Agglomeration and Informality: Evidence from Peruvian Firms

Maria Bernedo Del Carpio[a]

Carlianne Patrick[b,*]

September 2017

[a] *Department of Economics, University of Maryland, Baltimore County*, 100 Hilltop Circle, Baltimore, MD 21250

[b] *Department of Economics, Georgia State University*, 427 Andrew Young School of Policy Studies, 14 Marietta Street, Atlanta, GA 30303, United States

[*] Corresponding author.

E-mail addresses: mbernedo@umbc.edu (Maria Bernedo Del Carpio), cpatrick@gsu.edu (Carlianne Patrick).

ABSTRACT

This paper contributes to the emerging literature investigating agglomeration externalities in developing countries. It is the first paper to estimate how agglomeration externality generation and reception varies for formal and informal firms by source. Using establishment level data, we find evidence of both significant agglomeration economies and diseconomies associated with the interaction of formal and informal firms. In spite of regulatory restrictions for formal and informal linkages, formal establishments benefit from an increase number of informal suppliers but not generally from an increase in formal supplier density. On the other hand, the number of workers in informal establishments demanding similar skills causes a negative effect on productivity. For establishments in some industries, increasing the number of formal and informal establishments in the same industry has a negative effect on productivity.

Keywords: Agglomeration; Productivity; Development; Informality; Input-Output Linkages; Labor pooling

JEL Codes: R12, O17, R30, D24

1. INTRODUCTION

Despite the link between economic growth and agglomeration, there is a dearth of evidence on agglomeration externalities in developing world contexts (Combes and Gobillon 2015; Duranton 2012; 2016). The few existing studies tend to use data on the formal sector; yet, the informal sector comprises a substantial share of economic activity in the developing world.[1] Research on informality suggests a large informal sector may slow economic growth, although others argue for positive growth effects associated with informal entrepreneurs (Loayza 1996; Sarte 2000; Dabla-Norris and Feltenstein 2005; Busso et al. 2012). This study bridges the disparate literatures on agglomeration and informality by investigating Marshallian agglomeration externality generation and reception between formal and informal Peruvian establishments.

We use detailed establishment level data to estimate production functions in which total factor productivity (TFP) is a function of formal and informal sources of agglomeration externalities. We construct measures to capture intermediate input-sharing and labor market interactions as well as other own-industry spillovers. We further differentiate these sources with separate measures for the formal and informal sectors. In order to distinguish between agglomeration externalities and other locational characteristics affecting firm productivity, we also estimate specifications that include measures for geographic advantages, local area skill levels, and province fixed effects. We present estimates pooling all industries as well as separately for manufacturing, telecommunications and information services, and business services. We address the potential endogeneity of our agglomeration measures by constructing instruments from Colombian data. We address the potential correlation between input choices and unobserved productivity shocks by repeating our estimates using labor productivity rather than TFP, and by using a version of the Levinsohn-Petrin instrumental variable estimator.

We find evidence of both significant agglomeration economies and diseconomies associated with the interaction of formal and informal establishments. In spite of regulatory restrictions for formal and informal linkages, our results show that formal establishments benefit from an increase in the number of informal suppliers. Presumably, savings from trading with cheaper and/or more

[1] In a study with 162 countries, Schneider at al. (2010) estimate that between 1999 and 2007 the shadow economy comprised an average value of 34.5% of official GDP.

innovative suppliers overcome the costs associated with their formality status. However, the number of workers in informal establishments demanding similar skills causes a negative effect on productivity, which supports the hypothesis that the informal sector hampers human capital accumulation. Moreover, establishments' productivity reduces from increasing the number of formal and informal firms in the same industry. We suspect that decreased establishment level incentives for productivity-enhancing actions caused by the reduction in the residual demand dominate any positive productivity spillovers from own industry firms.

This research makes several important contributions. We add much-needed evidence on agglomeration economies in the developing world from a heretofore unstudied country. We are the first to consider how agglomeration externality reception *and* generation varies for formal and informal establishments. We simultaneously investigate multiple potential sources of agglomeration externalities, improving upon the existing literature that largely examines either one source in isolation or broad measures meant to capture multiple potential sources. We therefore provide novel evidence on transmission of agglomeration externalities between formal and informal establishments by source. We also go beyond the sectoral scope of existing studies and estimate effects for all industries and separately, for manufacturing, telecommunication and information services, and business services.

We proceed by providing some background, outlining our empirical implementation, presenting our results and robustness checks, and then providing some concluding remarks.

2. BACKGROUND

Agglomeration externalities can be thought of as generating localized increasing returns to scale – decreasing average costs as output increases in the location. The "sources" of agglomeration were first discussed by Marshall (1890; 1920) and the literature generally refers to three Marshallian "sources": labor-market interactions, linkages between intermediate and final good suppliers, and knowledge spillovers. As noted by Duranton and Puga (2004), the Marshallian "sources" may be the result of underlying mechanisms that produce observationally equivalent outcomes. They explain the theoretical microfoundations of agglomeration - the mechanisms

through which agglomeration economies cause increasing returns to scale in the aggregate production of a local economy. There are three types of mechanisms: (i) sharing, (ii) matching and (iii) learning. When firms cluster, they can share infrastructure, a greater variety of input suppliers (love-of-variety effects), a bigger pool of specialized and unspecialized workers, etc. They can also improve the quality of input or worker matches or the chances to find a good match. Proximity to individuals with greater skills and knowledge also facilitates the acquisition of skills and innovation. These mechanisms either directly affect productivity (technological externalities) or reduce the average costs of production (pecuniary externalities). As noted by the authors and demonstrated more recently by Combes and Gobillon (2015), the empirical challenge is distinguishing between the mechanisms.

Our data are not well-suited to uncovering underlying mechanisms. Thus, we follow much of the empirical literature and focus on classic Marshallian "sources" of agglomeration economies. We focus on measures related to industry size in the district, industry input suppliers in the district, and the size of the labor force with skills demanded by the industry. These allow us to consider different Marshallian sources and any between industry effects stemming from shared inputs and labor. Province level fixed effects allow us to control for any agglomeration externalities associated with overall size. We also include measures for geographic advantages, such as roads, topography, and skills to control for locational characteristics that may cause outcomes that are observationally equivalent to agglomeration externalities.

Much of the agglomeration literature suggests productivity benefits dominate any negative effects from competition. It also suggests significant productivity enhancements associated with the other Marshallian sources of agglomeration. However, agglomeration of firms also leads to competition, which in turn could affect productivity negatively. Moreover, most empirical agglomeration studies employ data from the developed world, where markets are (arguably) less distorted and the informal sector is small compared to the formal sector.

Syverson (2011) notes that increasing the number of competitor firms may affect individual firm productivity in three ways: spillover effects, selection effects, or by directly affecting the incentives to raise productivity levels. Spillover effects include the classic sources of

agglomeration. The selection effect occurs when more competition pushes market share towards lower cost, and therefore presumably more efficient, producers – pushing out less efficient firms and raising the productivity bar for entry. The literature is mixed concerning the effect of competition on individual firm incentives to raise productivity. Some models suggest that an increase in the number of competitors induces firms to incur adoption costs associated with productivity-enhancing actions, while others suggest that the reduction in residual demand reduces the incentives for such actions (Dasgupta and Stiglitz 1980; Spence 1984; Vives 2008; Syverson 2011). Vives (2008) tests competing theories across a broad range of market structures and finds evidence of both effects. In his paper, he finds evidence that an increase in the number of firms caused by a reduction in the entry costs reduces output at each firm as well as productivity investments. He does not consider a market structure where some firms face higher entry costs (formal firms) than others (informal); however, his results hint that such a market may have perverse effects on individual firm incentives to enhance productivity.

The nascent literature on agglomeration in developing world contexts is reviewed most recently by Combes and Gobillon (2015). They note that most studies examine the relationship between some measure of productivity, most commonly wages, and some measure of market size (city size, market potential, etc.) as the local determinant of agglomeration economies. With a few noteworthy exceptions, the agglomeration effect is estimated without addressing potential endogeneity concerns arising from reverse causality and omitted variables. The focus on wages as the productivity measure has advantages in the developing world context, but wages can be affected by a number of confounding factors, such as housing prices, amenities, local monopsony power, etc. Examining measures of productivity such as labor productivity or TFP avoids some of these issues (at the expense of creating new ones that we discuss later). Combes and Gobillon note only two studies of developing country TFP and a local determinant of agglomeration other than market size, both of which examine Chilean establishments and do not address the endogeneity of agglomeration measures. The research herein fills this gap by simultaneously estimating the TFP effects of multiple local determinants, or sources, of agglomeration as well as using instruments to address endogeneity concerns.

Only a handful of studies consider the informal sector when researching agglomeration economies in developing countries. Yet, the informal sector comprises a large portion of economic activity and the labor force in developing countries. Duranton (2016) highlights this issue when he notably includes wages paid by informal firms in his analysis of agglomeration in Colombia. His findings suggest higher returns to agglomeration in the informal sector. Garcia and Monroy (2013) study Colombian establishments' location decisions rather than wages. They find that formal and informal firms of similar size seem to locate close to each other. Their results also indicate that firms from industries that share inputs and labor cluster only with firms of their same type (formal or informal). Similarly, Mukim (2011) analyzes how Indian firms that belong to the unorganized sector choose their location[2]. Although the Indian informal sector is actually a subset of unorganized sector[3], Mukim (2011) finds evidence that agglomeration economies influence informal sector firms' location decisions. These studies suggest differences in agglomeration externalities by formality status. However, to our knowledge, no one has investigated how agglomeration externality generation and reception varies between formal and informal establishments by source.

The body of work on the determinants and consequences of informality gives some insight into why agglomeration externalities might vary depending on whether the establishments are in the formal or informal sectors. For example, De Paula and Scheinkman (2010) find that formal firms have the incentive to buy inputs only from other formal firms and sell their products to formal buyers under a value-added tax (VAT) system. Thus, VAT directly influences the extent to which agglomeration externalities may be transmitted through intermediate input suppliers of different types (informal or formal). Agglomeration externalities may also differ between types because productivity levels differ. Although some researchers argue that informal firms enhance growth through innovation free of burdensome government regulations and taxes, the literature suggests several reasons for lower productivity in informal firms (Busso et al. 2012). Informal firms lack access to credit which may lead to less productive input mixes than their formal counterparts

[2] The National Commission for Enterprises in the Unorganised Sector (NCEUS 2012) defines the unorganized sector as "…all unincorporated private enterprises owned by individuals or households engaged in the sale or production of goods and services operated on a proprietary or partnership basis and with less than ten total workers".
[3] Some firms with more employees than what the unorganized sector considers might not register and remain hidden.

(Dabla-Norris et al. 2008, Straub 2005). Informal firms tend to be small, limiting potential benefits from internal economies of scale. A priori less productive entrepreneurs may self-sort into the informal sector. Research indicates that informal firms are managed by less capable managers, have a lower capital-labor ratios, lower wages and lower profits (De Paula and Scheinkman 2011; La Porta and Schleifer 2008). Docquier et al. (2016) find that the informal sector hampers human capital accumulation because it reduces the skill premium, suggesting that agglomeration externalities operating through the shared labor and knowledge accumulation mechanisms may be affected. We let the data speak and discuss our findings in light of this literature.

The next section discusses our empirical implementation in more detail.

3. EMPIRICAL IMPLEMENTATION

3.1. Data

The research uses the following databases: the 2007 National Economic Census (CENEC), the 2007 Input-Output Matrix, the 2006 Annual Economic Survey (EEA) and the 2007 National Household Survey ENAHO[4]. All these databases come from the National Institute of Informatics and Statistic (INEI). To create some of the geographic advantages proxies we use the Shuttle Radar Topography Mission (SRTM) digital elevation data and the Global Roads Open Access Data Set.

The CENEC data includes all industry sectors except for the sectors "agriculture and public administration and defense". It contains information on establishments' production, location, 4 digit-ISIC (4th revision) industry, value added, employment, expenses, tax payments, inventories, etc. Establishments are located in districts. In Peru there are 1846 districts that form 195 provinces grouped in 25 regions. The size of a district is comparable to the size of a US county which is the geographic unit in various papers of agglomeration economies (Rosenthal and Strange 2004). We use 2007 CENEC to construct the proxies for the sources of agglomeration economies and to obtain measures on establishments' total value added, capital and labor.

[4] We did not use the 2007 National Census of Population and Households because it uses the international standard industrial classification code ISIC 3rd revision while CENEC and ENAHO use the 4th revision.

In order to differentiate effects between the formal and informal sectors, we require a working definition of informality. As noted in the informality literature, it is very difficult to have an unbiased indicator of informality because informal firms are often reluctant to report their legal situation to survey administrators. Studies on informality usually work with approximate measures of informality. In the CENEC, the establishments are asked their tax identification number, called RUC. Registration in the National Superintendency of Tax Administration (SUNAT) is mandatory for every firm and registered firms are assigned a RUC. In the study, we define informal firms as those firms that produce legal goods but do not provide their RUC when asked for it[5]. The RUC number is not a confidential number and it is provided to firms' customers in every sale receipt. One possible problem with this definition of informal firms is that firms could have provided a false RUC but this is a possible bias in all survey questions. If that is the case, the study underestimates the number of informal firms. The fact that it is well known that INEI does not provide establishment-level information to enforcement agencies reduces the potential severity of this issue in our data.

Even though the CENEC identifies industries using the ISIC code (4th revision), the Input-Output matrix uses the Peruvian economic activity code. The most extended version of this code includes 101 groups of industries. In order to merge CENEC and the Input-Output matrix, we applied a correspondence table. In a few cases, the correspondence was not a one-to-one mapping so that we had to aggregate some groups. Because of this and the fact that some sectors are not part of CENEC, we work with 92 groups of industries.

The Input-Output matrix contains information on intermediate demand and final demand for the 92 industries. We use the Input-Output matrix to construct the input-sharing measure. Specifically, the input-sharing measure is defined for each industry and district as the number of firms from major supplier industries. Major supplier industries are industries with significant vertical linkages to the industry for which the measure is constructed. Major suppliers are determined by choosing the top 25% of industry-supplier pairs for each industry. For every

[5] Business registration is an objective criterion for classifying businesses as formal or informal that is commonly used in the literature (Schneider 2005; Tanaka and Hashiguchi 2015). This is also the informality definition used by Moron et al. (2011), who make the caveat that it does not include other forms of informality such as keeping accounting books or benefits payment.

industry and district, we created two variables: the number of establishments in formal major supplier industries in the district and the number of establishments in informal major supplier industries in the district. We also create a supplier access measure that is the distance-weighted sum of the industry-districts' major suppliers in other districts.[6]

We identify the set of industries that share similar labor inputs using the 2007 National Household Survey ENAHO. This survey contains information on Peruvians' occupation and industry. As in Ellison et al. (2010), similarity is determined through the correlations of occupation's shares in each industry. Using the survey, we create a matrix of correlations and select the 25% of industries with the highest correlation to the industry for which the measure is constructed. Then, we calculate the labor-market interactions for each establishment in a certain industry and district as the number of workers in the establishments of the similar industries in terms of occupations.

We create our proxies for geographic and local cost advantages using district information on infrastructure, education and topography. We use the length of roads per district, the average altitude per district, the average slope per district and whether the district is located in the coast. We use two education measures: i) the interaction of the percentage of uneducated workers (do not have a high-school degree) per district times the percentage of uneducated workers that work in each economic activity and ii) the interaction of the percentage of educated workers (bachelor or more) per district times the percentage of educated workers that work in each economic activity.

3.2. *Methodology*

In this section we specify how the sources of agglomeration affect the firms' level of productivity, explicitly considering that the size of the spillover depends on the type of firm and the type of other firms with which the firm interacts.

[6] Using the same methodology, we construct a distance-weighted customer access measure. We estimate specifications using both the supplier and customer access variables. Although we sometimes gain statistical significance for the market access measure, inclusion of the customer access measure does not otherwise change our findings. Results are available upon request.

In the spirit of Greenstone, Hornbeck, and Moretti (2010), we consider a model of plant production technology in the presence of agglomeration spillovers where plants use the following Cobb-Douglas production technology:

$$(1)\ \mathrm{Y_{ijd}} = \mathrm{A_{ijd}}\left(\mathrm{L_{ijd}}\right)^{\rho_1}\left(\mathrm{K_{ijd}}\right)^{\rho_2}$$

where i, j, d stand for establishment, industry and district, respectively; $\mathrm{Y_{ijd}}$ is establishment i´s total value added, $\mathrm{A_{ijd}}$ is TFP, $\mathrm{L_{ijd}}$ is labor used in production measured as the number of paid and unpaid workers, and $\mathrm{K_{ijd}}$ is the value of land, building capital and equipment. $\mathrm{K_{ijd}}$ is the average of the book value at the beginning and at the end of year 2007 minus the depreciation in 2007.

Agglomeration theory suggests that TFP is a function of local plant density, $\mathrm{A_{ijd}} = A\left(N_{\mathrm{ijd}}\right)$ where N_{ijd} the number of firms in industry j in the district d where firm i is located. More specifically, agglomeration theory suggests that the effect on plant productivity will vary based on the source of the agglomeration externality and economic distance between plants. We consider the intermediate input-sharing and labor market interactions sources using the proxy measures described above. We also consider own-industry externalities by constructing measures of own-industry plant density in the firms' districts. We further differentiate these sources with separate measures for the formal and informal sectors. Allowing for differential productivity across plants in different industries and locations as well as idiosyncratic shocks to plant productivity suggests (log) TFP may be written as:

$$(2)\ \ln\left(\mathrm{A_{ijd}}\right) = \beta_1 ln\left(\mathrm{N_{F_{jd}}}\right) + \beta_2 ln\left(\mathrm{N_{Inf_{jd}}}\right) + \beta_3 ln\left(\mathrm{Input_{F_{jd}}}\right) + \beta_4 ln\left(\mathrm{Input_{Inf_{jd}}}\right) + \beta_5 ln\left(\mathrm{Labor_{F_{jd}}}\right) + \beta_6 ln\left(\mathrm{Labor_{Inf_{jd}}}\right) + \alpha_{\mathrm{j}} + \varepsilon_{\mathrm{ijd}}$$

where α_{j} are industry dummies, $\varepsilon_{\mathrm{ijd}}$ is an idiosyncratic productivity shock, $\mathrm{N_{F_{jd}}}$ $\left(\mathrm{N_{Inf_{jd}}}\right)$ is the number of establishments in industry j and in the formal (informal) sector in district d; $\mathrm{Input_{F_{jd}}}$ ($\mathrm{Input_{Inf_{jd}}}$) is the number of formal (informal) establishments in district d from the major industries that provide inputs to the firm according to the input-output matrix. $\mathrm{Labor_{F_{jd}}}$ ($\mathrm{Labor_{Inf_{jd}}}$) is the number of workers in formal (informal) establishments of similar industries, i.e. industries that are correlated with industry j in labor characteristics.

Taking logs of (1) and using equation (2) to replace $\ln(A_{ijd})$, the log of output *Y* in establishment *i*, in industry *j*, in district *d* is:

$$(3)\ \ln(Y_{ijd}) = \rho_1 \ln(L_{ijd}) + \rho_2 \ln(K_{ijd}) + \beta_1 ln\left(N_{F_{jd}}\right) + \beta_2 ln\left(N_{Inf_{jd}}\right) + \beta_3 ln\left(Input_{F_{jd}}\right) + \beta_4 ln\left(Input_{Inf_{jd}}\right) + \beta_5 ln\left(Labor_{F_{jd}}\right) + \beta_6 ln\left(Labor_{Inf_{jd}}\right) + \alpha_j + \varepsilon_{ijd}$$

As discussed above, we are interested in differential effects of agglomeration externalities on formal and informal firms. Thus, one empirical strategy is to estimate Equation (3) separately for the set of formal and informal firms.

Alternatively, we pool all firms and estimate the following equation:

$$(4)\ \ln(Y_{ijd}) = \rho_1 \ln(L_{ijd}) + \rho_2 \ln(K_{ijd}) + \beta_1 ln\left(N_{F_{jd}}\right) + \beta_2 ln\left(N_{Inf_{jd}}\right) + \beta_3 ln\left(Input_{F_{jd}}\right) + \beta_4 ln\left(Input_{Inf_{jd}}\right) + \beta_5 ln\left(Labor_{F_{jd}}\right) + \beta_6 ln\left(Labor_{Inf_{jd}}\right) + \varphi Inf_{ijd} + \left[\gamma_1 ln\left(N_{F_{jd}}\right) + \gamma_2 ln\left(N_{Inf_{jd}}\right) + \gamma_3 ln\left(Input_{F_{jd}}\right) + \gamma_4 ln\left(Input_{Inf_{jd}}\right) + \gamma_5 ln\left(Labor_{F_{jd}}\right) + \gamma_6 ln\left(Labor_{Inf_{jd}}\right)\right] * Inf_{ijd} + \alpha_j + \varepsilon_{ijd}$$

One draw-back to estimating Equation (4) on the pooled sample is that it assumes the same production parameters for formal and informal firms, with the only differences coming from differential TFP. This may be a strong assumption, particularly given the limited access to capital implied by credit constraints for informal firms. The solution is, of course, to include interaction terms for the production parameters as well. Adding an interaction term with geographic advantages makes this strategy equivalent to running (3) on the separate samples. Our estimates suggest that the assumptions in Equation (4) are indeed inappropriate for our data, and thus we present estimates from separate regressions of Equation (3)[7]. We also present results in which we include controls for geographic advantages in a district, supplier access, and province fixed effects.

[7] Equation (4) estimates are available upon request.

The literature suggests two sets of important potential types of endogeneity that should be considered when estimating equations (3) or (4). Given that our primary interest is estimating the effect of agglomeration measures, we are most concerned with the potential endogeneity stemming from input-output and labor relationships that are driven by unobserved productivity determinants or reverse causality issues. While our controls for geographic advantages and province fixed effects address obvious omitted variables, it is possible that other endogeneity issues remain. We therefore follow the intuition laid out in Ellison, Glaeser, and Kerr (2010) and instrument for input-output and labor relationships using data from a similar country, Colombia. The idea behind this instrument is that the Colombian industrial relationships should reflect actual technological relationships, but not any productivity determinant unique to Peru. Using the Colombian 2005 Input-Output matrix and 2007 Integrated Household Survey (Gran Encuesta Integrada de Hogares 2007) provided by the National Administrative Department of Statistics (DANE) as well as a mapping between Colombian and Peruvian national accounts, we identify for each Colombian industry the major suppliers and the main industries with which it shares labor. We apply these relationships to our firm level data to create instruments for our input-sharing and labor market interaction variables.

Another concern is that input choices are correlated with unobserved productivity shocks, causing biased productivity estimates. Our value-added specification removes concerns over the correlation between materials inputs and unobserved productivity; however, it is possible that labor and capital input choices are also correlated with unobserved productivity. We address this issue in two ways. First, we use a common approach in the literature and estimate a labor productivity specification of (3) separately for formal and informal firms in which the dependent variable is value-added per employee. In addition to removing potentially endogenous input choices from the estimation, this specification therefore has the advantage of focusing on, arguably, the most important part of TFP. The second way we address this issue is by implementing a variant of the Levinsohn-Petrin instrumental variable estimator first introduced by Sivadasan (2009), which is designed to explicitly address endogeneity in production function estimates in the case of repeated cross-sectional data.

3.3. *Summary statistics*

The census took place in 2008 and the production questions correspond to the economic year 2007. The original census data is composed of 864,030 establishments and exclude the sectors "agriculture" and "public administration and defense". Many of the establishments did not provide production information and the reason for that in 78% of these cases is that they started operations in 2008. We discard these observations. Some establishments provided a negative value added and/or provided a negative capital value. These observations are considered missing values when we take logarithms and therefore dropped. Finally, we follow Ghani, Kerr, and O'Connell (2011) and exclude establishments located in districts with fewer than 50 total formal and informal establishments.[8]

As we mentioned before, the CENEC uses the ISIC 4th revision to identify industries but the input and output matrix uses the Peruvian industry code. Using the latter we identify 92 industries. These industries are further aggregated into 12 major industries, or industrial sectors, for descriptive purposes. Table 1 shows the distribution of establishments in the final sample according to this industry classification as well as formality status. Approximately 62% of all sample establishments belong to the sector "trade, maintenance and repair service of vehicles". The next biggest sectors are "lodging and restaurants", "other services" and "manufacturing".

[Insert Table 1 approximately here]

Table 1 reveals that nearly 40% of sample establishments are informal, but the relative proportion of formal and informal firms varies substantially across major industries. Nearly all establishments in the "fishing and aquaculture" sector are informal firms. The "trade, maintenance and repair service of vehicles" sector has the second highest share of establishments that did not report a tax-id at just over 40%. The sectors with the highest proportion of formal firms are "construction" and "financial services, insurance and pensions". While we retain these sectors in our pooled results, we focus on the manufacturing, telecommunications and information, and

[8] Our primary findings are similar when we use all establishments.

business service sectors in our sectoral analyses to increase comparability with previous research that examines manufacturing and business services.

Table 2 reports the estimation variables by formality status categorized into three groups: inputs, agglomeration source measures, geographic and locational cost advantages proxies. Value added and capital are in Nuevos Soles[9]. On average, value added in the economic year 2007 was around US$7,619 and the capital value was US$5,607. The level of capital and value added are much higher for formal establishments than for informal ones, as expected. In our estimation sample, formal firms report an average value added of US$11,641 (34,690 Nuevos Soles) compared to US$3,128 (9,322 Nuevos Soles) for informal firms. The difference between formal and informal establishment capital stocks are even more pronounced, at US$9,940 (29,622 Nuevos Soles) and US$987 (2,940), respectively. Establishments are small in terms of employees with an average of 2 employees per establishment for both types of firms. The reason for that is that most firms in Peru are small or microenterprises with one or two workers in formal and informal sectors.

[Insert Table 2 approximately here]

Considering our measures of agglomeration sources in Table 2, the concentration of similarly skilled workers employed by formal firms in the same district is significantly higher for formal establishments than for informal ones. Interestingly, though, exposure to shared skill workers employed by informal firms does not vary significantly between formal and informal Peruvian establishments. Regardless of formality status, the number of formal firms in the industries supplying the top twenty-five percent of a firm's intermediate inputs is higher than the number of informal firms. Formal firms tend to be located in districts with higher concentrations of both formal and informal intermediate input suppliers than their informal counterparts, suggesting greater potential for agglomeration externalities through the intermediate input channel for formal firms.

District geographic advantages are captured by a measure of the length of roads, the average slope of the district, the average altitude of the district and two measures of skills availability.

[9] In 2007, the exchange rate was 2.98 Nuevos Soles per US dollar.

Establishments tend to locate in almost flat areas, as expected, but not necessarily in low altitude areas. The skills variables show a low match between skills requirements per sector and labor availability per district.

To have a better understanding of the distribution of establishments across the country, Figure 1 maps the number of establishments per district in Peru. There are a few important economic areas that are difficult to identify in the complete map. We focus on the province of Lima and map the formal and informal establishments in the manufacturing sector in Figures 2A and 2B. We can see that formal and informal manufacturing establishments tend to locate in the same districts. Analysis for other major sectors reveals that this is true in most of the cases.

4. PRODUCTION FUNCTION ESTIMATES

Tables 3-6 present the results of separate regressions by formality status for firms in all industries, manufacturing, telecommunications and other information services, and business services, respectively. Again, we focus on these sectors to increase comparability with previous research that examines manufacturing and services. Panels A and B present the results for formal and informal establishments, respectively. Column (1) presents results from the base specification. Column (2) adds the controls for geographic advantages to the base specification. Column (3) includes the weighted supplier access measure and Column (4) includes the geographic advantage measures as well as province fixed effects. All specifications include 92-level industry fixed effects. Standard errors are clustered at the Department level.

4.1 Pooled Industries

Table 3 contains the results for all establishments in the trimmed sample regardless of industry. The pooled results (as well as those for each major industry) show that production is very labor-intensive in Peru, as expected for labor intensive technologies and overall level of development in Peru. Our measures of geographic and locational cost advantages affect both formal and informal establishment productivity in the expected directions.

Table 3 suggests that formal and informal firms benefit from agglomeration externalities generated through intermediate input suppliers in their district that are also informal. Panel A indicates an elasticity of formal firm productivity with respect to the number of informal intermediate input suppliers in the top twenty-five percent of activity codes that provide inputs to the firm's industry of 0.07-0.14. This is similar to estimates of the elasticity of developed world (manufacturing) productivity with respect to intermediate input supplier density, but lower than the 0.4-.0.5 range of entry-input elasticities estimated by Ghani, Kerr, and O'Connell (2011) for the organized and unorganized manufacturing sectors in India. Similar to Duranton's (2016) finding of greater returns to agglomeration in the informal sector, our results also indicate a higher elasticity with respect to informal suppliers for informal establishments than for formal ones. Table 3 Panel B reports elasticities around 0.2, suggesting increasing the number of informal intermediate input suppliers in the top twenty-five percent of activity codes that provide inputs to the firm's industry by ten percent increases informal firm productivity by two percent. These productivity gains could be due to decreasing average costs of inputs as supplier density increases, love-of-variety effects in intermediate inputs, or knowledge transmission through supplier relationships.

Interestingly, productivity gains through intermediate input channels appear to be limited to informal suppliers. Informal firms do not appear to benefit from having more formal suppliers in their district. Formal establishment productivity seems to actually decline with the density of formal suppliers – although the increases associated with informal supplier density more than offsets this estimated decrease. The fact that increasing the number of formal suppliers does not have similar effects on productivity as increasing the number of informal suppliers suggests declining average costs as the primary mechanism driving the informal supplier result. Pure love-of-variety effects should exist for the variety of both formal and informal intermediate inputs – rather than just informal intermediate inputs. Presumably, the average cost of informal intermediate inputs is lower than that of intermediate inputs produced by formal firms that incur additional production costs in the form of taxes and formal labor. If there are not substantial differences in the quality of intermediate inputs produced by formal and informal suppliers, then firms may substitute lower cost informal intermediate inputs for formal intermediate inputs – particularly if the cost savings are higher than the costs associated with dealing with an informal

firm, such as the loss of VAT discounts. The negative coefficient on formal suppliers suggests this might be the case. The fact that formal firm elasticity with respect to informal intermediate inputs is about half of the informal firm elasticity suggests VAT or other restrictions may discourage informal intermediate inputs purchases by formal firms, but not enough to prevent all informal intermediate input purchases.

Formal firm productivity is also increasing in the number of employees with similar skills working at formal firms, but decreasing in the number of workers with similar skills working at informal firms. If both formal and informal firms compete for the same labor pool, then increasing the number of employees demanded by other firms should put upward pressure on wages and increase production costs. However, a larger pool of similarly skilled labor may also increase worker productivity through risk sharing, matching, or learning mechanisms. Such increases in worker productivity can reduce production costs, thereby offsetting the increases due to competition. The results suggest that there are productivity benefits to a larger pool of workers in formal firms that outweigh the upward pressure on wages. However, the competition effect appears to dominate for informal firm workers. Moreover, an additional effect in line with Docquier et al. (2016) could take place in this context. The supply of informal workers in industries with similar occupation needs might reduce the premium of high skilled workers whose labor market is mainly the formal economy. The fall in the high skilled premium affects the incentives to gain more skills, which ultimately reduces the productivity of formal firms.

[Insert Table 3 approximately here]

4.2 Manufacturing, Telecommunications and Information, and Business Services

Tables 4-6 demonstrate that the source of agglomeration externalities and competition effects varies across major industry groups as expected. Table 4 presents the results for establishments in manufacturing industries. Both informal and formal manufacturing productivity are increasing in informal supplier density, although the increase is only statistically significant for formal establishments. Unlike the pooled results, formal manufacturing plant productivity is increasing in the number of formal suppliers as well. Sharing similarly skilled workers with informal firms

reduces productivity for formal manufacturers – again, suggesting that informal employment hampers human capital accumulation.

Table 5 contains the results for informal and formal firms engaged in telecommunications and other information services. Again, increasing the number of informal firms in the top twenty-five percent of activity codes that provide inputs to the firm's industry has significant positive productivity benefits for both informal and formal firms. Our results indicate an elasticity of formal and informal firm productivity with respect to informal intermediate input suppliers of approximately 0.15 and 0.3, respectively. These are much higher than the manufacturing elasticities with respect to informal intermediate input suppliers in Table 4, suggesting stronger spillovers. Increasing the number of formal establishments supplying intermediate inputs in the district is also positively associated with the productivity of both types of establishments, but the effects are not generally statistically significant. The smaller productivity gain for formal firms also suggests that there are frictions in formal-informal establishment relationships such as VAT that may discourage formal firms from using informal intermediate input suppliers and thereby reduce potential gains from these cost reductions.

As with the results from pooled industry groups and manufacturing industries, formal firm productivity declines with the number of workers employed by informal firms demanding similarly skilled workers. Again, this is consistent with the idea that the informal labor pool reduces the skill premium and, thus, the skilled labor supply, which in turn affects formal firms' productivity.

Table 5 also suggests competition effects offset any productivity benefits associated with increasing own-industry-formality status establishments. Both formal and informal telecommunications and information service firms experience decreased value-added as the number of formal and informal own-industry firms increases, respectively. The elasticity of formal establishment productivity with respect to own-industry formal establishments is between -0.14 and -0.17. Interestingly, increasing the number of informal own-industry competitors has no effect for formal establishments. On the other hand, the elasticity of informal establishment productivity with respect to own-industry informal competitors is approximately -0.12.

Table 6 contains the estimated effects for formal and informal establishments providing services to other businesses. The results echo the previously estimated effects for formal establishments in other industries. Formal business service firm productivity is increasing in the number of similarly skilled employees in formal firms, with an estimated elasticity between 0.18-0.29. However, the elasticity of value-added with respect to workers employed by informal firms is around -0.19 and -0.24. Again, this suggests that workers employed in formal firms increase their human capital in ways that generate positive externalities; however, workers in informal firms do not, even worse they seem to affect formal firms' productivity by reducing the skill premium.

While informal establishments providing services to other businesses appear unaffected by the number of intermediate input suppliers in their districts, formal establishments experience significant changes in value-added as the number of establishments in the top twenty-five percent of activity codes that provide inputs to the establishments' industry increases. However, the effects vary substantially by the formality status of supplier firms. Panel A of Table 6 indicates an elasticity of formal firm productivity with respect to informal suppliers of 0.2; while the elasticity with respect to formal suppliers ranges from -0.2 to -0.3. These estimates suggest that formal business services benefit from lower cost inputs from the informal sector despite the frictions in formal and informal trade such as VAT that discourage buying from informal firms. However, similar benefits do not accrue to from more formal suppliers – perhaps, due to competition for a limited supply of inputs from the formal sector and VAT-induced preferences for formally supplied inputs.

It is also interesting to note that the productivity of informal firms providing business services is increasing in the number of informal firms in their own industry in the sector, with an estimated elasticity of 0.15-0.19. There is no such effect for formal firms in the same 92-level industry in the district. Thus, it appears that informal business service establishments generate positive externalities for informal firms in the same industry. This is different than other major industry sectors in which we either a negative or no effect from increasing own-industry informal establishment density.

Taken together, the results in Tables 3-6 indicate that the formality status interacts with agglomeration economies and diseconomies. The positive externalities generated through large

labor pools with shared skills dominate any upward pressure on wages when those workers are employed in formal firms. However, employment in informal firms does not appear to generate the same positive externalities, indicating less human capital accumulation for workers in informal firms and/or reduced incentives for skill accumulation by formal workers. Despite the decrease in productivity generally associated with more workers being employed by informal firms, the presence of more informal suppliers tends to have productivity benefits for both formal and informal firms. Formal suppliers, however, do not appear to generate positive externalities. These results suggest that the productivity effects associated with intermediate input suppliers are due to cost differences between the types of supplier firms (rather than knowledge spillovers or love-of-variety effects) and substitution of informally-supplied inputs for formally-supplied ones.

5. ROBUSTNESS AND SENSITIVITY

5.1. Labor productivity estimates

Table 7 reports the results of estimating labor productivity as a function of our measures of agglomeration sources by formality status. The results confirm our findings in Section 4 and are of similar magnitude. Increasing the density of informal suppliers increases labor productivity for formal and informal establishments in all industries. The effects are generally stronger for informal establishments (Columns 2) and are strongest in the service sectors (Panel C and D). Productivity across industries and formality status is also decreasing in the number of shared-skill workers employed by informal establishments.

[Insert Table 7 approximately here]

5.2. Instrumenting for agglomeration metrics

As we mentioned earlier in the paper, the input-output and labor relationships could be driven by unobserved productivity determinants. It could be that industries coagglomerate due to some randomly proximate unobserved advantages. Once industries are located in the same area, they start sharing workers and suppliers. These unobserved advantages could also directly affect firms' productivity. We address this potential threat to identification by using input-output and labor similarity relationships observed in Colombia. It is unlikely that if two industries in Peru coagglomerate because they use unobserved geographic advantages that are randomly close to

each other, they will coagglomerate for the same reasons in Colombia. We map Peruvian and Colombian industries and use Colombian input-output measures and labor-market interactions per industry as instruments for Peruvian variables. Table 8 presents the results of this exercise. Again, the results confirm our previous findings, although the point estimates suggest even stronger elasticities.

[Insert Table 8 approximately here]

5.3. Levisohn-Petrin instruments

To check for possible endogeneity of variable inputs, we follow the Levisohn-Petrin methodology modified in Sivadasan (2009). The author adapted the methodology to the case of a repeated cross-section database. Unfortunately, the previous census of firms has been questioned because of the quality of the data. Thus, we follow Sivadasan's method usingthe 2006 Annual Economic Survey (EEA 2006) for the Manufacturing sector and formal firms with net sales greater than 300 000 Nuevos Soles (or $100 000). We focus on this economic sector and firm size because these are the ones for which we have data on energy expenditure. All the firms that participate in the EEA survey are formal according to our definition. Results are shown in Table 9. Results correspond only to the formal firms. The impact of the agglomeration variables show the same direction as in Panel A of Table 4.

[Insert Table 9 approximately here]

6. CONCLUSION

This paper contributes to our understanding of informality and agglomeration externalities by estimating the effect of formal and informal sources of agglomeration on formal and informal establishment productivity. Specifically, we estimate production functions using Peruvian establishment-level data that allow us to identify formal and informal establishments based on their registration status. We specify establishment TFP as a function of formal and informal local major supplier plant density, density of plants sharing similar labor, own-industry plant density,

geographic advantages and industry and province fixed effects. To address the potential endogeneity, we instrument for our sources of agglomeration using Colombian input-output and labor relationships. For robustness, we also compare our primary results with estimates of labor productivity as a function of our measures of agglomeration sources as well as estimates using a modified version of the Levinsohn-Petrin estimator. To our knowledge, this is the first study to explicitly consider how agglomeration externality generation and reception varies for formal and informal establishments, to do so while considering multiple sources simultaneously, and to address the potential endogeneity of agglomeration measures in a developing world context.

We find evidence of both significant agglomeration economies and diseconomies associated with the interaction of formal and informal establishments. We find evidence that in spite of regulatory restrictions for formal and informal linkages, formal firms benefit from an increase number of informal suppliers. Informal suppliers augment productivity either because they are more innovative or because savings from trading with cheaper suppliers overcome the costs associated with their formality status. Our results for most industries suggest substitution of informally-supplied intermediate inputs for formally-supplied ones, which is more consistent with the cost-savings mechanism than innovation. Manufacturing appears to the exception to this substitution pattern. These results indicate that isolating the effects through the potential cost-savings and innovation channels is an important area for future research.

While we find that the size of the labor pool demanded by formal establishments generates positive externalities on firms' productivity, the number of workers in informal firms demanding similar skills causes a negative effect on productivity which supports the hypothesis that the informal sector hampers human capital accumulation. Additional research is required to determine the mechanism by which informality discourages human capital accumulation.

For some establishments in our sample, increasing the number of formal and informal firms in the same industry has a negative effect on productivity. More competition is not necessarily productivity enhancing, as part of the literature advocates, but the opposite. We suspect that this is due to reductions in residual demand dominating any productivity-enhancing own-industry spillovers. These results suggest the need for additional research into productivity effects of competition in markets characterized by large informal sectors with low entry costs and lower operating costs that do not result from lower productivity.

REFERENCES

Aghion, P. and P. Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60: 323-351.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In *Handbook of Regional and Urban Economics, Vol. 5*, edited by J.V. Henderson, G. Duranton, and W. Strange. Amsterdam: North Holland, 247-348.

Dabla-Norris, Era and Andrew Feltenstein. 2005. "The underground economy and its macroeconomic consequences." *Journal of Policy Reform* 8:153-174.

Dabla-Norris, Era, Mark Gradstein, Gabriela Inchauste. 2008. What causes firms to hide output? The determinants of informality. *Journal of Development Economics 85 (1,2): 1-27.*

Dasgupta, P. and Joseph Stiglitz. 1980. "Industrial Structure and the Nature of Innovative Activity." *The Economic Journal* 90: 266-293.

De Paula, Áureo and José A. Sheinkman. 2010. "Value-Added Taxes, Chain Effects, and Informality". American Economic Journal: Macroeconomics, 2(4): 195-221.

Duranton, Gilles and Henry G. Overman. 2008. "Exploring the Detailed Location Patterns of U.K. Manufacturing Industries Using Detailed Microgeographic Data." *Journal of Regional Science* 48(1): 213-243.

Duranton, Gilles. 2012. "Agglomeration and jobs in developing countries." *Background paper for the World Development Report 2013.*

Duranton, Gilles. 2016. "Agglomeration effects in Colombia." *The Journal of Regional Science* 56(2): 210-238.

Docquier, Frederic, Tobias Muller and Joaquin Naval. 2014. "Informality and Long Run Growth". IZA DP No. 7883

Ellison, Glenn; Edward L. Glaeser and William R. Kerr. 2010. "What causes industry coagglomeration? Evidence from coagglomeration patterns?" *American Economic Review*, 100(3): 1195-1213.

Ellison, Glenn and Edward L. Glaeser. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *The American Economic Review*. 89(2): 311-316.

Fujita, M. and J. Thisse. 2002. *Economics of Agglomeration.* Oxford: Oxford University Press.

Fajnzylber, Pablo, William F. Maloney and Gabriel V. Montes-Rojas. 2011. "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program". Journal of Development Economics. 94 (2): 262–276

Garcia, Gustavo A. and Ana I. Moreno Monroy. 2012. "Intra-metropolitan Clustering of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia" University of Groningen, Groningen, The Netherlands.

Gal, Peter N. 2013. "Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS." Organization for Economic Co-operation and Development Economics Department Working Paper N. 1049.

Ghani, Ejaz, William R. Kerr, and Stephen D. O'Connell. 2011. "Spatial Determinants of Entrepreneurship in India," NBER Working Paper 17514.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *The Journal of Political Economy* 118 (3) (June): 536-598.

Grossman, Gene .M. and Elhanan Helpman. 1994. *Innovation and Growth in the Global Economy*. Cambridge and London: MIT Press.

Henderson, J. V. (1986), "Efficiency of Resource Usage and City Size,"*Journal of Urban Economics* 19: 47-70. La Porta, R., and A. Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity*, 39(2), pp. 275-363.

Levinshohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70(2): 317-341.

Loayza, Norman. 1996. "The economics of the informal sector: a simple model and some empirical evidence from Latin America." Carnegie-Rochester Conference Series on Public Policy 45:129-162.

McDonald, John F. and Daniel P. McMillen. 2011. *Urban Economics and Real Estate, Theory and Policy*, 2nd Edition. John Wiley & Sons, Inc.

Nakamura, R. (1985), "Agglomeration economies in urban manufacturing industries: a case of Japanese cities,"*Journal of Urban Economics* 17: 108-124.

Morón, E., Salgado, E., & Seminario, C. (2011). Productividad, aglomeración y asignación de recursos en el Perú: ¿qué dice el Censo Económico Nacional 2008?. Documento de Discusión 11/09. Universidad del Pacífico.

Olley, G. Steven and Ariel Pakes. 1996. The Dynamics of Productivity in the Telecommunications and Equipment Industry. *Econometrica* 64(6): 1263-1297.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98:71-102.

Rosenthal, Stuart S. and William C. Strange. 2004. Evidence on the Nature and Sources of Agglomeration. In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.

Sarte, Pierre-Daniel. 2000. "Informality and rent-seeking bureaucracies in a model of long-run growth." *Journal of Monetary Economics* 46(1): 173-197.

Schneider, Friedrich. 2005. "Shadow economies around the world: what do we really know?" *European Journal of Political Economy* 21(3): 598-642.

Schneider, Friedrick, Andreas Buehn, and Claudio E. Montenegro. 2010. "Shadow economies all over the world: new estimates for 162 countries from 1999 to 2007." *Policy Research Working Paper* 5356, World Bank.

Sivadasan, Jagadeesh. 2009. "Barriers to Competition and Productivity: Evidence from India. *The B.E. Journal of Economic Analysis and Policy* 9(1).

Spence, Michael 1984. "Cost Reduction, Competition, and Industry Performance." *Econometrica* 52(1): 101-121.

Straub, Stephane. 2005. "Informal sector: the credit market channel". *Journal of Development Economics* 78: 299-321.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 43(2): 326-365.

Tanaka, Kiyoyasu and Yoshihiro Hashiguchi. 2015. "Agglomeration Effects of Informal Sector: Evidence from Cambodia." *Discussion Paper No.* 495, Institute of Developing Economies.

Vives, Xavier. 2008. "Innovation and Competitive Pressure." *The Journal of Industrial Economics* LVI: 419-469.

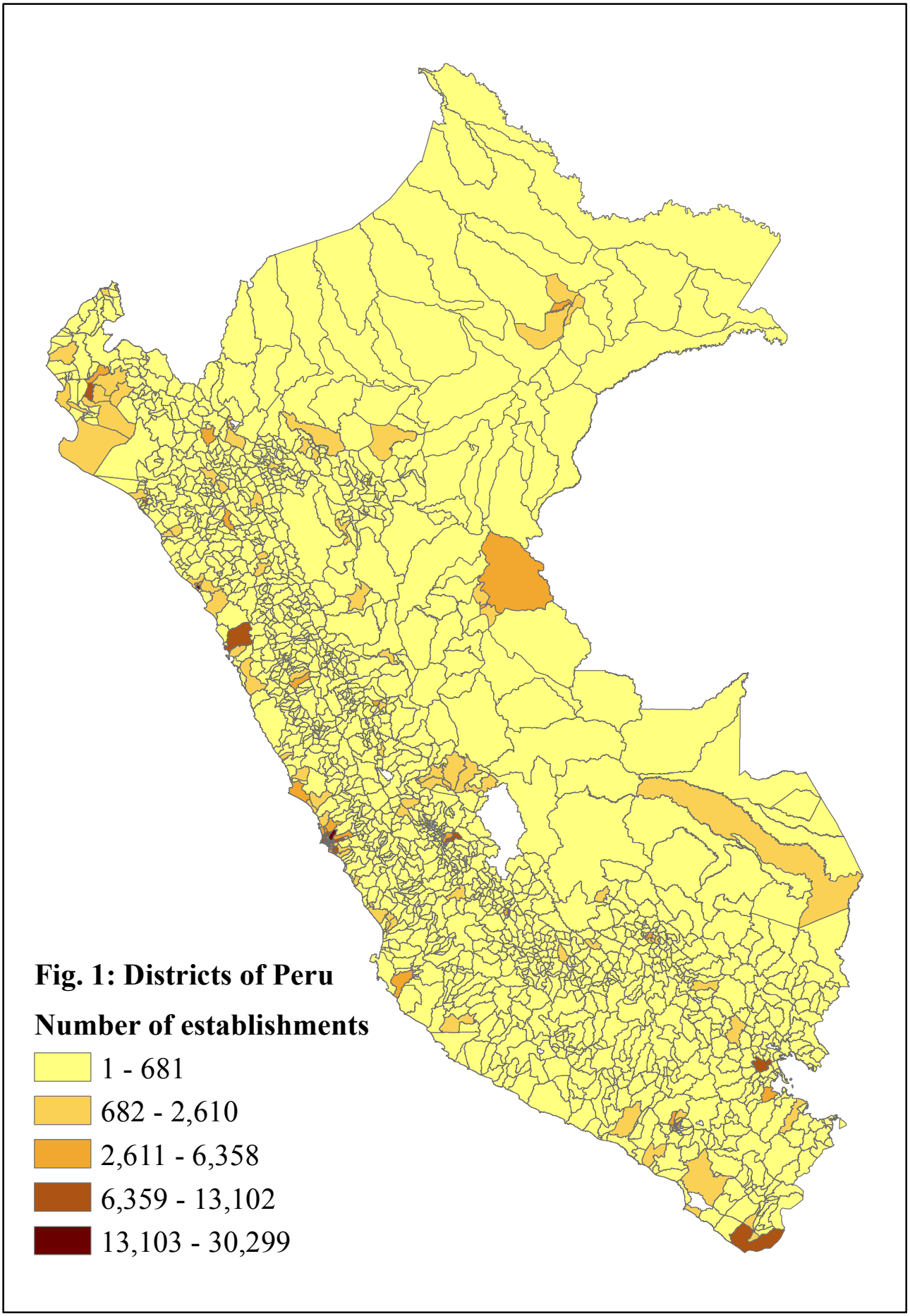
Fig. 1: Districts of Peru
Number of establishments
1 - 681
682 - 2,610
2,611 - 6,358
6,359 - 13,102
13,103 - 30,299

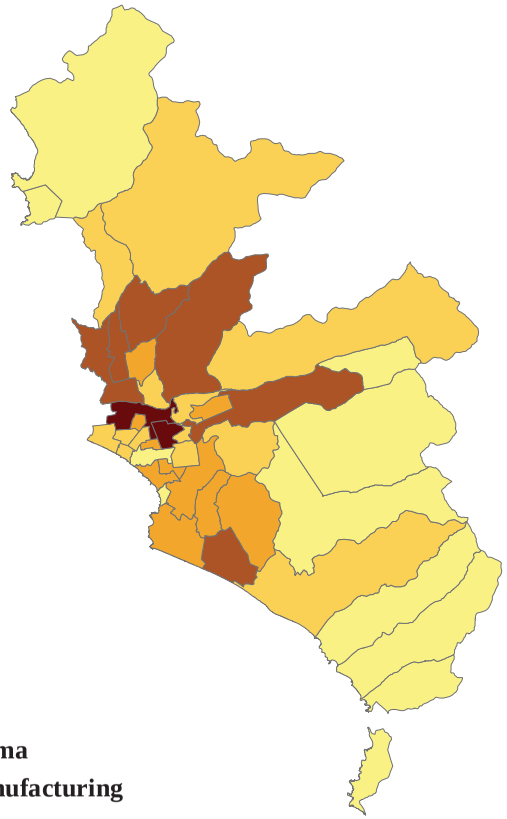

Fig. 2A

Province Lima

Formal Manufacturing

- 4 - 83
- 84 - 220
- 221 - 380
- 381 - 939
- 940 - 2,829

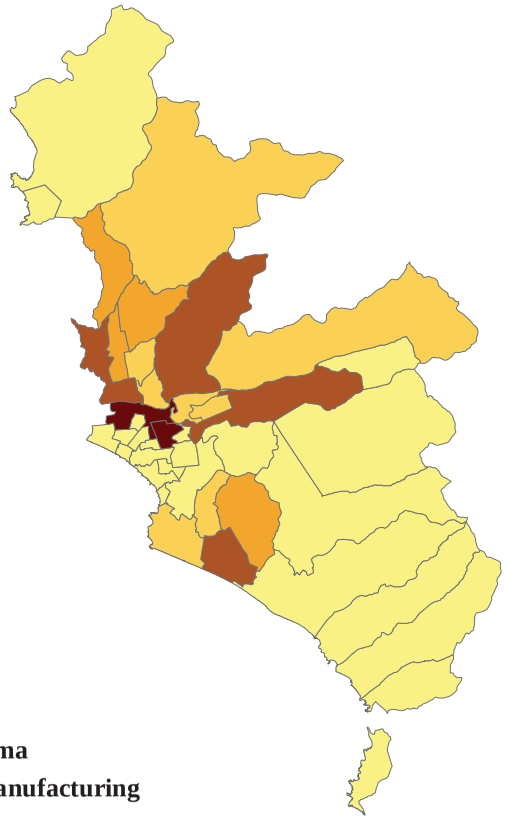

Fig. 2B

Province Lima

Informal Manufacturing

- 0 - 41
- 42 - 120
- 121 - 223
- 224 - 394
- 395 - 822

Table 1: Distribution of Establishments by Major Industry Classification and Formality

Major Industry	Total	% Total	Formal	Informal	% Informal
Fishing and Aquaculture	5,969	1.33	82	5,887	98.63%
Oils, Minerals, and Natural Gas Extraction	65	0.01	57	8	12.31%
Manufacturing	37,943	8.43	26,906	11,037	29.09%
Electricity, Gas, and Water	308	0.07	236	72	23.38%
Construction	1,103	0.24	1,003	100	9.07%
Trade, Maintenance, and Repair Services	280,591	62.32	164,927	115,664	41.22%
Transportation, Storage, and Mail	3,017	0.67	2,593	424	14.05%
Lodging and Entertainment	43,708	9.71	29,158	14,550	33.29%
Telecommunications and Other Information Services	17,759	3.94	12,472	5,287	29.77%
Financial Services, Insurance, and Pensions	657	0.15	602	55	8.37%
Business Services	19,018	4.22	16,205	2,813	14.79%
Other Services	40,109	8.91	27,689	12,420	30.97%
Total	450,247		281,930	168,317	37.38%

Table 2: Summary Statistics by Formality Status

		Formal		Informal	
	Variables	Mean	Sd	Mean	Sd
Production Variables	Value added (in Nuevos Soles)	34,690	246,499	9,322	27,237
	Capital (in Nuevos Soles)	29,622	610,209	2,940	76,888
	Number of paid and unpaid employees	2.411	5.554	1.561	1.260
Agglomeration Variables	Number of Formal Own Ind.	1,783	3,001	1,222	2,352
	Number of Informal Own Ind.	841.7	1,280	884.5	1,357
	Shared-Skills Labor in Formal Firms	5,422	8,102	3,459	6,402
	Shared-Skills Labor in Informal Firms	2,170	3,105	2,027	3,047
	Formal suppliers	1,268	1,934	776.5	1,737
	Informal suppliers	959.3	1,541	694.4	1,340
Geographic Advantages	Roads (km/km2)	143.2	163.9	179.3	177.9
	Slope (degrees)	7.293	6.462	8.954	6.624
	Altitude (m)	1,092	1,423	1,487	1,581
	% Dist. x % Ind. Low Education	0.036	0.029	0.057	0.037
	% Dist. x % Ind. High Education	0.020	0.051	0.006	0.020

Table 3: Formal and Informal Establishments in All Industries

	Panel A: Formal Establishments				Panel B: Informal Establishments			
VARIABLES	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Capital	0.278***	0.277***	0.278***	0.273***	0.254***	0.253***	0.254***	0.245***
	(0.00738)	(0.00699)	(0.00736)	(0.00589)	(0.0105)	(0.00908)	(0.0105)	(0.00605)
Labor	0.726***	0.725***	0.726***	0.728***	0.528***	0.528***	0.528***	0.526***
	(0.0130)	(0.0123)	(0.0122)	(0.0124)	(0.0372)	(0.0371)	(0.0361)	(0.0308)
Formal Suppliers	-0.0713**	-0.0659**	-0.0781***	-0.0568***	-0.0524	-0.0348	-0.0535	-0.0402
	(0.0339)	(0.0296)	(0.0247)	(0.0144)	(0.0525)	(0.0531)	(0.0612)	(0.0350)
Informal Suppliers	0.120***	0.134***	0.141***	0.0740***	0.218***	0.224***	0.221***	0.125**
	(0.0324)	(0.0316)	(0.0338)	(0.0227)	(0.0620)	(0.0586)	(0.0575)	(0.0472)
Shared-Skills Labor	0.177***	0.174***	0.179***	0.188***	-0.00337	-0.0429	-0.00415	0.0267
in Formal Firms	(0.0371)	(0.0390)	(0.0275)	(0.0263)	(0.0874)	(0.105)	(0.0922)	(0.0718)
Shared-Skills Labor	-0.159***	-0.164***	-0.165***	-0.0930***	-0.105*	-0.0841	-0.104*	0.0152
In Informal Firms	(0.0256)	(0.0251)	(0.0247)	(0.0261)	(0.0573)	(0.0620)	(0.0565)	(0.0808)
Formal Own Ind.	-0.0574**	-0.0403	-0.0390	-0.0312*	0.0394	0.0341	0.0422	0.00617
	(0.0264)	(0.0295)	(0.0263)	(0.0157)	(0.0615)	(0.0545)	(0.0514)	(0.0379)
Informal Own Ind.	0.0343	0.0202	0.0210	-0.0076	-0.00816	-0.0221	-0.0112	-0.0206
	(0.0236)	(0.0246)	(0.0254)	(0.0117)	(0.0496)	(0.0493)	(0.0402)	(0.0358)
Formal Supplier Access			0.170				0.0319	
			(0.156)				(0.378)	
Informal Supplier Access			-0.221				-0.0399	
			(0.171)				(0.397)	
Industry FEs	Y	Y	Y	Y	Y	Y	Y	Y
Geographic Advantages	N	Y	N	Y	N	Y	N	Y
Province FEs	N	N	N	Y	N	N	N	Y
Observations	269,528	269,087	269,528	269,087	159,565	158,370	159,565	158,370
R-squared	0.297	0.298	0.297	0.311	0.188	0.189	0.188	0.226

Notes: Panels A and B contain the results for separate regressions for formal and information establishments, respectively, in all industries. The dependent variable is (logged) value-added. Capital, Labor, the number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, the number of informal firms with the same activity code in the district, the distance-weighted sum of formal firms of the industry-districts' major suppliers in other districts, the distance-weighted sum of informal firms of the industry-districts' major suppliers in other districts are measured in logs. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 4: Manufacturing Establishments

	Panel A: Formal Establishments				Panel B: Informal Establishments			
VARIABLES	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Capital	0.264***	0.264***	0.263***	0.260***	0.236***	0.234***	0.235***	0.224***
	(0.0069)	(0.0068)	(0.0071)	(0.0070)	(0.0166)	(0.0174)	(0.0168)	(0.0160)
Labor	0.734***	0.734***	0.733***	0.737***	0.600***	0.606***	0.601***	0.601***
	(0.0302)	(0.0300)	(0.0303)	(0.0274)	(0.0287)	(0.0279)	(0.0290)	(0.0263)
Formal Suppliers	0.0014	0.0333	0.0351	0.0768***	-0.0002	0.0132	0.0044	-0.0575
	(0.0331)	(0.0334)	(0.0319)	(0.0229)	(0.0546)	(0.0555)	(0.0558)	(0.0478)
Informal Suppliers	0.0725***	0.0569**	0.0458	-0.0313	0.0784	0.0777	0.0737	0.0150
	(0.0239)	(0.0252)	(0.0268)	(0.0291)	(0.0618)	(0.0590)	(0.0658)	(0.0549)
Shared-Skills Labor	0.0792	0.0464	0.0455	0.0876*	-0.0284	-0.111*	-0.0322	0.0420
in Formal Firms	(0.0636)	(0.0654)	(0.0558)	(0.0469)	(0.0458)	(0.0574)	(0.0440)	(0.0514)
Shared-Skills Labor	-0.0766***	-0.0642**	-0.0606**	-0.0452**	0.00192	0.0360	-0.000704	0.0412
In Informal Firms	(0.0245)	(0.0259)	(0.0230)	(0.0204)	(0.0470)	(0.0432)	(0.0463)	(0.0424)
Formal Own Ind.	-0.0266	-0.0323	-0.0371	-0.0488**	0.0465	0.0556*	0.0450	0.0380
	(0.0298)	(0.0264)	(0.0248)	(0.0213)	(0.0327)	(0.0291)	(0.0281)	(0.0283)
Informal Own Ind.	0.0196	0.0212	0.0248	0.0237	-0.0014	-0.0009	0.0009	0.0069
	(0.0181)	(0.0173)	(0.0181)	(0.0154)	(0.0281)	(0.0323)	(0.0277)	(0.0260)
Formal Supplier Access			-0.0047				-0.0603	
			(0.0850)				(0.233)	
Informal Supplier Access			0.0401				0.0776	
			(0.0899)				(0.247)	
Industry FEs	Y	Y	Y	Y	Y	Y	Y	Y
Geographic Advantages	N	Y	N	Y	N	Y	N	Y
Province FEs	N	N	N	Y	N	N	N	Y
Observations	22,565	22,535	22,565	22,535	9,330	9,261	9,330	9,261
R-squared	0.316	0.317	0.317	0.329	0.214	0.218	0.214	0.271

Notes: Panels A and B contain the results for separate regressions for formal and information establishments, respectively, in all industries. The dependent variable is (logged) value-added. Capital, Labor, the number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, the number of informal firms with the same activity code in the district, the distance-weighted sum of formal firms of the industry-districts' major suppliers in other districts, the distance-weighted sum of informal firms of the industry-districts' major suppliers in other districts are measured in logs. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 5: Telecommunications and Other Information Services

	Panel A: Formal Establishments				Panel B: Informal Establishments			
VARIABLES	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Capital	0.293***	0.294***	0.293***	0.290***	0.277***	0.286***	0.277***	0.288***
	(0.0211)	(0.0206)	(0.0211)	(0.0224)	(0.0173)	(0.0194)	(0.0180)	(0.0215)
Labor	0.411***	0.410***	0.411***	0.407***	0.188***	0.173***	0.184***	0.178***
	(0.0220)	(0.0233)	(0.0221)	(0.0214)	(0.0351)	(0.0372)	(0.0371)	(0.0355)
Formal Suppliers	0.0584	0.0549	0.0519	0.110*	0.0113	0.0052	0.0287	-0.0006
	(0.0561)	(0.0581)	(0.0588)	(0.0627)	(0.0526)	(0.0555)	(0.0492)	(0.0999)
Informal Suppliers	0.152***	0.149***	0.156***	0.0710	0.326***	0.311***	0.305***	0.134
	(0.0511)	(0.0517)	(0.0528)	(0.0469)	(0.0901)	(0.0756)	(0.0997)	(0.0986)
Shared-Skills Labor	0.0678	0.0695	0.0722	0.0499	-0.0475	-0.0579	-0.0447	0.0181
in Formal Firms	(0.0430)	(0.0455)	(0.0446)	(0.0572)	(0.0691)	(0.0690)	(0.0643)	(0.121)
Shared-Skills Labor	-0.125***	-0.127***	-0.126***	-0.0724	-0.105	-0.0652	-0.101	-0.00366
In Informal Firms	(0.0392)	(0.0410)	(0.0400)	(0.0541)	(0.0813)	(0.0703)	(0.0732)	(0.0754)
Formal Own Ind.	-0.153***	-0.139***	-0.152***	-0.174***	-0.0269	-0.0300	-0.0494	-0.0200
	(0.0409)	(0.0447)	(0.0405)	(0.0530)	(0.0539)	(0.0340)	(0.0495)	(0.0722)
Informal Own Ind.	0.0002	0.0030	-0.0030	0.0241	-0.125*	-0.117*	-0.116*	-0.0495
	(0.0373)	(0.0389)	(0.0376)	(0.0539)	(0.0609)	(0.0603)	(0.0649)	(0.113)
Formal Supplier Access			-0.0577				-0.0137	
			(0.0967)				(0.295)	
Informal Supplier Access			0.0639				0.0430	
			(0.110)				(0.326)	
Industry FEs	Y	Y	Y	Y	Y	Y	Y	Y
Geographic Advantages	N	Y	N	Y	N	Y	N	Y
Province FEs	N	N	N	Y	N	N	N	Y
Observations	11,725	11,710	11,725	11,710	5,062	5,032	5,062	5,032
R-squared	0.153	0.154	0.153	0.174	0.148	0.152	0.148	0.196

Notes: Panels A and B contain the results for separate regressions for formal and information establishments, respectively, in all industries. The dependent variable is (logged) value-added. Capital, Labor, the number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, the number of informal firms with the same activity code in the district, the distance-weighted sum of formal firms of the industry-districts' major suppliers in other districts, the distance-weighted sum of informal firms of the industry-districts' major suppliers in other districts are measured in logs. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 6: Business Services

	Panel A: Formal Establishments				Panel B: Informal Establishments			
VARIABLES	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Capital	0.321***	0.324***	0.321***	0.321***	0.331***	0.328***	0.329***	0.310***
	(0.00926)	(0.00928)	(0.00927)	(0.00751)	(0.0301)	(0.0282)	(0.0303)	(0.0252)
Labor	0.619***	0.618***	0.618***	0.622***	0.338***	0.354***	0.348***	0.361***
	(0.0255)	(0.0264)	(0.0253)	(0.0257)	(0.0560)	(0.0551)	(0.0537)	(0.0598)
Formal Suppliers	-0.221***	-0.292***	-0.210***	-0.273***	-0.0514	0.104	0.0151	0.328**
	(0.0724)	(0.0662)	(0.0671)	(0.0345)	(0.202)	(0.134)	(0.173)	(0.147)
Informal Suppliers	0.214**	0.248***	0.204**	0.129**	0.0898	0.0633	0.0951	-0.197**
	(0.0786)	(0.0797)	(0.0735)	(0.0556)	(0.172)	(0.129)	(0.137)	(0.0919)
Shared-Skills Labor	0.188***	0.215***	0.183***	0.288***	0.0915	0.0541	0.0625	0.000839
in Formal Firms	(0.0659)	(0.0638)	(0.0645)	(0.0443)	(0.0973)	(0.0856)	(0.0909)	(0.0706)
Shared-Skills Labor	-0.231***	-0.243***	-0.224***	-0.192***	-0.125	-0.169	-0.113	-0.134**
In Informal Firms	(0.0486)	(0.0509)	(0.0469)	(0.0295)	(0.0880)	(0.0993)	(0.0854)	(0.0646)
Formal Own Ind.	0.0572	0.0520	0.0573	0.0429	-0.0872	-0.106*	-0.0890	-0.0893
	(0.0377)	(0.0398)	(0.0391)	(0.0406)	(0.0559)	(0.0569)	(0.0585)	(0.0617)
Informal Own Ind.	0.00094	0.0133	0.0020	0.0019	0.192***	0.193***	0.172***	0.149***
	(0.0233)	(0.0230)	(0.0232)	(0.0291)	(0.0624)	(0.0517)	(0.0562)	(0.0527)
Formal Supplier Access			0.0894				0.142	
			(0.135)				(0.328)	
Informal Supplier Access			-0.104				-0.243	
			(0.150)				(0.395)	
Industry FEs	Y	Y	Y	Y	Y	Y	Y	Y
Geographic Advantages	N	Y	N	Y	N	Y	N	Y
Province FEs	N	N	N	Y	N	N	N	Y
Observations	13,781	13,775	13,781	13,775	2,527	2,519	2,527	2,519
R-squared	0.385	0.389	0.385	0.401	0.159	0.170	0.161	0.238

Notes: Panels A and B contain the results for separate regressions for formal and information establishments, respectively, in all industries. The dependent variable is (logged) value-added. Capital, Labor, the number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, the number of informal firms with the same activity code in the district, the distance-weighted sum of formal firms of the industry-districts' major suppliers in other districts, the distance-weighted sum of informal firms of the industry-districts' major suppliers in other districts are measured in logs. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 7: Labor Productivity Estimates

	Panel A: All		Panel B: Manufacturing		Panel C: Telecom.		Panel D: Business Services	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Formal Suppliers	-0.0372	-0.0137	-0.0458	-0.0050	0.107**	0.0612	-0.269***	-0.177
	(0.0235)	(0.0608)	(0.0602)	(0.0633)	(0.0481)	(0.0730)	(0.0924)	(0.202)
Informal Suppliers	0.122***	0.235***	0.109***	0.112*	0.236***	0.377***	0.244***	0.305*
	(0.0288)	(0.0547)	(0.0273)	(0.0618)	(0.0658)	(0.0825)	(0.0719)	(0.173)
Shared-Skills Labor	0.183***	-0.0137	0.109	0.000713	0.0112	-0.0408	0.313***	0.163*
in Formal Firms	(0.0370)	(0.0766)	(0.0909)	(0.0620)	(0.0541)	(0.0824)	(0.0721)	(0.0847)
Shared-Skills Labor	-0.204***	-0.183***	-0.0842***	-0.0434	-0.210***	-0.245***	-0.322***	-0.304***
In Informal Firms	(0.0325)	(0.0608)	(0.0261)	(0.0477)	(0.0609)	(0.0558)	(0.0458)	(0.0774)
Formal Own Ind.	-0.0548**	0.00504	0.0312	0.0571	-0.158***	-0.130***	0.0339	-0.0646
	(0.0251)	(0.0588)	(0.0372)	(0.0495)	(0.0410)	(0.0349)	(0.0366)	(0.0414)
Informal Own Ind.	0.0208	0.0466	-0.0455*	-0.0734	-0.00181	-0.0120	-0.0178	0.0967
	(0.0148)	(0.0500)	(0.0236)	(0.0528)	(0.0386)	(0.0602)	(0.0223)	(0.0584)
Observations	345,854	227,199	28,429	14,308	15,575	8,297	16,929	4,204
R-squared	0.026	0.044	0.020	0.047	0.011	0.014	0.071	0.051

Notes: Panels A-D contain the results for separate regressions using establishments in all, manufacturing, telecommunications and other information services, and business service industries, respectively. Within in each panel, Columns (1) and (2) contain the results for formal and informal establishments, respectively. The dependent variable is labor productivity measured as value-added per employee. The number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, the number of informal firms with the same activity code in the district are measured in logs. All specifications include 92 level industry fixed effects. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 8: Instrumenting for Agglomeration Sources

	Panel A: All		Panel B: Manufacturing		Panel C: Telecom.		Panel D: Business Services	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Formal Suppliers	-0.140***	-0.113	0.0830	0.0353	-0.0321	0.0116	-0.347***	-0.259
	(0.0472)	(0.141)	(0.237)	(0.179)	(0.186)	(0.265)	(0.126)	(0.360)
Informal Suppliers	0.133**	0.0639	0.506**	0.147	0.162**	0.556***	0.302***	0.0680
	(0.0611)	(0.130)	(0.207)	(0.150)	(0.0731)	(0.201)	(0.112)	(0.218)
Shared-Skills Labor	0.283***	0.154	-0.0848	-0.0877	0.158	-0.111	0.265***	0.237
in Formal Firms	(0.0530)	(0.177)	(0.300)	(0.166)	(0.181)	(0.263)	(0.0667)	(0.253)
Shared-Skills Labor	-0.176***	0.000870	-0.437***	-0.0228	-0.139**	-0.271	-0.260***	-0.00423
In Informal Firms	(0.0501)	(0.165)	(0.154)	(0.130)	(0.0671)	(0.181)	(0.0666)	(0.156)
Formal Own Ind.	-0.105***	-0.0105	-0.0272	0.0440	-0.158***	-0.0144	0.0626	-0.0751
	(0.0201)	(0.0606)	(0.0385)	(0.0326)	(0.0488)	(0.0607)	(0.0419)	(0.0741)
Informal Own Ind.	0.0470***	-0.00718	0.0227	-0.0132	0.00742	-0.130*	-0.00961	0.132
	(0.0179)	(0.0702)	(0.0269)	(0.0377)	(0.0351)	(0.0665)	(0.0353)	(0.0923)
F-Stat	120.127	8.795	0.581	14.619	9.606	3.555	11.196	23.397
Observations	269,528	159,565	22,565	9,330	11,725	5,062	13,781	2,527
R-squared	0.251	0.126	0.281	0.181	0.151	0.142	0.347	0.122

Notes: Panels A-D contain the results for separate regressions using establishments in all, manufacturing, telecommunications and other information services, and business service industries, respectively. Within in each panel, Columns (1) and (2) contain the results for formal and informal establishments, respectively. The dependent variable is (logged) value-added. The number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, and the number of informal firms with the same activity code in the district are measured in logs. Supplier and labor measures have been instrumented using Colombian data and the Kleibergen-Paap rk F statistic is reported at the bottom of each column. All specifications include 92 level industry fixed effects. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.

Table 9: Levinsohn-Petrin Instrumental Variable Estimator
Manufacturing, Formal Establishments

VARIABLES	(1)	(2)	(3)	(4)
Capital	0.0768***	0.0769***	0.0769***	0.0760***
	(0.00319)	(0.00304)	(0.00328)	(0.00360)
Formal Suppliers	-0.272**	-0.190	-0.225*	-0.145
	(0.114)	(0.136)	(0.126)	(0.138)
Informal Suppliers	0.0962**	0.0503	0.0712*	-0.0270
	(0.0415)	(0.0453)	(0.0350)	(0.0542)
Shared-Skills Labor	0.303**	0.258	0.270*	0.407***
in Formal Firms	(0.131)	(0.160)	(0.152)	(0.101)
Shared-Skills Labor	-0.0989**	-0.0887*	-0.0874*	-0.124***
In Informal Firms	(0.0417)	(0.0481)	(0.0444)	(0.0367)
Formal Own Ind.	-0.0151	-0.0229	-0.0191	-0.0512
	(0.0521)	(0.0484)	(0.0537)	(0.0440)
Informal Own Ind.	0.00413	0.00851	0.0135	0.00885
	(0.0507)	(0.0434)	(0.0468)	(0.0423)
Formal Supplier Access			0.302**	
			(0.135)	
Informal Supplier Access			-0.228**	
			(0.108)	
Industry FEs	Y	Y	Y	Y
Geographic Advantages	N	Y	N	Y
Province FEs	N	N	N	Y
Observations	11,388	11,388	11,388	11,388
R-squared	0.063	0.064	0.064	0.069

Notes: Columns 1 to 4 contain the results for separate regressions for the second stage of the LP estimator using formal establishments in the manufacturing sector. Capital, Labor, the number of district formal firms in the top twenty-five percent of activity codes that provides inputs to the firm's industry, the number of district informal firms in the top twenty-five percent of activity codes that provides to the firm's industry, the number of workers in formal firms in the district that shares similar skills, the number of workers in informal firms in the district that shares similar skills, the number of formal firms with the same activity code in the district, and the number of informal firms with the same activity code in the district are measured in logs. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the Department level are in parentheses.