

Dapena, Alberto Díaz; Vázquez, Esteban Fernández; Morollón, Fernando Rubiera

**Conference Paper**

## Labor density and productivity in Spain: Evidence from geographically disaggregated data

55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Dapena, Alberto Díaz; Vázquez, Esteban Fernández; Morollón, Fernando Rubiera (2015) : Labor density and productivity in Spain: Evidence from geographically disaggregated data, 55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124614>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Labor density and productivity in Spain: evidence from geographically disaggregated data

Alberto Díaz Dapena  
Esteban Fernández Vázquez\*  
Fernando Rubiera Morollón

REGIOlab and Department of Applied Economics

University of Oviedo (Spain)

*\*Corresponding author*

## Abstract

In this paper we estimate agglomeration economies in Spain in 2009 basing on Ciccone's (2002) model, which explains average labor productivity in one spatial unit on employment density and other controls. The novelty of our analysis is that the empirical model is estimated at a highly disaggregated spatial scale, oppositely of taking as unit of analysis NUTS-2 or NUTS-3 regions. We argue that working at this level of spatial disaggregation is more coherent with the idea of externalities from agglomerations, which are generated at a local scale. This is especially relevant in the case of Spain because the NUTS-3 (provinces) are large regions in comparison with other cases in Europe. From a sample of income-taxpayers published by the Spanish Fiscal Studies Institute (IEF, *Instituto de Estudios Fiscales*) we derive figures on average wages by worker at the scale of Local Labor Markets (LLMs). The empirical analysis bases on several estimation strategies; namely, Ordinary Least Squares (OLS), Two-Stage Least Squares (2SLS), Quantile Regressions (QR) and Instrumental Variable Quantile Regressions (IVQR) estimators, all they finding a significantly positive effect of agglomeration in the conditional mean of labor productivity. Additionally, the QR and IVQR estimators find a progressively decreasing, but still positive, effect of employment density along the conditional distribution of labor productivity.

**Keywords:** Agglomeration economies, labor productivity, density, spatially disaggregated data and Spain.

**JEL Classification:** D24, J31, R10 and R12.

## 1. Introduction

A well-established empirical fact is that firms and workers are more productive on average in denser areas. Basing on the classic Marshall's ideas of agglomeration economies derived by the location of one firm close to others, recent literature has paid attention to the quantification of the impact of agglomeration economies on productivity (see, for example, Rosenthal and Strange, 2004, for an extensive review or Ciccone and Hall, 1996; Combes, 2000, Combes et al., 2008 or Artis et al., 2012).

The estimation strategy followed by Ciccone (2002) may be considered as one of the most popular ways of measuring agglomerations economies effect on local/regional productivity by estimating the effect of employment density on the generation of spatial externalities. More specifically, in Ciccone (2002) a model on which average labor productivity in one area depends on labor density –defined as labor units by unit of land– is derived. The empirical estimations of this model for the cases of Germany, UK, France, Italy and Spain find that the elasticity of labor productivity with respect to employment density is within the range of 4.5 and 5 percent, under several specifications at the scale of NUTS-3 regions.<sup>1</sup>

One potential issue in the measurement of spatial externalities on productivity is the geographical scale at which the empirical estimation of agglomeration economies is conducted. Ciccone (2002) argues that NUTS-3 regions is an appropriate spatial scale, since their median size in the set of countries studied is 1,511 km<sup>2</sup>, which is slightly smaller than the median size of U.S. counties. However, administrative NUTS-3 division can be considered a highly aggregated spatial scale for the case of some countries. This is the case of Spain, which is divided into 50 NUTS-3 regions (Spanish *Provincias*) with sizes ranging from less than 2,000 km<sup>2</sup> to more than 20,000 km<sup>2</sup>–the median size is 9,998 km<sup>2</sup>–. This relatively large size of the regional units can be hiding a potential heterogeneity within the regions that can be conditioning the results of empirical model: by assuming the same average productivity and density figures the potential intra-regional heterogeneity is neglected.

---

<sup>1</sup> NUTS is the acronym of Nomenclature of Territorial Units for Statistics in French.

The literature on the empirical quantification of agglomeration economies in Spain is not vast. Alonso-Villar et al. (2004) measured agglomeration economies for the manufacturing industries between 1993 and 1999 at the level of NUTS-2 and NUTS-3 regions, finding significant inter-industry differences in the scope of agglomerations and a positive correlation between their size with the technological intensity of the industries. Martinez-Galarraga et al. (2010) studied the productivity of industrial labor in Spain during the period 1860-1999, basing on the same estimation strategy as in Ciccone (2002) and taking NUTS-3 regions as units of analysis. They found that the elasticity with respect to employment density ranged between slightly less than 2% to more than 8%, depending on the time period and the estimator applied. More recently Jofre-Monseny (2009) conducted a similar exercise but for the specific case of Catalonia –a NUTS-2 Spanish region– for the period 1995-2002, finding agglomeration elasticities ranging between insignificant to more than 7% depending on the specific branch of the manufacturing industry. Oppositely to the previously mentioned papers, they base on information highly disaggregated at the spatial level –microdata at the scale of establishments– from data not publically available on registers in the Spanish National Social Security Registry. Alañón-Pardo and Arauzo-Carod (2013) also study the agglomeration effect, but they focus on the effect over the locations decisions. Their analysis highlights the agglomeration effects, accessibility and the spatial interactions between municipalities in the locations decisions.

In this paper we estimate agglomeration economies in the spirit of Ciccone (2002), explaining average labor productivity in one spatial unit on its employment density. The novelty of the paper is that, instead of estimating our empirical model at the level of NUTS-2 or NUTS-3 regions, we base our analysis on more disaggregated spatial data. Specifically, we take Income-tax microdata compiled by the Spanish Fiscal Studies Institute (*Instituto de Estudios Fiscales*) for calculating average compensations by worker at the scale of Local Labor Market (LLM), as defined in Boix and Galleto (2008). Our claim is that taking highly disaggregated geographical units allows for considering an appropriate spatial scale to measure agglomeration economies, since

spatially aggregated data implies assuming a high level of intra-regional homogeneity.

The rest of the article is organized as follows. Section 2 presents a brief overview of the approach developed in Ciccone (2002) to measure agglomeration economies, which will be later applied to our database. Section 3 details the characteristics of the database, as well as describe the procedure followed to classify the information at the scale of LLMs. Section 4 presents the empirics of the model under different estimation strategies and section 5 discusses the results. Finally, section 6 concludes the paper.

## 2. The model: density and spatial externalities on productivity

Ciccone (2002), basing on Ciccone and Hall (1996), proposes a model with spatial externalities on productivity caused by the economic density of the territory. The point of departure is the following equation explaining the production by unit of land in the geographical unit or sub-region  $s$  that belongs to a larger region  $c$ :

$$q = \Omega_s f(nH, k, Q_{sc}, A_{sc}) = \Omega_s ((nH)^\beta k^{1-\beta})^\alpha \frac{Q_{sc}}{A_{sc}}^{\frac{\lambda-1}{\lambda}} \quad [1]$$

where  $q$  stands for the output per unit of land,  $\Omega_{sc}$  is an index of total factor productivity in the area,  $A_{sc}$  is the total surface,  $n$  denotes economic density,  $H$  is the average level of human capital of workers per unit of land, and  $k$  stands for the density of physical capital. Parameter  $\alpha$  captures the returns of capital and labor, whereas  $\beta$  is a distribution parameter. The empirical specification of [1] assumes that spatial externalities are driven by the density of production in the area  $Q_{sc} A_{sc}$ , where  $\frac{\lambda-1}{\lambda}$  represents the elasticity of output per unit of land with respect to economic density. In this specification, there are spatial externalities when  $\lambda > 1$ .

Some transformations are required in order to have an estimable version of [1]. Assuming that the distribution of labor and capital is uniform within each spatial unit

s across  $c$ , aggregate production may be written as  $Q_{sc} = A_{sc}q$ . Defining  $N_{sc}$  and  $K_{sc}$  as the levels of employment and physical capital in  $s$ , respectively, and assuming that the demand function of capital follows the expression:

$$K_{sc} = \frac{\alpha(1 - \beta)}{r_c} Q_{sc} \quad [2]$$

where  $r_c$  stands for the price of capital that is assumed constant in every sub-region  $s$  within the large region  $c$ . Under this assumption, labor productivity ( $Q_{sc} / N_{sc}$ ) is given by:

$$\frac{Q_{sc}}{N_{sc}} = \Lambda_c \Omega_{sc}^\omega H_{sc} \frac{N_{sc} H_{sc}}{A_{sc}}^\theta \quad [3]$$

In [3],  $\omega$  is a constant and  $\Lambda_c$  depends on the rental price of capital and is assumed common for all the geographical units within  $c$ . Moreover,  $\theta$  is defined as:

$$\theta = \frac{\alpha\lambda - 1}{1 - \alpha\lambda} \frac{1}{1 - \beta} \quad [4]$$

Parameter  $\theta$  measures the effect on labor productivity of the density of employment in sub-region  $s$ . The value of  $\theta$  in equation [3] can be estimated from data on production, employment density and human capital by assuming that differences in  $\Lambda_c$  across large regions are captured by dummy variables at the level of these larger regions.

By taking logarithms in [3] and including dummy variables for large regions, the final equation to be estimated is:

$$\log \frac{Q_{sc}}{N_{sc}} = \text{Large region dummies} + \theta \log \frac{N_{sc}}{A_{sc}} + \gamma \log H_{sc} + u_{sc} \quad [5]$$

Equation [5] relates labor productivity ( $Q_{sc}$   $N_{sc}$ ) to employment density ( $N_{sc}$   $A_{sc}$ ) in the spatial unit  $s$ , controlling by the effect of the stock of human capital ( $H_{sc}$ ) by means of parameter  $\gamma$ , dummies that account for differences in total factor productivity and rental prices of capital between large regions and a disturbance term  $u_{sc}$ .

### **3. Database: fiscal data for Local Labor Markets in Spain (2009)**

The empirical work draws on data on employment density and indicators of human capital and labor productivity at a spatial scale more disaggregated than NUTS-3 regions. Estimating equations like [5] from data collected at the scale of NUTS-3 administrative regions can imply working at a too highly aggregated scale, since average indicators of labor productivity or employment density can be hiding large intra-regional heterogeneity. This could be an issue, especially for those NUTS-3 regions on which the largest Spanish cities are located.<sup>2</sup>

With the purpose of avoiding this problem, the data to estimate [5] are taken from a database that allows a more detailed spatial disaggregation. In particular, we base on microdata at the individual level in a cross-sectional sample of income-taxpayers published on a yearly basis by the Spanish Fiscal Studies Institute (*Instituto de Estudios Fiscales*), an institution dependent on the Spanish Ministry of Economy. This database provides information on wages reported on their income-tax declarations by the sample of individuals. The micro-data released from the sample in 2009 of approximately 549,000 individuals has been analyzed and taken as the main source of information for this study. One disadvantage of the database is that it does not provide data on variables as education level or years of tenure, for example, which could be useful when controlling for individual characteristics. On the other hand, it allows for deriving average indicators of labor productivity and employment density at a highly disaggregated spatial scale, more specifically, at the level of Spanish ZIP

---

<sup>2</sup> As an example, the NUTS-3 province of Madrid is divided into 179 municipalities. Data from the 2011 census on population density (employment figures at municipal scale were not published in the 2011 census) show huge disparities on this variable: the average population density in the province was approximately 800 inhabitant per km<sup>2</sup>, but at municipal level population densities ranged from less than 2 to more than 7,000.

codes. Since the model takes as unit of analysis spatial sub-regions  $s$ , this database is specially convenient for estimating models like [5] for Spain.

Wage reported in the sample is taken as indicator of labor productivity, and average wage figures can be derived at the scale of Zip codes.<sup>3</sup> Similarly, it is possible to derive employment figures at the same level and then they can be aggregated at the desired spatial scale. Indicators of human capital, however, are not available at this scale, which prevents the use of ZIP codes areas as the spatial unit of analysis. The most detailed spatial classification to estimate [5] is at the scale of municipalities, since the Housing and Population Census publishes information on the academic level of workers at municipal level.<sup>4</sup> Information on the surface of municipalities is available in the Housing and Population Census as well.

Even when our databases will allow us to take municipalities as the spatial sub-regions  $s$  in our model, we opted for aggregating these areas into larger spatial units for several reasons. One is the huge number of municipalities present in the Spanish spatial configuration –more than 8,000–. Consequently, for many of them the number of individuals sampled is too small to have reliable estimates of the variables of interest. Secondly, the sample of income tax-payers provides information on their place of residence, not the place where these individuals work, and the labor density should be referred to the place where economic activity is located. For these reasons, ZIP codes are aggregated at the level of Local Labor Markets (LLMs).

LLMs are analytical areas resulting from aggregating municipalities among which the commuting flows are especially intense. A LLM is a group of municipalities designed to maximize flows of commuting intra-LLMs and, conversely, commuting flows

---

<sup>3</sup> Ciccone (2002) bases his study on data on value added, which can be considered a better indicator of labor productivity. However, this variable is not observable at the desired spatial scale and wages are taken instead. See Combes et al. (2011) or Melo et al. (2009) for examples of previous research when this approach is followed.

<sup>4</sup> The 2011 census has not released information on educational levels of workers for all the municipalities, which prevents using 2011 data. The census conducted in 2001, however, released this information and will be the basis for recovering indicators of human capital in our estimations.



between LLMS are minimized. The specific procedure for defining LLMS applied in this paper corresponds to the definition given by the Italian Statistical Agency (ISTAT) and applied later by Boix and Galleto (2008) for Spain. This technique groups contiguous municipalities with the condition that at least 75% of people living within a LLM work there as well.<sup>5</sup> Consequently, the individuals in the sample are assigned to some of the 763 LLMS on which the Spanish territory is classified. The 763 areas do not cover all the Spanish territory because individuals paying their taxes in Basque Country and Navarra are not sampled, since these NUTS-2 regions have their own fiscal system, the so-called *Haciendas Forales*. The full set of variables, their definitions and sources and a summary of descriptive statistics is set out in Table 1.

<<Insert Table 1 about here>>

#### 4. Estimation

Equation [5] is estimated from information contained in the previously described databases. Average wages by LLM ( $Q_{sc}$   $N_{sc}$ ) are derived from the sample of income-taxpayers; human capital ( $H_{sc}$ ) defined as the fraction of workers with a secondary or a university degree is extracted from the 2001 Housing and Population Census; and the indicator of labor density ( $N_{sc}$   $A_{sc}$ ) is derived by combining both statistical sources. Each and every LLM is assigned to one larger region  $c$  defined at the scale of the 15 NUTS-2 regions sampled in the survey.<sup>6</sup> This could be problematic if LLMS were formed by grouping municipalities belonging to different NUTS-2 regions. In practical terms, however, this is not an issue since there are only a few cases –41 out of 763 LLMS– where this problem happens. Anyway, it has been solved by assigning that specific LLM to the NUTS-2 region on which most of its population is located.

---

<sup>5</sup> Details on the specific algorithm used by these authors can be found in Boix and Galleto (2008).

<sup>6</sup> The dummies variables are introduced at NUTS-2 level because this is the level in which the autonomous government (Spanish Autonomous Communities) works. At this level the Autonomous Communities have independency to carry out social, educational and health care policies among others. Provinces (NUTS-3 level) are in Spain an administrative division without political independency.

With these considerations in mind, the final equation to be estimated is:

$$\log \frac{Q_{sc}}{N_{sc}} = NUTS-2 \text{ dummies} + \theta \log \frac{N_{sc}}{A_{sc}} + \gamma \log H_{sc} + u_{sc} \quad [6]$$

In this section we describe several approaches to estimate parameter  $\theta$  in [6]. The simplest procedure consists of an ordinary least-squares (OLS) estimation from the 763 LLMs. However, an endogeneity problem caused by reverse causality can emerge if more productive LLMs attract more workers by unit of land (see Graham et al., 2010). Consequently, the OLS estimator of [6] would become inconsistent. Ciccone (2002) address endogeneity by adopting a Two Stage Least Squares (2SLS) estimator, where employment densities of the European regions analyzed are instrumented by their total land area. The argument for this choice is that land area is a variable historically predetermined and not conditioned by current productivities. This instrument would not be valid in our estimation, since LLMs are constructed grouping municipalities strongly interconnected by commuting flows, so the total land area of one LLM is not exogenous but determined by the economic characteristics of the municipalities.

Alternatively, we follow the approach of Ciccone and Hall (1996), Rice et al. (2006), Graham and Kim (2008) or Artis et al. (2012), where current levels of density are instrumented by long lags of density. The justification is that modern densities are conditioned by past densities, being these not correlated with current productivities. Applying this approach to our problem requires data on historical densities at the spatial scale of LLMs. From the 1950 Spanish Housing and Population Census, which is the oldest one providing information on population densities at a municipal level, we recover the data necessary to define our instrument and [6] is estimated by 2SLS.<sup>7</sup> Both OLS and 2SLS approaches are estimated in their respective versions robust to heterocedasticity.

---

<sup>7</sup> Data on labor density is not available at this geographical scale until the 1981 census. This forces us to take as instrument population density instead.

Additionally to OLS and 2SLS estimations, Quantile Regression (QR) estimations of [6] have been obtained as well. This estimation strategy has been previously applied in the context of quantifying agglomeration economies, as in Combes et al. (2009) or Brian (2010). Developed by Koenker and Basset (1978), the QR approach allows for estimating a coefficient for each conditional quantile of the dependent variable, not only for its conditional mean like in OLS and 2SLS estimators (see Koenker, 2005, for a more recent overview). The coefficient estimates by QR show the reaction at different points in the conditional distribution of the dependent variable to changes in the regressors. Estimating a modified version of [6] by QR will assess the conditional effect of labor density at several quantiles of the distribution of labor productivity. Denoting as  $\tau$  these quantiles, the QR equations to be estimated will be:

$$\log \frac{Q_{sc}}{N_{sc}}_{\tau} = NUTS-2 \text{ dummies}_{\tau} + \theta_{\tau} \log \frac{N_{sc}}{A_{sc}} + \gamma_{\tau} \log H_{sc} + u_{sc\tau} \quad [7]$$

where the coefficients  $\theta_{\tau}$  and  $\gamma_{\tau}$  measure respectively the effect to labor density and human capital at the  $\tau$ th quantile of labor productivity. QR estimates can be affected by the same endogeneity problems commented for the case of OLS. The solution to this issue lies in applying the Instrumental Variable Quantile Regression (IVQR) estimator developed by Chernozhukov and Hansen (2005, 2006). The IVQR estimation of [7] is based on the same instrument –1950 population densities– for current labor densities as in the 2SLS estimation.

## 5. Results

Tables 2 and 3 summarize the estimates of the four estimation strategies described in the previous section.<sup>8</sup> Table 1 reports the OLS estimation of [6] in the first column,

---

<sup>8</sup> Ciccone (2002) extends its model –see page 219– by including externalities across regions derived from a spatial autoregressive process: higher productivities in neighboring regions could increase the own productivity in one area. The self-contained nature of the LLMs in our analysis excludes theoretically the presence of these spatial effects, since commuting flows between two different LLMs are close to zero. However, a Moran’s test has been conducted to test for spatial autocorrelation in labor productivity, basing on a distance based and a binary contiguity matrix among LLMs. The respective Moran’s-I statistics were -0.001 and 0.01, not rejecting in neither case the null hypothesis of absence of spatial autocorrelation.

together with QR estimates of [7] at quantiles 10th, 25th, 50th, 75th and 90th in columns 2 to 6.

**<<Insert Table 2 about here>>**

The OLS estimate of  $\theta$  is significantly different from zero, being the point estimate equal to 3.23 percent. This result is similar to the estimates found by Artis et al. (2012) for the British counties in the period 2001-2005, being their results ranging between 3.9 and 4.2 percent depending on the estimator used. Human capital seems to contribute significantly to labor productivity as well, with an impact on the dependent variable of approximately 1.6 percent. The estimates of coefficients associated to NUTS-2 dummies suggest that regional differences in productivity are significant: taking the region of Madrid as reference and controlling for the effect of density and human capital, LLMs located in the rest of NUTS-2 regions have in mean labor productivities significantly smaller than those located in Madrid, ranging these differences from around 8% –in the case of Castilla La-Mancha, a region contiguous to Madrid– to more than 25% for the case of Cantabria –a more rural and peripheral region–.

Estimates from QR estimation in columns 2 to 6 show the impact of labor density along the conditional distribution of labor productivity. Estimates of  $\theta_\tau$  quantify the change in the conditional labor productivity quantile caused by a shift in LLM employment density. A first remarkable result is that the effect of employment density is estimated to be significantly positive at any of the quantiles reported. Additionally, the effect of changes in labor density on productivity is progressively decreasing with quantiles: lower quantiles of labor productivity are more sensitive to changes in density, ranging its impact from more than 4.5 percent in the 10th quantile to less than 2.5 percent in the 90th. This result can be interpreted as a signal that LLMs at the upper-end of the conditional distribution of labor productivity benefit less than those at the lower-end of the distribution from a shift in employment density.

The effect of human capital on productivities of LLMs is significantly positive as well, being its effect remarkably stable along the conditional distribution in the range of 1.4 to 1.7 percent. The estimates of NUTS-2 dummies in QR show that regional differences in productivity are much more sizeable in the lower-end of the conditional distribution than in the upper-end: the negative effect for a LLM of being located out of Madrid region is much bigger in the lower part of the distribution of productivity –more than 40% for several cases in the 10th quantile– than in the upper part, where this effect becomes statically non-significant for one half of the regions in the 90th quantile.

To avoid potential endogeneity problems in the OLS and QR estimation, Instrumental Variable (IV)-based versions of the previous estimators have been applied as well, taking as instrument of the modern levels of density population densities in 1950. The auxiliary regression of employment density on the rest of regressors and the log of 1950 population density is useful to test the quality of this instrument. The partial  $R^2$  statistic measuring the correlation between  $\log(Q_{sc} N_{sc})$  and the log of 1950 population density after partialling out the effect of the other regressors is larger than 75%, being the F-statistic for testing the significance of this instrument significant at 1%.

The results of the 2SLS and the IVQR estimators are reported in Table 3. Similar to Table 2, the 2SLS estimation of [6] is reported in the first column, being the IVQR estimates of [7] at quantiles 10th, 25th, 50th, 75th and 90th reported in columns 2 to 6.

**<<Insert Table 3 about here>>**

Results in Table 3 are in concordance with the non IV-based estimates in Table 1. The effect of labor density on the mean of productivity is estimated to be significantly different from zero, being the point estimate 3.11 percent very close to the OLS result. The coefficients measuring the effects of human capital and NUTS-2 dummies estimated in the first column are very similar as well to those in Table 1. Additionally, the results from the IVQR estimator show the same picture as the QR estimates: the

effect of labor density is steadily diminishing along the conditional distribution of productivity, with point estimates similar to those found in the QR estimator in Table 1. The main differences between these IV-based estimators are in the significance – not the size– of the regional dummies, being these dummies insignificant from the 50th quantile upward.

## **6. Conclusions**

In this paper we have estimated a model to quantify agglomeration economies in Spain based on highly disaggregated spatial data. In particular, we follow the model developed in Ciccone (2002), but oppositely to previous empirical research that takes NUTS-2 or NUTS-3 regions as spatial units of analysis, we base our analysis on Local Labor Markets (LLMs). We use microdata on wages reported in a sample of income taxpayers in 2009 that is disaggregated at the level of ZIP codes, together with information from the 2001 census, to calculate indicators of labor productivity, labor density and human capital at the desired spatial scale.

The estimable equation on which the empirical analysis bases is estimated by four different approaches. First, Ordinary Least Squares (OLS) and Quantile Regressions (QR) estimators are obtained for quantifying, respectively, the effect of employment density in the conditional mean and along the conditional quantiles of labor productivity. Additionally, in order to avoid endogeneity problems, Instrumental Variable (IV) versions of these estimators are applied as well. More specifically, a Two-Stage Least Squares (2SLS) and Instrumental Variable Quantile Regressions (IVQR) estimators are obtained by using as instrument of current densities population densities taken from the 1950 Spanish census.

Our empirical analysis finds a significantly positive effect of agglomeration in any of the approaches described. The effect of employment density in the mean is around 3 percent, with small differences found between the OLS and 2SLS estimators. The two quantile regression approaches, the ordinary QR and the IVQR estimator, show a similar pattern of the effect of employment density on the conditional distribution of

labor productivity. Both estimators reveal a decreasing effect –but always significantly positive– of density along the conditional quantiles of labor productivity: QR and IVQR estimates of this effect at the 10th quantile are respectively 4.56 and 3.49 percent, whereas they are estimated in 2.44 and 2.83 percent respectively at the 90th quantile.

This paper remarks the importance of the choice of the spatial unit of analysis in the empirical estimation of agglomeration economies. The combination of the two data sources described in the paper allows for estimating Ciconne's (2002) model at a highly disaggregated geographical scale, which implies theoretical and empirical advantages. From the theoretical viewpoint, taking these small spatial areas as the units of analysis is more in concordance with the recent developments of NEG, where not administrative areas as NUTS-2 or NUTS-3 regions but cities within these regions play a key role. From an empirical viewpoint, working at a more disaggregated spatial scale -when the data required is available- increases the number of units of analysis and allows for studying differential responses to shifts in density along the distribution of labor productivity, being this analysis practically impossible at the scale of NUTS-2 or NUTS-3 regions for a country like Spain.

## References

- Alañón-Pardo, Á., Arauzo-Carod, J. M., 2013, Agglomeration, accessibility and industrial location: Evidence from Spain. *Entrepreneurship and Regional Development*, 25(3-4), 135-173.
- Artis, M.J., Miguelez, E., Moreno, R., 2012. Agglomeration economies and regional intangible assets: an empirical investigation. *Journal of Economic Geography*, 12(6), 1167-1189.
- Blien, U., Hermann, G., Stüber, H., Wolf, K., 2009, Regional price levels and the agglomeration wage differential in western Germany. *The Annals of Regional Science*, 43, 71-88.
- Boix, R., Galletto, V., 2008, Marshallian industrial districts in Spain. *Scienze Regionali*, 7(3).
- Briant, A., (2010) Marshall's scale economies: a quantile regression approach. PhD dissertation, Paris School of Economics.
- Chernozhukov, V., Hansen, C., 2005, An IV model of quantile treatment effects. *Econometrica* 73: 245–261.

- Chernozhukov, V., Hansen, C., 2006, Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2), 491-525.
- Ciccone, A, 2002, Agglomeration effects in Europe. *European Economic Review*, 46, 213-227.
- Combes, P.P., 2000, Economic Structure and Local Growth: France, 1984-1993. *Journal of Urban Economics*, 47, 329-355.
- Combes, P.P., Duranton, G., Gobillon, L., 2008, Spatial wage disparities: Sorting matters!. *Journal of Urban Economics*, 63(2), 723-742.
- Combes, P.P., Duranton, G., Gobillon, L., 2011, The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253-266.
- Combes, P.P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2009. The productivity advantages of large markets: Distinguishing agglomeration from firm selection. C.E.P.R. Discussion Papers 7191, Centre for Economic Policy Research.
- Graham, D.J., Kim, H.Y., 2008, An empirical analytical framework for agglomeration economies. *The Annals of Regional Science*, 42, 267-289.
- Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., Noland, R. B., 2010, Testing for causality between productivity and agglomeration economies, *Journal of Regional Science*, 50, 935-95
- Jofre-Monseny, J. 2009, The scope of agglomeration economies: Evidence from Catalonia, *Papers in Regional Science*, 88(3), 575-590
- Koenker, R., Basset, G., 1978, Regression Quantiles. *Econometrica*, 46 (1), 33-50.
- Koenker, R., 2005, Quantile Regression, *Econometric Society Monographs*.
- Martinez-Galarraga, J., Paluzie, E., Pons, P., Tirado-Fabregat, D.A., 2008, Agglomeration and labour productivity in Spain over the long term. *Cliometrica*, 2, 195-212.
- Melo, P.C., Graham, D.J., Noland, R.B., 2009, A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39 (3), 332-342.
- Rice, P., Venables A.J, Patacchini, E., 2006, Spatial determinants of productivity: Analysis for the regions of Great Britain, *Regional Science and Urban Economics*, 36, 727-752.
- Rosenthal S., Strange W., 2004, Evidence on the nature and sources of agglomeration economies. In: Henderson V, Thisse J-F (eds) *Handbook of urban and regional economics*, vol. 3. North-Holland, Amsterdam.



**Table 1. Variable definition, sources of information and descriptive statistics**

	$Q_{sc}$	$N_{sc}$	$N_{sc}$	$A_{sc}$	$H_{sc}$
<b>Definition</b>	Average wage (€) by LLM		Number of jobs by km <sup>2</sup>		Percentage (%) of workers with secondary or university education
<b>Source</b>	Sample of income-taxpayers		Sample of income-taxpayers; Housing and Population Census		Housing and Population Census
<b>Mean</b>	14,529.08		36.30		49.65
<b>Median</b>	14,153.31		13.32		48.89
<b>Standard dev.</b>	3,499.68		85.06		9.70
<b>Min.</b>	5,883.20		0.96		28.77
<b>Max.</b>	27,369.53		1640.78		78.00

**Table 2. OLS and QR estimations**

	(1)	QR				
		(2)	(3)	(4)	(5)	(6)
		<b>OLS</b>	<b>0.10</b>	<b>0.25</b>	<b>0.50</b>	<b>0.75</b>
Labor density ( $\theta$ )	0.0332***	0.0456***	0.0405***	0.0327***	0.0238***	0.0244**
Human capital ( $\gamma$ )	0.0157***	0.0162***	0.0172***	0.0166***	0.0165***	0.0143***
NUTS-2 dummies (ref: Madrid)						
Extremadura	-0.2147***	-0.4324***	-0.2832***	-0.1813***	-0.1204**	-0.1009*
Galicia	-0.1956***	-0.4474***	-0.2570***	-0.1533***	-0.1273***	-0.1337***
Aragon	-0.2315***	-0.4189***	-0.2924***	-0.2220***	-0.1580***	-0.0850
Com. Valenciana	-0.2489***	-0.4314***	-0.3148***	-0.2120***	-0.1865***	-0.1511**
Andalusia	-0.2176***	-0.4038***	-0.2986***	-0.2019***	-0.1620***	-0.1000**
Castilla-Leon	-0.2150**	-0.3734***	-0.2773***	-0.1917***	-0.1612***	-0.1300***
Castilla-La Mancha	-0.0812***	-0.2428***	-0.1116***	-0.0501**	-0.0367	-0.0216
Canary islands	-0.1705***	-0.3755***	-0.2478***	-0.1596***	-0.1280	-0.0675
Catalonia	-0.1322***	-0.2724***	-0.2003***	-0.1187**	-0.0787**	-0.0557
Asturias	-0.1395***	-0.2545**	-0.1869***	-0.0893*	-0.1094***	-0.0566
Rioja	-0.2493***	-0.4714***	-0.2968***	-0.2305***	-0.2087**	-0.1183
Cantabria	-0.2579***	-0.3617***	-0.3030***	-0.2371***	-0.2156***	-0.2042***
Balearic Islands	-0.2264***	-0.3563***	-0.2909***	-0.2625***	-0.1823***	-0.1506**
Murcia	-0.1562***	-0.2330***	-0.1889***	-0.1501***	-0.1468***	-0.1295*
Constant	4.3562***	4.3177***	4.2610***	4.3015***	4.3687***	4.5203***
N	763	763	763	763	763	763
R <sup>2</sup>	0.5528					
Pseudo R <sup>2</sup>		0.3143	0.3462	0.3722	0.3979	0.3419

Note: \*, \*\* and \*\*\* represent estimates significantly different of zero at 10%, 5% and 1%, respectively.

Table 3. 2SLS and IVQR estimations

	(1)	IVQR				
		(2)	(3)	(4)	(5)	(6)
		2SLS	0.10	0.25	0.50	0.75
Labor density ( $\theta$ )	0.0311***	0.0349**	0.0391***	0.0339***	0.0270**	0.0283**
Human capital ( $\gamma$ )	0.0158***	0.0168***	0.0176***	0.0162***	0.0161***	0.0138***
NUTS-2 dummies (ref: Madrid)						
Extremadura	-0.2167***	-0.4282**	-0.2762**	-0.1835	-0.1182	-0.1017
Galicia	-0.1955***	-0.3845**	-0.2259*	-0.1470	-0.1219	-0.1352
Aragon	-0.2358***	-0.4303**	-0.3004**	-0.2242*	-0.1521	-0.0747
Com. Valenciana	-0.2486***	-0.4302**	-0.3092**	-0.2243*	-0.1886	-0.1470
Andalusia	-0.2182***	-0.4025**	-0.2968**	-0.2053*	-0.1586	-0.0986
Castilla-Leon	-0.2184***	-0.3837**	-0.2784**	-0.1951*	-0.1532	-0.1225
Castilla-La Mancha	-0.0833***	-0.2559	-0.1052	-0.0566	-0.0345	-0.0147
Canary islands	-0.1710***	-0.3866**	-0.2463*	-0.1749	-0.1299	-0.0656
Catalonia	-0.1337***	-0.2810*	-0.2073	-0.1188	-0.0729	-0.0595
Asturias	-0.1406***	-0.2338	-0.1823	-0.0912	-0.1093	-0.0558
Rioja	-0.2508***	-0.4632**	-0.2949**	-0.2326*	-0.1977	-0.1153
Cantabria	-0.2598***	-0.3559*	-0.3029**	-0.2409*	-0.2055	-0.2003
Balearic Islands	-0.2271***	-0.3511**	-0.2948**	-0.2613**	-0.1800	-0.1496
Murcia	-0.1559***	-0.2779	-0.1892	-0.1320	-0.1347	-0.1232
Constant	4.3577***	4.2878***	4.2350***	4.3223***	4.3855***	4.5434***
N	763	763	763	763	763	763
R <sup>2</sup>	0.5530					
F(1,746)	1198.48***					

Note: \*, \*\* and \*\*\* represent estimates significantly different of zero at 10%, 5% and 1%, respectively. F(1,746) represents the F-statistic for the auxiliary regression of  $\log(Q_{sc} N_{sc})$  on the log of 1950 population density and the rest of regressors in [6].