The impact of Brazilian clusters on local development: a propensity score matching approach

Abstract: The theme of the article is the impact of Brazilian clusters on local development: a propensity score matching approach. The objective is to determine if there are improvements on the local development of municipalities with clusters compared to those without clusters. At the same time, the objective is to provide the explanatory factors of clustering in a territory. The methodology is Propensity Score Matching technique (ROSEMBAUM; ROBIN, 1983). The object of study are the non-metropolitan municipalities of the São Paulo State, in Brazil. Data is collected from the SEADE Foundation and the United Nations Program for the years 2009 and 2010. The results obtained show statically significant differences in the average local development (measured by well-known indexes) between municipalities with and without clusters. Municipalities with clusters achieve a higher local development than municipalities without clusters. The present paper contributes to the scant quantitative research on the impact of clusters on local development.

Keywords: cluster; local development; propensity score matching.

Resumen: El tema de este trabajo es el impacto de las Aglomeraciones Productivas Locales (APL) de Brasil en el desarrollo local: una aproximación basada en la Técnica de Emparejamiento. El objetivo es determinar si existen mejoras en el desarrollo local de los municipios con APL en comparación con los municipios sin APL. Al mismo tiempo, analizar los factores explicativos de la conformación de un APL en un territorio. La metodología consiste en la Técnica de Emparejamiento, Matching (ROSEMBAUM; ROBIN, 1983). El objeto de estudio son los municipios no metropolitanos del estado de São Paulo, Brasil. Los datos se obtienen de la Fundación SEADE y del Programa de las Naciones Unidas para los años 2009 y 2010. Los resultados obtenidos muestran que existen diferencias estadísticamente significativas en el desarrollo local promedio (medido por indicadores reconocidos) entre los municipios con y sin APL. El desarrollo local es en promedio superior en los municipios con APL. El presente estudio brinda un aporte a la escasa investigación de carácter cuantitativo en torno al impacto de los APL en el desarrollo local.

Palabras clave: APL; desarrollo local, técnica de emparejamiento.
1 INTRODUCTION

Academics and policy makers have evidenced an increased interest in the study of clusters during the 1990s. One of the main reasons is the presumed impact of clusters on firm performance, regional economic development, and country competitiveness. Several studies have stressed the positive impact of clusters on firm performance as enough evidence of their contribution to development (CARVALHO; CARRARO; SHIKIDA, 2016; KUKALIS, 2010; FOLTA; COOPER; BAIK, 2006; ROCHA, 2004; MASKELL; KEBIR 2004); by assuming that firm-level outcomes directly lead to regional or national outcomes. On the other side, only a few studies have identified a non-clear (and even negative) effect of clustering on productivity (DURANTON et al., 2011).

However, important clusters in terms of productivity have not always led to increases in local development (JACOMETTI et al., 2016). This finding can be understood by distinguishing between growth and development. Productivity growth is far from being a sufficient condition to achieve the development goals. The Latin-American case, among others, has demonstrated that growth per se does not guarantee equity or development (FAJNZYLBER, 1992). Local development means a better-living society and the improvement of quality of life. A better quality of life is achieved by taking into account other factors such as schooling, health, labour market conditions, and social mobility opportunities (ASHEIM; COOKE; MARTIN, 2006; GALLICHIO, 2004; FRANCO, 2000; ALBURQUERQUE, 1997).

The Brazilian version of a cluster is known as APL - Local Productive Arrangement. In Brazil, the Research Network for Local Productive and Innovative Systems (RedeSist) has developed the concepts of “Local Productive Arrangement” (APL) and “local productive and innovative system” in order to study the relationship between groups of firms and their linkages with other agents (economic, political and social) within a given territory (CASSIOLATO; LASTRES, 2003). APL is a local production system of small and medium-sized enterprises (SME). By definition, APLs include enterprises of all sizes, although they usually refer to small and medium-sized enterprises that are not vertically integrated. These enterprises, in turn, attract suppliers and other similar or supporting enterprises that are import for the local systems and markets. APLs are based on a set of economic, institutional, social, and political actors that jointly perform a production activity to achieve a collective goal (TEIXEIRA, 2009). In addition, APLs give rise to many local institutions and supporting enterprises (LÜBECK; WITTMANN; SILVA, 2012; SUZIGAN et al., 2004). Therefore, APL building has been a long-term historical process that has already taken 40 years in Brazil.

This study focuses on the São Paulo State due to the historic supremacy of that territory. This supremacy has emerged from three socio-economic processes: the coffee complex building, the emergence of the industrial economy (until 1939) and the effective industrialization of the state which was dominated by the coffee-industry relationship (SELINGARDI-SAMPAIO, 2009). The São Paulo State has a surface of 267,8 thousand km$^2$ (2,9% of Brazil), a population of nearly 40 million of inhabitants (approximately a fifth of the national population), an income that represents a third of Brazil’s GDP and a urbanization rate of 93,4 percent. Not only does the state have the highest productive and industrial pole of the country, but also it has the largest network of commerce and services. The state is compound of 645 municipalities distributed throughout the 42 governmental regions, 14 administrative regions and three metropolitan regions: São Paulo, Baixada Santista and Campinas.
The São Paulo state, as well as other administrative units in Brazil, cannot be treated as a homogenous entity. Selingardi-Sampaio (2009) argues there is no uniformity or homogeneity in the industrial space. For instance, the unemployment rate at municipal level is a more adequate indicator for the evaluation of poverty’s consequences than the regional or national levels (MARTINELLI; JOYAL, 2004). In this sense, Albuquerque (2004) expresses it is of main importance to focus on local data and production networks. Lack of information is a barrier to develop strategies towards local economic development. In fact, conventional statistics are usually based on aggregate economic activity data (or sectorial data) which limits a local production networks approach.

The secretary of Economic Development, Science, Technology and Innovation of the Sao Paulo State (SD), through the Federal Programme for APLs Promotion, has detected 38 APLs and 25 productive agglomerations (APs) distributed in more than 120 municipalities (SD, 2018). Since the year 2006, the Secretary of Development of the São Paulo State (SD) has been responsible for the promotion of clusters and strengthening of SME. This Secretary, the Brazilian Service of Support for Micro and Small Enterprises (SEBRAE), and the Federation of the Industries of the State of São Paulo (FIESP) are the third most important APL’s support institutions. The APL policy is part of the policy to enhance the competitiveness of SME. These policies have led to the São Paulo’s economy growth and, consequently, to better employment and income level, and lower regional disparities. In 2003, an APL program was built to support the promotion of municipalities (including rural ones) hosting agglomerations of SME.

The number of studies on APLs has increased in the academic literature due to ALP’s contribution to economic and regional development (MARINI; SILVA; NASCIMENTO, 2016; FERREIRA et al., 2016; OLIVEIRA; MARTINELLI, 2014; MACEDO; DIAZ MERINO; DALLAZEN CAMILLO, 2014). In Brazil, most studies have mainly focused on detecting and mapping clusters; and some others on their effects on GDP or employment (which are only one dimension of local development). However, there only a few studies on cluster impact evaluation, especially in Latin America. In this sense, the present paper contributes to the scant quantitative research on the impact of cluster on local development.

The objective of this paper consists in examining whether clusters are promoters of local development. Do municipalities hosting a cluster of firms exhibit a better local development than municipalities without? To achieve this goal, a Propensity Score Matching Technique (PSM) is employed; a quasi-experimental method to control the impact of clusters on local development from the self-selection problem. This methodology, widely used for impact evaluation, artificially builds a “match” for each municipality having identical characteristics except for one: cluster building.

This study excludes metropolitan areas and examines non-metropolitan municipalities with just one or a few dominant clusters. The reasons that explain this object of study are: a) cluster building and local development will more likely occur in metropolitan areas; b) metropolitan regions centralize many productive activities throughout different clusters; c) it is more likely that the reverse causality, from development to cluster building, occurs in metropolitan regions; d) the geographic concentration of industries in metropolitan areas and big cities ensures the presence of external economies (location and urbanization economies).

The paper is structured as follows. Firstly, the paper offers a theoretical framework about the concepts of cluster, local development and their relationship. Besides, it describes the
explanatory factors of cluster building in some municipalities. Secondly, the Propensity Score Matching (PSM) methodology is explained. Thirdly, a descriptive analysis is presented and the model is estimated. Lastly, results obtained and concluding remarks are discussed.

2 THEORETICAL FRAMEWORK

2.1 Local development and clusters

Many years ago, the international literature analysed the relationship between industrial districts and local development (BECATTINI, 1979; BELLANDI, 2003; BECATTINI et al., 2003). Several local development theories had emerged but only a few of them, which were an extension of the industrial location theories, had envisaged a positive relationship between agglomeration and development. Among this approach, the ideas of Perroux (1955), Hirschman (1958) and Myrdal (1957) had prevailed. These authors had observed a causal relationship between the spatial agglomeration of productive activities, especially industrial activities, and inequalities inherent in the development of capitalist countries.

Most schools of thought have explained local development through the presence of clusters; where local development consists of an endogenous development process based on the interdependence of local actors in the territory. A sustainable cluster depends on an efficient and effective relationship between the actors. This approach assumes local productive systems are important for growth and structural changes (NARVAEZ; FERNÁNDEZ; SENIOR, 2008). Besides, clusters play a critical role to promote local development (FARINA et al., 2017; MARINI et al., 2012; LÜBECK et al., 2012).

There is consensus about the concept of territory as a social construction which is explained by several factors: economic, social, environmental, and institutional, among others. Albuquerque (1997) states that local development is a transformation process to overcome obstacles and challenges coming from the economy and local society. By seeking a more efficient and sustainable use of resources, local socioeconomic agents (public and private) can improve inhabitants’ quality of life. However, there is a lack of consensus about the theoretical model of local development (COCCO; GALVÃO, 2001).

Clusters lead to employment and wealth in the region, and therefore, promote economic and social development (CORRAL et al., 2006). A main strategy for local economic development, have been the improvement of territories with competitive advantages. As a result, these territories can develop and reorganize their production systems, increase employment and improve the standard of living of the local population (VÁZQUEZ BARQUERO, 1986). Cluster firms can directly or indirectly achieve these local goals by developing social activities in their territory (FERRARO; CORNICK, 2018; FARINA et al., 2017; LÜBECK; WITTMANN; SILVA, 2012; ROMIS, 2008). However, there is concern whether clusters can increase employment for the poor, reduce vulnerability and risks among firms, and promote labour, social, and environmental standards and corporate social responsibility (ROMIS, 2008). To sum up, whether clusters can enhance the poor’s quality of life.

Although there are several studies on clusters impact evaluation on local development (MARINI; SILVA; NASCIMENTO, 2016; OLIVERIA; SANTANA, 2016; OLIVEIRA; MARTINELLI, 2014; MACEDO; DIAZ MERINO; DALLAZEN CAMILLO, 2014), most of them are case studies or qualitative studies. Studies based on a quantitative approach are only a few in Brazil (CARVALHO; CARRARO;
SHIKIDA, 2016; JACOMETTI et al., 2016; ALDERETE; BAGIC, 2016; LEITE FILHO; ANTONIALLI, 2011; LEITE FILHO, 2010; BRITO et al., 2010; SANTOS, 2009). JACOMETTI et al. (2016) analyses the impact of public policies on the economic development of the Arapongas, PR, furniture-focused APL. Based on an institutional perspective, they show the institutional environmental conditions were weaker than expected. Hence, although there was economic growth in Arapongas, there was no economic development.

On the other side, Alderete and Bacic (2016) suggest municipalities with clusters have a better local development than municipalities without cluster by utilizing data from Sao Paulo. Although results are interesting, the methodology employed is not appropriate to explain causality between cluster and local development. On the other side, Leite Filho and Antonialli (2011) and Leite Filho (2010) study Minas Gerais’s APL. Their results indicate young APLs have a significant influence on the local development of small and medium sized cities. Lastly, Santos (2009) examines the clusters supported by the PROMOS Project of SEBRAE (Brazilian Service of Support for Micro and Small Sized Enterprises) and The Interamerican Development Bank (IDB). The author wonders if cluster policies are effective to reduce regional inequalities. Other authors have only focused on some aspect of local development. For instance, Carvalho, Carraro and Shikida (2016) evaluate the impact of APLs public policies on per capita GDP in the municipalities of Rio Grande do Sul. The authors find a statistically significant and positive impact of the farm-machinery and implements cluster on GDP per capita. Similarly, Brito et al. (2010) study the relationship between agglomerations of industrial firms and local employment. Based on a multilevel analysis, they show a positive relationship between APL and local employment.

2.2 Why are Brazilian clusters built in some municipalities?

The RedeSist states that whenever there is production of any product or service, there will always be a local productive arrangement throughout it, entailing activities and actors in the acquisition of raw materials, machines and other resources. Even though exceptions are rare, this argument corresponds to regional and macro-level analysis (LASTRES; CASSIOLATO, 2003). However, it is likely to find activities not involved in cluster at municipal level.

APLs are characterized by two important features: production specialization of firms and spatial proximity. Based on this criterion, any production agglomeration in a region, city, neighbourhood or even a street can become an APL (NORONHA; TURCHI, 2005). Even though APLs are usually built spontaneously; some of them are supported by local institutions (such as municipalities, universities) or national institutions (such as SEBRAE, BNDES- National Bank for the Economic and Social development and federal government). In this paper, we utilize all APLs from the Sao Paulo State, whether they are supported or not.

Municipalities, which are the territorial base of a cluster, affect the process of clustering by determining the infrastructure conditions (access to motorways, telecommunications, among others), urbanization norms, and other critical factors. Hence, it is expected that municipalities with adequate transport and telecommunications infrastructure conditions, high levels of urbanization, industrial activity, geographic density and qualified labour; and a suitable institutional environment more likely will host a cluster.

According to the Ministry of Development, Industry and External Commerce (MDICEB) (BRASIL, 2004), a productive arrangement can be detected by a set of variables. Studies on
mapping clusters use the number of establishments and employees based on the national classification of economic activity (CNAE 4 digits). On the other side, FIESP (2003) identifies some critical variables to detect clusters such as the number of establishments, number of employees, coefficient of location, employment growth (1996-2000), international market dynamism (1996-2000), sales growth (1996-2000), and SME participation.

The Location coefficient QL measures the relative degree of concentration of a given activity in a given region. QL is a quotient between the productive specialization of a municipality in a certain sector/activity, relative to the productive specialization of the state in the same sector/activity (SUZIGAN et al., 2004). Although the location coefficient can detect places where there is concentration of a specific activity, it is not sufficient condition to clustering. Suzigan et al. (2004) states that a cluster is of high importance for local development if QL is higher than five, employment is above one percent and the number of establishments is larger than five. Hence, the location coefficient QL is only one of the predictors of clustering. The higher the QL coefficient, the greater the probability that a municipality will host a cluster.

External economies from clustering play a main role to increase the local producers’ competitiveness. External economies emerge from a wide contingent of specialized labour and specific skills (i.e. a set of specialized suppliers of raw materials, knowledge, skills and information). The existence of a specialized labour force in some industrial sectors, such as the sugar-alcohol and metal mechanic, has explained industrial location. Thus, inhabitants’ education level is a critical factor for clustering. Apart from that, the urbanization rate and demographic density are proxies of the presence of suppliers of specialized raw materials, and dissemination of knowledge, skills and information. Since most densely populated and urbanized areas have access to services, telecommunications, and transports, they are more likely to host a cluster.

As soon as the quality of human resources becomes a strategic resource for building competitive advantages, the institutional and social innovative environment is crucial for local development (MARTINELLI; JOYAL, 2004). Institutions and norms jointly contribute to build clusters in a territory. Martinelli and Joyal (2004) describes how the Community Development Corporations (CDEC) in Montreal, Canada has provided direct support to entrepreneurs and managers both in training and education, leading to an adequate environment for economic and local development. In this vein, the existence of industrial parks\(^3\) and the application of territorial laws are useful to explain clustering in a territory.

Furthermore, the increasing penetration of new technologies has led to a high interdependence of activities among clustered firms. The interrelation among producers, suppliers, sellers and service-suppliers firms and other institutions requires of equipment and innovative and operative methodologies. The performance of a cluster depends on new information and communication technologies (ICT), and information and knowledge exchanges. In part, these technologies permit the spread of Industries across the interior of São Paulo, a relevant fact for the de-concentration process (AZZONI, 1985). Therefore, municipalities with computers in the Public Administration are more likely to host a cluster.

---

\(^3\) We should not confuse industrial parks with industrial districts. Industrial parks are portions of a city that are zoned for industrial use. Some industrial parks offer tax incentives. On the contrary, industrial districts are the Italian clusters, geographically closed productive systems, characterized by small and medium sized enterprises and embedded in a social and cultural atmosphere.
Moreover, the socioeconomic changes of the last 50 years in São Paulo have pursued a deep redistribution process, leading to demographic concentration (GONÇALVES DA SILVA; BACIC; LANNA, 2010). It is expected that regions with a high demographic density will concentrate a large amount of small and medium firms. Since Brazil is a huge country, a high demographic concentration also implies a large number of firms and of production volume (BARCELLOS DA COSTA, 2010).

Demographic concentration and urbanization are interrelated processes, which result in a high population concentration of urban areas. Based on 2005 data, in the São Paulo State there are 37 million inhabitants living in urban areas, reaching an urbanization rate of 93.7 percent. On the contrary, rural areas are spaces with a lower demographic density and more natural resources than urban areas (VEIGA, 2002). The higher the demographic density, the higher the social networking and the fewer the number of rural areas will be. In Europe, the demographic density has promoted the development of local productive systems more extensively than in North America (KAGEYAMA, 2004). Hence, the higher the demographic density of a municipality, the greater the probability of holding a cluster will be.

3 METHODOLOGY

The best way to evaluate the impact of clusters on local development would be an experiment where clustering were an exogenous process. That is, where the decision to cluster were independent of the territory/municipality characteristics or conditions. However, this experiment is not feasible. Due to the multidimensional nature of local development, the relationship between clustering and local development suffers from endogeneity. Hence, a quasi-experimental design seems adequate when is not feasible to control the treatments assignment and the other factors under study.

The objective of this paper consists in analysing what would have happened in terms of local development if clusters were not build in a municipality. It is necessary to measure the development improvements of municipalities with clusters conditional to the pre-treatment development; where the development improvement is the difference between the development levels of a municipality with and without clusters. To define a statistic group of comparison to the group of municipalities with clusters, the Rosembaum and Robin (1983) Propensity Score Matching (PSM) technique is applied.

The PSM technique consists of the following steps: firstly, the probability that a municipality receives the treatment (holding a cluster) is estimated; this probability is the municipality score. Secondly, the sample is divided in two sub-samples: the treatments (municipalities receiving the treatment) and the controls (those not receiving the treatment), and both samples are ordered in a descendent manner. Then, a treatment is assigned a control with similar score, and matches are built (the same control can be matched with more than one treatment).

The next step consists in estimating the local development difference of each match. Then, the average difference of the whole sample is determined. This result is known as “the average treatment of the treated” (ATE). The standard error of the difference between each match leads to a t-significance test to contrast the null hypothesis that ATE equals to zero. If the hypothesis is rejected, it can be asserted that ATE is statically different from zero. Otherwise, it cannot be confirmed that ATE is different from zero.
In analytical terms, the average effect of a binary treatment under a continuous scalar product is estimated. For each municipality $i$, $1 = 1, ..., N$, with all interchangeable units, $(Y_i(0), Y_i(1))$ are defined as the two potential products such that $Y_i(0)$ represents municipality $i$ local development if it does not hold a cluster (not exposed to treatment) and $Y_i(1)$ is municipality $i$ local development if it holds a cluster. In spite of the multidimensional nature of local development, there are some suitable development indexes for São Paulo. These composite indexes are useful to compare regions and time periods. Composite indicators are also used for benchmarking and focusing on some specific research area, even the expense of losing regional specificities (STALLIVIERI, 2011).

If both local development levels (when a municipality holds a cluster, $Y_i(1)$, and when it does not hold a cluster, $Y_i(0)$) were observable, the treatment effect (holding a cluster) of municipality $i$ would be the difference $Y_i(1)-Y_i(0)$. However, a problem arises since only one of these states or products is observable.

$$Y_i = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases}$$

Where $D_i$ indicates whether a municipality receives the treatment or not (holding a cluster or not). Based on Roy (1957), Quandt (1972) and Rubin (1978) models, the local development levels assumed are:

\begin{align*}
(2) & \quad Y_1 = \mu_1(X) + U_1 \\
(3) & \quad Y_0 = \mu_0(X) + U_0
\end{align*}

The local development gain is expressed as $\Delta = Y_1-Y_0$. If $Y_1$ and $Y_0$ were observable in each municipality, the impact evaluation of the treatment would not be a problem. However, it is not possible to observe both states at the same time. To solve this problem, policies are evaluated using different versions of means of variations of the population under study. One of the methods is the Average Treatment Effect (ATE).

A comparison of average local development explains something about the potential developments but it does not necessarily explain the phenomenon. The comparison of the average local development levels conditioned on holding a cluster is formally related to the average causal effect though the following equation:

$$E(Y_i | D_i = 1) - E(Y_i | D_i = 0) = [E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1)] + [E(Y_{01} | D_i = 1) - E(Y_{00} | D_i = 0)]$$

Equation (4) means:

Observed difference in the average local development levels = ATT (average effect of the APL presence in the treated) + Selection Bias.

Under the treatment of the treated, the problem arises because data does not usually include observations of the $Y_0$ local development index for municipalities with cluster ($D=1$). The selection bias problem (HECKMAN, 1990) emerges as a result of this unobservable data. The matching technique solves this selection bias problem replacing the randomization with the regressors’ conditioning. The selection bias is removed only if the treatment has been purely random between municipalities with the same propensity score.

Matching with propensity score can reduce the selection bias and, therefore, enhance the treatment effect estimation with observable data. To achieve this goal, a logit or probit model is...
estimated where the most relevant information is the maximum likelihood function, more than the estimators’ significance (HECKMAN; LALONDE; SMITH, 1999).

There are different methods to estimate the ATE based on the propensity score. These methods differ in their definition of the distance between the treated and the control as follows:

Nearest Neighbour Matching, that matches treated and control municipalities taking a treated unit for each control according to the nearest propensity score estimated. A non-treated municipality j is selected as the counterfactual of municipality i such that the control group of municipality i, Co(pi) with propensity score pi, is a municipality j that accomplishes: C(i)=min j||Pi-Pj||. This estimator uses a municipality from the control group to compare with each municipality from the treated group.

Kernel estimator matches treated municipalities with a weighted average of all control municipalities. Weights are inversely proportional to the distance between the propensity scores of treated and untreated municipalities.

Stratification allows matching based on a variable that contains the layer number in the area of common support. Municipalities used in the ATT estimation belong to the minimum maximum range of the propensity scores of the treated group. Thus, the defined area includes positive density values for both treated and control municipalities (SMITH; TODD, 2005).

3.1 Variables and data sources

The main variable of interest is Cluster, a binary variable that takes value 1 if a municipality holds an APL (cluster) and zero otherwise. Hence, the treatment is assigned to those municipalities that in 2010 were holding an APL. Clusters detected correspond to the year 2010. This clustering process is the result of building and strengthening processes that in many cases had emerged in 2004. This study is based on 138 clusters (APL) detected by Goncalves da Silva, Bacic and Lanna (2010). The authors have detected and mapped clusters in the municipalities of São Paulo; since already existing mapping of clusters are incomplete, often incompatible among each other and are focused on industrial activities.

The object of study is the non-metropolitan municipalities (excluding metropolitan municipalities and their neighbours)⁴ of the Sao Paulo State in Brazil. Metropolitan municipalities are not examined due to their complexity and specific nature: a) Metropolitan regions centralize and perform several productive activities, they have infrastructure, resources and technology, and they are characterized by a larger demographic concentration than other regions. A recurrent process is built where the city creates its own increasing demand as a result of the agricultural and agro-industrial dynamism. Hence, the reverse causality (from local development to cluster building) is more likely to occur in metropolitan regions. b) The de-concentration process of the metropolitan region towards the interior of the São Paulo state. This process is evidenced by the decrease of the industrial transformation value from 74,7 to 53,3 percent. Nowadays, the industrial dispersion is one of the key elements for the São Paulo’s restructuring (LENCIONI, 1998).

Since clustering is the result of building and strengthening processes, it is reasonable to evaluate the cluster impact on the local development indexes of 2010, instead of 2004 in which the process had begun (local development might take time to occur). Some composite indexes

⁴ Neighbours of metropolitan municipalities are those municipalities that share any of the metropolitan municipalities’ geographic boundaries.
are used as proxy of local development. The Gross Domestic Product (GDP) alone is not an adequate measure to assess the social progress of a territory. The United Nations Program (PNUD) has developed the Municipal Human Development Index (IDHM) for the São Paulo state that combines income, educational attainments and life expectancy. The IDHM is a methodological adaptation for municipal data of the Human Development Index (IDH) built by Mahbubul Haq and Amartya Sen.

The IDH’s three fundamentals are health, education and income. These variables are measured as follows: a) The health dimension is assessed by life expectancy at birth; b) The education component of the HDI is measured by mean of years of schooling for adults aged 25 years and expected years of schooling for children of school-entering age; and c) The standard of living dimension is measured by gross national income per capita (in purchasing power parity 2005 constant dollars)\(^5\).

On the other side, the SEADE- Foundation Data Analysis State System- publishes two alternative indexes of development: the São Paulo Social Responsibility Index (IPRS) and the São Paulo Social Vulnerability Index (IPVS). Both indexes include socio-economic and demographic dimensions of development with some methodological differences. According to SEADE, the São Paulo’s Social Responsibility Index or *Indice Paulista de Responsabilidade Social* (IPRS) synthetizes each municipality situation based on wealth, schooling and longevity indicators. Each indicator consists of a linear combination of four variables which are expressed on a zero to 100 scale in which 100 represents the best situation and zero the worst. The combination of these indicators enables a typology of municipalities in five groups (from one to five) from the best to the worst development level.

Lastly, the São Paulo Social Vulnerability Index (IPVS) published by the SEADE Foundation, identifies municipalities’ areas where families are exposed to different levels of social vulnerability (percentage of population). The IPVS is based on two assumptions: 1) a social vulnerability study must consider the multiple dimensions of poverty; 2) São Paulo’s urban centres are characterized by spatial segregation which explains the social inequality patterns. The 2010 IPVS built a typology of social vulnerability by adding to the income indicators other indicators related to schooling and family life cycle. Hence, it combines the socio-economic and demographic dimensions. The index classifies a municipality resident population in seven groups of social vulnerability, from group 1 (the lowest vulnerability group) to group 7 (the highest vulnerability group).

These composite indexes, which are general and synthetic measures of human development, are used as indicators of local development. Data corresponds to the year 2010. The IDH and IPVS are more integral indexes than the IPRS (Table 1).

---

\(^5\) Even though the IDH approaches the human development perspective, it does not include all the development aspects. It is neither a representation of people’s “happiness” nor an expression of “the best place in the world to live” (PNUD, 2013). Other aspects not considered in the index are democracy, participation, equity and sustainability.
Studies on mapping and detecting clusters have brought information about the factors that explain clustering in certain territories. Most of the data in this study comes from the Geography and Statistics Brazilian Institute (IBGE) through the SEADE Foundation, except as otherwise specified\(^6\).

**Demographic density**: Measured as inhabitants/km\(^2\). It is defined as the number of inhabitants in a geographic unit area at a certain time.

**Urbanization rate**: Urban population as percentage of total population. It is calculated from census data. Data refers to municipalities defined by the administrative division of the São Paulo state currently in use. Until 1997, there were 645 municipalities.

**Area**: Total territorial surface (urban and rural) of a municipality. We add this variable as a control variable, to distinguish between small and large municipalities.

Both the urbanization rate and demographic density are proxies of the presence of suppliers of specialized raw materials, and dissemination of knowledge, skills and information; since the most densely populated and urbanized areas have access to services, telecommunications, transports, among others.

**Education**: Secondary school approval rate. Students that accomplished the final evaluation and the minimum frequency and approval requirements as percentage of total students enrolled at the end of school year. Data comes from the Education State Secretary (SEE), Educational Information Centre (CIE), Ministry of Education (MEC) and National Institute on Educational Studies and Research (INEP). Published by SEADE Foundation.

**Industry participation on total aggregate value**: Aggregate value of the Industrial sector as percentage of the region’s total aggregate value.

**Industrial parks**: Dummy variable that takes value 1 if a municipality has any industrial district. It is based on the number of industrial districts in a municipality. Data comes from SEADE Foundation, Unified Municipal Research (PMU). Industrial districts are considered industrial parks.

**Location Coefficients** (QL agriculture; QL industry; QL commerce; QL service): Based on SEADE data, this coefficient is calculated as:

\[
\frac{\text{A sector participation (i.e: agriculture) in a municipality’s total employment (i.e. Ubatuba)}}{\text{A sector participation (i.e: agriculture) in the State’s total employment (i.e: São Paulo)}}
\]

The numerator is the agriculture participation in total aggregate value (in %) (Municipal data). The denominator is calculated based on SEADE data.

**Computers (PC) in Public Administration**: Number of computers in the local government the Municipal Prefecture.

\(^6\) Unfortunately, SEADE does not publish continuous data. For many variables, information is missing. In 2009, data is complete for the selected variables leading to a balanced data.
Urban plot laws: It expresses the existence or not of a legal instrument that determines the legal limit or boundary between a municipality’s urban and rural areas.

4 DESCRIPTIVE ANALYSIS

This study is based on 540 non-metropolitan municipalities from the São Paulo state. Municipalities with clusters (or holding a cluster) are 107 and represent nearly 20 percent of the sample. These municipalities, on average, have achieved a better local development than municipalities without clusters (Table 2).

Table 2 – Average indexes by group of municipalities (before matching)

<table>
<thead>
<tr>
<th>Municipalities</th>
<th>Percent of total municipalities</th>
<th>IDH2010</th>
<th>IPRS</th>
<th>IPVS2</th>
<th>IPVS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>With clusters</td>
<td>19,82</td>
<td>0,7428692</td>
<td>3,2050</td>
<td>42,92</td>
<td>11,96</td>
</tr>
<tr>
<td>Without clusters</td>
<td>80,18</td>
<td>0,7391155</td>
<td>3,5150</td>
<td>32,19</td>
<td>17,62</td>
</tr>
<tr>
<td>ANOVA</td>
<td></td>
<td><strong>Ns</strong></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Total municipalities</td>
<td>100</td>
<td>0,7398593</td>
<td>3,45</td>
<td>34,55</td>
<td>16,29</td>
</tr>
</tbody>
</table>

Source: The authors. Note: ns non-significant, *** significant at 1%.

Since the IPRS varies from one (best situation) to five (worst situation), it is observed that municipalities with clusters, on average, show a better IPRS level than municipalities without clusters. Such difference is statistically significant (F = 10,458, p value < 0,01). Similarly, differences between both groups of municipalities regarding the IPVS are statically significant (F = 28,56, p value < 0,000 in the IPVS2, and F = 15,11, p value < 0,000 in the IPVS5). Therefore, municipalities with clusters have a larger percentage of population under low vulnerability and a lower percentage under high vulnerability than municipalities without clusters.

However, this methodology is not adequate to explain causality between clustering and local development, if local development differences among municipalities arises from other observable characteristics than clustering. Table 3 describes the main characteristics of both groups, with and without clusters, before matching is estimated.
Table 3 – Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Groups of municipalities</th>
<th>N</th>
<th>Mean</th>
<th>Standard Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic density</td>
<td>With cluster</td>
<td>107</td>
<td>123.774</td>
<td>158.7698</td>
<td>11.63</td>
<td>944.11</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>46.20603</td>
<td>59.14157</td>
<td>3.73</td>
<td>590.41</td>
</tr>
<tr>
<td>Urbanization</td>
<td>With cluster</td>
<td>107</td>
<td>89.77692</td>
<td>8.925384</td>
<td>54.44</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>81.04762</td>
<td>14.57788</td>
<td>24.92</td>
<td>100</td>
</tr>
<tr>
<td>Area</td>
<td>With cluster</td>
<td>107</td>
<td>517.71</td>
<td>368.6685</td>
<td>48.6</td>
<td>1826.75</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>373.9326</td>
<td>295.7892</td>
<td>3.64</td>
<td>1656.73</td>
</tr>
<tr>
<td>Education</td>
<td>With cluster</td>
<td>107</td>
<td>85.24299</td>
<td>5.763976</td>
<td>69</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>86.49654</td>
<td>7.273469</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Industry participation in aggregate value</td>
<td>With cluster</td>
<td>107</td>
<td>25.21346</td>
<td>13.08161</td>
<td>5.78</td>
<td>72.89</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>21.86376</td>
<td>15.16331</td>
<td>4.6</td>
<td>84.5</td>
</tr>
<tr>
<td>Industrial parks</td>
<td>With cluster</td>
<td>107</td>
<td>.4018692</td>
<td>.492583</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>.4018476</td>
<td>.4908386</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>QL agriculture</td>
<td>With cluster</td>
<td>107</td>
<td>4.979832</td>
<td>5.046638</td>
<td>.192</td>
<td>22.873</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>6.954097</td>
<td>5.049232</td>
<td>.026</td>
<td>28.744</td>
</tr>
<tr>
<td>QL commerce</td>
<td>With cluster</td>
<td>107</td>
<td>.927486</td>
<td>.3996479</td>
<td>.184</td>
<td>1.911</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>.7223002</td>
<td>.4210959</td>
<td>.012</td>
<td>2.752</td>
</tr>
<tr>
<td>QL industry</td>
<td>With cluster</td>
<td>107</td>
<td>1.35471</td>
<td>.6783922</td>
<td>.013</td>
<td>2.957</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>422</td>
<td>.9453886</td>
<td>.8386359</td>
<td>.007</td>
<td>4.102</td>
</tr>
<tr>
<td>QL service</td>
<td>With cluster</td>
<td>107</td>
<td>.6699626</td>
<td>.1880338</td>
<td>.23</td>
<td>1.306</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>.8318637</td>
<td>.3483721</td>
<td>.088</td>
<td>1.722</td>
</tr>
<tr>
<td>PC in Public Administration</td>
<td>With cluster</td>
<td>107</td>
<td>60.50467</td>
<td>129.9327</td>
<td>0</td>
<td>1055</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>433</td>
<td>56.08314</td>
<td>124.5535</td>
<td>0</td>
<td>1452</td>
</tr>
<tr>
<td>Urban plot law</td>
<td>With cluster</td>
<td>99</td>
<td>.3030303</td>
<td>.461907</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without cluster</td>
<td>404</td>
<td>.1782178</td>
<td>.3831705</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: The authors based on SEADE Foundation.

Municipalities with cluster, on average, show a higher demographic density, a higher degree of urbanization, a larger area or surface, a lower percentage of inhabitants with secondary school year calendar approved, a higher participation of the industry in total aggregate value, and a higher locational coefficients for the commerce and industry sectors than municipalities without clusters. All these differences are statistically significant based on the ANOVA test, except for education.

5 RESULTS

The objective of this paper consists in determining the average effect of holding a cluster in the local development of the São Paulo’s municipalities. In particular, it is of interest to compare the development levels of municipalities holding clusters against their counterfactual, the development levels if these municipalities would have not held clusters. Since this counterfactual is never observed, it must be estimated.

As a first step in the Propensity Score Matching, a Probit model is estimated. By using probit, the propensity of a municipality to hold a cluster is obtained, conditioned on explanatory variables of clustering in a territory. Only those variables that simultaneously affect the participation...
decision (to hold a cluster or not) and the outcome levels (local development) must be included in the probability estimation (BERNAL; PEÑA, 2011). The dependent variable is clustering in a municipality which is an observable variable (a municipality holds a cluster) based on Gonçalves da Silva et al. (2010). The Probit model estimated using STATA 12 emerges from a latent or non-observable variable, \( y^* \), a latent variable that means a municipality’s propensity to hold a cluster. This variable is explained by means of a set of independent variables (observable) through the following structural equation:

\[
y^* = \beta_0 + x \beta + e. y = 1\{y^* > 0\}
\]

The relation between the observable variable \( y \) (if a municipality holds a cluster) and the latent variable \( y^* \) (the propensity to attract or hold a cluster) is observed through the following equation:

\[
(5) \begin{cases} 
y = 1 & \text{if } y^* > 0 \\
y = 0 & \text{if } y^* \leq 0
\end{cases}
\]

where \( y \) is a dummy variable equal to 1 if the municipality holds a cluster and zero otherwise.

The propensity score estimation leads to the following results using the Stata 12 software (Table 4).

Table 4 – Explanatory factors of cluster building in a municipality

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Density</td>
<td>.0044914</td>
<td>.0010265</td>
<td>***</td>
</tr>
<tr>
<td>Urbanization</td>
<td>.0244691</td>
<td>.008278</td>
<td>***</td>
</tr>
<tr>
<td>Area</td>
<td>.0008237</td>
<td>.0002338</td>
<td>***</td>
</tr>
<tr>
<td>Education</td>
<td>-.0029428</td>
<td>.0113137</td>
<td>ns</td>
</tr>
<tr>
<td>Industry participation in aggregate value</td>
<td>-.0131997</td>
<td>.0062165</td>
<td>**</td>
</tr>
<tr>
<td>Industrial park</td>
<td>-.0390361</td>
<td>.1504793</td>
<td>ns</td>
</tr>
<tr>
<td>QL agriculture</td>
<td>-.031755</td>
<td>.05158</td>
<td>ns</td>
</tr>
<tr>
<td>QL commerce</td>
<td>-.2544621</td>
<td>.4006782</td>
<td>ns</td>
</tr>
<tr>
<td>QL industry</td>
<td>-.0194444</td>
<td>.3731702</td>
<td>ns</td>
</tr>
<tr>
<td>QL service</td>
<td>-1.435.287</td>
<td>.8858238</td>
<td>ns</td>
</tr>
<tr>
<td>PC in Public Administration</td>
<td>-.000232</td>
<td>.0006982</td>
<td>ns</td>
</tr>
<tr>
<td>Urban Plot Law</td>
<td>.3341615</td>
<td>.1705481</td>
<td>**</td>
</tr>
<tr>
<td>Cons</td>
<td>-1.650.161</td>
<td>2.002.165</td>
<td>ns</td>
</tr>
</tbody>
</table>

\( N=493 \)

LR chi2(10) = 107.82; Prob> chi2 =0.0000; Pseudo R2=0.2180

Source: The authors. Note: ***, ** significant at 1 and 5 percent respectively, ns= non-significant.

The PSM obtained is valid under the belief that the observables variables selected determine participation (to hold a cluster). Thus, there is no unobservable variables bias, or unobservable/ unavailable variables are not the main determinants either of participation or of the potential outcome (BERNAL; PEÑA, 2011). On the other side, the method focuses on the treatment assignment process and not on the more complex processes determining the outcomes.

Afterwards, the PSM is used to define the common support area where the distribution of the estimated propensity scores of the treated and the controls overlaps. This common support area emerges from the ATT estimation. In this area, there are municipalities that belong to the
The impact of Brazilian clusters on local development: a propensity score matching approach

The range delimited by the minimum and maximum propensity score of the treated group. Therefore, the defined area includes positive density values for both treated and control municipalities (SMITH; TODD, 2005). The common support area has been [.03796161, .99922907]. Figure 1 depicts the kernel density estimates of the propensity scores and the selected region.

Figure 1 - Kernel density estimate

![Kernel density estimate](image)

Source: The authors, based on Stata 12.

Figure 1 shows the common support area. From the 540 non metropolitan municipalities, 402 belong to this common support area. Among these 402 municipalities, 99 are municipalities with clusters (Table 5).

Table 5 – Frequency table

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (Without cluster)</td>
<td>303</td>
<td>75.37</td>
<td>75.37</td>
</tr>
<tr>
<td>Treated (With cluster)</td>
<td>99</td>
<td>24.63</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>402</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

Source: The authors.

Results obtained indicate that the demographic density, the urbanization rate, the area and the presence of a legal instrument of urban division into lots significantly and positively affect the probability of holding a cluster. On the other side, the industry participation in the total aggregate value has a significant but negative incidence. Hence, municipalities with a higher participation of industry in the total aggregate value are less likely to hold a cluster than the rest. This result could be explained since most municipalities of São Paulo are from the agriculture (farming) sector, followed by industry, commerce and services respectively. Later on, the propensity score is estimated by using the matching techniques. Differences on local development levels estimated are (Table 6).
The hypothesis that there are no differences between treated and control municipalities is rejected. Firstly, there is a statistically significant difference in the Municipal Human Development Index (IDHM) between municipalities with cluster (treated) and without cluster (control). Based on the stratification method, the average treatment effect (ATT) is significant (Table 6). Differences on the average local development between municipalities with cluster and without cluster can be expected with 95 percent significance. However, this average effect is not significant using the other matching methods (nearest neighbour or Kernel).

Even though this difference is statistically significant, its amount is lower than 0.0010 (and IDHM varies from zero to one). Therefore, results obtained with IDH are compared with other indexes to verify the robustness of the model (Table 6). The IPVS can identify areas where families exposed to different levels of social vulnerability prevail. The IPVS is based on a typology/classification that emerges from the combination of socio-economic and demographic indicators. This classification leads to seven groups or categories according to the social vulnerability degree of the resident population. As a result of municipalities’ distribution among each group, this study focuses on IPVS Group 2 and IPVS Group 5 which are the greatest groups and where significant differences in the treatment are observed.

At last, the average treatment effect was estimated using the IPRS index. Although the difference between the average IPRS of the treated and control groups exist, it was not statistically significant according to the matching techniques. Municipalities with cluster present a larger percentage of population under lower social vulnerability than municipalities without cluster. Hence, they achieve better socioeconomic and demographic indicators than municipalities without cluster. This is evidenced by using both the IPVS2, where municipalities with clusters have a larger population that belongs to group 2, and the IPVS5, where municipalities with clusters have a lower percentage of population in group 5.

### 6 DISCUSSION AND CONCLUSIONS

The embeddedness of local development drives the need to think over the role of clusters. This paper focuses on the role of clusters in improving local development. We analyse whether municipalities can improve their local development by holding clusters in their territory. The relationship between clustering and local development may suffer from endogeneity since municipalities with clusters (or holding clusters) are most likely different from those without, and these differences may be correlated with local development. Therefore, a raw comparison of average test scores between the groups of municipalities with and without clusters would
be biased due to the presence of other (observable and non-observable) factors affecting this relationship.

A good exercise would be doing a random assignment of clusters among municipalities and then compare the average development indexes of both groups (with and without cluster). Since a controlled randomization experiment is not feasible, non-experimental methods are employed to replicate this experiment under reasonable conditions. The present paper estimates the impact of clusters on the local development of non-metropolitan municipalities of São Paulo, Brazil. To achieve this goal, some matching techniques were applied to estimate the local development gains from clusters. The matching technique intends to control the endogeneity problem arising from the inverse causality and self-selection problems of the relationship between cluster and local development.

This paper pretends to deal with the multidimensional nature of local development by using different local development indexes, such as the São Paulo Human Development Index (IDHM), the Social Responsibility Index (IPRS), and the São Paulo Social Vulnerability Index (IPVS). The IDHM is a widely accepted index; the IPRS offers information on economic and social development at municipal level, but it does not study the inequality and poverty concentration issues. On the other side, the IPVS take into account all the dimensions. As Cassiolato and Lastres (2001) state that available data and indicators are not a perfect measure of the qualitative development changes.

Results obtained show that cluster improves local development of non-metropolitan municipalities in the Sao Paulo State, in Brazil. On the one side, these results confirm previous research in Brazil (JACOMETTI et al., 2016; ALDERETE; BACIC, 2016; LEITE FILHO; ANTONIALLI, 2011; LEITE FILHO, 2010) while they contradict others (Santos, 2009). On the other side, these findings validate cluster policies due to APL contribution to local development, and hence, to the society welfare. In particular, the APL programs in Brazil could be a useful policy tool to bring solutions towards local development.

There are several local development experiences which differ in terms of history, evolution, institutional organization, social and cultural frames. Besides, they impact on the productive structure, production management, learning processes and local governance. This paper employs a quantitative methodology to complement already existent case studies about the impact of clusters on local development. In this sense, the present paper contributes to the scant quantitative research on the impact of cluster on local development. Future research could be done to examine whether this results is sustainable along time.

REFERENCES


LEITE FILHO, Geraldo A; ANTONIALLI, Luis M. Proposta de classificação de arranjos produtivos locais por indicadores de identificação: um estudo multivariado. Interações, Campo Grande, MS, v. 12, n. 1, p. 53-64, jan./jun. 2011.


LÜBECK, Rafael M.; WITTMANN, Milton L.; SILVA, Marcio S. Afinal, quais variáveis caracterizam a existência de cluster arranjos produtivos locais (APLS) e dos sistemas locais de produção e inovação (SLPls)? Revista IberoAmericana de Estratégia, São Paulo, v. 11, n. 1, p. 120-51, Jan./Apr. /2012.


TEIXEIRA, Francisco. Políticas públicas para o desenvolvimento regional e local: o que podemos aprender com os arranjos produtivos locais (APLs)? Organizações & Sociedade, Salvador, BA, v. 15, n. 46, p. 57-75, July/Sept. 2009.


About the authors:

**María Verónica Alderete**: PhD in Economics (Universidad Nacional del Sur [UNS], Bahía Blanca, Argentine, 2010). Adjunct researcher of CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), at IIESS (Instituto de Investigaciones Económicas y Sociales del Sur) – CONICET-UNS. Besides, she is an adjunct professor at the Department of Economics (UNS). Her research interests are Information Society, Economics of Innovation, and Local Development. Her more recent publications are “Broadband adoption in Latin American countries: does geographic proximity matter?” in the Revista Problemas del Desarrollo; “Mobile broadband: a key enabling technology for entrepreneurship?” in Journal of Small Business Management; and “An approach to the broadband effect on Latin American growth: a structural model”, in Cuadernos de Economía, among others. **E-mail**: mvalderete@iiess-conicet.gob.ar, **Orcid**: https://orcid.org/0000-0002-9617-7526

**Miguel Juan Bacic**: Holds a Certified Public Accountant degree- Universidad Nacional Del Sur, (1976), Master’s Degree in Economical Sciences from the Universidade Estadual de Campinas (State University of Campinas, 1990) and a Doctor’s Degree in Business Administration – Universidad Nacional Del Sur (1998). Full Professor in Economics (field of Business Economics) of the Universidade Estadual de Campinas (2002). Currently an MS6 professor (head professor) at Universidade Estadual de Campinas. Has experience in Business Economics and Administration, acting mainly with the following subjects: Cost Management, Corporate Strategy, Quality Management, Economy of Small Businesses, Local Development Policies, Entrepreneurship and Solidarity Economy **E-mail**: bacic@unicamp.br, **Orcid**: http://orcid.org/0000-0002-5645-9422