SPATIAL CONCENTRATION AND FIRM-LEVEL INNOVATION EVIDENCE FROM GHANA

Anthony Krakah and Gonzague Vannoorenberghe

LIDAM Discussion Paper IRES
2023 / 05

UCLouvain

IRES | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

Spatial concentration and firm-level innovation
Evidence from Ghana

Anthony Krakah* Gonzague Vannoorenberghe†

February 13, 2023

Abstract

We analyze how the spatial concentration of economic activity affects innovation among firms in Ghana. We use the 2014 census of all establishments to map economic activity at a precise geographic level and the responses to a detailed survey of more than 5000 firms to capture measures of innovation and firm-level characteristics. We find a strong positive effect of the overall density of economic activity on innovation (urbanization economies) but a negative effect of the density of employment in an establishment's sector (localization economies). Several questions in the survey allow us to address the issue of endogeneity and shed some light on the mechanisms. We control for many firm characteristics and confirm our results on a subsample of establishments declaring that their location is that of their founder's origin, i.e. firms with a plausibly exogenous geographic location. We find that firms in regions with denser economic activity report less problems to access funding and knowledge, while the presence of firms in the same sector is associated with more uncertainty about the gains from innovating.

Keywords: innovation, development, localization, urbanization, externalities, Ghana.

JEL Classification: R10, R11, R12, O14, O18

*Ghana Statistical Service

†LIDAM/IRES, Université catholique de Louvain, email: gonzague.vannoorenberghe@uclouvain.be.

We thank Joris Knoben and Florian Mayneris for very helpful comments. This work was supported by the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche Scientifique – FNRS under EOS Project No. 30784531 "Winners and losers from globalization and market integration".

1 Introduction

A large body of research has pointed to a positive effect of economic density on different measures of performance, such as firm productivity (see for example Combes and Gobillon (2015) and Duranton and Puga (2020) for surveys). Such positive effects, which can help spur development (Duranton (2015)), have a well-established economic rationale. They notably rely on different types of spillovers that firms benefit from when locating close to other firms. The literature has distinguished intra-industry (*localization*) agglomeration economies in the spirit of Marshall (1890) from inter-industry (*urbanization*) agglomeration economies, more in the line of Jacobs (1969). Empirical studies often find that the localization and the urbanization economies play in the same, positive, direction. With a few exceptions, however, the existing evidence ovewhelmingly relies on data from developed countries, partly for reasons of data availability.

In the present paper, we seek to quantify urbanization and localization economies on innovation in the context of a major Sub-Saharan African economy: Ghana. We use a recent census collecting data on the universe of firms in the country to obtain a precise mapping of the location of economic activity in Ghana. The census contains information on the employment, the industry classification and the locality (there are more than 15,000 localities in Ghana) for each firm, formal or informal, in the country. We then combine this mapping with a detailed survey for a representative sample of more than 5,000 manufacturing firms, which includes balance-sheet data as well as detailed questions about their innovative behavior. In this context, we ask whether firms are more likely to innovate in regions with a higher overall density of employment (urbanization) and with a higher density of same-industry employment (localization). Our main results are that urbanization and localization have opposite effects: we find a positive and robust effect of urbanization and a mostly negative effect of localization economies on the probability to innovate.

In contrast to most of the economic literature on agglomeration economies, which studies productivity or wages, we focus on innovation as an outcome variable. We define a firm as innovating if it introduced in the last year a product or a process innovation[1]. In the context of Ghana, such innovations are often far from the technological frontier: less than 10% of product innovators in our sample report an innovation that is "new to the world". In this sense, our concept of innovation is very different from studies looking at the effect of agglomeration economics on

[1]Following the Oslo manual, a product innovation is defined as the "introduction to the market of a new or significantly improved good or service". A process innovation is the "use (implementation) of new or significantly improved methods for the production or supply of goods and services". The concepts of product and process innovation have widely been used in the literature, both from a growth theory perspective (e.g. Jaimovich (2021)) or as an empirical measure of innovation in developing countries (e.g. Gorodnichenko et al. (2010), Bos and Vannoorenberghe (2019) or Cirera and Muzi (2020)).

R&D spending or patenting in developed countries. Our concept of innovation is rather a way through which firms seek to improve their performance in an incremental way.

As is well-known in the literature (see e.g. Combes and Gobillon (2015)), identifying a causal effect of spatial concentration on firm performance is particularly challenging. Part of the problem comes from unobserved factors, which may affect both the density of economic activity and the innovation decision by firms. For example, a region offering better infrastructure may both attract more workers and be more conducive to innovation by firms. Another issue arises from the mobility of firms. If innovative firms move to denser economic environments, denser places will appear as having a disproportionate share of innovators. To the extent that those firms would also have innovated in their region of origin, this causes an upward bias of the effect of spatial concentration on innovation[2]. To deal with unobserved heterogeneity, the literature often relies on the time dimension of the data, typically available in developed countries. In this case, using lagged density as an instrument for current density (in the spirit of Ciccone and Hall (1996)), or comparing the evolution of plants after a natural experiment (e.g. Greenstone et al. (2010)) allow capturing some of the endogeneity problems. We cannot use these approaches here as we only have a cross-section of firms, on which our density measures (urbanization and localization) are based. To tackle the endogeneity issues, we leverage on a unique feature of our data: all firms participating to the survey were asked for the reason why they located in a particular locality. We replicate our analysis on a subset firms that declare being close to where their founder comes from, which should dampen the mobility issue. We also control for the fraction of firms in a locality that mention infrastructure as a determinant of their location to control for the quality of infrastructure, which is usually unobservable. Our main conclusions - a positive urbanization and a negative localization effect - still hold, even if the precision of our estimates slightly decreases.

In disentangling the effects of urbanization and localization economies, we put particular care in analyzing at which level of industrial aggregation and at which distance the urbanization and localization effects materialize. There is *a priori* no strong theoretical argument for whether spatial concentration should refer to economic activity within 10, 25 or 50km from the firm, and the empirical literature on the question is mostly focusing on developed countries (Rosenthal and Strange (2003), Rosenthal and Strange (2020)). Similarly, localization effects could happen between firms within broad sectors or manufacturing or only within precisely defined industries. We follow the recommendation of Beaudry and Schiffauerova (2009) and test for a range of different distances and industry definitions. Urbanization effects are strongest and most significant

[2]A third type of issue could be a simultaneity bias, which would arise if our left-hand side variable - the firm's innovation - would attract additional firms or workers, our right hand side variable. As argued below, we do not see this as a major issue in our case.

when defined 25km to 60km around the establishment. Localization effects negatively impact innovation at most distances when considering relatively precise definitions of industries (2 or 3 digit ISIC).[3] These negative coefficients suggest that competition effects may have a stronger detrimental effect on innovation than the positive Marshallian spillovers in the context of Ghana.

We use a number of questions in the survey to shed some light on the mechanisms behind these results. All firms in the innovation survey report the factors that were hampering innovation, whether or not they innovated. Using these factors as outcome variables, we find that urbanization is associated with lower difficulties to find funding or to access sources of knowledge. Firms located in places with dense industry-specific employment on the other hand report the lack of information or the uncertainty about their market as factors hampering their innovation. Among firms innovating, those with a high value of the localization variable on average report that their suppliers or buyers were both a stronger motivation to innovate, and a more important source of external knowledge. These point to the potential existence of positive Marshallian externalities, even if they seem too weak to generate positive localization economies.

The rest of the paper is structure as follows. Section 2 places our work in the extensive related literature. Section 3 presents our data sources as well as some basic descriptive statistics. We show our main empirical strategy and results in Section 4, as well as a number of robustness checks in Section 5. Section 6 elaborates on the mechanisms and Section 7 concludes.

2 Literature

A long tradition in the urban economics literature estimates how the density of economic activity affects a number of outcomes at the local level, such as nominal wages, productivity or, to a lesser extent, innovation. The idea that the agglomeration of economic activity can raise productivity dates back to the concept of Marshallian externalities. Marshall (1890) argues that agglomeration economies arise from interactions on the labor market, the access to more specialized inputs and the existence of knowledge spillovers. As pointed out by Duranton and Puga (2004), the actual microeconomic mechanisms within each of these "markets" (labor, inputs and knowledge) can be of three kinds. Agglomeration of economic activity allows firms to *share* indivisible factors of production, raise the *quality of the match* between different agents and fosters *learning* through more frequent contacts. Marshallian externalities, as well as Porter (1990), posit that these mechanisms are stronger if agglomeration happens between firms pro-

[3]The literature summarized in Beaudry and Schiffauerova (2009) identifies stronger urbanization effects when using detailed industry or geographic classifications, and positive localization effects with more aggregate industry or geographic definitions. In our case, urbanization is strongest at intermediate aggregation levels and localization is most negative for precise industry definitions.

ducing similar goods. On the other hand, Jacobs (1969) emphasizes the diversity of industries within a city as a an important source of spillovers fostering growth.

Agglomeration economies can also affect innovation. Duranton and Puga (2001) develop a model where young entrepreneurs can learn about their ideal process of production by drawing from the experience of firms located around them, putting forward a learning mechanism. Helsley and Strange (2002) argue that having access to a denser network of input suppliers may make it less costly to implement new ideas, emphasizing a sharing mechanism. Within an industry, a higher spatial concentration affects the intensity of competition, potentially acting as a driver of innovation in the spirit of Porter (1990). Desmet and Parente (2010) propose a framework in which competition increases the size of firms, giving them additional incentives to conduct process innovation. From a more Schumpeterian perspective, however, competition can be detrimental to innovation and Aghion et al. (2005) point to an inverted U-shape relationship between competition and innovation. Those arguments could imply less innovation in places with more spatial concentration of an industry, generating a negative localization effect. Heblich et al. (2022) show that the opening of large plants may, in the long run, limit the incentives for firms that interact with them to innovate. When a regional economy becomes too specialized around a major player, this create a long-run "lock-in" effect that stifles local innovation. Such mechanisms could also account for a negative localization effect.

The empirical estimation of agglomeration effects has been the subject of a very broad empirical literature (see for example Combes and Gobillon (2015) and Duranton (2015) for surveys). According to Duranton (2015), the elasticity of productivity with respect to population density is typically of the order of 2 to 4%, but can be substantially higher in developing countries (Combes et al. (2013)). The type of spillovers behind such effects has also been studied extensively. Glaeser et al. (1992) exploit variation in the growth of industries in US cities and identify the existence knowledge spillovers between industries à la Jacobs (urbanization), rather than within industries à la Marshall (localization). Exploiting firm-level data, Rosenthal and Strange (2003) identify a robust localization effect for most of the industries that they study and a rather unstable urbanization effect. They use variation at the zipcode level in the US, and show that the localization economies dissipate very quickly within the first miles around the centroid of a zipcode and much less quickly beyond 5 miles. In the case of France, Martin et al. (2011) identify a positive effect of localization economies on productivity but no evidence of urbanization economies at least in the short run.

Closer to us in terms of outcome variable, Baptista and Swann (1998) shows that firms in the UK innovate much more if there is a lot of employment in their own sector, but not in general. A large literature has concentrated on the link between agglomeration and innovation (see Carlino and Kerr (2015) for a review). A large chunk of this literature focuses on research activity or patents, i.e. on innovations that are new to the world.[4] As suggested by Duranton (2015) and many others, however, innovation in developing country is less often associated with R&D spending, patents or the introduction of major product innovations. It is typically more incremental and relies on the absorption of existing knowledge, with potentially more scope for spillovers (Siba et al. (2012)). This, as well as the lack of available data, may explain the striking fact that Carlino and Kerr (2015)'s survey on agglomeration and innovation contains almost no reference to studies on developing countries.

Since innovation in developing countries is more incremental, a large literature has focused on spillovers from Foreign Direct Investment to local firms in developing countries (see Javorcik (2004), Gorodnichenko et al. (2010) or Görg and Greenaway (2004) for a survey), on learning by exporting (Atkin et al. (2017)) or on benefits from arising from the use of imported inputs (Amiti and Konings (2007) and Goldberg et al. (2010)). While most of these studies look at the effect of such spillovers on the productivity of local firms in developing countries, and how they can spread through local spillovers (Bos and Vannoorenberghe (2018)), only few examine their effect on the incentives to innovate by those firms. Several empirical studies of spillovers via FDI have explored the hypotheses that the incidence of spillovers may be conditional upon their distance to the technology frontier and on their capacity to absorb external knowledge (see e.g. Kokko et al. (1996), Damijan et al. (2013) or Blalock and Gertler (2009)).

Works studying the link between spatial concentration and firm performance in the context of developing countries are still rare, but have started emerging in the last decade. Siba et al. (2012) use a census of firms of more than 10 employees in Ethiopian manufacturing over 10 years. They find no urbanization effect but a significant localization effect that is positive on physical productivity, but negative on prices (competition effect). The two cancel out, giving little incentives for firms to agglomerate. Howard et al. (2014), in the case of a census of large Vietnamese manufacturing firms, find positive urbanization and localization effects on productivity. In a cross-country sample of Sub-Saharan African firms, Sanfilippo and Seric (2016) show that urbanization is positively linked to firm productivity while localization is negatively correlated with productivity, a result in line with the study of Chhair and Newman (2014) in Cambodia. Compared to these works, we use data on the universe of establishments in Ghana, giving a much broader coverage than previous studies except for Chhair and Newman (2014), which is similar in size. Having a complete coverage of small firms is a key advantage in a

[4] A number of studies argue that research activities are even more concentrated than population or than other production activities, suggesting that agglomeration forces are even stronger for knowledge-related activities. Carlino et al. (2007) show for example that, among metropolitan areas in the U.S., a higher density of employment substantially raises the number of patents per capita. Other studies confirm using patent citations that knowledge spillovers decrease at a very high rate with distance (see e.g. Murata et al. (2014)). It is not only the size of the city which may matter but also the diversification of its production base, see Feldman and Audretsch (1999).

developing country where firms are typically small (Hsieh and Olken, 2014) and given the previous evidence that agglomeration economices are larger for small firms (e.g. Rosenthal and Strange (2010)). We look at innovation rather than productivity and can use a number of detailed questions that were asked to firms on their motives, which were previously unavailable. Close to our work in terms of question, Zhang (2015) finds a positive impact of urbanization on product innovation, but no effect of localization in the Chinese context. The positive effect of urbanization and negative impact of localization that we identify on innovation are in line with the results of Knoben et al. (2022) in a sample of four South-East Asian countries. The variety of empirical results in the literature is not necessarily suprising given the different theoretical mechanisms at stake, which can be positive (Jacobs or Marshall spillovers) or negative (congestion, competition, lock-in). The relative size of the channels may furthermore depend on the context or sample, in view of the heterogeneous effects that spatial concentration may have on different firms (Knoben et al. (2016)) or in different contexts.

3 Data

Our study relies on a unique establishment-level data collected in 2014 by the statistical institute of Ghana. The data collection consisted of two phases.

Integrated Business Establishment Survey I In the first phase, which took place in 2014, Statistics Ghana established a list of the universe of establishments in Ghana, both formal and informal. The resulting data contains a number of basic indicators on all establishments such as their sales, employment, ownership structure and detailed information on their industry and geographic location. Phase 1 surveyed around 630.000 establishments, employing more than 3.3. million people. Among these, around 100.000 are manufacturing establishments, employing close to 430.000 people, or 13% of employment. The largest 2-digit industry is retail trade, with 18% of employment. Among manufacturing, the manufacture of wearing apparels is the largest sector, accounting for 34% of manufacturing employment, followed by the manufactures of food products (19%), metal products (8%), wood and wood products (6%) and furniture (5%). Phase 1 contains information about the locality of each establishment, allowing to construct a very precise map of economic activity in Ghana. Ghana is divided into over 15.000 localities, which are part of about 200 districts and 10 large regions (see the map in Figure 3). As shown in Table 1, localities vary hugely in size, from hamlets with no or one establishment and employee to Accra Central, for which Phase 1 reports more than 14.000 establishments and 300.000 employees. We observe at least one establishment in Phase 1 for 12.463 localities. We also compute the number of employees and of establishments per square kilometer in each location to capture the density of economic activity. Figure 4 maps the density of employment in Ghana.

Table 1: Descriptive statistics for localities present in Phase 1

	Mean	Min	p10	p50	p90	p99	Max
Employment	270	1.000	4.00	31	285	4,252	347,139
N. Estab.	51	1.000	1.00	8	66	824	14,329
Area (sq. km)	14	0.010	0.34	2	32	168	1,197
Emp./sq. km	196	0.002	0.28	19	291	2,044	386,385
Estab./sq. km	38	0.001	0.10	5	71	438	41,512

Integrated Business Establishment Survey II In the second phase, Statistics Ghana conducted in-depth face-to-face interviews with around 5400 manufacturing establishments, randomly using stratified sampling[5]. About 5400 establishments are part of that second phase, accounting collectively for more than 90.000 employees, one fifth of the total manufacturing employment captured in Phase 1. The map in Figure 5 in the appendix shows the location of those establishments. The industries most represented in the sample of phase 2 are the manufacturing of food products (15% of establishments), wearing apparels (14%), and fabricated metal products (13%). Enumerators in phase 2 conducted in-depth interviews to collect data on different balance sheet items of establishments as well as a number of quantitative and qualitative questions on their innovative behavior. Table 2 summarizes a number of characteristics of these manufacturing establishments. These are on average small, with a median employment of 5 and a mean of 21. The average age of establishments, is 12 years and only 4% are foreign-owned. The density of employment in the localities within 25km of the establishment's locality is on average 185 employees per square km.

The data on innovation include a set of standard questions, such as whether the establishment introduced product or process innovation, or how much it spent on innovation activities. The prevalence of innovation appears very low among Ghanaian firms, with 10% of firms reporting some kind of innovation[6], with only 8% of establishments reporting a product innovation, defined as "the introduction to the market of a new of significantly improved good or service with respect to its capabilities, such as improved user-friendliness, components, software or sub-systems". The prevalence of process innovation, defined as "the use (implementation) of

[5]Some details and reference

[6]These numbers are quite unusual for surveys of manufacturing firms, even in African countries. The 2015 innovation report of Kenya (Nzau et al., 2016) for example finds that 45% of firms implemented product and 40% of firms implemented process innovations in a representative sample of 700 firms.

new or significantly improved methods for the production or supply of goods and services", is even lower, with 5% of establishments reporting one. We also have a number of more detailed questions on the reasons why the firm innovated and on the sources of knowledge that it had access to, as well as qualitative descriptions about the type of innovations performed by the firm.

Table 2: Descriptive statistics for firms in Phase 2. Nb of obs: 5.285

	Mean	Min	p10	p50	p90	p99	Max
Employment	20.73	1.00	1.00	5.00	29.00	318.00	3,695.00
Innovate	0.10	0.00	0.00	0.00	1.00	1.00	1.00
Age	11.59	0.00	3.00	9.00	23.00	50.00	96.00
Foreign owned	0.04	0.00	0.00	0.00	0.00	1.00	1.00
Empl./ $km^2 \leq$ 25km	185.06	0.12	5.31	23.55	847.09	1,123.17	1,224.55
Empl./ $km^2 \leq$ 50km	75.39	0.32	5.70	15.52	299.83	343.54	348.74

Ghana Population Sampling Frame Data on the geographic coordinates of localities are obtained from the Ghana population sampling frame. We compute the centroids of enumeration areas and average their coordinates at the locality (localities can consist of several enumeration areas) level to obtain an estimated center of the locality. We compute the bilateral distance between localities using the great circle distance between the centers of localities. We use fuzzy matching methods to map the 2014 Ghana Business Sampling Frame to the Population Sampling Frame. The detailed locality reported by the interviewers in the Business Sampling frame were sometimes reported with an error, most of the time simple spelling misstakes. To match those names with the official list of localities in Ghana, we use fuzzy matching methods and match localities with the best match in the official list in the same district, as long as the similarity score is higher than 10%. We show that our results are insensitive to that threshold in the robustness section.

4 Empirics

4.1 Localization and Urbanization

Marshallian externalities are more likely to hold for firms belonging to the same industry if they are close geographically. This is one of the main reasons why the literature predicts that some industries are geographically concentrated. Such externalities could happen at a relatively broad level of industry definition or only within very narrowly defined industries. Similarly, the geographic distance at which Marshallian externalities propagate is unclear, in particular in a developing country context. In this context, we follow the recommendation of Beaudry and Schiffauerova (2009) and present our results at different levels of industrial and geographical aggregation. For a given locality l, we define the set of localities $\mathcal{L}_l^d$ as those which are situated less than d km away from l. We define the *localization* of a firm in locality l and industry s as the density of employment in that industry in and around the location, excluding the firm's own employment, which we compute as:

$$Loc_i^{sd} = Log\left(\frac{1 + \sum_{l' \in \mathcal{L}_l^d} Emp_{l's} - Emp_i}{\sum_{l' \in \mathcal{L}_l^d} Area_{l'}}\right). \tag{1}$$

The exclusion of the own firm's employment from the construction of the localization guarantees that there is no mechanical correlation between the size of the firm itself and the measure of localization and is standard in the literature (see Martin et al. (2011)). We let d vary in steps of 5 km between 0 km, considering only employment in the own locality l, and 100km, including all localities of which the geographical center is less than 100km from the center of l. Similarly, we define sectors at 3 different levels of aggregation: a very broad category consisting of 7 manufacturing sectors[7], as well as the 2-digit and 3-digit ISIC classifications. We experiment with alternative measures of localization in the robustness section. As an alternative potential source of spillovers, we consider more general measures of the density of economic activity in a region, that is not specific to an industry, in the spirit of Jacobs (1969). We define the *urbanization* of a locality l and surrounding areas as:

$$Urb_i^{sd} = Log\left(\frac{1 + \sum_{l' \in \mathcal{L}_l^d} Emp_{l'} - Emp_{l's}}{\sum_{l' \in \mathcal{L}_l^d} Area_{l'}}\right), \tag{2}$$

where we exclude the employment of the industry s of firm i to avoid that the localization measure be mechanically related to the urbanization measure.

4.2 Baseline specification

Our baseline specification is:

$$Innov_i = \beta_0 + \beta_1 Loc_i^{sd} + \beta_2 Urb_i^{sd} + \beta_2 \mathbf{X}_i + \delta_{S(i)} + \delta_{R(i)} + \epsilon_i \tag{3}$$

[7] These are: Food, beverages and tobacco (ISIC 10-12), Textile, wearing apparel and leather (ISIC 13-15), Wood, Paper and paper products, Publishing and printing (ISIC 16-18), Refined petroleum, Chemicals and pharmaceuticals, Rubber and plastics, and Non-metallic mineral products (ISIC 19-23), Basic metals, Fabricated metals, Electronics and computers, Machinery and equipment, Transport and motor vehicles (ISIC 24-30), Manufacture of furniture (ISIC 31) and Other manufacturing (ISIC 32).

where $Innov_i$ is a binary measure of innovation by establishment i. In our baseline, we define it as one if firm i conducts either process or product innovation and zero otherwise. We report in the robustness section separate results for process and product innovations. Loc_i^{sd} and Urb_i^{sd} are the measures of localization and of urbanization for firm i as defined in (1) and (2), with d indexing the spatial range considered and s the level of aggregation for industries. $\mathbf{X}_i$ is an establishment-level vector of characteristics typically thought to affect innovation, such as the log age, the log number of employees, or whether the establishment is foreign-owned. $\delta_{S(i)}$ denotes the coefficient on a dummy that takes value one if firm i is in the 2-digit industry $S(i)$ and zero otherwise. $\delta_{R(i)}$ is a regional dummy which takes value 1 if firm i is located in the region R, one of the 10 macro regions of Ghana (see map in Figure 3). We estimate (3) by a complemetary log-log regression and cluster standard errors at the locality level. The choice of the complementary log-log model stems from the binary nature of the left hand side variable, where the probabiliy of innovation in the data is small (10%), making this approach preferable to a standard logit regression. We replicate our analysis with OLS and different levels of clustering of standard errors in the robustness section and show that the results are very similar.

Table 3 reports our baseline regression for three different definitions of distance (25, 50 and 100km) and all three levels of aggregation of industries. Our localization measure enters with a negative and significant coefficient at all 3 distances when defined at the 2-digit or 3-digit ISIC level, while it is mostly insignificant when using broad industry levels. The urbanization variable on the other hand appears positive and significant at 25 or 50km distance, but not at 100km. All firm-specific control variables have very stable coefficients throughout. The estimated effects are economically large. Taking the 25km radius and the 2-digit industry definition, a one standard deviation increase in localization decreases the probability to innovate by about 3 percentage points for a firm with an average value of all the covariates[8]. Considering that 10% of the firms innovate, this is a strong effect. Similarly, a one standard deviation increase in urbanization raises the probability to innovate by 4 percentage points for an average firm. Firm size as measured by the log of employment is a very positive and significant determinant of the probability to innovate. A doubling in size for the average firm raises the probability to innovate by close to 5 percentage points. The other firm-level controls such as age, the legal status or the foreign ownership do not appear as strong predictors of the decision of firms to innovate. The legal status and the foreign ownership enter with a positive and negative coefficient respectively, but only sometimes significant at the 10% level.

In Figure 1, we run a separate regression for each distance range between 0 and 100km and plot the coefficients on localization and urbanization as well as their 95% confidence interval.

[8]The marginal effects of localization, urbanization and log size for a firm with an average value of all the covariates are respectively -0.013, 0.023, and 0.073.

This confirms that the positive and significant effect of urbanization identified in Table 3 for a radius of 25 or 50km extends to distances up to 80km depending on the precise industry definition, and is strongest when considering a radius of 25 to 60km. The negative localization effect is stable and consistent at all distances up to 100km.

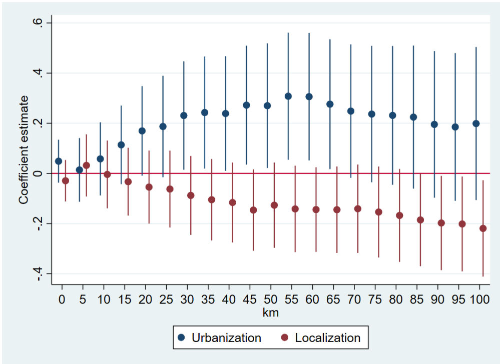
Table 3: Urbanization, localization and innovation

	25km range			50 km range			100 km range		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Broad	2-digits	3-digits	Broad	2-digits	3-digits	Broad	2-digits	3-digits
Localization	-0.06	-0.16***	-0.11***	-0.13	-0.18***	-0.11***	-0.22**	-0.20***	-0.12***
	(-0.80)	(-3.39)	(-3.20)	(-1.46)	(-3.76)	(-3.38)	(-2.24)	(-3.45)	(-3.21)
Urbanization	0.19*	0.29***	0.23***	0.27**	0.34***	0.25***	0.20	0.20	0.09
	(1.81)	(3.72)	(3.36)	(2.13)	(3.45)	(2.89)	(1.28)	(1.45)	(0.76)
Log(age)	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.01
	(0.19)	(0.14)	(0.12)	(0.25)	(0.22)	(0.15)	(0.30)	(0.26)	(0.16)
Log(empl)	0.89***	0.92***	0.90***	0.90***	0.93***	0.91***	0.92***	0.94***	0.93***
	(5.95)	(6.15)	(6.06)	(6.09)	(6.26)	(6.17)	(6.18)	(6.29)	(6.24)
Foreign owned	-0.29	-0.27	-0.28	-0.29	-0.26	-0.29	-0.28	-0.26	-0.28
	(-1.49)	(-1.42)	(-1.50)	(-1.50)	(-1.39)	(-1.52)	(-1.44)	(-1.35)	(-1.50)
Incorporated	0.24	0.26*	0.24*	0.25*	0.26*	0.24*	0.27*	0.27*	0.25*
	(1.64)	(1.81)	(1.68)	(1.71)	(1.86)	(1.72)	(1.83)	(1.90)	(1.78)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5272	5272	5272	5272	5272	5272	5272	5272	5272
Log-Likelihood	-1607.7	-1602.3	-1603.4	-1607.9	-1602.6	-1604.6	-1608.3	-1604.7	-1606.7

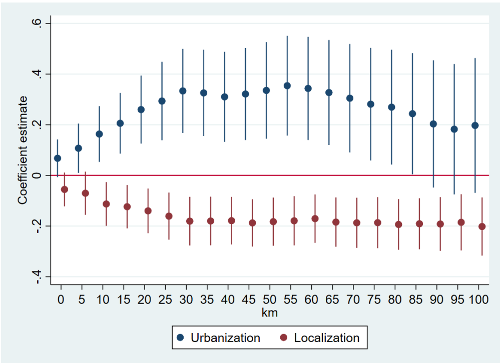
All regressions are complementary log-log models. The dependent variable is a dummy for whether a firm innovates. Urbanization is the number of people employed per square kilometer (see eq. (2)) in the localities within 25km (col. 1-3), 50km (col. 4-6) or 100km (col 7-9) of the firm's locality. Localization is defined as the number of people employed in the same industry per km (see eq. (2)), where an industry follows the broad classification in col. 1, 4 and 7, is defined at 2-digits in col. 2, 5 and 8, and at 3-digits in col. 3,6 and 9. Standard errors clustered at the locality level. Industry dummies are at the level used for the localization variable and region dummies are based on the 10 regions of Ghana (see Figure 3).

Figure 1: Coefficient estimates and 95% confidence intervals for Urbanization and Localization with different spatial lags

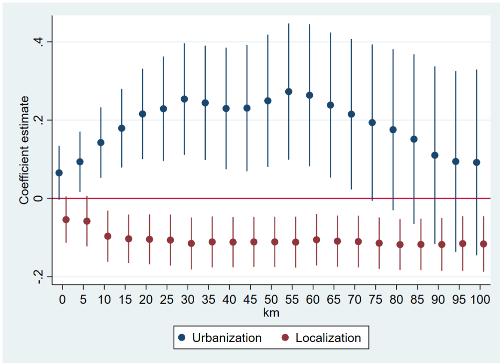
(a) Localization defined at the broad level



(b) Localization defined at the 2-digit level



(c) Localization defined at the 3-digit level



4.3 Endogeneity

The issue of endogeneity is a major challenge for the identification of a causal effect of geography on firm-level outcomes (see e.g. Combes and Gobillon (2015)). To the extent that firms choose their geographic location endogenously, observing a correlation between the economic activity in a region and the innovation of firms in this region may reflect different forces. First, firms may benefit from the presence of consumers, suppliers or workers in their region (Marshallian externalities), or have higher incentives to innovate when facing a stronger local competition. This is the type of causal effect that we want to identify. Second, there can be joint determinants of innovation and of firm location at the local level. Some local amenities may affect both the level of innovation and the choice of location of firms. The presence of a university, the quality of infrastructure or the efficiency of local public institutions could for example all contribute simultaneously to innovation and to the density of economic activity. If these local amenities are a result of the density of economic activity itself, their effect on innovation should be taken into account as part of the total effect of local economic activity on innovation. If they are not, or only partially so, however, they will cause a bias in our estimate. Third, the correlation may be due to reverse causality. The most innovative firms may for example be more mobile and may endogenously decide to locate in regions with a higher density of economic activity[9].

Identifying the direction of causality is essential to determine the strength of externalities and to draw appropriate policy recommendations. This problem has been widely recognized in the literature (see e.g. Combes and Gobillon (2015) and Baum-Snow and Ferreira (2015) for a review) and different studies apply different approaches to tackle the issue. First, to address a potential omitted variable bias due for example to local amenities, the literature typically controls for local fixed effects in a panel regression. This strategy, which would not solve the problem of reverse causality and of time-varying amenities, is not available in our case due to the cross-sectional nature of our data. A second route followed by the literature is to instrument for the density of local economic activity, using either historical or geographical instruments. Carlino et al. (2007) use for example geographic variable, such as the temperature or the presence of water as instruments for population density. Ciccone and Hall (1996) on the other hand instruments for current population density by the historical location of economic activity. Finding a credible and strong instrument for an establishment's location in a cross-sectional dataset in Ghana however seems a real challenge.

We take an alternative strategy and exploit the answer to a question of the survey about the reasons why the establishment is located in the current location ("What were the reasons for locating at the present address?"). Figure 2 replicates Figure 1 using the subsample of firms from Phase 2 that report being located "close to where the founder was born, grew up or has family". We argue that the concerns resulting from the endogenous choice of location by firms should be dampened in this sample consisting of 1894 firms, 8.2% of which innovate. While the confidence intervals naturally become wider, the main conclusions remain unchanged: localization economies remain mostly negative and urbanization positively affects innovation when defined at a range of up to 60km, with close to 5% significant at most intermediate distances.

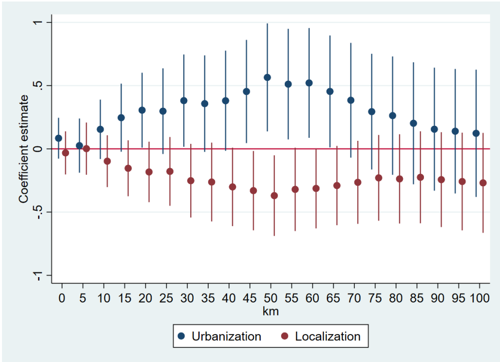
To proxy for the quality of infrastructure in a locality, a variable that is typically unobservable, we exploit again the response to the question on the reasons for locating at the present address. We use the share of firms in a locality that mention infrastructure as a reason to proxy

[9]The existing evidence on the mobility of entrepreneurs mostly relies on developed economies (see e.g. Figueireido et al. (2002) for Portugal and Michelacci and Silva (2007)for the US and Italy. These studies show that entrepreneurs are typically staying in their home region, suggesting that mobility is not a strong issue. However, these effects may be different in Ghana and even low mobility rates may bias our estimates if innovators have a different propensity to move than non-innovators. In a similar vein, Glaeser and Saiz (2004) find that skilled workers sort into larger cities, an effect which may be picked up by estimates of the urban scale effect on wages.

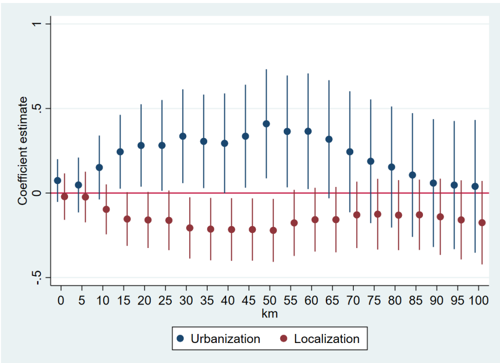
for a locality's quality of infrastructure and add it as a control. Reproducing Figure 2 with this additional control (not reported) has virtually no effect.

Figure 2: Coefficient estimates and 95% confidence intervals for Urbanization and Localization - subsample close to founder's origin

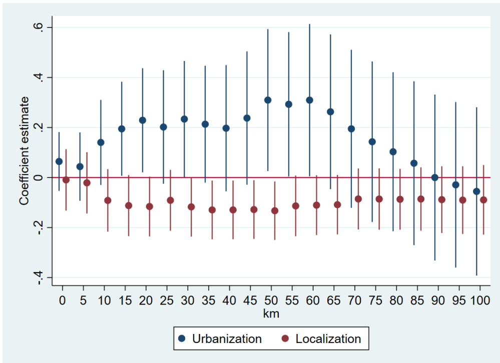
(a) Localization defined at the broad level



(b) Localization defined at the 2-digit level



(c) Localization defined at the 3-digit level



5 Robustness

In this section, we report a number of robustness checks. To simplify the exposition, we only present the results for Localization defined at the 2-digit level and for Urbanization and Localization based on a distance of 25km around the locality. We choose this specification as it has the highest log-likelihood in Table 3 but the robustness of our results does not hinge on that particular choice. The first column of Table 4 reports our baseline specification (column 2 of Table 3) to ease comparison.

Estimator and standard errors. Column 2 of Table 4 estimates our baseline with OLS instead of the complementary log-log specification and shows that results are very similar in a linear probability model. Column 3 explicitly takes into account the potential correlation of error terms in space and uses Conley standard errors based on a 25km radius around the locality. Our standard errors are barely affected by this exercise.

Measures of urbanization and localization. We experiment with two alternative measures of urbanization and localization. In column 4, and in line with Martin et al. (2011), we replace the *density* of employment per square kilometer in localities around the firm by the *number* of employees in those localities, which corresponds to using (1) and (2) without the area in the denominator. In column 5, we compute the densities based on the number of establishments, and not the number of employees, per square kilometer in and around a locality. Both exercices yield very similar results to our baseline, with slightly stronger results when using establishments. In column 6, we compute measures of urbanization and localization based on the density of employment within a district, and cluster our standard errors at the district level. The coefficients are smaller in magnitude, as our results rely on less variation, but point to similar qualitative results. The negative coefficient on localization however turn insignificant.

Putting urbanization and localization separately. Our measures of urbanization and localization are strongly correlated. At the 2-digit level and with a 25km radius, for example, the correlation coefficient is 0.7, even though our measure of urbanization excludes the employment of the industry considered (see the definition of urbanization in (2)). Such a high correlation is not surprising. A high density of employment in an industry is more likely in places with denser economic activity. Taking out the greater Accra region for example reduces this correlation to 0.54. Columns 7 and 8 of Table 4 introduce our two measures separately. Urbanization remains positive and significant, while localization alone becomes insignificant. The negative localization effect that we estimate arises when fixing the degree of urbanization. Not including urbanization means that the localization mixes both Jacobs externalities on the one hand, and Marshallian externalities or competition effects within industries on the other, thereby turning insignificant.

Different sets of fixed effects. Columns 9 to 11 of Table 4 experiment with different sets of fixed effects. Column 9 does not include any fixed effects. Column 10 contains both region and industry fixed effects, with industries defined at the 3-digit ISIC level. The results remain very similar in both cases. Column 11 replaces region fixed effects by district fixed effects. The coefficient on the localization variable remains negative and significant while the coefficient on urbanization remains of a similar size but turns insignificant. Including district dummies eliminates too much of the spatial variation for urbanization - defined in a 25km radius - to remain statistically significant.

Table 4: Robustness

	(1) Base	(2) OLS	(3) Conley	(4) Level Emp.	(5) Estab.	(6) Distr.	(7) Loc. only	(8) Urb. only	(9) No FE	(10) 3-dig. FE	(11) Distr. FE
Localization	-0.16***	-0.01***	-0.01***	-0.16***	-0.32***	-0.07	-0.02		-0.09***	-0.16***	-0.13**
	(-3.39)	(-3.10)	(-3.08)	(-3.40)	(-4.59)	(-1.56)	(-0.68)		(-3.38)	(-3.30)	(-2.56)
Urbanization	0.29***	0.02***	0.02***	0.34***	0.45***	0.11**		0.12**	0.18***	0.28***	0.24
	(3.72)	(3.75)	(3.77)	(4.22)	(4.80)	(2.00)		(2.08)	(4.69)	(3.38)	(1.34)
Log(age)	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.04	0.01	-0.00
	(0.14)	(0.50)	(0.57)	(0.17)	(0.18)	(0.10)	(0.22)	(0.17)	(0.73)	(0.14)	(-0.08)
Log(empl)	0.92***	0.07***	0.07***	0.91***	0.90***	0.92***	0.92***	0.88***	0.67***	0.92***	0.98***
	(6.15)	(6.46)	(5.92)	(6.08)	(6.03)	(7.29)	(6.13)	(5.89)	(5.04)	(6.07)	(6.02)
Foreign owned	-0.27	-0.02	-0.02	-0.26	-0.33*	-0.29	-0.28	-0.29	-0.34*	-0.26	-0.36*
	(-1.42)	(-1.07)	(-0.88)	(-1.34)	(-1.74)	(-1.16)	(-1.46)	(-1.51)	(-1.83)	(-1.20)	(-1.92)
Incorporated	0.26*	0.03*	0.03**	0.26*	0.23	0.25**	0.25*	0.23	0.07	0.31**	0.21
	(1.81)	(1.96)	(2.33)	(1.84)	(1.60)	(1.97)	(1.72)	(1.60)	(0.48)	(2.15)	(1.39)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Observations	5272	5280	5280	5272	5272	5272	5272	5272	5280	5218	4305
Log-Likelihood	-1602.3			-1600.4	-1595.5	-1608.7	-1610.6	-1608	-1678.4	-1578.7	-1456

The regressions in columns 1, as well as all columns 4 to 11 are complementary log-log models. Columns 2 and 3 are estimated by OLS. The dependent variable is a dummy for whether a firm innovates. Standard errors are clustered at the locality level except in column 3 (Conley standard errors, 25km) and column 6, at the district level. By default, Urbanization and Localization are computed as in (1) and (2) based on localities within 25km and 2-digit industry. In column 4, Urbanization and Localization are based on the number of employees (instead of density), while they are are based on establisments per square km in column 5. Column 6 computes a measure of urbanization and localization at the district level. Columns 7 and 8 introduce localization and urbanization separately. Columns 9 to 11 vary the set of dummies included. In column 9, there are no industry dummies and no regional dummies. Industry dummies in all other columns are at the 2-digit ISIC level except for column 10, where they are at the 3-digit ISIC. Regional dummies in all columns are based on the 10 regions of Ghana (see Figure 3) except for column 11, where they are defined at the district level.

Data construction. As described in section 3, our procedure requires to match the detailed locality reported by the interviewer to the official list of localities in Ghana, for which we have geographic coordinates. We use fuzzy matching methods and match localities with the best match in the official list, as long as the similarity score is higher than 10%. While this generates a very strong matching for most observations, the result is more speculative in some cases. In columns 1 and 2 of Table 5, we discard matches that are respectively below 50% or below 80% in terms of similarity score. This means that those observations are dropped both from the estimation, but also from the construction of the localization and urbanization measures. In both cases, our results are virtually unchanged.

Process and product innovation. Columns 3 and 4 of Table 5 show separate results for product innovation and process innovation respectively as dependent variables. The results are very close to our baseline in both cases.

Heterogeneity across industries, regions and firms. To test whether the effects differ across types of manufacturing activities, we group all industries that pertain to ISIC codes 19 to 30, combining goods typical though of as more advanced, such as chemicals, electronics, pharmaceutical, metal products and motor vehicles ("Advanced"). We group all other manufacturing industries, pertaining to food, textile, wood paper or furniture in a second categroy ("Basic"). Column 5 of Table 5 reports our baseline result only for advanced industries while column 5 reports results only for basic ones. The coefficients are in line with our baseline estimation for both sub-groups. We then split our sample between firms that have less than 10 employees in column 7 and firms with 10 employees or more in column 8. Again, our main results are confirmed for both subgroups and appear stronger for large firms. Finally, we exclude firms based in the region of Greater Accra in column 9 and the 3 regions of the North (Northern, Upper East and Upper West) in column 10. Our results continue to hold in both cases and appear stronger when excluding the capital region.

Table 5: Robustness - continued

	(1) Fuzzy > .5	(2) Fuzzy > .8	(3) Prod. Inn.	(4) Proc. Inn.	(5) Adv. ind	(6) Basic ind.	(7) Small firms	(8) Large firms	(9) Excl. Accra	(10) Excl. North
Localization	-0.15***	-0.14***	-0.16***	-0.17**	-0.21***	-0.14**	-0.11*	-0.24***	-0.23***	-0.10*
	(-3.07)	(-2.71)	(-2.91)	(-2.44)	(-2.69)	(-2.21)	(-1.65)	(-3.10)	(-3.84)	(-1.88)
Urbanization	0.30***	0.28***	0.33***	0.26**	0.33**	0.29***	0.21**	0.44***	0.38***	0.24**
	(3.55)	(3.19)	(3.62)	(2.42)	(2.41)	(2.96)	(2.17)	(3.47)	(3.92)	(2.46)
Log(age)	0.01	0.02	0.07	-0.12	0.00	0.00	0.11	-0.12	-0.01	0.03
	(0.10)	(0.36)	(1.06)	(-1.32)	(0.04)	(0.05)	(1.29)	(-1.19)	(-0.21)	(0.40)
Log(empl)	0.91***	0.93***	0.90***	1.41***	0.80***	0.97***	0.97***	1.33***	0.87***	0.83***
	(5.90)	(5.71)	(5.24)	(6.72)	(2.69)	(5.78)	(4.86)	(2.81)	(5.34)	(4.88)
Foreign owned	-0.29	-0.25	-0.30	-0.45*	-0.21	-0.31	-0.62	-0.39*	-0.01	-0.32*
	(-1.47)	(-1.11)	(-1.29)	(-1.82)	(-0.74)	(-1.13)	(-0.66)	(-1.80)	(-0.02)	(-1.65)
Incorporated	0.24	0.18	-0.00	0.58***	0.44*	0.16	0.66***	-0.09	0.17	0.19
	(1.62)	(1.20)	(-0.01)	(2.96)	(1.78)	(0.93)	(2.98)	(-0.45)	(0.75)	(1.21)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5121	4300	5272	5256	1553	3719	3679	1511	4342	4120
Log-Likelihood	-1557.2	-1330.8	-1352.8	-948.56	-482.56	-1116.3	-1027.6	-537.02	-1215.4	-1263.4

All regressions are complementary log-log models. The dependent variable is a dummy for whether a firm innovates except in columns 3 and 4, where it is respectively a dummy for whether the firm conducts product or process innovation. Standard errors are clustered at the locality level. Urbanization and Localization are computed as in (1) and (2) based on localities within 25km and 2-digit industry. Industry dummies are at the 2-digit ISIC level and regional dummies in all columns are based on the 10 regions of Ghana (see Figure 3). In Columns 1 and 2, we recompute the data using different thresholds of similarity scores in the fuzzy matching (see section 3). Columns 5 to 10 split the sample in different ways. Colums 5 and 6 split firms in advanced and basic industries. Columns 7 and 8 look separately at small and large firms. Columsn 9 and 10 exclude respectively firms in the Greater Accra region and firms in the North (Upper East, Upper West and Northern regions).

6 Channels

In this section, we provide some tentative mechanisms explaining the strong positive effect of urbanization on innovation and the negative or zero effect of localization. For this, we use the response to additional questions in the dataset regarding the difficulties that firms face when innovating, their motivation or their sources of knowledge.

Factors hampering innovation. All firms in the sample were asked about the factors "hampering your innovation activities or influencing your decisions not to innovate". This question has the advantage that it gives a precise answer about difficulties faced by firms whether they innovate or not. We construct for each factor a dummy that takes value one if the firm declares a factor to be very important. The identified factors are (i) the lack of access to funds, internal or external to the firm, (ii) the high costs of innovating, (iii) the lack of knowledge (lack of qualified personnel, of information about technology or of innovatin partners), (iv) the lack of information about markets, or (v) the market being dominated by established firms. Firms can also declare that they did not perceive the need to innovate. Firms can identify several factors as important in hampering innovation and they often do. The median firm declares 2 out of the 5 factors as very important[10]. Table 8 in the appendix gives some descriptives statistics about the different factors. Those factors that firms mention most often as very important pertain to access to funding, to knowledge or the high costs of innovation. Table 6 shows how our measures of localization and urbanization correlate with each of the factors hampering innovation, conditional on the same variables as in the previous section. Firms are significantly less likely to mention access to funding or access to knowledge as problematic in denser economic places (urbanization). These negative coefficients are consistent with the positive effect of urbanization on innovation, and give a hint as to the reason behind that effect. In places with more employment in the same industry (localization), on the other hand, firms complain significantly more about the lack of information and uncertainty about market conditions as reasons not to innovate. The presence of many other firms in the sector may make the environment more difficult to navigate. Interestingly, a higher localization does not come with firms mentioning other established firms as a reason for not innovating. The negative effect of the presence of other firms in the sector (conditional on urbanization) may thus not be a perceived increase in competition per se but a decrease in the visibility of the market, with a resulting negative impact on innovation. While not implausible, this interpretation remains of course speculative at that stage.

[10]The median firm declares 4 out of 5 factors as important. This is the reason why we concentrate on those factors that firms define as very important as they appear more discriminating.

Table 6: Factors hampering innovation

	(1) Funding	(2) Costs	(3) Knowledge	(4) Market info	(5) Competitors	(6) No need
Localization	-0.00	0.01	0.02	0.08**	0.03	0.03
	(-0.01)	(0.32)	(0.91)	(2.34)	(0.79)	(0.80)
Urbanization	-0.05*	-0.01	-0.10***	-0.05	0.08	0.01
	(-1.73)	(-0.37)	(-2.62)	(-1.05)	(1.52)	(0.20)
Log(age)	-0.07**	-0.05	-0.08**	-0.08*	-0.08*	-0.08*
	(-2.50)	(-1.55)	(-2.45)	(-1.87)	(-1.73)	(-1.91)
Log(empl)	-0.07	-0.02	0.01	0.09	-0.12	0.18**
	(-1.24)	(-0.29)	(0.21)	(1.26)	(-1.48)	(2.01)
Foreign owned	-0.53***	-0.21	-0.45**	-0.50**	0.15	-0.19
	(-3.60)	(-1.23)	(-2.53)	(-2.19)	(0.76)	(-1.18)
Incorporated	-0.19**	-0.02	-0.35***	-0.17	-0.32***	-0.17
	(-2.45)	(-0.17)	(-4.01)	(-1.64)	(-2.60)	(-1.51)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5277	5015	5272	5273	5021	5272
Log-Likelihood	-3401.9	-3299.5	-3435.5	-2971.9	-2529.8	-2657.1

All regressions are complementary log-log models. The dependent variable is a dummy for whether a firm declares a factor as hampering innovation. Urbanization is the number of people employed per square kilometer (following (2)) within 25km of the firm's locality. Localization is defined as the number of people employed in the same industry (following (1)) per square km, with 2-digit industries. Standard errors clustered at the locality level. Industry dummies are at the 2-digit level, region dummies are based on the 10 regions of Ghana (see Figure 3). The detailed description of factors hampering innovation is available in Table 8.

Motivation to innovate and sources of knowledge. The survey contains a question, asked to all innovating firms, about the reasons that motivated them to engage in innovating activities. About 300 firms answered whether customers, competitors, suppliers or other firms that bought their products ("buyers") were important in motivating their decision to engage in innovative activites. Among the respondents, 88% mention customers as a very important reason for the decision to engage in innovation activities. 59% of innovators mention competitors and 53% identify firms that buy their output as very important reasons for innovation. Only 30% say that suppliers were very important in their decision. The first four columns of Table 7 presents a similar specification as in the previous section where the dependent variable is a dummy that takes value one if the firm mentions one particular factor as a very important motivation. The interpretation is of course different as we are conditioning on a sample of firms that innovate. The purpose of the regression is to identify whether, among innovating firms, the stated motivation to innovate differs across firms in more or less dense areas. Table 7 shows in particular that firms located close to other firms in the sector are more likely to mention buyers or suppliers as a motivation to innovate. Considering that many firms will buy or sell inputs to firms in the same industry, this may be seen as a hint for the existence of some positive Marshallian externalities. As an additional piece of evidence, we use another item in the survey, which asks all firms conducting innovating activities whether they "used [...] external sources of information or ideas for any innovation activity". Columns 5 to 7 rerun our analysis and suggest that innovating firms with a higher value of localization are more likely to report having used knowledge from other firms (parents, suppliers or buyers). They show no difference for external knowledge from other types of institutions (Academic, consultancies or business associations), which we denote as "Research links" or from other sources. Conditional on innovating, firms close to other firms in the same sector are thus more likely to report buyers and sellers not only as a motivation to innovate but also as a source of external knowledge, in line with at least some role for Marshallian externalities.

7 Conclusion

This paper analyzes how the spatial concentration of economic activity affects innovation among firms in Ghana. We use the 2014 census of all establishments to map economic activity at the level of localities, combined with a detailed survey on innovation administered to more than 5000 firms. We find a positive and robust effect of the density of economic activity on innovation (urbanization economies), which is strongest when defining the density within 25 to 60km of an establishment's locality. Conditional on urbanization, we also identify a negative effect of the density of employment in an establishment's sector (localization economies). We find that

Table 7: Motivation to innovate

	Motivations to innovate				External knowledge		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Consumers	Competitors	Buyers	Suppliers	Firm links	Research links	Other links
Localization	0.11	0.07	0.22^{**}	0.25^{**}	0.31^{**}	0.19	0.15^{*}
	(1.11)	(0.71)	(2.12)	(2.23)	(2.40)	(1.41)	(1.66)
Urbanization	-0.11	-0.05	-0.15	-0.27	-0.25	-0.36	-0.16
	(-0.81)	(-0.32)	(-0.84)	(-1.29)	(-1.17)	(-1.64)	(-1.03)
Log(age)	-0.03	-0.00	-0.21^{*}	0.01	0.19	0.01	0.06
	(-0.32)	(-0.03)	(-1.65)	(0.08)	(1.17)	(0.06)	(0.56)
Log(empl)	-0.16	0.45	0.35	0.33	0.59	1.23^{**}	0.60^{**}
	(-0.61)	(1.44)	(1.12)	(0.78)	(1.52)	(2.36)	(2.24)
Foreign owned	0.10	-0.35	-0.26	-0.17	0.67^{*}	-0.08	-0.13
	(0.28)	(-0.91)	(-1.01)	(-0.32)	(1.67)	(-0.17)	(-0.36)
Incorporated	0.47^{*}	0.11	1.01^{***}	0.40	0.61^{*}	0.64^{**}	0.10
	(1.82)	(0.43)	(4.19)	(1.21)	(1.95)	(2.04)	(0.45)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	246	295	292	292	296	294	293
Log-Likelihood	-86.677	-182.85	-171.79	-154.61	-169.76	-157.38	-175.32

All regressions are complementary log-log models. The dependent variable is a dummy for whether a firm declares as very important a motivation to innovate (columns 1 to 4) or a source of external knowledge (columns 5 to 7). Urbanization is the number of people employed per square kilometer (following (2)) within 25km of the firm's locality. Localization is defined as the number of people employed in the same industry (following (1)) per square km, with 2-digit industries. Standard errors clustered at the locality level. Industry dummies are at the 2-digit level, region dummies are based on the 10 regions of Ghana (see Figure 3).

both effects are economically large, with the probability to innovate for the average establishment increasing by 4 percentage points or decreasing by 3 percentage points for a one standard deviation increase in urbanization and in localization respectively. Our results are conditional on many establishment-level characteristics as well as on industry and region fixed effects. To tackle the well-known issue of endogeneity in such regressions, we replicate our analysis using a subsample of establishments declaring that their location is close to where the founder was born or grew up, i.e. those with a plausibly exogenous geographic location. We also control for some measures of the quality of infrastructure at the municipality level to further reduce the risk of an omitted variable bias. Turning to the mechanisms behind our results, we show that firms in regions with denser economic activites report less problems to access funding and knowledge, while the presence of firms in the same sector is associated with more uncertainty about the gains from innovating. It is worth noting that the negative coefficient on localization does not necessarily mean the absence of Marshallian externalites. Among firms that innovate, those with a higher value of localization typically disproportionately report their suppliers or buyers as a motivation to innovate, and identify them as sources of external knowledge. These externalities however seem too weak to generate positive localization economies on average.

Taken together, our results add to the nascent literature on the agglomeration innovation nexus in developing countries, with very precise measures of spatial concentration in an African context. We find a large positive effect of urbanization on innovation even at a 50km range. This is a relatively longer distance than typically found in developed countries (Rosenthal and Strange (2020)), where both the type of innovation and the channels through which urbanization matters may differ. Our use of detailed survey questions gives new indications of the channels through which urbanization and localization effects act in a country like Ghana. Interestingly, access to knowledge seems indeed facilitated by urbanization economies, consistent with the view of Jacobs. Beyond knowledge, a facilitated access to finance appears as a key component of the positive effect of urbanization, a channel that may be stronger in developing countries. The uncertain returns to innovation for firms with a higher localization is in our view an interesting result, which deserves further investigation and points to the need for further disentangling the specific mechanisms to developing countries in the relationship between agglomeration and innovation.

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and innovation: an inverted-U relationship. *Quarterly Journal of Economics 120*, 701–728.

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. *American Economic Review 97*, 1611–1638.

Atkin, D., A. Khandelwal, and A. Osman (2017). Exporting and firm performance: Evidence from a randomized experiment. *Quarterly Journal of Economics 132*(2), 551–615.

Baptista, R. and P. Swann (1998). Do firms in clusters innovate more? *Research Policy 27*(5), 525 – 540.

Baum-Snow, N. and F. Ferreira (2015). Causal inference in urban economics. In G. Duranton, V. Henderson, and W. Strange (Eds.), *Handbook of Urban and Regional Economics*, Volume 5A, pp. 3–68.

Beaudry, C. and A. Schiffauerova (2009). Who's right, marshall or jacobs? the localization versus urbanization debate. *Research Policy 38*(2), 318 – 337.

Blalock, G. and P. J. Gertler (2009). How firm capabilities affect who benefits from foreign technology. *Journal of Development Economics 90*(2), 192 – 199.

Bos, M. and G. Vannoorenberghe (2018). Total factor productivity spillovers from trade reforms in india. *Canadian Journal of Economics 51*(2), 549 – 606.

Bos, M. and G. Vannoorenberghe (2019). Imported input varieties and product innovation: evidence from five developing countries. *Review of International Economics 27*, 520 – 548.

Carlino, G., S. Chatterjee, and R. Hunt (2007). Urban density and the rate of invention. *Journal of Urban Economics 61*(3), 389–419.

Carlino, G. and W. Kerr (2015). Agglomeration and innovation. In G. Duranton, V. Henderson, and W. Strange (Eds.), *Handbook of Urban and Regional Economics*, Volume 5A, pp. 349–404.

Chhair, S. and C. Newman (2014). Clustering, competition, and spillover effects: evidence from cambodia. WIDER Working Paper 2014/065.

Ciccone, A. and R. Hall (1996). Productivity and the density of economic activity. *American Economic Review 86*(1), 54–70.

Cirera, X. and S. Muzi (2020). Measuring innovation using firm-level surveys: evidence from developing countries. *Research Policy 49*(103912).

Combes, P.-P., S. Démurger, and S. Li (2013). Urbanisation and migration externalities in china. *CEPR Discussion Paper 9352*.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In G. Duranton, V. Henderson, and W. Strange (Eds.), *Handbook of Urban and Regional Economics*, Volume 5A, pp. 247–348.

Damijan, J. P., M. Rojec, B. Majcen, and M. Knell (2013). Impact of firm heterogeneity on direct and spillover effects of fdi: Micro-evidence from ten transition countries. *Journal of comparative economics 41*(3), 895–922.

Desmet, K. and S. Parente (2010). Bigger is better: market size, demand elasticity and innovation. *International Economic Review 51*(2), 319–333.

Duranton, G. (2015). Growing through cities in developing countries. *The World Bank Research Observer 30*(1), 39–73.

Duranton, G. and D. Puga (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review 91*(5), 1454–1477.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In V. Henderson and J. Thisse (Eds.), *Handbook of Urban and Regional Economics*, Volume 4, pp. 2063–2117.

Duranton, G. and D. Puga (2020). The economics of urban density. *Journal of Economic Perspectives 34*(3), 3–26.

Feldman, M. and D. Audretsch (1999). Innovation in cities: science-based diversity, specialization and localized competition. *European Economic Review 43*(2), 409–429.

Figueireido, O., P. Guimarães, and D. Woodward (2002). Home-field advantage: location decisions of portuguese entrepreneurs. *Journal of Urban Economics 52*(2), 341–361.

Glaeser, E., H. Kallal, J. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of Political Economy 100*(6), 1126–1152.

Glaeser, E. and A. Saiz (2004). The rise of the skilled city. *Brookings-Wharton papers on urban affairs 5*, 47–94.

Goldberg, P., A. Khandelwal, N. Pavcnik, and P. Topalova (2010). Imported intermediate inputs and domestic product growth: evidence from india. *Quarterly Journal of Economics 125*(4), 1727–1767.

Görg, H. and D. Greenaway (2004). Much ado about nothing? Do domestic firms really benefit from Foreign Direct Investment? *World Bank Research Observer 19*(2), 171–197.

Gorodnichenko, Y., J. Svejnar, and K. Terrell (2010). Globalization and innovation in emerging markets. *American Economic Journal: Macroeconomics 2*(2), 194–226.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Heblich, S., M. Seror, H. Xu, and Y. Zylberberg (2022). Industrial clusters in the long run: evidence from million-rouble plants in china. NBER Working paper 30744.

Helsley, R. and W. Strange (2002). Innovation and input sharing. *Journal of Urban Economics 51*(1), 25–45.

Howard, E., C. Newman, J. Rand, and F. Tarp (2014). Productivity-enhancing manufacturing clusters: Evidence from Vietnam. WIDER Working Paper 2014/071.

Hsieh, C.-T. and B. Olken (2014). The missing "missing middle". *Journal of Economic Perspectives 28*(3), 89–108.

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

Jaimovich, E. (2021). Quality growth: From process to product innovation along the growth path. *Economic Theory 71*, 761–793.

Javorcik, B. (2004). Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. *American Economic Review 93*(3), 605–627.

Knoben, J., A. Arikan, F. van Oort, and O. Raspe (2016). Agglomeration and firm performance: one firm's medicine is another firm's poison.˙ *Environment and Planning A 48*(1), 132–153.

Knoben, J., D. Speldekamp, and H. Hulshof (2022). Managing where you are: agglomeration economies, managerial industry experience and innovation in south-east asia. *Industry and innovation*.

Kokko, A., R. Tansini, and M. C. Zejan (1996). Local technological capability and productivity spillovers from fdi in the uruguayan manufacturing sector. *The Journal of Development Studies 32*(4), 602–611.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Martin, P., T. Mayer, and F. Mayneris (2011). Spatial concentration and firm-level productivity in france. *Journal of Urban Economics 69*, 182–195.

Michelacci, C. and O. Silva (2007). Why so many local entrepreneurs? *Review of Economics and Statistics 89*(4), 615–633.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014). Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics 96*(5), 967–985.

Nzau, C., T. Ingana, R. Mavisi, C. Kariuki, B. Mulatya, F. Ndakala, R. Kongani, B. Obasi, and B. Mworia (2016). Kenya innovation indicators survey. Working paper, Kenya ministry of Education, Science and Technology.

Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press.

Rosenthal, S. and W. Strange (2003). Geography, industrial organization and agglomeration. *Review of Economics and Statistics 85*, 377–393.

Rosenthal, S. and W. Strange (2010). Small establishments/big effects. In E. Glaeser (Ed.), *Agglomeration Economics*, Chapter 9, pp. 277–302. Chicaco: The University of Chicago Press.

Rosenthal, S. and W. Strange (2020). How close is close? the spatial reach of agglomeration economies. *Journal of Economic Perspectives 34*(3), 27–49.

Sanfilippo, M. and A. Seric (2016). Spillovers from agglomeration and inward fdi: a multilevel analysis on sub-saharan african firms. *Review of World Economics 152*, 147–176.

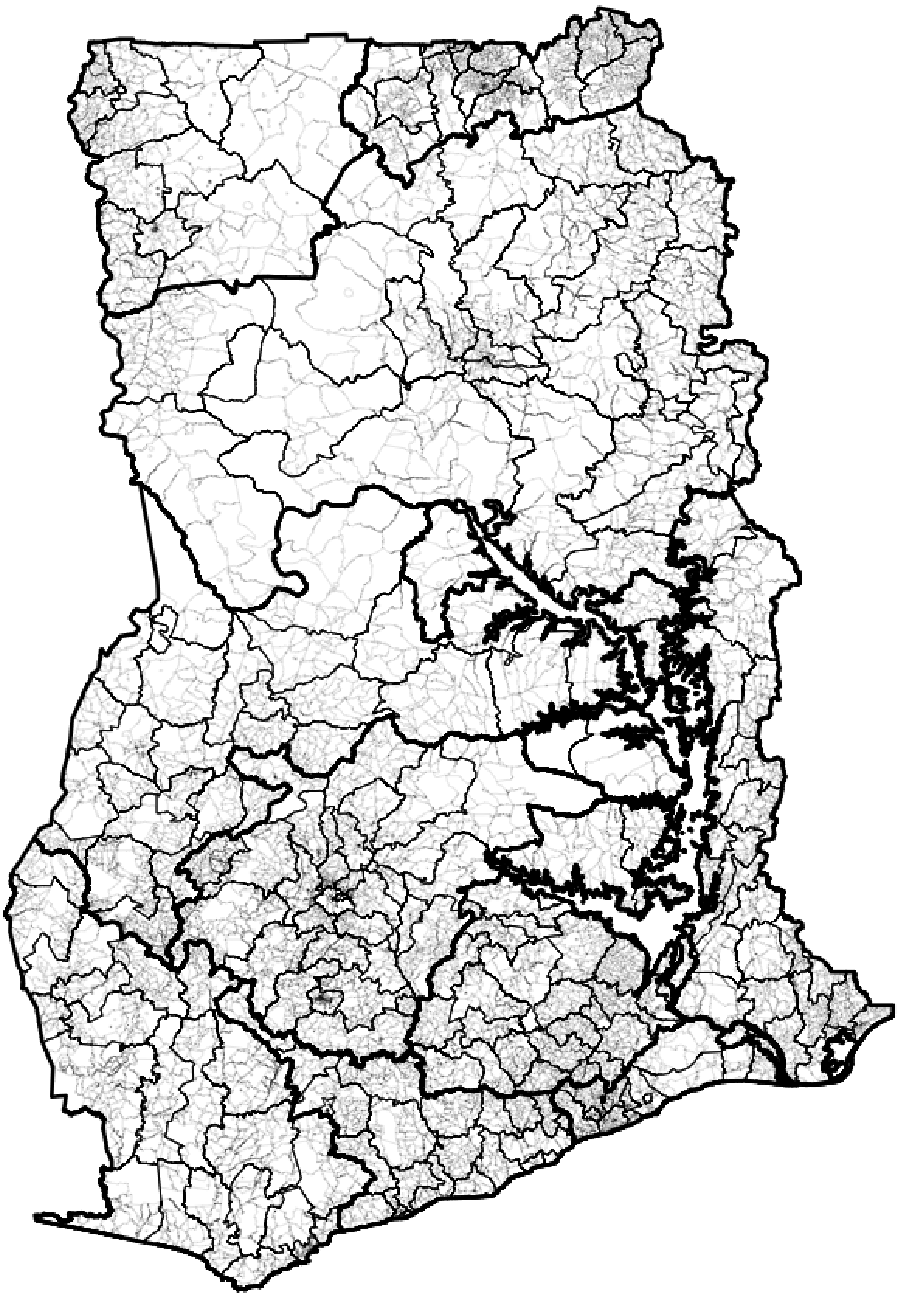
Siba, E., M. Soderbom, A. Bigsten, and M. Gebreeyesus (2012). Enterprise agglomeration, output prices and physical productivity: firm-level evidence from ethiopia. WIDER Working Paper 2012/085.

Zhang, H. (2015). How does agglomeration promote the product innovation of chinese firms? *China Economic Review 35*, 105 – 120.

Appendix

Maps

Figure 3: Administrative map of Ghana



The thinnest lines represent localities, the intermediate lines show districts and the thickest lines the regions.

Figure 4: Employees per square km by locality - Phase 1)

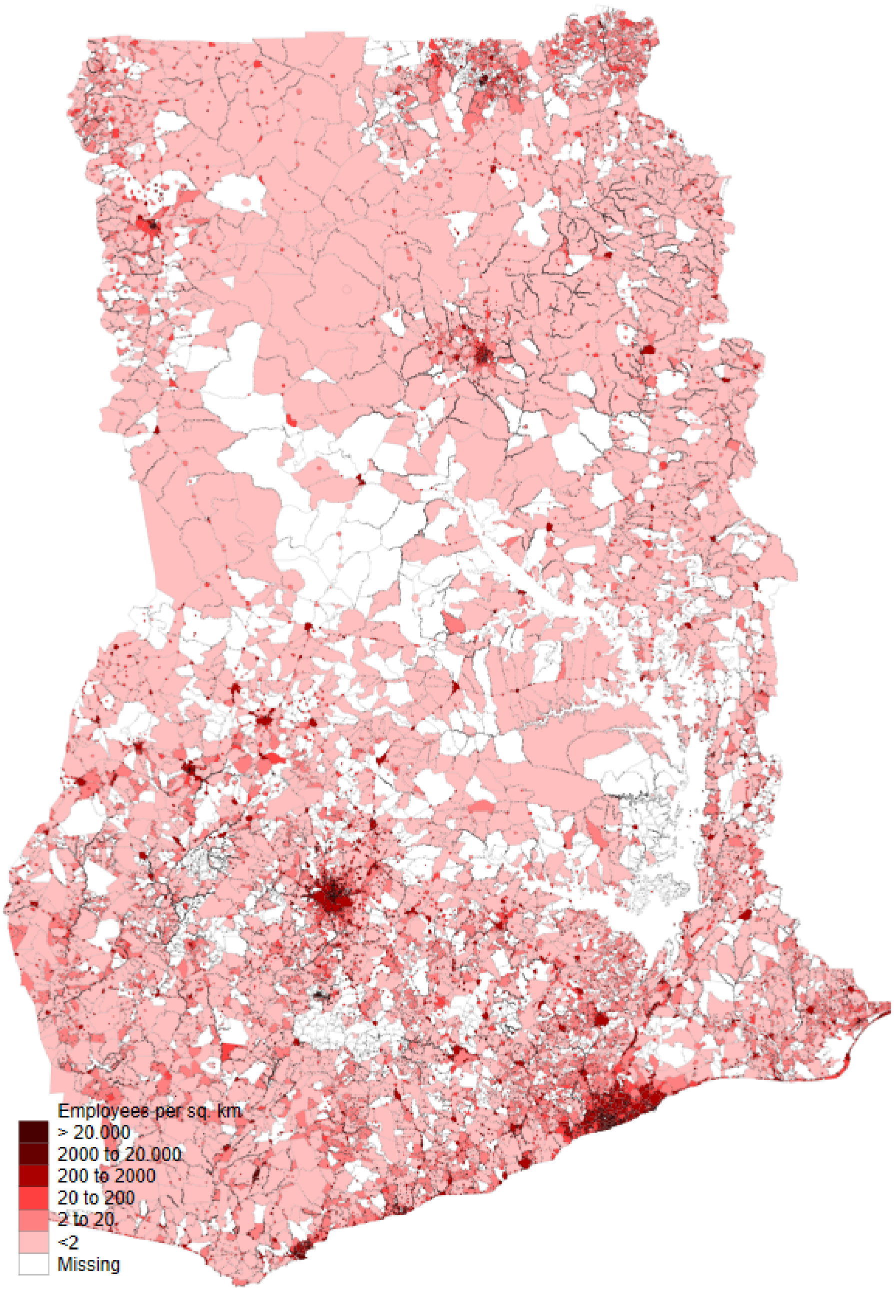
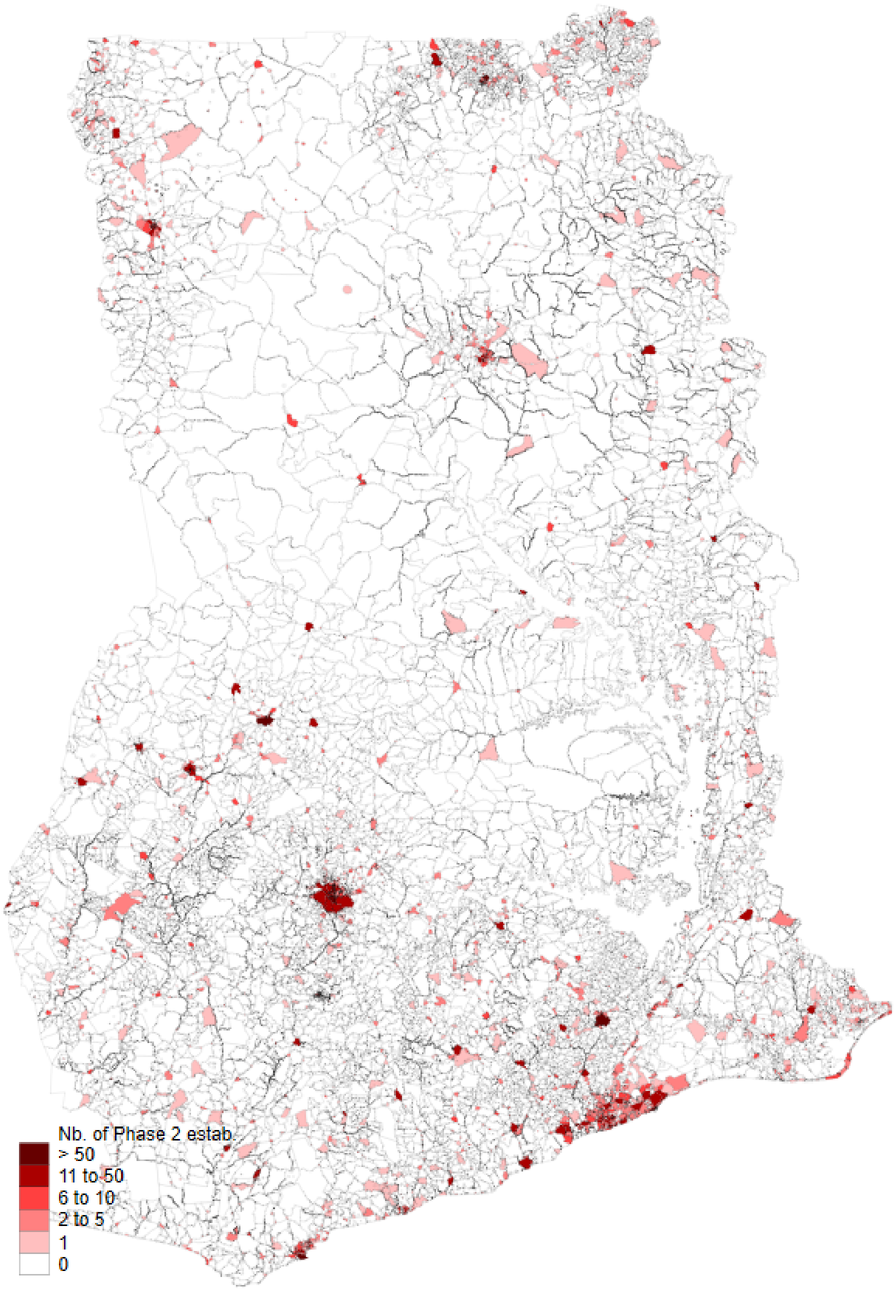


Figure 5: Location of phase 2 establishments



Tables and Figures

Table 8: Factors that hampered innovation

	Non-innovators		Innovators	
	Yes	No	Yes	No
Funding	2721	2029	333	198
Costs	1740	2768	206	309
Knowledge	1968	2782	216	315
Market info	1227	3523	156	375
Established firms	934	3575	125	391
No need	1025	3725	74	457

Answers to the question "How important were the following factors in hampering your innovation activities or influencing your decisions not to innovate". "Yes" means "very important" and "No" to "important" or "not important". Yes to Funding means that the firm answered very important to "Lack of funds within your firm or group" or to "Lack of external sources of funding". Costs refers to the answer "Innovation costs too high". Yes to Knowledge means very important to "Lack of qualified personnel", "Lack of information on technology" or "Difficult in finding co-operation partners for innovation". Yes to Market info means very important to "Lack of information on markets" or to "Uncertain demand for innovative goods or services". Yes to No need means very important to "No need due to prior innovation" or "No need because of no demand for innovation". Innovators are those firms that self-report having introduced a process or a product innovation in 2013.

INSTITUT DE RECHERCHE
ÉCONOMIQUES ET SOCIALES

Place Montesquieu 3

1348 Louvain-la-Neuve

ISSN 1379-244X D/2023/3082/05