_{il} \ln(a_{il})$$

where a_{il} is the share of total income for household i and income generating activity l. The two measures are calculated to reflect diversification with respect to: i) diversification beyond agriculture $H_{nonagri}$ using $l = D.,,,G$ and ii) diversification within agriculture H_{agri} using $l = A.,,,C$. When the indexes take on high values it means that a given household i has a high degree of diversity with regard to either the total number of income generating activities, activities beyond agriculture or within agriculture, and when the value approaches the lower bound zero it implies increases in the extent of income concentration.

Following the approach in Owoo and Naude, (2016), Figure 1a-1d provide the results of an exploratory analysis of the spatial patterns observed in the data, for example, with regard to the clustering of non-agricultural and agricultural activities in rural Rwanda. The purpose of these preliminary and descriptive analyses is to provide a visual representation of the patterns and to identify areas with significant clusters in non-farm activities. To produce these maps the author relied on the Getis-Ord spatial statistical tool provided in ArcView and household level data was aggregated to the sector level.

Figure 1a-1d show returned GiZ scores from this analysis, classified using standard deviations (Getis and Ord, 1996). Areas in black (above 2.58 Std. Dev) denote the significant clusters evaluated based on the values of neighbouring areas and in relation to the national average. The figure shows that non-farm activity is significantly above the national average in areas that are located near the borders of the Democratic Republic of the Congo and around city but are near the border with Burundi. Clustering within agriculture seems to have different geographical patterns.

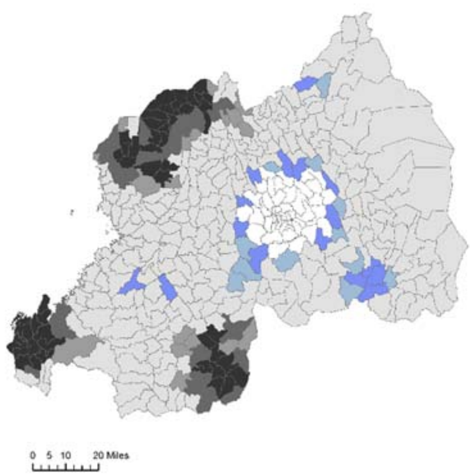

Figure 1a. Off-farm clusters (D-G)

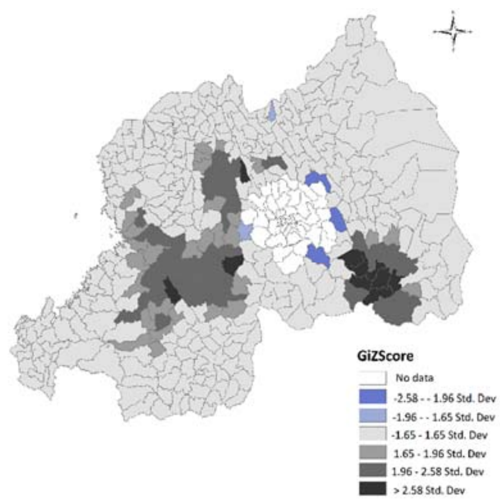

Figure 1b. Agricultural (total) clusters (A-C)

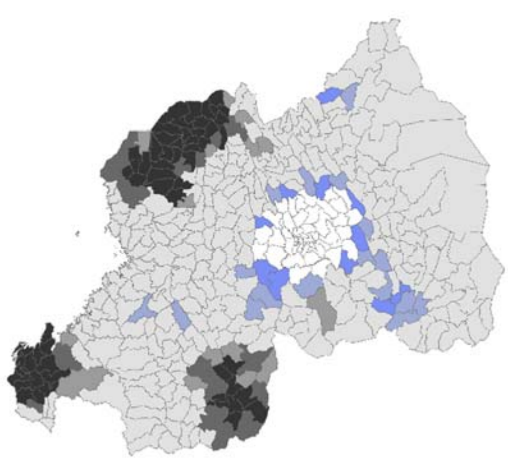

Figure 1c. Non-farm clusters (D,E)

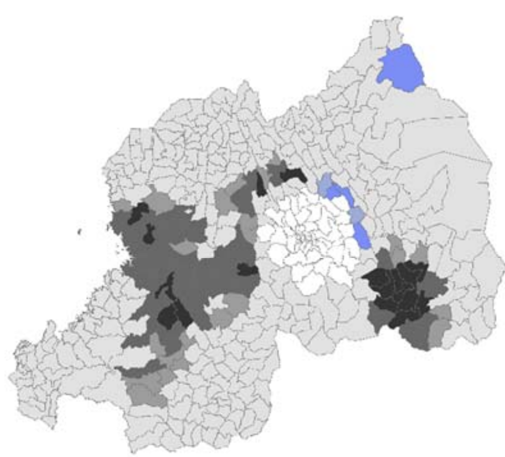

Figure 1d. Agricultural clusters (A)

Location and agglomeration

A location quotient (LQ) was used to measure specialization in the non-farm sector at the local level. The measure was calculated to reflect local specialization relative to the nation as a whole and with respect to the number of workers in the non-farm sector. The data used to calculate this measure came from the Housing and Population Census. The location quotient was calculated as:

(6) $$LQ_j = \frac{e_{s,j}/e_j}{e_s/e_n}$$

where $e_{s,j}$ denote the number of non-farm workers s in local area j, e_j denotes the total number of employees (regardless of industry) and e_s the number of non-farm workers in the nation k. Moreover, e_n denote the total number of workers in the nation n. If LQ is larger than one, the local area has a larger share of workers within the non-farm sector as compared to the national average, indicating that the area is more specialized than average in the non-farm sector. Population density at the local level is introduced to control for pure size effects and two (Euclidean) distance measures are calculated to capture the effects of market access to both the nearest small town and to Kigali. This differentiation may provide evidence on the importance of small towns as opposed to large cities in improving households' capacity to diversify (Davis et al., 2014).[4]

Household and location specific controls

A number of household and location specific control variables were included in the estimations, for example, measures of human capital (literacy, educational attainment and age), asset endowments, access to capital (through credit or remittances) and locational factors that improve the potential for agricultural production (climate, land, precipitation and soil quality). Together these determinants were hypothesized to improve households' capacity to access non-agricultural activities as they lower transaction costs and information barriers which provide access to financial capital and natural pre-conditions for agriculture (cf. Barrett et al., 2001). Control variables are summarized and defined in Table B1 and B2 in Appendix B.

4. Estimation results

The model in Eq 2 is estimated including the variables described earlier. Before introducing the explanatory variables in the model, the hierarchical structure was examined by estimating variances for random intercepts at the local, regional and temporal levels. This preliminary estimation provides information on how the proposed hierarchical structure relates to

[4] These variables are calculated using Geodata (road network layers) from the Rwanda Transport Development Agency and network analysis in ArcView.

dependent variables and validates the use of a multi-level model for these data. In a first step, the following unconditional model was estimated:

(7) $$y_{ijkt} = \alpha + u_{jkt} + u_{kt} + u_t + \varepsilon_{ijkt}$$

where y_{ijkt} is the share of income obtained from the different categories of households i in sector j, district k and year t and α is the overall constant. Moreover, u_{jkt} is the random intercept for sector j within district k and year t, u_{kt} is the random intercept of the regions and u_t is the random intercept of the years. The estimation results are given in Table 4; these show that the sample of 8,044 households is nested in 415 sectors, 30 districts and two years.[5] The between-level heterogeneity at the sector, district and yearly levels is significant at the 5 per cent level for each of the geographical units, suggesting that they add important information on locational and temporal aspects that are unobservable.

The same phenomenon can also be described by the intra-class correlation coefficient (ICC) that measures the degree of correlation among observations and how much of the total variance in the dependent variable can be assigned to the different levels. ICC ranges from zero to one, that is, it ranges from a grouping bearing no information to all units in a group being identical. Estimation results show that the sector level is able to explain most of the variances such that households located in the same sector resemble each other more as compared to those that share the same higher levels (districts or years). ICC for sectors ranges between 8 to 9 per cent, while ICC for districts and years is around 1 and 0.1 per cent respectively.

Although the district and year levels are shown to explain less of the variance in the dependent variables their random intercepts still explain a significant share of the variation (5 per cent or lower) in both the variables reflecting non-agricultural activities and will therefore be included in the analyses that follow.

Table 4. Estimated variances in unconditional models

	Coeff.	Std. Err.	ICC
Dependent: share off-farm income			
Fixed effects:			
Intercept	0.264*	0.015	
Random effects:			
Household	0.100*	0.001	
Sector	0.089*	0.001	0.147
District	0.010*	0.002	0.041
Year	0.001*	0.000	0.001
Dependent: share non-farm income			
Fixed effects:			
Intercept	0.240*	0.013	
Random effects:			

[5] See Appendix A for maps showing the size and distribution of district and sector units.

Household	0.095*	0.001	
Sector	0.076*	0.001	0.160
District	0.011*	0.001	0.039
Year	0.001	0.000	0.001

Note: * indicates significance at the 5 per cent level or lower.

4.1 Determinants of off-farm, non-farm and agricultural incomes

In a second step, explanatory variables were introduced in Eq 7 in accordance with Eq 4 and cluster means were included as instruments and to distinguish among within-level and between-level effects (Eq 2). The empirical approach is to select which clustered means to include on the basis of the Durbin-Wu-Hausman test statistic and using Wald tests rejected at the 5 per cent level. For comparison the models were also estimated both by including and excluding the clustered means, indicating a bias as a result of omitted endogenous covariates.[6]

Income shares, described in Table 1, were used as dependent variables and estimated in separate models, while controlling for key factors at the household, local and regional levels.[7] The results are presented in Tables 5 and 6. The focus is on the share of income from off-farm and non-farm activities, and the share of income from agricultural production (as indicated by category A in Table 6) is included mainly for comparison.

Starting with the results reported in Table 5, the main variables of interest are the measures of location and agglomeration. The results show that the share of income from off-farm and non-farm activities is positively associated with increases in the degree of local specialization in the non-farm sector, as indicated by the location quotient. Based on the magnitude of its clustered covariate (measured at the more aggregated district level), the within-effect appears to be stronger as compared to the between-effect. This indicates that spatial proximity to non-farm activities is important in explaining households' involvement in the non-farm sector in rural Rwanda. This may be due to non-market interactions in terms of knowledge spill-overs, collective learning and network effects as argued in the theoretical section (Duranton and Puga, 2004). A relatively large degree of local specialization in non-farm activities may also reflect improved opportunities for market interactions to take place through thicker markets and better connectivity between customers and suppliers (Charlot and Duranton, 2004).

These results are robust to the inclusion of population density, indicating that it is local specialization in non-farm activities as such that gives rise to the positive association. Population density is also a variable of interest because of its association with scale in economic activities and in its interpretation as a measure of urbanization economies (Frenken et al., 2007). The results show that population density is positively associated with non-agricultural activities. This may indicate that households can take advantage of being

[6] The Hausman test statistic is not significant at the 5 per cent level when the clustered covariates are included.
[7] Categories A-C not reported. Since (A-C) + (D-G) = 1 the categories A-C and D-G are exact opposites, for example, they have the same absolute value but opposite signs.

located in dense areas as this provides them access to a range of shared services and infrastructure, which are independent of non-farm activities. There is no significant association between these agglomeration measures and households' degree of involvement in agricultural activities as indicated by the share of income from category A (agricultural production) in Table 6. Rather, there seems to be a negative association between population density and share of income from agricultural production. This indicates that higher levels of density are associated with a higher competition for land (Boserup, 1965).

Results show negative coefficients for distance to the nearest town, indicating that income from off-farm, non-farm and agricultural activities diminishes at a greater distance from local markets. The coefficient for distance to Kigali is only significant in the model with non-farm income as the dependent variable, indicating that distance to both the nearest town and the largest town is important for diversification into wage employment and self-employed businesses (Davis et al., 2014). A particular interest in constructing these market access variables is to disentangle if the effects vary when one considers distance to the nearest small town, as compared to the distance to Kigali. The results show that the size of the market does matter for the degree of involvement in non-agricultural activities in that distance to nearest town seems to be a more relevant measure.

Table 5. Determinants of off-farm and non-farm incomes

	Off-farm (D-G)		Non-farm (D,E)	
Fixed effects	Coeff.	Std.Err.	Coeff.	Std.Err.
Household level predictors				
Household size (ln)	-0.014	0.009	0.016**	0.009
Age of head (ln)	-0.100***	0.013	-0.132***	0.012
Literacy (head)	0.012	0.009	0.007	0.007
Educational attainment (head)	0.171***	0.030	0.178***	0.029
Female head	-0.015	0.009	-0.024**	0.009
Credit and asset endowments				
Access to credit	0.014***	0.003	0.020***	0.009
Access to remittances	0.096***	0.017	0.086***	0.016
Agricultural assets	-0.001***	0.000	-0.001	0.000
Land (ha)	-0.006***	0.001	-0.005***	0.001
Crop diversity	-0.199***	0.008	-0.168***	0.008
Electricity	0.119***	0.024	0.133***	0.023
Transportation assets	0.063	0.043	0.052	0.030
ICT assets	0.042***	0.011	0.040***	0.010
Location and agglomeration				
LQ_j	0.023**	0.001	0.037***	0.001
Population density	0.033***	0.009	0.027***	0.008
Distance small town	-0.061**	0.004	-0.082***	0.001
Distance Kigali City	-0.001	0.004	-0.020**	0.001
Average precipitation	0.004	0.005	0.003	0.004
Soil quality	-0.041	0.087	-0.015	0.080
Clustered covariates				
LQ_K	0.009**	0.000	0.037***	0.001

Population density *(k)*	0.001**	0.000	0.002	0.003
Constant	0.671***	0.101	0.714***	0.094
Random effects				
u_{jkt}	0.088***	0.001	0.085***	0.001
u_{kt}	0.010**	0.009	0.007***	0.000
u_t	0.003	0.002	0.002	0.001
Sample size	8044		8044	

Note: ***, ** indicate statistical significance at the 1 and 5 per cent levels respectively. Sample weights are included in the estimations. The difference between the within and between effects w.r.t. the clustered covariates is confirmed using a Wald test of the null hypothesis that the coefficient of the cluster mean is zero is rejected at the 5 per cent level, the Wald statistic is 51.07 with df =7.

Table 6. Determinants of agricultural income

	Agricultural (A)	
Fixed effects	Coeff.	Std.Err.
Household level predictors		
Household size (ln)	-0.026***	0.009
Age of head (ln)	0.081***	0.013
Literacy (head)	0.009	0.008
Educational attainment (head)	-0.147***	0.031
Female head	0.023**	0.010
Credit and asset endowments		
Access to credit	-0.015**	0.008
Access to remittances	-0.095***	0.008
Agricultural assets	0.010***	0.001
Land (ha)	0.007***	0.001
Crop diversity	0.176***	0.008
Electricity	-0.131***	0.024
Transportation assets	-0.069	0.045
ICT assets	-0.057***	0.011
Location and agglomeration		
LQ_j	0.040	0.047
Population density	-0.021***	0.009
Distance small town	-0.011**	0.001
Distance Kigali City	0.030	0.021
Average precipitation	0.003**	0.000
Soil quality	0.027**	0.000
Clustered covariates		
LQ_K	0.001	0.001
Population density *(k)*	-0.023	0.020
Constant	0.331***	0.100
Random effects		
u_{jkt}	0.098***	0.001
u_{kt}	0.088***	0.001
u_t	0.003	0.010
Sample size	8044	

Note: ***, ** indicate statistical significance at the 1 and 5 per cent levels respectively. Sample weights are included in the estimations. The difference between the within and between effects w.r.t. the clustered covariates is confirmed using a Wald test of the null hypothesis that the coefficient of the cluster mean is zero is rejected at the 5 per cent level, the Wald statistic is 42.89 with df =8.

In summarizing the effects it seems that the degree of involvement in agricultural activities is driven by a different set of locational factors as compared to non-agricultural activities. The share of income from agricultural production seems to be determined by land access, agricultural asset endowments and natural pre-requisites for agriculture (soil quality and weather conditions).

The household level predictors are in line with expectations in that educational attainment and access to capital (through credit or remittances) are all positively associated with non-agricultural activities. These results point to evidence that is broadly consistent with theory and with empirical evidence from other sub-Saharan countries. The results also add to existing knowledge in showing a positive association between external economies of scale and households' degree of involvement in the non-farm sector.

4.2 Agglomeration and income diversification

Next the dependent variables were substituted for measures of diversification, calculated to reflect the dispersion of activities' shares in the total as defined by Eq 5. The main variables of interest are the measures of location and agglomeration and the results are presented in Table 7. The results are in line with those presented and discussed earlier, showing a positive association between off-farm diversification and local specialization in non-farm activities. Based on the coefficient of its clustered covariate, the between-effect appears to be stronger here also indicating that spatial proximity to non-farm activities in the immediate local area is more important in explaining households' degree of diversification. The results show a consistent and positive association between population density and income diversification with regard to diversification across off-farm activities. Altogether, the results show some sign of attenuation in geography in that the within-effect appears significant and positive throughout the estimations, whereas the clustered covariates have lower magnitudes and are in some cases insignificant. This indicates that spatial spill-overs related to agglomeration economies are very much place-based and critical predictors within rather than between regions (Andersson et al., 2014).[8]

Turning to contextual variation as indicated by intra-class correlation coefficients, ICC at the local level is 10 per cent, suggesting that the most disaggregated geographical unit is able to explain most of the unobserved heterogeneity in the dependent variable. The regional level explains around 3 per cent and variation as a result of temporal factors is insignificant. A possible concern still is the presence of level-1 endogeneity that the clustered means are unable to mitigate. This may result from measures of households' involvement in non-farm activities being positively correlated with aspects of human capital that are unobservable (for

[8] The full set of robustness estimations are available on request.

example, ability). Hence, a high degree of diversification into non-farm activities may also reflect a high level of income (Barrett et al., 2001), an empirical regularity that has been confirmed in prior studies and that also holds for these data. However, given the positive correlation between asset endowments, educational attainment and income, it is likely that the control variables are able to pick up a significant part of this unobserved heterogeneity.

Table 7. Determinants of agricultural and off-farm diversification

	Off-farm diversification $H_{nonagri}$		Agricultural diversification H_{agri}	
Fixed effects	Coeff.	Std. Err.	Coeff.	Std.Err.
Location and agglomeration				
LQ_j	0.016**	0.002	0.002	0.002
Population density	0.031***	0.001	-0.014	0.019
Distance small town	-0.006**	0.000	-0.001***	0.000
Distance Kigali City	-0.020**	0.010	0.010	0.011
Clustered covariates				
LQ_k	0.001***	0.000	0.020	0.029
Population density (k)	0.166***	0.048	-0.015	0.031
ICC				
Year	0.003	0.004	0.002	0.003
District | Year	0.029**	0.001	0.020**	0.001
Sector | District | Year	0.098***	0.009	0.087***	0.008

Note: ***, ** indicate statistical significance at the 1 and 5 per cent levels respectively. Sample weights are included in the estimations. For brevity, household level predictors are not presented in the table but are included in all the estimations, their coefficient estimates are in line with those presented in Table 5.

5. Conclusions

Increases in off-farm income generating activities are widely understood as a central strategy for lifting farmers out of subsistence agriculture and obtaining development in the rural areas of Africa (Barrett et al., 2001). An important challenge to be addressed in Rwanda is how to design policies that are supportive of rural non-farm labour creation. Increasing an understanding of spatial issues is important as it may indicate if there exist scale efficiencies associated with the expansion of local markets and the role played by the local business climate. Studying the degree of attenuation in space may also indicate the relevant definition of place as in the concept of place-based policies (Andersson et al., 2014).

Although literature exists on the determinants of diversification, there is a gap when it comes to an understanding of spatial spill-overs related to external economies of scale in the context of rural Africa. Though there are studies with similar motives these focus exclusively on firms and in explaining how agglomeration externalities alter enterprise productivity (Ali and Peerlings, 2010; Owoo and Naudé, 2016; Rijkers et al., 2010). This paper argued for a broader understanding of the concept of agglomeration economies and the framing of agglomeration economies in terms of rural individuals or households, rather than for firms. Hence, the contribution of this paper is its focus on individual households and on the type of knowledge spill-overs that occur at the level of individuals, which are related to learning processes (Duranton and Puga, 2004).

Hypotheses were tested using data from two rounds of the Comprehensive Food Security and Vulnerability Analysis (CFSVA) in Rwanda for 2006 and 2009. CFSVA is a nationwide household survey that provides quantitative data on key perspectives; 8,100 households were included in these rounds. The measure in focus is a location quotient, calculated with respect to the share of non-farm workers in a local area in relation to the nation as a whole. A three-level multi-level model was used to assess its relevance in explaining households' degree of involvement in the non-farm sector, distinguishing among within-level and between-level effects.

This study found that spatial proximity to non-farm activities was positive and significant in explaining households' involvement in the non-farm sector in rural Rwanda. This indicates that the type of market and non-market interactions that take place at the individual level are important as they may provide a breeding ground for local spill-over effects to take place and create opportunities for rural households to engage in the non-farm sector. The results also show signs of attenuation in geography in that the within (local) effect appears to be more relevant as compared to the between (regional) effect. This indicates that spatial spill-overs related to agglomeration economies are place-based and more relevant predictors within rather than between regions. This points to the need of considering local conditions in the formation of rural growth policies, as a particular policy is unlikely to fit different regions or even within different regions. These results are thus supportive of the increasing awareness that one-size-fits-all regional policy models should be reformulated into policies that are both place-based and knowledge-based (Naldi et al., 2015). The finding that small and large urban areas are likely to have different influences on the rural economy may also lead to the conclusion that improved connectivity between urban and rural areas and among small towns in rural areas may improve the potential for diversification for a larger set of rural households.

References

Abdulai, A. and A. CroleRees (2001). "Determinants of income diversification amongst rural households in Southern Mali", *Food Policy*, 26(4), 437-452.

Ali, D. A., K. Deininger, and M. Duponchel (2014). "Credit constraints and agricultural productivity: evidence from rural Rwanda", *Journal of Development Studies*, 50(5), 649-665.

Ali, M. and J. Peerlings (2011). "Value added of cluster membership for micro enterprises of the handloom sector in Ethiopia", *World Development*, 39(3), 363-374.

Andersson, M., J. Klaesson, and J.P. Larsson (2014). "How local are spatial density externalities? Neighborhood effects in agglomeration economies", *Regional Studies*, 1-14.

André, C. and J.P. Platteau (1998). "Land relations under unbearable stress: Rwanda caught in the Malthusian trap", *Journal of Economic Behavior and Organization*, 34(1), 1-47.

Artz, G. M., Y. Kim, and P.F. Orazem (2016). "Does Agglomeration Matter Everywhere? New Firm Location Decisions in Rural and Urban Markets", *Journal of Regional Science*, 56(1), 72-95.

Arzaghi, M. and J.V. Henderson (2008). "Networking off Madison Avenue", *The Review of Economic Studies*, 75(4), 1011-1038.

Barrett, C.B., T. Reardon and P. Webb (2001). "Nonfarm income diversification and households livelihood strategies in rural Africa: concepts, dynamics, and policy implications", *Food Policy*, 26(4), 315-331.

Barrett, C. B., M.B. Clark, D.C. Clay, and T. Reardon (2005). "Heterogeneous constraints, incentives and income diversification strategies in rural Africa", *Quarterly Journal of International Agriculture*, 44(1), 37-60.

Bigsten, A. (1983). *Income Distribution and Development: Theory. Evidence, and Policy*. London: Heinemann Educational.

(1996). "The circular migration of smallholders in Kenya", *Journal of African Economies*, 5(1), 1–20.

Bigsten, A. and S. Tengstam (2011). "Smallholder diversification and income growth in Zambia", *Journal of African Economies*, 20(5), 781-822.

Boserup, E. (1965). The Conditions of Agricultural Growth. *The Economics of Agrarian Change under Population Pressure. Chi-cago: Aldine Boserup The Conditions of Agricultural Growth. The Economics of Agrarian Change under Population Pressure1965*.

Browne, W., and Goldstein, H. (2010). MCMC sampling for a multilevel model with nonindependent residuals within and between cluster units. *Journal of Educational and Behavioral Statistics*, *35*(4), 453-473.

Capozza, D. R. and R.W. Helsley (1989). "The fundamentals of land prices and urban growth", *Journal of Urban Economics*, 26(3), 295-306.

Charlot, S. and G. Duranton (2004). "Communication externalities in cities", *Journal of Urban Economics*, 56(3), 581-613.

Collier, P. and S. Dercon (2014). "African Agriculture in 50Years: Smallholders in a Rapidly Changing World?" *World Development*, 63, 92-101.

Dabalen, A., S. Paternostro, and G. Pierre (2004). *The returns to participation in the non-farm sector in rural Rwanda* (Vol. 3462). World Bank Publications.

Davis, B., S. Di Giuseppe, and A. Zezza (2014). "Income diversification patterns in rural sub-Saharan Africa: reassessing the evidence", World Bank Policy Research Working Paper, 7108.

Duranton, G. and D. Puga (2004). "Micro-foundations of urban agglomeration economies", Handbook of regional and urban economics, 4, 2063-2117.

Dorosh, P. and J. Thurlow (2014). "Can cities or towns drive African development? Economywide analysis for Ethiopia and Uganda", *World Development*, 63, 113-123.

Ellis, F. (1998). "Household strategies and rural livelihood diversification", *Journal of Development Studies*, 35(1), 1–38.

Ellis, F. (2000). "The Determinants of Rural Livelihood Diversification in Developing Countries", *Journal of Agricultural Economics*, 51(2), 289–302.

Frenken, K., F. Van Oort and T. Verburg (2007). "Related variety, unrelated variety and regional economic growth", *Regional Studies*, 41(5), 685-697.

Fujita, M., P.R. Krugman, and A. Venables (2001). *The spatial economy: Cities, regions, and international trade*. MIT Press. Cambridge Massachusetts, London, England.

Fielding, A., Yang, M., and Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Statistical modelling*, *3*(2), 127-153.

Getis, A., and Ord, J. K. (1996). Local spatial statistics: an overview. *Spatial analysis: modelling in a GIS environment*, *374*.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *The American Economic Review*, *104*(5), 165-170.

Haggblade, S., Hazell, P. B., and Reardon, T. (Eds.). (2007). *Transforming the rural nonfarm economy: Opportunities and threats in the developing world*. Intl Food Policy Res Inst.

Henderson, J. V. and H.G. Wang (2005). "Aspects of the rural-urban transformation of countries", *Journal of Economic Geography*, 5(1), 23-42.

Hoover, E.M. (1937). *Location Theory and the Shoe Leather Industries*. MA: Harvard University Press.

Isaksson, A. S. (2013). "Manipulating the rural landscape: Villagisation and income generation in Rwanda", *Journal of African Economies*. Vol. 0, number 0, pp. 1-43.

Jacobs, J. (1969). *The economy of cities*. New York: Vintage.

Kim, T. Y. and A. Heshmati. (Eds.) (2014). "Introduction to and Summary of Economic Growth: The New Perspectives for Theory and Policy", *in Economic Growth*. Berlin Heidelberg: Springer, pp. 1-20.

Malmberg, A. and P. Maskell (2002). "The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering", *Environment and Planning A*, 34(3), 429-449.

Marshall, A. (1920). *Principles of Economics* (8th edition). London, UK: Macmillan.

McCormick, D. (1999). "African enterprise clusters and industrialization: theory and reality", *World Development*, 27(9), 1531-1551.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69-85.

Naldi, L., P. Nilsson, H. Westlund, and S. Wixe (2015). "What is smart rural development?" *Journal of Rural Studies*, 40, 90-101.

Ohlin, B. (1933). *Interregional and International Trade*. Cambridge, MA: Harvard University Press.

Owoo, N. S. and W. Naudé (2016). "Spatial Proximity and Firm Performance: Evidence from Non-Farm Rural Enterprises in Ethiopia and Nigeria", *Regional Studies*, 1-13.

Plantinga, A. J., R.N. Lubowski and R.N. Stavins (2002). "The effects of potential land development on agricultural land prices", *Journal of Urban Economics*, 52(3), 561-581.

Porter, M.E. (2000). "Location, competition, and economic development: Local clusters in a global economy", *Economic Development Quarterly*, 14(1), 15-34.

Reardon, T. (1997). "Using evidence of household income diversification to inform study of the rural nonfarm labor market in Africa", *World Development*, 25(5), 735-747.

Reardon, T., J.E. Taylor, K. Stamoulis, P. Lanjouw, and A. Balisacan (2000). "Effects of nonfarm employment on rural income inequality in developing countries: an investment perspective", *Journal of Agricultural Economics,* 51 (2), 266–288.

Rijkers, B., M. Söderbom, and J. L. Loening (2010). "A rural–urban comparison of manufacturing enterprise performance in Ethiopia", *World Development*, 38(9), 1278-1296.

Rijkers, B. and M. Söderbom (2013). "The effects of risk and shocks on non-farm enterprise development in rural Ethiopia", *World Development*, 45, 119-136.

Rosenthal, S. S. and W.C. Strange (2004). "Evidence on the nature and sources of agglomeration economies", *Handbook of regional and urban economics*, 4, 2119-2171.

Shannon, C. E. (1948). A "note on the concept of entropy", *Bell System Technology*. J, 27, 379-423.

Siba, E., M. Söderbom, A. Bigsten, and M. Gebreeyesus (2012). "Enterprise agglomeration, output prices, and physical productivity: Firm-level evidence from Ethiopia", (No. 2012/85). WIDER Working Paper.

Smith, A., K. Gordon, K. Meadows, and Zwick, K. (2001). "Livelihood diversification in Uganda: patterns and determinants of change across two rural districts", *Food Policy*, 26 (4), 421–435.

Snijders, T.A.B. and J. Berkhof (2006). "Diagnostic checks for multilevel models", in Jan de Leeuw (ed.), *Handbook of Multilevel Analysis*. New York: Springer.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC. London

Skrondal, A and S. Rabe-Hesketh (2008). *Multilevel and longitudinal modeling using Stata*. STATA Press.

Storper, M. and A.J. Venables (2004). "Buzz: face-to-face contact and the urban economy", *Journal of Economic Geography*, 4(4), 351-370.

Weldegebriel, Z. B., G. Folloni, and M. Prowse (2015). "The determinants of non-Farm Income Diversification in Rural Ethiopia", *Journal of Poverty Alleviation and International Developmen*t, 6(1). Pp, 109-130.

Wixe, S. and M. Andersson (2015). "Which Types of Relatedness Matter in Regional Growth? Industry, Occupation and Education", *Regional Studies*, 1-14.

APPENDIX A

Figure A1. District units

Figure A2. Sector units

APPENDIX B

Table B1. Summary statistics of household and locational controls

Variable	Type	Mean	Std. Dev.
Household size	Number	5.55	2.12
Age of head	Number	41.92	13.51
Literacy (head)	Categorical	0.64	0.48
Educational attainment (head)	Categorical	0.01	0.12
Female head	Categorical	0.18	0.38
Credit and asset endowments			
Access to credit	Categorical	0.35	0.47
Access to remittances	Categorical	0.04	0.21
Agricultural assets	PCA score	0.67	0.34
Land	Hectares	2.76	3.77
Crop diversity	Entropy measure	0.86	0.44
Electricity	Categorical		
Transportation assets	PCA score	0.22	0.42
ICT assets	PCA score	0.01	0.11
Average precipitation	ml	4.02	1.42
Soil quality	Kg/ha	84354	4088
Altitude	meters	1746	273.81

Table B2. Definitions of variables

Variable	Definition
Literacy (head)	Equals 1 if the head can read and write a simple message in any language, zero otherwise.
Educational attainment (head)	Equals 1 if the head has completed the secondary level, zero otherwise.
Access to credit	Equals 1 if the household has been granted credit during the last year, zero otherwise.
Access to remittances	Equals 1 if the household has a member who works away from home and sends back money, zero otherwise.
Agricultural assets	Principal component calculated w.r.t. households' ownership of agricultural assets (e.g., plough, ax, donkey cart).
Crop diversity	Entropy measure calculated w.r.t. to the number of crops cultivated by the household. $H = -\sum c_k \ln(c_k)$ where c_{ik}denote the share of total crops for household I and crop k ($k = 1.,,,.26$).
Transportation assets	Principal component calculated w.r.t. households' ownership of transportation assets (e.g., motorized vehicles, bicycles).
ICT assets	Principal component calculated w.r.t. households' ownership of information and communication technology assets (e.g., mobile phone, radio, TV, computer).