Extended abtract

EXTENDED ABSTRACT

The Marshallian industrial district as a living innovation machine: modelling technological innovation in space and time variable-geometry units using big data and machine learning

Rafael Boix-Domenech
Email: rafael.boix@uv.es
Departament d'Estructura Econòmica
Facultat d'Economia, Universitat de València

Vittorio Galletto
Email: vittorio.galletto@uab.cat
Institut d'Estudis Regionals i Metropolitans de Barcelona
Universitat Autònoma de Barcelona

Fabio Sforzi
Email: fabio.sforzi@unipr.it
Dipartimento di Scienze Economiche e Aziendali
Università degli Studi di Parma

Subject area: *12.- Economy of knowledge, creativity and geography of innovation*

Keywords: Marshallian industrial districts; technological innovation; iMID effect; variable-geometry units; change and evolution; regional innovation policy
JEL codes: O14, O31, R12

Abstract: In this paper we investigate how the iMID effect changes using dynamic territorial units that modifies their geographical boundaries and typologies over time. The article raises two questions: How does the iMID effect evolve when space-time dynamic units are used? How can we explain and predict the evolution of the iMID effect in space-time? The paper focuses on the evolution of the iMID effect in Spain during the period 1991–2014.

The paper makes three contributions. First, it is one of the few researches on the geography of innovation that uses a variable and adaptive geometry of territorial units and typologies that changes over time, closing a relevant gap in the empirical literature. Second, we propose a methodology that allows to decompose the changes in the innovative intensity of the MIDs according to their endogenous characteristics and according to their geographical and typological transformations. Third, we use big data and machine learning methods to explain and predict the evolution of the innovative intensity in the LPSs.

A population of firms specialized in different phases of the same production process (i.e. phases of processing, parts of a product or products) and embedded in a given local

community is what economists call the 'Marshallian industrial district' (MID). Conceptualized in the 1980s by Giacomo Becattini (see Becattini, 2001 and 2004), the MID is now, as then, the theoretical benchmark for explaining the economic competitiveness of small and medium-sized enterprises (SMEs).

The segmentation of production into independent firms of phase means that each phase a) has a specific technical culture and b) develops its own market. This plexus of markets also extends to the subsidiary industries that supply the main industry of the district, with implements, specialized machinery or chemical products for textile processing (Becattini, 2004, pp. 45–46). As argued by Sebastiano Brusco – a leading economist of the district research in the 1980s – phase entrepreneurs can switch from one to another of the numerous production processes that take place in a MID, and subsidiary entrepreneurs can make or modify their products on demand, thus fostering the circulation and sharing of innovations (Brusco, 1986, pp. 87–88). The continuous recombination of production relations that occurs within the MID production system breeds a constant stream of innovations and stimulates the tendency to innovate (Becattini, 2004, p. 46).

The term 'district effect' was coined by Signorini (1994) to explain the high efficiency rates of firms localised in Marshallian industrial districts (MIDs). Dei Ottati defined the district effect as the 'set of competitive advantages derived from a strongly related collection of economies external to the individual firms but internal to the district' (Dei Ottati, 2006, p. 74).

The empirical research on the district effect has been especially intense in regard to the so-called static efficiency – that is, efficiency in costs, productivity and exports-comparative advantages (Boix, Galletto, & Sforzi, 2018).

However, the competitive advantage of the district lies in its dynamic efficiency. The introduction of the concept of external economies, 'which arise out of the collective organization of the district as a whole', is due to Alfred Marshall (Marshall, 1930, p. XIII). In the Marshallian theoretical framework, these external economies are nothing but economies of knowledge, and, as such, they support innovation. The dynamic district effect is associated with the production of knowledge and innovation (Becattini, 2004; Bellandi, 1992).

The innovation-Marshallian industrial district (iMID) thesis defines the existence of dynamic efficiency in the Marshallian industrial district (MID) in the form of a positive differential of innovation of MIDs compared to the average of the national economy (Boix & Galletto, 2009). The MID is a Marshallian innovation machine (Boix, Galletto & Sforzi, 2019) and its long-term competitive advantage lies in dynamic efficiency (technological progress and innovation) and not in the static efficiency (costs allocative and productive efficiency).

Previous research as proven the existence of the iMID effect in Spain during a long time period (Boix, Galletto & Sforzi, 2019). In that research, the unit of analysis - the local production systems (LPSs) - has been considered constant as defined at a point of time. However, MIDs and other LPSs are constantly evolving. Sforzi and Boix (2019) have showed how, between 1991 and 2011, the number MIDs and other LPSs has changed in

number, size and specialization, due to the redefinition of the boundaries of the local labour systems (LLSs) and internal socio-economic changes.

Boix & Galletto (2009) have stated that the measurement of innovation is a widely discussed topic in the literature although there is no agreement about which indicator is the most appropriate. In this paper, we follow the line started in the previous analyses, so to measure innovation we use indicators based on instruments for the protection of intellectual property related mainly to technology, such as patents and utility models.

As long as patents imply novelty and utility, and an economic expenditure for the applicant, it is supposed that patented innovation has economic value (Griliches, 1990). Furthermore, patent documents contain highly useful data, such as the inventor's name and address as well as the invention's date and technological classification. For these reasons, patent indicators are the most widely employed indicators of innovation (Khan & Dernis, 2006). Therefore, the use of patents offers the additional advantage of allowing one to discuss the results regarding the most extended empirical line. There are two additional reasons for its use: patent microdata cover the entire population and not just a sample and allow for exact georeferencing, which is fundamental when working at a detailed territorial level. The validity and convenience of the use of patents as indicators of technological innovation in MIDs and other LPSs have been profusely discussed in previous research (Boix & Galletto, 2009; Boix & Trullén, 2010; Galletto & Boix, 2014; Boix, Galletto & Sforzi, 2018).

However, in our study we are not interested in patents per se but as (technological) innovation indicators. For this reason, patent data are not restricted to a single register or intellectual property office (IPO), as is the usual practice, but rather cover several IPOs to produce a more precise assessment of innovation and the characteristics of different types of LPSs: the OEPM, the EPO and the USPTO. Furthermore, they cover applications with at least one inventor with an address in Spain.

The complete patent database includes 143,229 documents from 1991 to 2014, consisting of the more comprehensive innovation database geocoded at the Spanish municipal level, at least to our knowledge. As is usual in the literature, in cases of multiple inventors a fractional assignation is made to the different municipalities of the addresses.

The selection of the period is marked by the availability of maps of MIDs for Spain since 1991 (Sforzi & Boix, 2019) and the fact that after 2014 the coverage of the innovation registers is not reliable as a result of delays in the publication of data due to secrecy.

To measure local technological innovation, the different sets of data from different IPOs are added to a single indicator related to each year and each municipality so that they can be aggregated by geographical scale and time periods.

In order to avoid yearly fluctuations and to take into account the lags in the outcome of innovation processes, the common practice is to show data on innovation in periods of four to five years (Griliches, 1992). In this research the data are divided into periods of four years. This will allow proper differentiation of the periods of growth and decline of the Spanish economy.

The measurement is made using big data geo-localised by municipality, including 143,229 patent registers (EPO, OEPM, USPTO, etc.), Social Security data (above 8 million registers by municipality, coming from Tesorería General de la Seguridad Social), and research and development (R&D) microdata (a panel of about 8,000 firms plus the central public sector endowments, coming from SABI and several departments of the Spanish Government) (See Boix, Galletto & Sforzi, 2019).

The LPSs identified for Spain by Sforzi & Boix (2019) for the years 1991, 2001 and 2011, provides a unit of analysis that changes over time and is used in this research. As in Boix, Galletto & Sforzi (2019), the period 1991-2014 is divided in 6 sub-periods of 4 years each. The optimal solution would be to use a delimitation of the LPSs for each sub-period. However, since census data is used as the basis for the delimitation procedure, we must use the LPSs for the years in which identification is possible. For each sub-period, we use the LPSs identified by Sforzi & Boix (2019) that are closest to the beginning of the period. That is: for 1991-1994 and 1995-1998, the LPSs identified in 1991; for 1999-2002 and 2003-2006, the LPSs identified in 2001; and for 2007-2010 and 2011-2014, the LPSs identified in 2011. The end result is that we use 3 delimitations of LPSs and not 6, but we still consider it sufficient to allow us to observe the changes in space-time dynamics.

According to their productive characteristics, the procedure allows the identification of up to nine categories of LPSs, which, for parsimony, we have aggregated into six homogeneous types of LPS

For the analysis of the evolution of the iMID effect, its causes and the factors that predict it, we use a mix of traditional methods and new machine learning methods. At first, the temporal dynamics of the iMID effect is analysed using descriptive statistics, box plots, density plots, transition arc plots, and GIS maps.

In a second stage, we use a Knowledge Production Function (KPF) to estimate the determinants and predictors of the innovative intensity of the LPSs. The KPF (Griliches, 1979; Pakes & Griliches 1984) relates innovation to R&D inputs. The KPF is modified to incorporate local economic characteristic (Anselin, Varga, & Acs, 2000) which are related to idiosyncratic effects associated to each typology of LPSs, denoted by δ (Boix & Galletto, 2009). Unlike Boix, Galletto & Sforzi (2019), we have space-time dynamic units, which allows us to take into account not only the LPS typology at present, but in previous periods. For this we introduce a variable δ_{t-1} that takes into account the typology of the LPS in the previous period t-1.

Since the effects of R&D on innovation are not immediate (Griliches, 1979; Pakes & Griliches, 1984), the input is lagged a period in the model. As the number of innovations of a place is directly related to the size of the place, the output and the input factors are divided by the total number of employees. The KPF takes the form

$$i_{t,j} = \gamma r_{t-1,j}^{\beta} \delta_{t,k} \delta_{t-1,k} \varepsilon \quad (1)$$

where i is the average innovation per employee; r is the average R&D per employee; t refers to the time period; j refers to the LPS; k refers to the typology of LPS; γ, β and δ are parameters and ε is a nuisance.

Taking logarithms, the KPF can be transformed into a log-linear expression

$$\log i_{t,j} = \gamma + \beta \log r_{t-1,j} + \delta_{t,k} + \delta_{t-1,k} + \varepsilon \quad (2)$$

The model, in its form (1) or (2), can be estimated to obtain the effect of the type of LPS on the innovative intensity.

For the estimation of the KPF, Boix & Galletto (2009), Boix & Trullén (2010) and Galletto & Boix (2014), used Heckman fixed effects, whereas Boix, Galletto & Sforzi (2019) used a nonparametric approach based on a quantile regression with non-additive fixed effects. Here, we propose the use of a flexible machine learning method - Conditional Regression Trees (CRT) - and a comparison with an estimation using quantile regression with fixed effects.

Regression trees are "machine-learning methods for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition" (Loh, 2011, p.14). There are several variants of the method. We use Conditional Regression Trees (Hothorn et al. 2006), which combine recursive binary partitioning with conditional inference based on permutation. Conditional Regression Trees have several advantages, since they are: nonlinear and nonparametric, with well-defined theoretical background (conditional inference), unbiased recursive partitioning and unbiased variable selection, robust against collinearity, avoids overfitting, and are a "white box model" that allows for transparent and easy interpretation. They also have some limitation: the results are not interpreted as in usual regressions (although the output is intuitive and provides a different approach), and they are less robust than other methods based on resampling (e.g. Random Forest or Boosting). CRT are more addressed to prediction than to causal explanations, although since we depart from a KPF, their results could be interpreted also as causal here. Since CRT results are compared with descriptive methods and quantile regression, they help to provide a more detailed explanation of the geography of innovation processes.

The results of the descriptive statistics show that the innovative intensity in MIDs has reduced from 2,089 patents per million employees in 1991-1994 to 1,783 in 2011-2014 The iMID effect (difference in innovative intensity of MIDs regarding the national average) is still positive although it has reduced from 29 to 11.

Conditional Regression Trees estimates suggest that the basic factor explaining the differences in innovative intensity between the LPSs is the present and past typology of LPS. Those LPSs that in the previous were MIDs, Manufacturing LPSs of large firm, or Business Services LPSs, show an innovative intensity much higher than the average, and falling between one of these categories in the current period also predicts a better performance. Firm R&D intensity is particularly important by predicting differences in innovative intensity between LPSs that were in one of these three categories in the

previous period but have transformed in Other Services LPSs. Public expenditures in R&D have slow predictive performance, due to its concentration on a few LPSs.

Quantile regression with fixed effects also suggest that the iMID effect maintains during all the period, with an average effect about 27% higher than the national average (with oscillations). The highest differential regarding other types of LPSs come from those LPSs that were MID in past and present period, Manufacturing LPS of large firm that have transformed in MIDs, and MIDs that become Large firm LPSs.

We conclude that, using dynamic units, the iMID effect is positive during all the period 1991-2014 - including two crises - although it has reduced over years. The decreasing seems to be related to endogenous factors. When the innovative intensity is broken down in R&D effect and territorial effect, using a territorial knowledge production function, we find that the part of the iMID effect explained by the territorial component does not decreases for the MIDs as a whole, which is similar to the results that Boix, Galletto & Sforzi (2019) obtained using static units. Since the use of dynamic units allows to break down the effect of the typological transformations, we can now observe that the highest differential regarding other types of LPSs come from those LPSs that were MID during all the period, Manufacturing LPS of large firm that have transformed in MIDs, and MIDs that become Large firm LPSs, although they follow different trends.

The use of new machine learning methods provides additional points of view and results complementing the traditional approaches.

These conclusions reveal a complex scenario for top-down innovation policies (national or regional), since suggests different responses from different types of local production systems. The spatial and typological evolution of LPSs in short and medium periods of time (e.g. 10 years) makes difficult to set the target for top-down innovation policy, leading to a discussion about the individualized and flexible responses against more general and stable policies.

References

Anselin, L., Varga, A. & Acs Z.J. (2000). Geographic and sectoral characteristics of academic knowledge externalities, Papers in Regional Science, 79 (4), 435–443.

Becattini, G. (2001). The Caterpillar and the Butterfly. An Exemplary Case of Development in the Italy of the Industrial Districts. Firenze: Le Monnier.

Becattini, G. (2004), Industrial Districts. A new Approach to Industrial Change, Cheltenham: Edward Elgar.

Bellandi, M. (1992). The incentives to decentralized industrial creativity in local systems of small firms, Revue d'Economie Industrielle, 59, 99–110.

Boix, R. & Galletto, V. (2009). Innovation and industrial districts: A first approach to the measurement and determinants of the I-district effect, Regional Studies, 43 (9), 1117–1133.

Boix, R. & Trullén, J. (2010). Industrial districts, innovation and I-district effect: Territory or industrial specialization?, European Planning Studies, 18 (10), 1707–1729.

Boix, R., Galletto, V., Sforzi, F., Llobet, J. (2015): "Sistemas locales de trabajo y distritos industriales en España en el año 2011", XLI Reunión de Estudios Regionales, Reus 18-20 Novembre.

Boix, R, Galletto, V. & Sforzi, F. (2018). Pathways of innovation: The I-district effect revisited. In F. Belussi & J.-L. Hervás-Oliver (Eds.), Agglomeration and Firm Performance, pp. 25–46. Springer.

Boix, R, Galletto, V. & Sforzi, F. (2019). Place-based innovation in industrial districts: The long-term evolution of the iMID effect in Spain (1991–2014). European Planning Studies. DOI 10.1080/09654313.2019.1588861.

Sforzi F, Boix R (2019): "Territorial servitization in Marshallian industrial districts: the industrial district as a place-based form of servitization", Regional Studies, 53(2), 398-409.

Brusco, S. (1986). Small firm and industrial districts: the experience of Italy, Economia internazionale, XXXIX (2-3-4), 85–97.

Dei Ottati, G. (2006). El 'efecto distrito': algunos aspectos conceptuales de sus ventajas competitivas, Economía Industrial, 359, 73–87.

Galletto, V. & Boix, R. (2014). Distritos industriales, innovación tecnológica y efecto I-distrito: ¿Una cuestión de volumen o de valor?, Investigaciones Regionales, 30, 27–51.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth, Bell Journal of Economics, 1979, 10 (1), 92–116.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey, Journal of Economic Literature, XXVIII, 1661–1707.

Griliches, Z. (1992). The search for R&D spillovers, Scandinavian Journal of Economics, 94, 29–47.

Hothorn T, Hornik K, Zeileis A (2006): Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651–674

Khan, M. & Dernis, H. (2006). Global Overview of Innovative Activities from the Patent Indicators Perspective, OECD Science, Technology and Industry Working Papers, 2006/03, Paris: OECD Publishing.

Loh WY (2011): "Classification and regression trees", WIREs Data Mining and Knowledge Discovery, 1, 14-23.

Marshall, A. (1930). Principles of Economics (Eighth edition). London: Macmillan.

Pakes, A. & Griliches, Z. (1984). Patents and R&D at the Firm Level: A First Look. In Z. Griliches (Ed.), R&D, Patents and Productivity, pp. 52–72. Chicago: University of Chicago Press.

Sforzi F, Boix R (2019): "Territorial servitization in Marshallian industrial districts: the industrial district as a place-based form of servitization", Regional Studies, 53(2), 398-409.

Signorini, L.F. (1994). The price of Prato, or measuring the industrial district effect, Papers in Regional Science, 73(4), 369–392.