Market Potential, Agglomeration Effects and the Location of French Firms in Africa

Alain Pholo Bala, Michel Tenikue and Baraka Nafari

ERSA working paper 774

March 2019

Market Potential, Agglomeration Effects and the Location of French Firms in Africa

Alain Pholo Bala[*,†] Michel Tenikue[‡] Baraka Nafari[*]

January 29, 2019

Abstract

This paper informs the debate on the existence of agglomeration effects in Africa. It uses a structural estimation approach to investigate the impact of agglomeration economies and forward linkages on the localization of French affiliates in Africa. Using a sample of French subsidiaries in Africa, we compare the theoretically derived measure of market potential with the standard form used by geographers and with a measure of local demand. Our results show that market potential matters for location choice. However, the semi-elasticity estimates suggest that the intensity of demand linkages in Africa is lower than what has been observed in the European Union. Moreover, their effects seem to be insignificant when we consider the spillover variables. These spillover effects have a positive and significant impact on location which suggests that agglomerations effects are at play throughout Africa.

Keyword: Agglomeration Economies, Location of Firms, Market Potential, Africa.

JEL Classification: F12, F15, R30, R32, R34.

[*]School of Economics, University of Johannesburg, APK Campus, PO Box 524, Auckland Park 2006, Johannesburg, South Africa.

[†]Public and Environmental Economics Research Centre (PEERC), University of Johannesburg, *Corresponding author*, e-mail address: `apholo@uj.ac.za`

[‡]Luxembourg Institute of Socio-Economic Research (LISER), Maison des Sciences Humaines, 11 Porte des Sciences, L-4366 Esch-sur-Alzette/Belval, Luxembourg.

1 Introduction

Foreign direct investments (FDI), one of the most powerful expression of the globalization process, have dramatically impacted the World economy in the last twenty years. While described as the poorest continent of the globe and representing from 1990 to 2017 only 1.39 to 2.92% of the total FDI inflows, Africa has not been set aside in these dynamics as the continent has experienced a substantial increase of the volume of FDI inflows: from US 2,845 millions in 1990 to US 41,772 millions of dollars in 2017.[1] There is a strong appeal throughout the World and even within Africa for capturing a significant share of these FDI inflows.

There is an extensive literature on what determines the ability to attract FDI inflows. Among several factors, the focus has been put on agglomerations effects. This is particularly true in the empirical literature based on the New Economic Geography (NEG) which has acquired quite a preferential space in the global FDI literature (Head and Mayer, 2004*a*,*b*; Amiti and Javorcik, 2008; Debaere et al., 2010). This paper pertains to this NEG strand of the FDI literature. Through this paper, we use the framework proposed by Head and Mayer (2004*a*) to explain location choices of French firms in Africa. To inform the debate about the existence and the intensity of agglomeration economies in Africa, there is a need for empirical studies providing new evidence on agglomeration effects in the African region. In this research, we aim to contribute to bridging this gap by investigating, through a structural estimation approach, the impact of agglomeration economies on the localization of French affiliates in Africa.

Consequently, we follow closely the approach of Head and Mayer (2004*a*) to perform an empirical investigation to determine whether these firms tend to locate "where the markets are". Hence, we resort to two different explanations to account for the location choices of French firms in the African continent. Forward linkages represent the first factor that we consider in our analysis. Hence, consistently with the New Economic Geography (NEG) literature, we assume that firms locate in sites where the demand is the highest and provide for smaller markets through exports (Head and Mayer, 2004*a*). The second factor that we put forward is the presence of other French firms in African regions. The economic geography literature has evidenced a statistical tendency of firms to make the same location decisions as previous firms with similar characteristics (such as national origin and industry).

The observation that, in the case of FDI, investors often agglomerate around investments from the same country (*agglomeration effects*) is well documented in the literature (Debaere et al., 2010). In their analysis of the location decision of US firms in Ireland, Barry et al. (2003) state

[1]Source: UNCTAD, FDI/MNE database (`http://www.unctad.org/fdistatistics`) (last access July 04, 2018).

that investors may show up a tendency to imitate each other's location choice due to uncertainty. Since foreign investors face greater uncertainty in the host country than local firms, they may interpret the presence of firms from their home country as a positive signal of the location's attractiveness. The fact that investors often agglomerate close to other investors from the same country of origin is one manifestation of the learning effects. Firms also cluster for sharing and matching purposes, i.e. to take advantage of the increasing availability of specialized labor and a growing pool of input providers (Duranton and Puga, 2004; Debaere et al., 2010).

This highlights the role of backward and forward linkages which underline the complementarities between firms of related production stages (Debaere et al., 2010; Head and Mayer, 2004*b*). The concentration of upstream firms indicates the accessibility to component suppliers in the region, whereas the concentration of downstream firms and final goods consumers shows the accessibility to markets (Du et al., 2012). As several of these different types of agglomeration economies might be simultaneously at play to incite a firm to opt for a specific location, it might be challenging to identify their separate effects. We will try to identify them by setting up an estimation strategy where we exclude the region of the affiliate from the calculation of the market potential variables. Tackling endogeneity is also a daunting task. Two causes of endogeneity may be at play here: simultaneous causality and omitted variables. Excluding the region where the affiliate is located from the computation of the market potential is also a way to alleviate the simultaneous causality bias. We further address this bias by using lagged information on French FDI and trade flows to compute the variables capturing agglomeration effects and the market potential variables. To reduce the omitted variable bias we include a range of country and firms fixed effects in our Linear Probability Models (LPM).

Through this paper we intend to contribute to the NEG literature on the determinants of firms' locations choice. While most of the insights of the NEG seem to have been well accounted for after the contribution of Amiti and Javorcik (2008),[2] and are well integrated in the empirical analysis of FDI determinants (Debaere et al., 2010), one can complain about the geographical bias of the empirical studies based on the NEG.

As correctly pointed by Hayakawa and Tsubota (2014), most existing studies consider de-

[2]While Head and Mayer (2004*a*) brought a significant contribution at that time, since then some studies have extended the existing literature. They consider the importance of supplier access in addition to market access (Amiti and Javorcik, 2008), and/or they design their measures of market and supplier access in order to take into account the varying degrees of inter-industry linkages (Debaere et al., 2010). The challenge with these contributions is that they used comprehensive data sets that cover a comprehensive set of manufacturing industries at a highly disaggregated level. Such a procedure cannot be replicated for an African continent characterized by the paucity of disaggregated data at a sector or industrial level. Moreover, computing supplier access would require to compile an input-output matrix which is quite challenging at the level of the African continent. Considering supplier access would only make sense if one considers a single country. But, we do not have enough data in our database to do so.

veloped countries, particularly European countries (Head and Mayer, 2004*a*,*b*; Crozet, 2004), and USA (Hanson, 2005; Redding and Venables, 2004). Nevertheless, thanks to the increasing availability of high quality firm-level data, a rising number of contributions have focused on Asia (with a strong concentration on China (Amiti and Javorcik, 2008; Debaere et al., 2010; Tokunaga and Jin, 2011) and with Hayakawa and Tsubota (2014) examining location choices in East Asian developing countries). Africa has been left aside by the empirical FDI literature based on the NEG; but not quite from the overall FDI literature. Since 2002, when Asiedu (2002) noticed that there is no published empirical study on FDI that focuses on Africa, there has been new developments in the empirical FDI literature concerning the African continent (Asiedu, 2002, 2004, 2006; Fedderke and Romm, 2006; Kandiero and Chitiga, 2014; Lederman et al., 2013; Naudé and Krugell, 2007; Sanfilippo, 2010). Most of these studies analyze the determinants of FDI with aggregated data by using the ratio of net FDI flows to GDP (Asiedu, 2002, 2004, 2006; Kandiero and Chitiga, 2014; Naudé and Krugell, 2007).

Other contributions use aggregate data on the stock of FDI. Fedderke and Romm (2006) performed structural analysis of the growth impact of foreign direct investment, as well as its determinants. To check the predictions of their model, they use data on real foreign direct investment liabilities (stock) in a VECM model. Sanfilippo (2010) also uses a stock variable, the stock of Chinese Outward Foreign Direct Investment to investigate the determinants of Chinese Outward FDI in 41 African countries over the period 1998-2007. None of these publications have used firm level data to analyze the determinants of the location of foreign affiliates in Africa at a sub-national level. Lederman et al. (2013) use firm level data to provide empirical evidence on the microeconomic consequences and macroeconomic causes of FDI in the Southern African Development Community (SADC). They do not analyze the impact of forward linkages and agglomeration effects on the location or the entry of foreign firms in Africa.

Some contributions investigate the impact of agglomeration on FDI: Anyanwu (2012) tests for agglomeration effects by checking how FDI inflows relates to past FDI inflows. However, they do so by using an ad-hoc specification and aggregate data that may not help to identify agglomeration economies at a regional level.

Finally, our contribution departs from most of the papers about FDI in Africa as they consider the impact of the endowment of natural resources on FDI inflows (Asiedu, 2002, 2004, 2006; Kandiero and Chitiga, 2014; Naudé and Krugell, 2007; Sanfilippo, 2010). Naudé and Krugell (2007) investigate the impact of geographical factors (climate, endowments of natural resources, disease burden etc.) on the level of FDI in Africa. These factors point to exogenously given characteristics of different locations, such as the type of climate, the presence of raw materials, the vicinity to natural ways of communication, etc. This is the 'first nature' mentioned in the geographical economics literature. Consistently with the NEG literature, we mainly focus on 'second nature'. 'Second nature' emerges as the outcome of human beings' actions to improve on the uneven distribution of natural resources (Ottaviano and Thisse, 2004).[3]

This paper is the first attempt to provide empirical evidence of forward linkages and agglomeration economies in Africa through a structural approach. To capture forward linkages, we use three measures of the market potential. The first is the Krugman Market Potential which is derived from theory (Krugman, 1992) and takes into account market accessibility and competition among firms located in the same region; the second is the Harris Market Potential which uses the inverse the physical distance between regions as a weight of the regional demand (Harris, 1954). The third measure is the regional GDP. As Head and Mayer (2004*a*), we use data on bilateral trade flow across Africa to estimate trade equations by industries. Based on the estimated set of parameters, we compute, for each industry, the Krugman Market Potential of the regions (Krugman, 1992). We then use data on French manufacturing firms extracted from the 2006 Survey of French Affiliates ("Enquêtes Filiales") to construct for each region the following proxies of agglomeration effects: the count of French firms from the same industry, the count of French firms having the same parent company, the total count of French firms located in the same region.

Our results show that the market potential of a given region affects the probability of firm to locate there. A 10% increase of the regional GDP yields an increase of the location probability from 1.6 to 1.8% , while the effect of a 10% increase of the Krugman (respectively Harris) Market Potential ranges from about 0.2 to 1.7% (respectively 1.1 to 2.1%). While significant, these effects are however weaker than those prevailing in the Europe Union where the marginal effects of the demand linkages varies from 3 to 11% according to Head and Mayer (2004*a*). Still, demand matters for the location choice of French firms in Africa.

This paper is organized as follows: in the next section we present the model. We start this section by deriving the profit equation of foreign affiliates from the consumer optimization problem. Through this derivation, we present the formal definition of the Krugman Market Potential. We derive our econometric specification by performing a monotone transformation of the profit equation. Then, we present the cross-country trade equation that we use to estimate the unknown parameters in the expression of the Market Potential. We describe the data and

[3]While 'first nature' is crucial to explain the location of heavy industries during the Industrial Revolution, because the proximity of raw materials was a critical factor, or why Cape Town remains attractive for so many tourists, it fails to provide a convincing explanation of many other clusters of activities, which are much less dependent on natural advantage (like the Silicon Valley) (Ottaviano and Thisse, 2004). 'Second nature' rather refers to the economic forces that determine the distribution of economic activities. It captures the cumulative causation process arising from the interaction between increasing returns and transport costs.

finally we present and discuss the empirical results afterwards.

2 Model and estimation strategy

The presentation of the model in subsection 2.1 follows Head and Mayer (2004a) closely. Then, in subsection 2.2 we present a modified version of the trade equation which is used in Head and Mayer (2004a) to estimate the unknown parameters that are present in the expression of the Market Potential.

2.1 The Profit equation for foreign affiliates

To derive the profit equation for foreign affiliates, Head and Mayer (2004a) assume a utility function "à la Dixit Stiglitz", with a CES specification. After maximizing such an utility function subject to budget constraint, they obtain a demand curve with a functional form similar to the following:

$$q_{ij} = \frac{p_{ij}^{-\sigma}}{\sum_{i=1}^{R} n_i p_{ij}^{1-\sigma}} Y_j \tag{1}$$

where Y_j denotes the expenditure in a representative industry in region j and q_{ij} represents the quantity of goods from region i consumed by a customer located in region j. They assume that each region produces unique product varieties and that each region $r = 1, \ldots, R$ produces n_r products. Yet, they also make the simplifying assumption of symmetric varieties k; as a consequence all products exported from region i sell for the same price p_{ij} in region j, $p_{ij} = p_i \tau_{ij}$. The trade cost factor τ_{ij} includes all transaction costs associated with moving goods across space and national borders, while $p_i = c_i \frac{\sigma}{\sigma-1}$ represents the mill price.

Substituting the expression of the mill price into equation (1), we obtain the quantity that a firm producing in region i will deliver to each destination j:

$$q_{ij} = \frac{\sigma - 1}{\sigma} \frac{(c_i \tau_{ij})^{-\sigma}}{P_j^{\sigma-1}} Y_j, \tag{2}$$

with $P_j = \left(\sum_{i=1}^{R} n_i (c_i \tau_{ij})^{1-\sigma} \right)^{\frac{1}{\sigma-1}}$. The gross profit earned in each importing region j for a firm producing in region i is

$$\pi_{ij} = (p_i - c_i) \tau_{ij} q_{ij} = \frac{(c_i \tau_{ij})^{1-\sigma}}{\sigma P_j^{\sigma-1}} Y_j. \tag{3}$$

Summing the gross profits earned in each market and subtracting the fixed costs F_r necessary to establish an affiliate k in region r, we obtain the aggregate net profit Π_r:

$$\Pi_r(k) = \frac{c_r(k)^{1-\sigma}}{\sigma} \sum_{j=1}^{R} \phi_{rj} \frac{Y_j}{P_j^{\sigma-1}} - F_r = \frac{c_r(k)^{1-\sigma}}{\sigma} \mathrm{MP}_r - F_r, \tag{4}$$

with

$$\mathrm{MP}_r = \sum_{j=1}^{R} \phi_{rj} \frac{Y_j}{P_j^{\sigma-1}}$$

MP_r is the Krugman Market Potential (KMP) (Krugman, 1992). The KMP aggregates the expenditures of all regions while adjusting for region r's access ϕ_{rj} and for competition from firms located in other regions, $P_j^{\sigma-1}$. HMP, Harris Market Potential (Harris, 1954), can be derived from MP_r by setting $P_j^{\sigma-1} = 1$ and $\phi_{rj} = 1/d_{rj}$: $\mathrm{MP}_r = \sum_{j=1}^{R} \frac{Y_j}{d_{rj}}$. The aggregate profit equation suggests that firms face a tradeoff between low production costs and high market potential. When a firm chooses its location, the only relevant information is the ordering of the profits. Invariant fixed costs do not affect the profit ordering of regions and can therefore be omitted. Thus, we suppose that fixed costs do not differ across locations ($F_r = F \ \forall\ r$)

We make the following transformation of the profit function:

$$V_r(k) = -\ln c_r(k) + (\sigma - 1)^{-1} \ln \mathrm{MP}_r \tag{5}$$

and we formalize the cost term as a Cobb-Douglas function with constant returns, using labor at cost w_r and other inputs (such as land and intermediates) at cost υ_r. Labor's share is α, and A_r represents total factor productivity. Therefore, log marginal costs are given by

$$\ln c_r(k) = \alpha \ln w_r(k) + (1 - \alpha) \ln \upsilon_r(k) - \ln A_r(k) \tag{6}$$

Substituting (6) into (5), we get

$$V_r(k) = -\alpha \ln w_r(k) + (\sigma - 1)^{-1} \ln \mathrm{MP}_r - (1 - \alpha) \ln \upsilon_r(k) + \ln A_r(k). \tag{7}$$

We observe wages w_r and will calculate MP_r using a trade equation. We do not observe υ_r and A_r, and they may be captured with several proxies. Therefore, we rather consider the following specification

$$V_r(k) = -\alpha \ln w_r(k) + (\sigma - 1)^{-1} \ln \mathrm{MP}_r - (1 - \alpha) \ln \nu_r(k) + \ln \Phi_r(k) + \varepsilon_r(k). \tag{8}$$

where $\nu_r(k)$ and $\Phi_r(k)$ represent observable proxies of other inputs and of TFP and $\varepsilon_r(k)$ is a random term capturing the effect of unobserved components of operating profits.

An implication of this set-up is that firms will choose to locate in the region r that offers the highest profit among the set Ω of all possible locations:

$$\mathrm{P}\left(\mathrm{aff}_r\left(k\right)=1\right)=\mathrm{P}\left(V_r\left(k\right)>V_s\left(k\right)\right), \qquad \forall r\neq s,\ s\in\Omega \tag{9}$$

where $\mathrm{aff}_r(k)$ is a dummy variable equal to 1 if the Multinational Entreprise (MNE) chooses to locate the affiliate k in region r and 0 otherwise.

We first estimate Equation (8) using a discrete choice model with a univariate extreme value marginal distribution of the $\varepsilon_r(k)$ errors. Decisions to implant an affiliate in a region are supposed to be independent from one another in this setting. Therefore, we can use a conditional logit model (CLM) to find out the probability for each region to host a French affiliate. The conditional logit model will assess how the features of the regions affect affiliates' likelihood of choosing them as a location. It is more appropriate than the multinomial logit which rather focuses on the role of individual characteristics in matching with certain categories of a dependent variable (Delbecque et al., 2014). In computing the likelihood though, the CLM will not take into account regions that do not host any French affiliates.

Then, we estimate Equation (8) with a mixed logit model. Unlike the CLM, the mixed logit uses the entire sample and is a highly flexible model that can approximate any random utility model (Train, 2009). The mixed logit avoids the limitations of the conditional logit by allowing for random taste variation, unrestricted substitution patterns (relaxation of the Independence of Irrelevant Alternatives (IIA) assumption), and correlation of unobserved factors over time. While the mixed logit model has been known for many years, it has only become fully applicable since the advent of simulation. Improvements in computer speed and in the understanding of simulation methods have allowed the full power of mixed logits to be utilized.

Finally, we will also estimate Equation (8) using a linear probability model (LPM). The major advantage of the linear probability model is its interpretability as opposed to the previous models. In the linear probability model the coefficients of regression represent the marginal effects of the regressors on the probability of a French subsidiary to locate in a African region. Over the logit models, and especially the mixed logit the LPM has also the advantage of the computing speed. Moreover, as the mixed logit, the LPM model uses the entire sample. The LPM model has the disadvantage that the predicted values may be less than zero or greater than one. However, we are less interested in this predicated probability than in the estimated effect of independent variables. This estimated effect from LPM is similar to the computed marginal effect obtained from a non linear estimated model.[4] Eventually, the LPM allows to control for country's observable and unobservable characteristics (with a country fixed effect) which capture the quality of institutions, common to all regions belonging to the same country.[5]

2.2 The Trade Equation

While, comparatively to the atheoretical Harris measure, the KMP has the advantage of being rigorously derived from theory, its use is quite challenging. Indeed, its calculation requires estimates of the unknown parameters ϕ_{rj} and $P_j^{\sigma-1}$. The problem is that we generally do not observe trade flows between regions. Inter-regional trade flows are only observed for a few countries. Statistics Canada provides estimates of shipments for each Canadian province to another Canadian province as well as shipments between each province and the rest of the World (imports and exports). Statistics Canada also provide estimates of exports from each province to each state, as well as imports into each province from each state (McCallum, 1995; Helliwell, 1996; Anderson and van Wincoop, 2003). The US Commodity Flow Survey (CFS) provides the total tonnage and value of commodity flows within the United States, within-state and cross-state shipments (Hillberry, 1998; Anderson and van Wincoop, 2003). Finally, the French Ministry of Transports' database on commodity flows includes both inter- and intra-regional flows and is originally available at a very detailed industry level. Its source and construction is similar to the US CFS (Combes et al., 2005).

This kind of detailed data of interregional trade flows is not available for the African continent. Thus, we follow Head and Mayer (2004*a*) by relying instead upon trade between nations to estimate the parameters that determine trade costs. Hence, we reinterpret equation (2) as giving the quantity exported by a representative firm in country I to country J. The aggregate value of country I's exports to country J, X_{IJ}, is given by the quantity exported by a representative variety firm from I multiplied by the price and the number of varieties from I:

$$X_{IJ} = p_{IJ} q_{IJ} n_I = n_I \frac{c_I^{1-\sigma} \phi_{IJ} Y_J}{P_J^{\sigma-1}}, \tag{10}$$

Taking natural logs and grouping variables according to subscripts,

$$\ln X_{IJ} = \ln \left(n_I / c_I^{\sigma-1} \right) + \ln \left(Y_J / P_J^{\sigma-1} \right) + \ln \phi_{IJ}. \tag{11}$$

[4]The computation of these marginal effects from non linear models may be somewhat cumbersome.

[5]The LPM also allows to capture firms and industry specific effects through a range of firms and industry fixed effects.

we estimate the first two terms using exporter and importer fixed effects, EX_I and IM_J. Bilateral market access (ϕ_{IJ}) is modeled with the following function:

$$\phi_{IJ} = d_{IJ}^{-\delta} \exp\left[-\left(\beta_J - \lambda L_{IJ} - \alpha_1 \mathrm{ETH}_{IJ} - \alpha_2 C_{IJ} - \alpha_3 \mathrm{COL}_{IJ}\right) B_{IJ} \epsilon_{IJ}\right] \tag{12}$$

.

We extend the specification of the market access function adopted by Head and Mayer (2004*a*) by accounting for social and cultural links that are relevant to explain trade connections within Africa. Hence, besides L_{ij} which takes the value of one for pairs of countries that share a common official language, and zero otherwise, we extend the expression : ETH_{IJ} takes the value of 1 when a language is spoken by at least 9% of the population in both countries I and J (ethnic language), $C_{ij} = 1$ if countries I and J are contiguous and $\mathrm{COL}_{ij} = 1$ if they had a common colonizer.

Thus, the estimated equation will be

$$\begin{aligned} \ln X_{IJ} &= \mathrm{EX}_I + \mathrm{IM}_J - \delta \ln d_{IJ} - \beta_J B_{IJ} + \lambda L_{IJ} B_{IJ} + \alpha_1 \mathrm{ETH}_{IJ} B_{IJ} \\ &+ \alpha_2 C_{IJ} B_{IJ} + \alpha_3 \mathrm{COL}_{IJ} B_{IJ} + \epsilon_{IJ}. \end{aligned} \tag{13}$$

The estimated parameters on trade costs and importers' fixed effects are then used to construct the market potential variable $\mathrm{MP}_i = \sum_{j=1}^{R} \phi_{ij} Y_j / P_j^{\sigma-1}$. The expressions of inter- and intraregional access are

$$\hat{\phi_{ij}} = d_{ij}^{-\hat{\delta}} \exp\left(-\hat{\beta}_J + \hat{\lambda} L_{IJ} + \hat{\alpha_1} \mathrm{ETH}_{IJ} + \hat{\alpha_2} C_{IJ} + \hat{\alpha_3} \mathrm{COL}_{IJ}\right)$$

when $i \in I$, $j \in J$, and $I \neq J$

$$\hat{\phi_{ij}} = d_{ij}^{-\hat{\delta}}$$

when i and j belong to the same country, and

$$\hat{\phi_{ii}} = d_{ii}^{-\hat{\delta}} = \left(\frac{2}{3}\sqrt{\mathrm{area}_i/\pi}\right)^{-\hat{\delta}}$$

for intraregional trade.

The other component of market potential calculation is regional-level competition-weighted expenditure. We set $Y_J/P_J^{\sigma-1} = \exp\left(\mathrm{IM}_J\right)$ as per equations (12) and (13). We compute $Y_j/P_j^{\sigma-1}$ for each region j of country J by allocating $Y_J/P_J^{\sigma-1}$ to the different regions in proportion to their share of national GDP, i.e. $Y_j/P_j^{\sigma-1} = (\text{GDP}_j/\text{GDP}_J)\exp(\text{IM}_J)$. Finally, we allocate national expenditure to regions according to GDP shares of regions $[Y_j = (\text{GDP}_j/\text{GDP}_J)\,Y_J]$. National expenditure is calculated using apparent consumption, in the considered industry.

2.3 Estimation strategy and data

As previously mentioned, we estimate equation (8) by using conditional logit, mixed logit models and LPMs. With these models we try to explain the location choices of 1,385 French-owned subsidiaries that were installed in 96 regions pertaining to 41 African countries in 2006.[6] The regions considered constitute the first level of administrative divisions of the different African countries except for Burundi, Comores, Cape Verde, Djibouti, the Gambia, Guinea Bissau, Equatorial Guinea, Liberia, Lesotho, Mauritius, Rwanda, Sao Tome and Principe, Swaziland, Seychelles which are small enough to be considered as a single region. As indicated by Table French affiliates are mostly concentrated in the Maghreb (Tunisia, Morocco, Algeria), which hosts about 67% of the total of affiliates located in Africa.

Data on French manufacturing firms are extracted from the 2006 Survey of French Affiliates ("Enquêtes Filiales") organized by French National Treasury. This survey covers two categories of entities: branches and representative offices of french firms (whose headquarters are located in France), and firms affiliated to French multinationals who own more than 10% of their capital. This survey lists more than 4,232 establishments with details on the date of creation, the postal and the physical address, the NAF ("Nomenclature d'activités française") code of the affiliate activity. With these data on French manufacturing firms, we construct for each region the following proxies to try to capture agglomeration effects: the count of french firms from the same industry, the count of French firms having the same parent company, the total count of French firms located in the same region.

We compute the market potential for 8 two digits industrial sectors in the ISIC classification Revision 2 by using the estimates obtained from the trade equation (13). To avoid any simultaneity issue between the contemporaneous location of French firms and the current forward linkages, we construct lagged market potential for 2005. To estimate equation (13) we use bilateral trade data and production data of the Tradeprod database constructed by de Sousa et al. (2012), and available on CEPII's website.[7] The database contains information on bilateral trade flows and industrial production for 26 industrial sectors in the ISIC (International Standard Industrial

[6] There is actually a total of 506 regions corresponding to 55 countries, but 410 of them do not experience the installation of any French affiliate. Tables 11 and 15 in Appendix B respectively provide the list of African 'Regions' and the list of African 'Regions' hosting a French affiliate.

[7] All data can be obtained at `http://www.cepii.fr/CEPII/en/bdd_modele/download.asp?id=5`

Classification) classification Revision 2 for the 1980-2006 period.

Tradeprod's information on trade flows is based on bilateral trade data and methods from BACI, the international trade database of Gaulier and Zignago (2010). BACI builds on the United Nations' COMTRADE data, and benefits from mirror trade flows (reports for both exporting and importing countries) in order to improve the coverage and accuracy of trade data at the most disaggregated international product-level, the 6-digit Harmonized System (HS6) classification. Tradeprod uses the UNIDO (United Nations Industrial Development Organization) database as its main source of manufacturing production data. UNIDO data sets provide worldwide information for the industrial production at the three and four digits levels. Tradeprod also uses STAN production data to fill some missing data. Extensive details on the Tradeprod database can be found in de Sousa et al. (2012).

To estimate trade equations we also rely on bilateral information on the prevalence of common languages,[8] contiguity and distances. These variables are obtained from CEPII's GeoDist database. A noteworthy contribution of the GeoDist database is that it provides internal and international bilateral distances computed in a totally consistent way (Mayer and Zignago, 2011). Indeed, Mayer and Zignago (2011) have computed the weighted distances using city-level data to assess the geographic distribution of population inside each nation.[9]

As for the market potential variable, we use 1-year lags for all the covariates to avoid any simultaneity issues. This is specially true for the spillover variables. Indeed, using contemporaneous information for them would imply explaining one particular investment by a variable that has been constructed using information on that investment (Delbecque et al., 2014). It is challenging to obtain data on GDP at a regional level. Only Nigeria and South Africa report such statistics. Therefore, we use regional GDP estimated with nighttime lights data archived by the National Geophysical Data Center (NGDC) (Ghosh et al., 2010; Henderson et al., 2012). The remaining variables are defined at the country level. To proxy real wage, we use data on minimum wage from the ILO (International Labour Organization) that we transform in real terms thanks to the PPP (Purchasing Power Parity) exchange rate from the Penn World Table. Data on the Corruption Perception Index are obtained from Transparency International, and those for the Property Rights are from the Heritage Foundation. Finally, we obtained data on the corporate tax rate from the World Development Indicators of the World Bank.

[8]One of the file of the GeoDist database, the dist_cepii dataset contains 2 variables indicating whether two countries, origin and destination, share a common official language, or a common ethnic language, i.e. a language that is spoken by at least 9% of the population in both countries.

[9]Details on the weighted distances formulas are given in Mayer and Zignago (2011, p. 11).

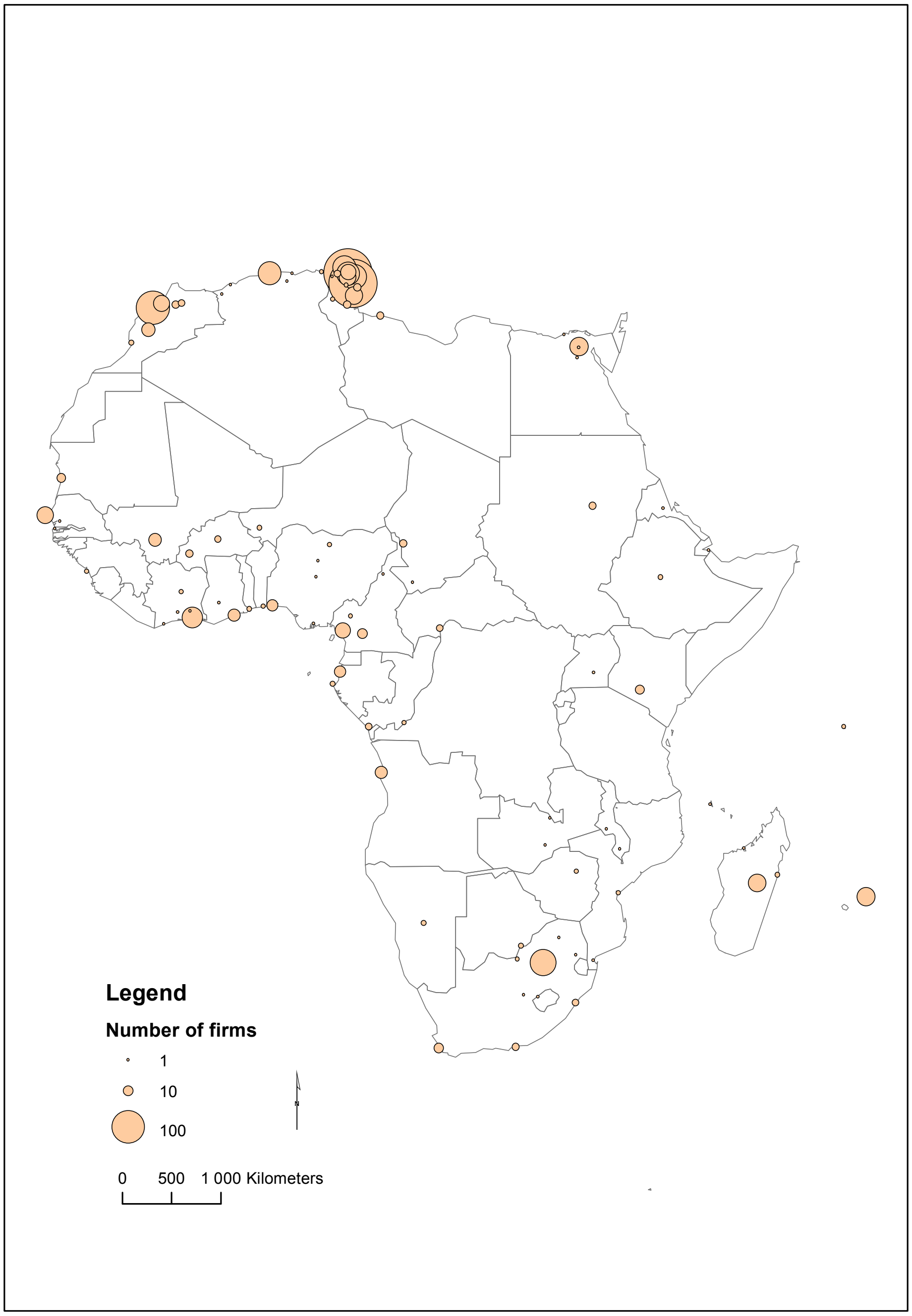

Figure 1: Distribution of French Affiliates in Africa in 2006 (Proportional Symbols)

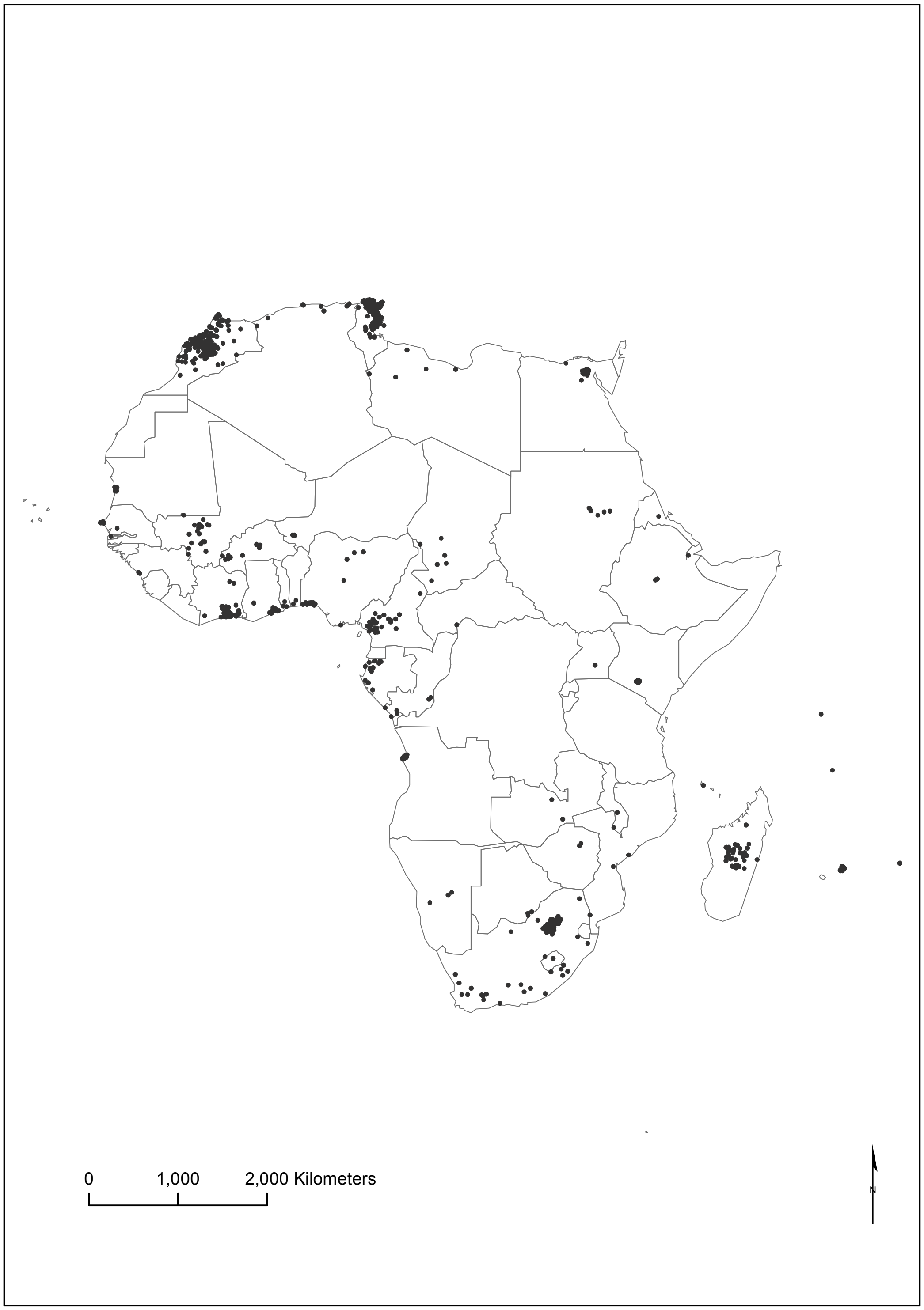

Figure 2: Distribution of French Affiliates in Africa in 2006 (Dot Density)

3 Results

3.1 Trade equations

Table 1: Estimates of border and distance effects: 1980-2006 averages.

Industry name	ISIC code	Border $(-\hat{\beta}_J)$	Distance $(-\hat{\delta})$	Official Language $(\hat{\lambda})$	Ethnic Language $(\hat{\alpha}_1)$	Contiguity $(\hat{\alpha}_2)$	Same Colonizer $(\hat{\alpha}_3)$
Manufacture of Food, Beverages and Tobacco	31	-2.0080	-0.7045	-0.0195	0.2205	1.1385	0.4897
Textile, Wearing Apparel and Leather Industries	32	-1.8882	-0.6306	0.1043	0.2865	0.7256	0.2832
Manufacture of Wood and Wood Products, Including Furniture	33	-2.2752	-0.3289	0.0444	0.1434	0.5776	0.1430
Manufacture of Paper and Paper Products, Printing and Publishing	34	-1.9932	-0.4405	0.1641	0.1988	0.5652	0.2436
Manufacture of Chemicals and Chemical Products Petroleum, Coal, Rubber and Plastic Products	35	-1.3079	-0.7916	0.0782	0.3340	1.0209	0.3577
Manufacture of Non-Metallic Mineral Products, except Products of Petroleum and Coal	36	-2.0122	-0.4115	0.0253	0.1083	0.8690	0.2447
Basic Metal Industries	37	-1.0267	-0.4733	0.0424	0.3063	0.8043	0.1963
Manufacture of Fabricated Metal Products, Machinery and Equipment	38	-1.2594	-0.7051	0.1340	0.2399	1.0204	0.4947

We start by estimating trade equations and then evaluate the impact of market potential of location choices. Table 1 summarizes the border, distance, language, contiguity and common colonizer effects estimated for each two-digit industry between 1980 and 2006. Border effects average 1.7214 for all intra-African trade. Expressed as the ratio of cross-border to within-border trade, this number is equivalent to 5.6 which is surprisingly close to the figure obtained by Head and Mayer (2004*a*) for Europe. Because of the higher transaction costs due to poor infrastructure and landlockedness, one might have indeed expected higher border effects for intra-African trade. Distance effects average -0.56; which has the expected sign but is quite low if one refers to the -0.9 average obtained by Disdier and Head (2008) in their meta-analysis of the effect of distance in bilateral trade. Among the two language effects, the official language effect is the weakest; it is even negative for the Food, Beverage and Tobacco industry. The ethnic language effect is stronger and has always the expected sign. The average of the ethnic language effect is 0.2297 which implies that countries sharing an ethnic language would trade 1.26 more. The contiguity and the common colonizer effects have the expected sign and suggest that contiguous countries trade 2.32 more while countries which had the same colonizer trade 1.36 more.

3.2 Location choice results

3.2.1 Conditional logit results

In line with most of the literature, we first estimate CLMs of location choice. CLMs are appropriate when the choice among alternatives is modeled as a function of the characteristics of the alternative. In addition, it is often used, as in our case with 506 regions, when the number of possible choices is large. It however relies on two restrictive assumptions: (1) IIA and (2) that firm preferences depend on firm's observable characteristics. We then turn to the estimation of a set of mixed logit models. The mixed logit model relaxes the IIA assumption and extends the CLM by allowing some of the parameters in the model to be randomly distributed across firms. Finally, we present results from LPMs. The LPM also relaxes the IIA assumption and has the advantage to provide an immediate indication of the marginal effects of the explanatory variables on the location choice .

Table 2 provides results for nine different conditional logit estimations of the location choice of French affiliates. In the first three columns we consider only demand variables and real wage as regressors. The real wage coefficient is positive: this seems to contradict the expectation that multinationals might seek low-wage regions. Yet, such a result is not uncommon in the literature. It may suggest that the wage also picks up the quality/education level of the labor force (Debaere et al., 2010). From columns (1) to (3) we use successively different measures of demand: regional GDP, Harris and Krugman Market Potentials. The explanatory power of the model with local demand (regional GDP) is the highest, while it seems weaker for the theoretically derived market potential (KMP). At first sight, this might indicate that "theory does not pay". In columns (4) to (9) we use unemployment rate, property rights and corporate tax rate and two count variables, capturing the spillover effects arising from the proximity to French affiliates, as additional regressors. From columns (4) to (6), we consider the count of french firms from the same industry and the count of French firms having the same parent company as spillover variables while in columns (7) to (9), we use the count of french firms from the same industry and the total count of French firms located in the same region as proxies of agglomeration effects. Results for the unemployment rate seem barely significant: positive but insignificant with the regional GDP, positive and significant at 10% significance for Harris and Krugman MP. Property rights coefficients are positive as expected and strongly significant. Results for the corporate tax rate are always insignificant, yet they have a positive sign when we consider regional GDP as the demand variable and a negative sign when we use Harris or Krugman MP. Therefore, even though the model with KMP has the lowest explanatory power, it has the advantage of providing a negative sign for the coefficient of the corporate tax rate that is consistent with intuition. As the reader will notice in the rest of the paper, this pattern will even be confirmed with different estimation strategies.

The coefficients of the count variables capturing the spillover effects are positive and significant in all the different specifications; which seem to provide a robust evidence of agglomeration effects. However, they have an adverse effect on other variables of interest: Harris and Krugman market potentials are no longer significant; the regional GDP coefficient is the only one that remains marginally significant. This can be explained by a correlation between these spillover effects and the market potential variables.[10]

[10] Table 9 in Appendix A reports the pairwise correlation between all the variables. It shows that while the correlation coefficients between the spillover variables and the market potential variables are moderate, they are all strongly significant.

Table 2: Conditional logit model for firms' regional location choice

	Dependent variable: choice of the region								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln regional area	-0.61794***	-0.58375***	-0.50095***	0.01494	0.02795	0.03619	0.02129	0.03539	0.04404*
	(0.01572)	(0.01453)	(0.01439)	(0.02367)	(0.02358)	(0.02544)	(0.02379)	(0.02373)	(0.02567)
ln regional GDP	0.88196***			0.09733**			0.08185*		
	(0.02535)			(0.04395)			(0.04437)		
ln Harris MP		0.29498***			0.00143			-0.00726	
		(0.01028)			(0.02489)			(0.02519)	
ln Krugman MP			0.10304***			-0.01047			-0.01402
			(0.00610)			(0.01538)			(0.01551)
ln Property right				0.77787***	0.66562***	0.59523***	0.71290***	0.58535***	0.52603**
				(0.19512)	(0.20769)	(0.20698)	(0.19541)	(0.20865)	(0.20808)
ln Corporate tax				0.31775	-0.01699	-0.14733	0.32169	-0.03088	-0.12099
				(0.22961)	(0.31254)	(0.27080)	(0.23136)	(0.31386)	(0.27280)
ln Unemployment				0.07642	0.17730*	0.16711*	0.08601	0.17925*	0.15663*
				(0.10247)	(0.09413)	(0.09283)	(0.10250)	(0.09435)	(0.09277)
ln Real wage	1.31399***	1.66505***	1.95803***	0.20696***	0.19279**	0.21708***	0.18657**	0.18068**	0.20394**
	(0.05134)	(0.06041)	(0.06464)	(0.07453)	(0.07957)	(0.08403)	(0.07490)	(0.07974)	(0.08383)
ln (1+French ind)				0.95700***	0.94932***	0.94377***	0.95118***	0.94222***	0.93658***
				(0.04393)	(0.04395)	(0.04429)	(0.04387)	(0.04392)	(0.04425)
ln (1+network)				0.45661***	0.49039***	0.50242***			
				(0.03783)	(0.03706)	(0.03720)			
ln (1+French count)							0.46992***	0.50499***	0.51529***
							(0.03831)	(0.03761)	(0.03760)
Observations	513,162	498,394	498,394	330,270	318,198	318,198	330,270	318,198	318,198
N regions	81	79	79	65	63	63	65	63	63
Pseudo R^2	0.228	0.202	0.172	0.4981	0.4978	0.4978	0.4985	0.4983	0.4983

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3.2.2 Mixed logit results

In Table 3, we consider a mixed logit model which has the advantage to relax the assumption of Independence of Irrelevant Alternatives (IIA) implied by the conditional logit model. The results arising from this model are not qualitatively different from Table 2. The first three columns of Table 3 seem to suggest once more that "theory does not pay". The coefficient of the measure of demand indeed decreases when we move from column (1) to column (3). It is higher when we use the purely local measure of demand; it reaches its lowest value when we use the measure derived from theory. Hence, once more atheoretical measures seem to convey a higher explanatory power. The last six columns of Table 3 yield results that are quite similar to Table 2. Once more, the incorporation of the spillover variables reduces the impact of the demand measures which lose their significance. With the two measures of market potential we get insignificant but negative coefficients for the corporate tax rate. Finally, for these columns we notice that the sign of the coefficient of the logarithm of the land area changes to positive. However, this coefficient is only significant for the last column when we use the KMP as the proxy for forward linkages. This is the expected result as we may presume that larger regions host more firms. We will discuss the intuition involved more in-depth with the results of the LPM model with countries FE.

Table 3: Mixed logit model for firms' regional location choice.

	Dependent variable: choice of the region								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln regional area	-0.62537***	-0.64488***	-0.58603***	0.01494	0.02784	0.03550	0.02129	0.03542	0.04330*
	(0.01593)	(0.01611)	(0.01620)	(0.02367)	(0.02366)	(0.02555)	(0.02379)	(0.02382)	(0.02577)
ln regional GDP	0.81920***			0.09721**			0.08172*		
	(0.02736)			(0.04419)			(0.04464)		
ln Harris MP		0.28615***			-0.00618			-0.01496	
		(0.01075)			(0.02663)			(0.02697)	
ln Krugman MP			0.12640***			-0.01456			-0.01817
			(0.00680)			(0.01808)			(0.01826)
ln Property right				0.77798***	0.65560***	0.59357***	0.71301***	0.57445***	0.52207**
				(0.19525)	(0.20843)	(0.21326)	(0.19557)	(0.20935)	(0.21443)
ln Corporate tax				0.31729	-0.10199	-0.20132	0.32120	-0.11584	-0.17196
				(0.23036)	(0.32763)	(0.28624)	(0.23216)	(0.32905)	(0.28856)
ln Unemployment				0.07621	0.15019	0.12883	0.08580	0.15247	0.12270
				(0.10275)	(0.10096)	(0.10342)	(0.10279)	(0.10108)	(0.10346)
ln Real wage	2.18325***	3.42865***	4.50196***	0.20783**	0.30245*	0.34165**	0.18749**	0.28977*	0.31879*
	(0.17000)	(0.21404)	(0.27424)	(0.08142)	(0.16361)	(0.16294)	(0.08228)	(0.16361)	(0.16278)
ln (1+French ind)				0.95699***	0.94610***	0.94030***	0.95117***	0.93896***	0.93355***
				(0.04394)	(0.04439)	(0.04487)	(0.04388)	(0.04436)	(0.04478)
ln (1+network)				0.45662***	0.49202***	0.50194***			
				(0.03784)	(0.03729)	(0.03778)			
ln (1+French count)							0.46994***	0.50675***	0.51475***
							(0.03832)	(0.03786)	(0.03818)
N observations	513,162	498,394	498,394	330,270	318,198	318,198	330,270	318,198	318,198

Standard errors in parentheses
*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

3.2.3 Linear probability models' results

We now turn to the LPM estimations. Table 4 presents LPM estimation results for specifications without countries fixed effects. However, it systematically includes firm fixed effects. Columns (1) to (3) show the results with only a small set of regressors: logarithms of the regional area, the demand measure, and the real wage. These results are somewhat in line with previous findings: among all the measures of demand, the KMP has the lowest impact on the probability of localization of French subsidiaries. Moreover, the specification with the 'atheoretical' Harris measure has a better fit than the one with KMP.

More generally the marginal effects are more precisely estimated in the LPM: a 10% increase in the regional GDP implies an increase in the location probability of 1.63%. The marginal effects of the Market Potential variables are even lower: for the Harris (respectively Krugman) Market Potential a 10% augmentation entails an increase in the probability of location of 1.09% (respectively 0.24%). This confirms the previous results where the KMP had the lowest impact of the probability of location. Moreover, once more the results confirm that these demand linkages are much lower than those prevailing in the European Union.

Including logarithms of the property rights index, the unemployment rate, the corporate tax rate and of the count of French affiliates from the same industry, as in columns (4) to (6), induces a boost in the explanatory power of the model as indicated by the increased R^2 values. Moreover, there are similarities with previous findings: as before the model with the Harris market potential has a better fit than the one with KMP. We get a positive and significant coefficient for the spillover variable, but again this comes at a price: due to multicollinearity, the coefficient of the logarithm of the demand variable becomes negative.

Eventually with columns (5) and (6), we get now negative and significant coefficients for the logarithm of the corporate tax rate for the two non-local measures of Market Potential.

Table 4: Linear probability model for firms' regional location choice without country FE

	Dependent variable: choice of the region					
	(1)	(2)	(3)	(4)	(5)	(6)
ln regional area	-0.00152***	-0.00172***	-0.00170***	-0.00018***	-0.00010*	-0.00012**
	(0.00004)	(0.00005)	(0.00005)	(0.00006)	(0.00006)	(0.00006)
ln regional GDP	0.00163***			-0.00101***		
	(0.00005)			(0.00007)		
ln Harris MP		0.00109***			-0.00078***	
		(0.00004)			(0.00005)	
ln Krugman MP			0.00024***			-0.00031***
			(0.00002)			(0.00003)
ln Property right				0.00030	-0.00067	-0.00169***
				(0.00044)	(0.00045)	(0.00045)
ln Corporate tax				0.00068**	-0.00145***	-0.00076**
				(0.00028)	(0.00037)	(0.00037)
ln Unemployment				0.00034**	-0.00021	-0.00046***
				(0.00014)	(0.00014)	(0.00015)
ln Real wage	0.00266***	0.00264***	0.00256***	-0.00049***	0.00003	0.00008
	(0.00008)	(0.00008)	(0.00008)	(0.00012)	(0.00013)	(0.00014)
ln (1+French ind)				0.02407***	0.02422***	0.02389***
				(0.00016)	(0.00016)	(0.00016)
Constant	-0.00761***	-0.00258***	0.00561***	0.00616**	0.01533***	0.01231***
	(0.00074)	(0.00067)	(0.00060)	(0.00245)	(0.00265)	(0.00266)
Observations	535,995	522,145	522,145	365,721	353,651	353,651
R-squared	0.00650	0.00652	0.00508	0.06914	0.06946	0.06918

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5 displays results of the LMP model with country fixed effects. Columns (1) to (3) of Table 5 provide results of the LMP without spillover variables. The inclusion of country fixed effects prevents the identification of the coefficient of the variables defined at the country level. Comparatively to Table 4, the inclusion of country FE induces an increase in the demand and Market Potential variables. The Harris MP has now the highest marginal effect: around 2% increase in probability for a 10% increase. The marginal effect arising from the regional GDP is very close to that figure, while with around 1.4% for a 10% increase, the Krugman MP has the lowest marginal effect. Columns (4) to (6) present results with the aforementioned spillover variable and countries fixed effect. In these three specifications the negative marginal effects of the local demand and of the market potential variables persist. In Table 4 and Table 5, we can notice that the semi-elasticity of the spillover effects range from 2.3 to 2.4%.

However, we find a result in columns (4) and (5) of Table 5 that is noteworthy. After controlling for country fixed and agglomeration effects, we find that the probability for a French affiliate to locate in an African region increases with the size of the region. That suggests that for a similar institutional and macroeconomic context, French firms have the tendency to locate in larger regions. This is consistent with earlier empirical results (Head and Mayer, 2004*a*) and consistent with intuition: we expect larger regions to host more affiliates even in the case where the location of subsidiaries would be purely random (Ellison and Glaeser, 1997). This result is not confirmed in column (6) with the KMP as a measure of demand while with the specification with spillover variables and KMP we have also obtained a positive coefficient (marginally significant) for the logarithm of the regional area through the conditional and mixed logit models (last columns of tables 2 and 3). However, that coefficient is systematically and significantly negative in most of the specifications without spillover variables.

The rationale behind this 'counterintuitive' result might be that, when we do not control for the agglomeration effects and for country fixed effects, that coefficient picks the adverse effect of poor infrastructure and lack of public services. Indeed, we might suspect that conversely to small capital regions that are favored in terms of infrastructure and public spending, large African regions struggle to get the funding needed to have proper public services and good infrastructure.

Table 5: Linear probability model for firms' regional location choice with country FE

	Dependent variable: choice of the region					
	(1)	(2)	(3)	(4)	(5)	(6)
ln regional area	-0.00114***	-0.00099***	-0.00098***	0.00027***	0.00020***	-0.00001
	(0.00004)	(0.00005)	(0.00005)	(0.00004)	(0.00005)	(0.00005)
ln regional GDP	0.00183***			-0.00122***		
	(0.00006)			(0.00007)		
ln Harris MP		0.00207***			-0.00108***	
		(0.00007)			(0.00007)	
ln Krugman MP			0.00137***			-0.00132***
			(0.00007)			(0.00007)
ln (1+French ind)				0.02339***	0.02348***	0.02350***
				(0.00013)	(0.00014)	(0.00014)
Constant	-0.00045	-0.00689***	0.00057	0.00566***	0.00690***	0.00999***
	(0.00208)	(0.00227)	(0.00228)	(0.00205)	(0.00223)	(0.00224)
Observations	700,810	655,105	655,105	610,742	570,911	570,911
R-squared	0.00267	0.00305	0.00224	0.05118	0.05155	0.05174
Number of countries	53	48	48	53	48	48
Country FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

3.2.4 A further attempt to address endogeneity concerns

As discussed in subsection 2.3 to avoid any simultaneity issue between the contemporaneous location of French firms and the current demand linkages, we construct lagged market potential. In this subsection we go a step further to address these endogeneity concerns by excluding the actual region of a firm from the calculation of the MP variables. This strategy also allows us to address another concern: the Tunisian law does not allow wholesale or retail marketing by foreign businesses (USA Department of Commerce, 2014). As FDI located in Tunisia represent more than half of the FDI of the whole sample, it is crucial to account for this specificity of FDI in that country. Tables 6, 7 and 8 provide results for respectively the conditional logit model, and the LPM without and with country FE. The results are qualitatively similar with the previous ones. From Table 8, we notice that the marginal effect of a 10% increase in the HMP (respectively KMP) is an increase of 2.1% (respectively 1.7%) of the probability that a firm locates in a given region. Finally, when we control for both the country FE and the agglomeration effects we get a positive coefficient for the logarithm of the regional area, which supports the intuition that in a similar institutional and macroeconomic framework, larger regions are likely to host more affiliates.

Table 6: Conditional logit model for firms' regional location choice: local region excluded

	Dependent variable: choice of the region					
	(1)	(2)	(3)	(4)	(5)	(6)
ln regional area	-0.60617***	-0.54037***	0.02878	0.03553	0.03680	0.04388*
	(0.01489)	(0.01467)	(0.02398)	(0.02613)	(0.02415)	(0.02638)
ln HMP non local	0.30294***		-0.00139		-0.01037	
	(0.01038)		(0.02491)		(0.02522)	
ln KMP non local		0.20471***		-0.01384		-0.01996
		(0.00897)		(0.02507)		(0.02537)
ln Property right			0.65494***	0.59554***	0.57511***	0.51934**
			(0.20537)	(0.21715)	(0.20623)	(0.21882)
ln Corporate tax			-0.04288	-0.14113	-0.05851	-0.12386
			(0.31030)	(0.28993)	(0.31159)	(0.29206)
ln Unemployment			0.18008*	0.17283*	0.18218*	0.16366*
			(0.09399)	(0.09174)	(0.09421)	(0.09168)
ln Real wage	1.65595***	1.98731***	0.19548**	0.21466**	0.18366**	0.20281**
	(0.06044)	(0.06421)	(0.07990)	(0.08546)	(0.08009)	(0.08530)
ln (1+French ind)			0.94879***	0.94507***	0.94165***	0.93794***
			(0.04391)	(0.04422)	(0.04388)	(0.04417)
ln (1+network)			0.49221***	0.50191***		
			(0.03706)	(0.03857)		
ln (1+French count)					0.50706***	0.51566***
					(0.03762)	(0.03905)
Observations	498,394	498,394	318,198	318,198	318,198	318,198

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: Linear probability model for firms' regional location choice without country FE: local region excluded

	Dependent variable: choice of the region			
	(1)	(2)	(3)	(4)
ln regional area	-0.00177***	-0.00173***	-0.00007	-0.00009
	(0.00005)	(0.00005)	(0.00006)	(0.00006)
ln HMP non local	0.00109***		-0.0008***	
	(0.00004)		(0.00005)	
ln KMP non local		0.00050***		-0.00053***
		(0.00003)		(0.00005)
ln Real wage	0.00264***	0.00250***	0.00003	0.00018
	(0.00008)	(0.00008)	(0.00013)	(0.00014)
ln Property right			-0.00064	-0.00179***
			(0.00045)	(0.00045)
ln Corporate tax			-0.00141***	-0.00123***
			(0.00037)	(0.00038)
ln Unemployment			-0.00022	-0.00051***
			(0.00014)	(0.00015)
ln (1+French ind)			0.02422***	0.02400***
			(0.00016)	(0.00016)
Constant	-0.00214***	0.00435***	0.01466***	0.01535***
	(0.00066)	(0.00061)	(0.00264)	(0.00272)
Observations	522,145	522,145	353,651	353,651
R-squared	0.00642	0.00530	0.06910	0.06893
Number of firms	1,385	1,385	1,207	1,207
Country FE	NO	NO	NO	NO
Firm FE	Yes	Yes	Yes	Yes

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Linear probability model for firms' regional location choice with country FE: local region excluded

	Dependent variable: choice of the region			
	(1)	(2)	(3)	(4)
ln regional area	-0.00110***	-0.00118***	0.0003***	0.00025***
	(0.00005)	(0.00004)	(0.0000)	(0.00005)
ln HMP non local	0.00206***		-0.0010***	
	(0.00007)		(0.0001)	
ln KMP non local		0.00166***		-0.00118***
		(0.00007)		(0.00007)
ln (1+French ind)			0.0235***	0.02348***
			(0.0001)	(0.00014)
Constant	-0.00561**	0.00131	0.0059***	0.00551**
	(0.00225)	(0.00222)	(0.0022)	(0.00218)
Observations	655,105	655,105	570,911	570,911
R-squared	0.00306	0.00255	0.0515	0.05164
Country FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Number of countries	48	48	48	48

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4 Conclusion

In their review of the literature on firms location choices, Combes and Gobillon (2015) assert that this literature can be considered as mostly descriptive. They also emphasize that a safer avenue for assessing the role of agglomeration effects on firm location choices would probably be to consider more structural approaches. This statement and this recommendation are particularly relevant for the literature on firm location choices throughout Africa, which has been so far dominated by contributions based on ad-hoc econometric specifications. In this contribution, we follow Combes and Gobillon (2015) recommendation by adopting a structural approach based on the methodology proposed by Head and Mayer (2004*a*).

Our results suggest that demand matters for the choice of location. However, the marginal effects of the demand and Market Potential strongly suggest demand linkages in Africa are weaker than the effects observed in the European Union. Furthermore, the measure of demand inspired by the theory seems to underperform comparatively to the ad-hoc Harris measure of Market potential and even with comparison to local GDP. Yet, estimations with the Krugman MP are noteworthy as they yield a nice result: the coefficient of the corporate tax rate has a negative sign which is consistent with intuition. The effects of the spillover variables are strong, positive and significant. This tends to support the idea that agglomeration effects are a strong driver for the location of French firms throughout African regions. Nevertheless, with the inclusion of these agglomeration effects, the effects of the demand and market potential variables seem to wane, possibly because of multicollinearity.

This multicollinearity problem is possibly caused by the way we modeled the spillover variables. Hence, these spillover effects might capture omitted exogenous location attributes. Improving the modeling of agglomeration effects would be the way forward; yet this strategy would face a daunting challenge: the dearth of reliable data at the micro-geographical level. Moreover, as pointed out by Head and Mayer (2004*a*) the spillover variables results might suggest that the forward linkages outlined by Krugman (1991) are not the only drivers of agglomeration. Including both forward and backward linkages as in Amiti and Javorcik (2008) would be an interesting way to reduce the impact of omitted exogenous location characteristic. This would imply reducing the scale of the analysis from a continent to a country level, as we would need for that to rely on a input-output matrix: South-Africa would be a natural candidate in this regard as this is one of the African countries with the most reliable regional data.

Eventually, estimations from the linear probability models outline an interesting result: if we control for country fixed and agglomeration effects, the probability that a french subsidiary locates in an African region increases with the geographical size of the region. This may indicate that for a similar institutional and geographical environment, French affiliates tend to locate in larger regions. This is in line with intuition: we would indeed expect larger regions to host more affiliates.

Acknowledgements. Alain Pholo Bala gratefully acknowledges financial support from Economic Research Southern Africa (ERSA), and for research grants provided by the University Research Committee (URC) of the University of Johannesburg, the Department of Higher Education and Training (DHET) and the National Research Foundation (NRF), South Africa. The views expressed in this paper and any remaining errors are ours.

References

Amiti, M. and Javorcik, B. S. (2008), 'Trade costs and location of foreign firms in China', *Journal of Development Economics* **85**(1-2), 129–149.

Anderson, J. E. and van Wincoop, E. (2003), 'Gravity with Gravitas: A Solution to the Border Puzzle', *American Economic Review* **93**(1), 170–192.

Anyanwu, J. C. (2012), 'Why Does Foreign Direct Investment Go Where It Goes?: New Evidence From African Countries', *Annals of Economics and Finance* **13**(2), 425–462.

Asiedu, E. (2002), 'On the determinants of foreign direct investment to developing countries: Is africa different?', *World Development* **30**(1), 107–119.

Asiedu, E. (2004), 'Policy reform and foreign direct investment in africa: Absolute progress but relative decline', *Development Policy Review* **22**(1), 41–48.

Asiedu, E. (2006), 'Foreign direct investment in africa: The role of natural resources, market size, government policy, institutions and political instability', *The World Economy* **29**(1), 63–77.

Barry, F., Görg, H. and Strobl, E. (2003), 'Foreign direct investment, agglomerations, and demonstration effects: An empirical investigation', *Review of World Economics* **139**(4), 583–600.

Combes, P.-P. and Gobillon, L. (2015), Chapter 5 - the empirics of agglomeration economies, *in* J. V. H. Gilles Duranton and W. C. Strange, eds, 'Handbook of Regional and Urban Economics', Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, pp. 247 – 348.

Combes, P.-P., Lafourcade, M. and Mayer, T. (2005), 'The trade-creating effects of business and social networks: evidence from france', *Journal of International Economics* **66**(1), 1 – 29.

Crozet, M. (2004), 'Do migrants follow market potentials? An estimation of a new economic geography model', *Journal of Economic Geography* **4**(4), 439–458.

de Sousa, J., Mayer, T. and Zignago, S. (2012), 'Market access in global and regional trade', *Regional Science and Urban Economics* **42**(6), 1037–1052.

Debaere, P., Lee, J. and Paik, M. (2010), 'Agglomeration, backward and forward linkages: evidence from South Korean investment in China', *Canadian Journal of Economics* **43**(2), 520–546.

Delbecque, V., Méjean, I. and Patureau, L. (2014), 'Labor market institutions and firms' location choices', *Review of World Economics* **150**(1), 115–148.

Disdier, A.-C. and Head, K. (2008), 'The Puzzling Persistence of the Distance Effect on Bilateral Trade', *The Review of Economics and Statistics* **90**(1), 37–48.

Du, J., Lu, Y. and Tao, Z. (2012), 'Institutions and FDI location choice: The role of cultural distances', *Journal of Asian Economics* **23**(3), 210 – 223. Technology, FDI and Asian Dynamism.

Duranton, G. and Puga, D. (2004), Micro-foundations of urban agglomeration economies, *in* J. V. Henderson and J. F. Thisse, eds, 'Handbook of Regional and Urban Economics', 1 edn, Vol. 4, Elsevier, chapter 48, pp. 2063–2117.

Ellison, G. and Glaeser, E. (1997), 'Geographic concentration in U.S. manufacturing industries: A dartboard approach', *Journal of Political Economy* **105**(5), 889–927.

Fedderke, J. and Romm, A. (2006), 'Growth impact and determinants of foreign direct investment into south africa, 19562003', *Economic Modelling* **23**(5), 738 – 760. Ninth Annual Conference on Econometric Modelling for Africa, School of Economics, University of Cape Town, 2004.

Gaulier, G. and Zignago, S. (2010), BACI: International Trade Database at the Product-Level. The 1994-2007 Version, Working Papers 2010-23, CEPII research center.

Ghosh, T., Powell, R. L., Elvidge, C. D., Baugh, K. E., Sutton, P. C. and Sharolyn, A. (2010), 'Shedding light on the global distribution of economic activity', *The Open Geography Journal* **3**, 147–160.

Hanson, G. (2005), 'Market potential, increasing returns and geographic concentration', *Journal of International Economics* **67**(1), 1–24.

Harris, C. D. (1954), 'The, market as a factor in the localization of industry in the United States', *Annals of the association of American geographers* **44**(4), 315–348.

Hayakawa, K. and Tsubota, K. (2014), 'Location choice in low-income countries: Evidence from Japanese investments in East Asia', *Journal of Asian Economics* **33**(C), 30–43.

Head, K. and Mayer, T. (2004*a*), 'Market Potential and the Location of Japanese Investment in the European Union', *The Review of Economics and Statistics* **86**(4), 959–972.

Head, K. and Mayer, T. (2004*b*), The empirics of agglomeration and trade, *in* J. V. Henderson and J. F. Thisse, eds, 'Handbook of Regional and Urban Economics', Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, chapter 59, pp. 2609–2669.

Helliwell, J. F. (1996), 'Do National Borders Matter for Quebec's Trade?', *Canadian Journal of Economics* **29**(3), 507–22.

Henderson, J. V., Storeygard, A. and Weil, D. (2012), 'Measuring economic growth from outer space', *American Economic Review* **102**(2), 994–1028.

Hillberry, R. (1998), 'Regional trade and the medicine line: The national border effect in u.s. commodity flow data', *Journal of Borderlands Studies* **13**(2), 1–17.

Kandiero, T. and Chitiga, M. (2014), 'Trade openness and foreign direct investment in africa', *South African Journal of Economic and Management Sciences* **9**(3), 355–370.

Krugman, P. (1991), Increasing returns and economic geography, Technical Report 3.

Krugman, P. (1992), A dynamic spatial model, Technical report, National Bureau of Economic Research.

Lederman, D., Mengistae, T. and Xu, L. C. (2013), 'Microeconomic consequences and macroeconomic causes of foreign direct investment in southern african economies', *Applied Economics* **45**(25), 3637–3649.

Mayer, T. and Zignago, S. (2011), Notes on CEPIIs distances measures: The GeoDist database, Working Papers 2011-25, CEPII research center.

McCallum, J. (1995), 'National Borders Matter: Canada-U.S. Regional Trade Patterns', *American Economic Review* **85**(3), 615–23.

Naudé, W. A. and Krugell, W. F. (2007), 'Investigating geography and institutions as determinants of foreign direct investment in africa using panel data', *Applied Economics* **39**(10), 1223–1233.

Ottaviano, G. and Thisse, J.-F. (2004), Agglomeration and economic geography, *in* J. V. Henderson and J. F. Thisse, eds, 'Handbook of Regional and Urban Economics', Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, chapter 58, pp. 2563–2608.

Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of international Economics* **62**(1), 53–82.

Sanfilippo, M. (2010), 'Chinese FDI to Africa: What Is the Nexus with Foreign Economic Cooperation?', *African Development Review* **22**(S1), 599–614.

Tokunaga, S. and Jin, S. (2011), 'Market potential, agglomeration and location of Japanese manufacturers in China', *Letters in Spatial and Resource Sciences* **4**(1), 9–19.

Train, K. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press.

USA Department of Commerce (2014), 'Doing Business in Tunisia: 2014 Country Commercial Guide for U.S. Companies', http://www.iberglobal.com/files/tunez_ccg(1).pdf. Online; accessed 31 July 2018.

Appendix A: Descriptive statistics

Pairwise correlation matrix

Table A1: Pairwise correlation matrix.

	ln reg. area	ln reg. GDP	ln HMP	ln KMP	ln HMP non local	ln KMP non local	ln Prop. right	ln Corp. tax	ln Unemp. rate	ln Real wage	ln (1+Fr. ind)	ln (1+netw.)	ln (1+Fr. count)
ln regional area	1.0000												
ln regional GDP	-0.0958*** (0.000)	1.0000											
ln HMP	0.0737*** (0.000)	0.6242*** (0.000)	1.0000										
ln KMP	0.0845*** (0.000)	0.2091*** (0.000)	0.7780*** (0.000)	1.0000									
ln HMP non local	0.1050*** (0.000)	0.6221*** (0.000)	0.9993*** (0.000)	0.7745*** (0.000)	1.0000								
ln KMP non local	0.1199*** (0.000)	0.2764*** (0.000)	0.8421*** (0.000)	0.9738*** (0.000)	0.8412*** (0.000)	1.0000							
ln Property right	-0.0777*** (0.000)	0.3291*** (0.000)	0.2232*** (0.000)	-0.0533*** (0.000)	0.2222*** (0.000)	-0.0387*** (0.000)	1.0000						
ln Corporate tax	-0.1595*** (0.000)	-0.0437*** (0.000)	-0.4864*** (0.000)	-0.4177*** (0.000)	-0.4886*** (0.000)	-0.4558*** (0.000)	-0.3227*** (0.000)	1.0000					
ln Unemp.	0.0167*** (0.000)	0.1983*** (0.000)	-0.1345*** (0.000)	-0.3433*** (0.000)	-0.1348*** (0.000)	-0.3335*** (0.000)	0.2999*** (0.000)	0.3176*** (0.000)	1.0000				
ln Real wage	-0.0068** (0.004)	0.0110*** (0.000)	0.0234*** (0.000)	0.1032*** (0.000)	0.0247*** (0.000)	0.1009*** (0.000)	-0.1631*** (0.000)	0.4698*** (0.000)	0.3979*** (0.000)	1.0000			
ln (1+French ind)	-0.1811*** (0.000)	0.2788*** (0.000)	0.2435*** (0.000)	0.1202*** (0.000)	0.2382*** (0.000)	0.1511*** (0.000)	0.1851*** (0.000)	0.0153*** (0.000)	0.1247*** (0.000)	0.1916*** (0.000)	1.0000		
ln (1+network)	-0.2028*** (0.000)	0.3961*** (0.000)	0.3642*** (0.000)	0.2223*** (0.000)	0.3561*** (0.000)	0.2533*** (0.000)	0.2011*** (0.000)	-0.0268*** (0.000)	0.0830*** (0.000)	0.1643*** (0.000)	0.7759*** (0.000)	1.0000	
ln (1+French count)	-0.2006*** (0.000)	0.4045*** (0.000)	0.3761*** (0.000)	0.2277*** (0.000)	0.3681*** (0.000)	0.2588*** (0.000)	0.2099*** (0.000)	-0.0308*** (0.000)	0.0912*** (0.000)	0.1719*** (0.000)	0.7786*** (0.000)	0.9987*** (0.000)	1.0000

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Distribution of French manufacturing affiliates in the Top 10 African countries

Country name	Number of firms
Tunisia	705
Morocco	160
South Africa	92
Algeria	58
Ivory Coast	45
Egypt	36
Cameroun	35
Madagascar	35
Mauritius	32
Senegal	29

Appendix B: List of African 'Regions'

Table B1: List of African 'Regions'

Code	Region	Code	Region	Code	Region
Algeria		**Angola**		**Burundi**	
DZA-ADF	Ain Defla	AO.BG	Benguela	BDI	Burundi
DZA-ADR	Adrar	AO.BI	Bié	**Côte d'Ivoire**	
DZA-ALG	Alger	AO.BO	Bengo	CI.AG	Agnéby
DZA-ANN	Annaba	AO.CB	Cabinda	CI.BF	Bafing
DZA-ATM	Ain Temouchent	AO.CC	Cuando Cubango	CI.BS	Bas-Sassandra
DZA-BAT	Batna	AO.CN	Cuanza Norte	CI.DE	Denguélé
DZA-BBA	Bourdj Bou Arrer	AO.CS	Cuanza Sul	CI.DH	Dix-Huit Montagnes
DZA-BCH	Bechar	AO.CU	Cunene	CI.FR	Fromager
DZA-BEJ	Bejaia	AO.HL	Huìla	CI.HT	Haut-Sassandra
DZA-BLI	Blida	AO.HM	Huambo	CI.LC	Lacs
DZA-BMR	Boumerdes	AO.LN	Lunda Norte	CI.LG	Lagunes
DZA-BOU	Bouira	AO.LS	Lunda Sul	CI.MC	Moyen-Comoé
DZA-BSK	Biskra	AO.LU	Luanda	CI.MR	Marahoué
DZA-CHL	Chlef	AO.ML	Malanje	CI.MV	Moyen-Cavally
DZA-CNS	Constantine	AO.MX	Moxico	CI.NC	N'zi-Comoé
DZA-DJL	Djelfa	AO.NA	Namibe	CI.SB	Sud-Bandama
DZA-EBY	El Bayadh	AO.UI	Uíge	CI.SC	Sud-Comoé
DZA-EOD	El Oued	AO.ZA	Zaire	CI.SV	Savanes
DZA-ETR	El Tarf	**Benin**		CI.VB	Vallée du Bandama
DZA-GHD	Ghardaia	BEN-ATK	Atakora	CI.WR	Worodougou
DZA-GUE	Guelma	BEN-ATL	Atlantique	CI.ZA	Zanzan
DZA-ILL	Illizi	BEN-BOR	Borgou	**Cameroon**	
DZA-JIJ	Jijel	BEN-MON	Mono	CMR-ADM	Adamoua
DZA-KHN	Khenchela	BEN-OUE	Oueme	CMR-CNT	Centre
DZA-LGH	Laghouat	BEN-ZOU	Zou	CMR-ENO	Extreme-Nord
DZA-MED	Medea	**Botswana**		CMR-EST	Est
DZA-MIL	Mila	BW.CE	Central	CMR-LTT	Littoral
DZA-MSC	Mascara	BW.GH	Ghanzi	CMR-NOR	Nord
DZA-MSL	M'Sila	BW.KG	Kgalagadi	CMR-NOU	Nord-Ouest
DZA-MST	Mostaganem	BW.KL	Kgatleng	CMR-OUE	Ouest
DZA-NAM	Naama	BW.KW	Kweneng	CMR-SOU	Sud-Ouest
DZA-OEB	Oum el Bouaghi	BW.NE	North-East	CMR-SUD	Sud
DZA-ORA	Oran	BW.NW	North-West	**Cape Verde**	
DZA-ORG	Ouargla	BW.SE	South-East	CPV	Cape Verde
DZA-RLZ	Relizane	BW.SO	Southern	**Central African Republic**	
DZA-SAH	Souk Ahras	**Burkina Faso**		CF.AC	Ouham
DZA-SAI	Saida	BFA.BO	Boucle du Mouhoun	CF.BB	Bamingui-Bangoran
DZA-SBA	Sidi bel Abbes	BFA.CA	Cascades	CF.BG	Bangui
DZA-SET	Setif	BFA.CE	Centre	CF.BK	Basse-Kotto
DZA-SKK	Skikda	BFA.CN	Centre-Nord	CF.HK	Haute-Kotto
DZA-TBS	Tebessa	BFA.CS	Centre-Est	CF.HM	Haut-Mbomou
DZA-TIA	Tiaret	BFA.CU	Centre-Sud	CF.HS	Mambéré-Kadéï
DZA-TLM	Tlemcen	BFA.ES	Est	CF.KB	Nana-Grébizi
DZA-TMN	Tamanghasset	BFA.HA	Hauts-Bassins	CF.KG	Kémo
DZA-TND	Tindouf	BFA.NO	Nord	CF.LB	Lobaye
DZA-TOU	Tizi Ouzou	BFA.OU	Centre-Ouest	CF.MB	Mbomou
DZA-TPZ	Tipaza	BFA.PL	Plateau-Central	CF.MP	Ombella-M'Poko
DZA-TSS	Tissemsilt	BFA.SA	Sahel	CF.NM	Nana-Mambéré
		BFA.SU	Sud-Ouest	CF.OP	Ouham-Pendé

Table B1: List of African ‘Regions’ (continued)

Code	Region	Code	Region	Code	Region
Central African Republic		**Egypt**		**Gabon**	
CF.SE	Sangha-Mbaéré	EG.DQ	Ad Daqahliyah	GA.OM	Ogooué-Maritime
CF.UK	Ouaka	EG.DT	Dumyat	GA.WN	Wouleu-Ntem
CF.VK	Vakaga	EG.FY	Al Fayyum	**Gambia**	
Chad		EG.GH	Al Gharbiyah	GMB	Gambia
TCD-BAT	Batha	EG.IK	Al Iskandariyah	**Ghana**	
TCD-BET	Borkou-Ennedi-Tibesti	EG.IS	Al Isma‘iliyah	GH.AA	Greater Accra
TCD-BLT	Biltine	EG.JS	Janub Sina’	GH.AH	Ashanti
TCD-CBG	Chari-Baguirmi	EG.JZ	Al Jizah	GH.BA	Brong Ahafo
TCD-GUE	Guera	EG.KS	Kafr ash Shaykh	GH.CP	Central
TCD-KAN	Kanem	EG.MF	Al Minufiyah	GH.EP	Eastern
TCD-LAC	Lac	EG.MN	Al Minya	GH.NP	Northern
TCD-LOC	Logone Occidental	EG.MT	Matruh	GH.TV	Volta
TCD-LOR	Logone Oriental	EG.QH	Al Qahirah	GH.UE	Upper East
TCD-MCH	Moyen-Chari	EG.QL	Al Qalyubiyah	GH.UW	Upper West
TCD-MKE	Mayo Kebi	EG.QN	Qina	GH.WP	Western
TCD-OUA	Ouaddai	EG.SJ	Suhaj	**Guinea**	
TCD-SLM	Salamat	EG.SQ	Ash Sharqiyah	GIN.BO	Boké
TCD-TND	Tandjile	EG.SS	Shamal Sina’	GIN.CO	Conarky
Comores		EG.SW	As Suways	GIN.FA	Faranah
COM	Comores	EG.WJ	Al Wadi al Jadid	GIN.KA	Kankan
Congo		**Equatorial Guinea**		GIN.KI	Kindia
COG-BOU	Bouenza	GNQ	Equatorial Guinea	GIN.LA	Labé
COG-CVT	Cuvette	**Eritrea**		GIN.MA	Mamou
COG-KOU	Kouilou	ER.AN	Anseba	GIN.NZ	Nzérékoré
COG-LEK	Lekoumou	ER.DK	Debubawi Keyih Bahri	**Guinea Bissau**	
COG-LIK	Likouala	ER.DU	Debub	GNB	Guinea Bissau
COG-NIA	Niari	ER.GB	Gash Barka	**Kenya**	
COG-PLT	Plateaux	ER.MA	Maekel	KE.CE	Central
COG-POO	Pool	ER.SK	Semenawi Keyih Bahri	KE.CO	Coast
COG-SNG	Sangha	**Ethiopa**		KE.EA	Eastern
Congo, DRC		ET.AA	Addis Ababa	KE.NA	Nairobi
ZAR-BAN	Bandundu	ET.AF	Afar	KE.NE	North-Eastern
ZAR-BZA	Bas-Zaire	ET.AM	Amhara	KE.NY	Nyanza
ZAR-EQT	Equateur	ET.BE	Benshangul-Gumaz	KE.RV	Rift Valley
ZAR-HZA	Haut-Zaire	ET.DD	Dire Dawa	KE.WE	Western
ZAR-KIV	Kivu	ET.GA	Gambela Peoples	**Lesotho**	
ZAR-KNS	Kinshasa	ET.HA	Harari People	LSO	Lesotho
ZAR-KOC	Kasai-Occidental	ET.OR	Oromia	**Liberia**	
ZAR-KOR	Kasai-Oriental	ET.SN	Southern Nations, Nationalities and People	LBR	Liberia
ZAR-SHA	Shaba	ET.SO	Somali	**Libya**	
Djibouti		ET.TI	Tigray	LBY.CI	Cirenaica
DJI	Djibouti	**Gabon**		LBY.FE	Fezzan
Egypt		GA.ES	Estuaire	LBY.TR	Tripolitania
EG.AN	Aswan	GA.HO	Haut-Ogooué	**Madagascar**	
EG.AT	Asyut	GA.MO	Moyen-Ogooué	MG.AS	Antsiranana
EG.BA	Al Bahr al Ahmar	GA.NG	Ngounié	MG.AV	Antananarivo
EG.BH	Al Buhayrah	GA.NY	Nyanga	MG.FI	Fianarantsoa
EG.BN	Bani Suwayf	GA.OI	Ogooué-Ivindo	MG.MA	Mahajanga
EG.BS	Bur Sa‘id	GA.OL	Ogooué-Lolo	MG.TL	Toliary

Table B1: List of African 'Regions' (continued)

Code	Region	Code	Region	Code	Region
	Madagascar		**Namibia**		**Nigeria**
MG.TM	Toamasina	NA.CA	Caprivi	NG.OG	Ogun
	Malawi	NA.ER	Erongo	NG.ON	Ondo
MWI-CNT	Central	NA.HA	Hardap	NG.OS	Osun
MWI-NRT	Northern	NA.KA	Karas	NG.OY	Oyo
MWI-STH	Southern	NA.KH	Khomas	NG.PL	Plateau
	Mali	NA.KU	Kunene	NG.RI	Rivers
ML.BA	Bamako	NA.OD	Otjozondjupa	NG.SO	Sokoto
ML.GA	Gao	NA.OH	Omaheke	NG.TA	Taraba
ML.KD	Kidal	NA.OK	Kavango	NG.YO	Yobe
ML.KK	Koulikoro	NA.ON	Oshana	NG.ZA	Zamfara
ML.KY	Kayes	NA.OS	Omusati		**Rwanda**
ML.MO	Mopti	NA.OT	Oshikoto	RWA	Rwanda
ML.SG	Ségou	NA.OW	Ohangwena		**Sao Tome**
ML.SK	Sikasso		**Niger**	STP	Sao Tome
ML.TB	Timbuktu	NE.AG	Agadez		**Senegal**
	Mauritania	NE.DF	Diffa	SEN-DKR	Dakar
MR.AD	Adrar	NE.DS	Dosso	SEN-DRB	Diourbel
MR.AS	Assaba	NE.MA	Maradi	SEN-FTC	Fatick
MR.BR	Brakna	NE.NI	Niamey	SEN-KLC	Kaolack
MR.DN	Dakhlet Nouadhibou	NE.TH	Tahoua	SEN-KLD	Kolda
MR.GD	Guidimaka	NE.TL	Tillabéry	SEN-LOU	Louga
MR.GO	Gorgol	NE.ZI	Zinder	SEN-STL	Saint-Louis
MR.HC	Hodh ech Chargui		**Nigeria**	SEN-THI	Thies
MR.HG	Hodh el Gharbi	NG.AB	Abia	SEN-TMB	Tambacounda
MR.IN	Inchiri	NG.AD	Adamawa	SEN-ZGN	Ziguinchor
MR.NO	Nouakchott	NG.AK	Akwa Ibom		**Seychelles**
MR.TG	Tagant	NG.AN	Anambra	SYC	Seychelles
MR.TR	Trarza	NG.BA	Bauchi		**Sierra Leone**
MR.TZ	Tiris Zemmour	NG.BE	Benue	SL.EA	Eastern
	Mauritius	NG.BO	Borno	SL.NO	Northern
MUS	Mauritius	NG.BY	Bayelsa	SL.SO	Southern
	Morocco	NG.CR	Cross River	SL.WE	Western
MAR-CEN	Centre	NG.DE	Delta		**Somalia**
MAR-CNR	Centre-North	NG.EB	Ebonyi	SO.AW	Awdal
MAR-CSO	Centre-South	NG.ED	Edo	SO.BK	Bakool
MAR-EST	Eastern	NG.EK	Ekiti	SO.BN	Banaadir
MAR-NWT	North-West	NG.EN	Enugu	SO.BR	Bari
MAR-SOU	South	NG.FC	Federal Capital Territory	SO.BY	Bay
MAR-TNF	Tensift	NG.GO	Gombe	SO.GA	Galguduud
	Mozambique	NG.IM	Imo	SO.GE	Gedo
MZ.CD	Cabo Delgado	NG.JI	Jigawa	SO.HI	Hiiraan
MZ.GA	Gaza	NG.KD	Kaduna	SO.JD	Jubbada Dhexe
MZ.IN	Inhambane	NG.KE	Kebbi	SO.JH	Jubbada Hoose
MZ.MN	Manica	NG.KN	Kano	SO.MU	Mudug
MZ.MP	Maputo	NG.KO	Kogi	SO.NU	Nugaal
MZ.NM	Nampula	NG.KT	Katsina	SO.SA	Sanaag
MZ.NS	Nassa	NG.KW	Kwara	SO.SD	Shabeellaha Dhexe
MZ.SO	Sofala	NG.LA	Lagos	SO.SH	Shabeellaha Hoose
MZ.TE	Tete	NG.NA	Nassarawa	SO.SO	Sool
MZ.ZA	Zambezia	NG.NI	Niger	SO.TO	Togdheer

Table B1: List of African 'Regions' (continued)

Code	Region	Code	Region	Code	Region
	Somalia		**Togo**		**Zambia**
SO.WO	Woqooyi Galbeed	TG.CE	Centre	ZMB-STH	Southern
	South Africa	TG.KA	Kara	ZMB-WST	Western
ZAF-ECP	Eastern Cape	TG.MA	Maritime		**Zimbabwe**
ZAF-GAT	Gauteng	TG.PL	Plateaux	ZW.BU	Bulawayo
ZAF-KNT	KwaZulu/Natal	TG.SA	Savanes	ZW.HA	Harare
ZAF-MPM	Mpumalanga		**Tunisia**	ZW.MA	Manicaland
ZAF-NCP	Northern Cape	TN.AN	Ariana	ZW.MC	Mashonaland Central
ZAF-NRT	Northern	TN.BA	Ben Arous (Tunis Sud)	ZW.ME	Mashonaland East
ZAF-NWS	North West	TN.BJ	Béja	ZW.MI	Midlands
ZAF-OFS	Free State	TN.BZ	Bizerte	ZW.MN	Matabeleland North
ZAF-WCP	Western Cape	TN.GB	Gabès	ZW.MS	Matabeleland South
	Sudan	TN.GF	Gafsa	ZW.MV	Masvingo
SDN.BA	Bahr el Ghazal	TN.JE	Jendouba	ZW.MW	Mashonaland West
SDN.BL	Blue Nile	TN.KB	Kebili		
SDN.DA	Darfur	TN.KF	Le Kef		
SDN.EQ	Equatoria	TN.KR	Kairouan		
SDN.KA	Kassala	TN.KS	Kassérine		
SDN.KH	Khartoum	TN.ME	Médenine		
SDN.KO	Kordofan	TN.MH	Mahdia		
SDN.NO	Northern	TN.MN	Manubah		
SDN.UP	Upper Nile	TN.MS	Monastir		
	Swaziland	TN.NB	Nabeul		
SWZ	Swaziland	TN.SF	Sfax		
	Tanzania	TN.SL	Siliana		
TZA-ARS	Arusha	TN.SS	Sousse		
TZA-DES	Dar es Salaam	TN.SZ	Sidi Bou Zid		
TZA-DOD	Dodoma	TN.TA	Tataouine		
TZA-IRN	Iringa	TN.TO	Tozeur		
TZA-KIG	Kigoma	TN.TU	Tunis		
TZA-KLM	Kilimanjaro	TN.ZA	Zaghouan		
TZA-LIN	Lindi		**Uganda**		
TZA-MAR	Mara	UGA-BUS	Busoga		
TZA-MBE	Mbeya	UGA-CNT	Central		
TZA-MRG	Morogoro	UGA-EST	Eastern		
TZA-MTW	Mtwara	UGA-KRM	Karamoja		
TZA-MWN	Mwanza	UGA-NIL	Nile		
TZA-PNR	Pemba North	UGA-NRB	North Buganda		
TZA-PSO	Pemba South	UGA-NRT	Northern		
TZA-PWA	Pwani	UGA-STB	South Buganda		
TZA-RUK	Rukwa	UGA-STH	Southern		
TZA-RUV	Ruvuma	UGA-WST	Western		
TZA-SHN	Shinyanga		**Zambia**		
TZA-SNG	Singida	ZMB-CNT	Central		
TZA-TAB	Tabora	ZMB-CPP	Copperbelt		
TZA-TAN	Tanga	ZMB-EST	Eastern		
TZA-ZCN	Zanzibar Central/South	ZMB-LUA	Luapula		
TZA-ZMG	Ziwa Magharibi	ZMB-LUS	Lusaka		
TZA-ZNR	Zanzibar North	ZMB-NRT	Northern		
TZA-ZUR	Zanzibar Urban/West	ZMB-NWS	North-Western		

Table B2: List of African 'Regions' hosting French firms

Region	Country	Cnt firms	Region	Country	Cnt firms
Luanda	Angola	15	Toamasina	Madagascar	3
Atlantique	Benin	2	Koulikoro	Mali	17
Centre	Burkina Faso	4	Nouakchott	Mauritania	8
Hauts-Bassins	Burkina Faso	6	Mauritius		28
South-East	Botswana	3	Central	Malawi	1
Bangui	Central African Republic	4	Southern	Malawi	1
Agnéby	Côte d'Ivoire	1	Maputo	Mozambique	1
Bas-Sassandra	Côte d'Ivoire	1	Sofala	Mozambique	2
Lagunes	Côte d'Ivoire	40	Khomas	Namibia	3
Sud-Bandama	Côte d'Ivoire	1	Niamey	Niger	3
Valle du Bandama	Côte d'Ivoire	2	Federal Capital Territory	Nigeria	1
Centre	Cameroon	9	Kaduna	Nigeria	1
Littoral	Cameroon	23	Kano	Nigeria	2
Nord	Cameroon	1	Lagos	Nigeria	13
Ouest	Cameroon	2	Rivers	Nigeria	1
Kouilou	Congo	4	Khartoum	Sudan	5
Pool	Congo	2	Dakar	Senegal	28
Comores		1	Kaolack	Senegal	1
Djibouti		1	Seychelles		2
Alger	Algeria	52	Chari-Baguirmi	Chad	5
Annaba	Algeria	2	Logone Occidental	Chad	1
Bourdj Bou Arrer	Algeria	1	Maritime	Togo	3
Bejaia	Algeria	1	Ariana	Tunisia	23
Oran	Algeria	1	Ben Arous (Tunis Sud)	Tunisia	50
Tlemcen	Algeria	1	Béja	Tunisia	4
Bani Suwayf	Egypt	1	Bizerte	Tunisia	52
Al Iskandariyah	Egypt	1	Gabs	Tunisia	5
Al Jizah	Egypt	1	Gafsa	Tunisia	2
Al Qahirah	Egypt	33	Jendouba	Tunisia	1
Maekel	Eritrea	1	Kairouan	Tunisia	2
Addis Ababa	Ethiopia	3	Mahdia	Tunisia	6
Estuaire	Gabon	12	Nabeul	Tunisia	57
Ogooué-Maritime	Gabon	3	Sfax	Tunisia	28
Greater Accra	Ghana	14	Sousse	Tunisia	222
Ashanti	Ghana	1	Tunis	Tunisia	219
Conarky	Guinea	2	Zaghouan	Tunisia	28
Gambia	Gambia	1	Central	Uganda	1
Nairobi	Kenya	8	Eastern Cape	South Africa	6
Tripolitania	Lybia	5	Gauteng	South Africa	67
Lesotho		1	KwaZulu/Natal	South Africa	4
Centre	Morocco	103	Mpumalanga	South Africa	1
Centre-North	Morocco	4	Northern	South Africa	1
Centre-South	Morocco	5	North West	South Africa	2
North-West	Morocco	26	Free State	South Africa	1
South	Morocco	3	Western Cape	South Africa	10
Tensift	Morocco	19	Copperbelt	Zambia	1
Antananarivo	Madagascar	31	Lusaka	Zambia	1
Mahajanga	Madagascar	1	Harare	Zimbabwe	2