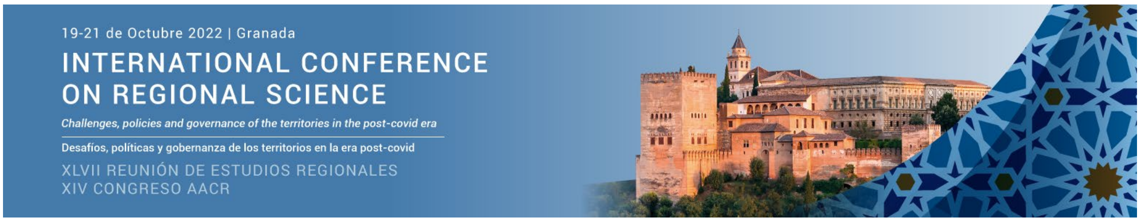

EXTENDED ABSTRACT

Title: Machine Learning to Measure the Effect of Cultural and Creative Industries on Wealth: an application to countries, regions and cities

Authors and e-mail of all: Rafael Boix Domenech (1), Jordi Sanjuan Belda (2), Pau Rausell Köster

Department: (1) Estructura Econòmica; (2) Economia Aplicada

University: Universitat de València

Subject area: *S06 - Artificial intelligence, creative industries and robotics*

Abstract: This paper focuses on the use of Machine Learning (ML) methods to measure the effects of Cultural and Creative Industries (CCI) on wealth and productivity, which are objectives of economic and development policies.

CCI can be defined as ""those industries that are based on cultural values, cultural diversity, individual and/or collective creativity, skills and talent with the potential to generate innovation, wealth and jobs through the creation of social and economic value, in particular from intellectual property" (European Parliament, 2016, p.10).

The problem is the following: during most of the 20th century, the cultural and creative industries (CCI) were a phenomenon fundamentally linked to the improvement in well-being, but not to improvements in productivity, wealth or growth (Potts and Cunningham, 2008; Boix and Rausell, 2018). Since the end of the 1990s, the intense growth of the CCIs in some countries with high levels of development called attention to their potential as an instrument and objective of growth and development policies. All this has sparked an intense debate on the contribution of the CCIs and their role in the structural processes of development and change in countries, regions and cities (UNCTAD, 2008; UNESCO, 2012; Boix, De Miguel and Rausell, 2021).

However, the real effects of the CCIs on such relevant aspects as the economy and well-being are far from being clear. The naivest version of economic policies interprets that the effects of the CCIs are always going to be positive. However, these effects can be both positive and neutral or negative, as shown by the recent works by Boix, Peiró and Rausell (2021) and Boix, De Miguel and Rausell (2021). These two papers also show how the measurement method or the way it is used affects the measured effect: for example, the violation of the linearity assumption in Ordinary Least Squares (OLS) produces upwardly biased estimates, and the use of the mean or median hide the heterogeneity of the effects and that these are negative for some places.

Given the economic dimension of creative industries in countries, regions and cities with high and low levels of income and development, the relevance that policies

related to CCI have in the public budgets of some places, and the limitations of the methods for measuring their impacts, it is convenient to develop new methodological approaches to measure the contribution of CCIs to the economy and well-being or improve existing ones.

This is where Machine Learning (ML) comes into play. In his book Introduction to Machine Learning, Alppayding (2010, p.3) defines ML as "programming computers to optimize a performance criterion using example data or past experience". That is, ML designates a set of methods based on algorithms that "learn" from the data, and that for each data set are capable of identifying patterns, making predictions or even elaborating causal explanations.

In practice, using the ML tag will lead to three qualitative improvements. First, the introduction of a new "culture" based on data science and computer modelling, as Leo Breiman explains in his well-known 2001 article Statistical modeling: The two cultures. That is, a change in the general philosophy under which we approach the statistical modelling using flexible algorithms (as opposed to rigid data models) to draw conclusions from the data. According to Breiman (2001), placing greater emphasis on the problem and the data than on the models themselves. This philosophy has been extraordinarily enhanced in recent years by the emergence of causal approaches in ML, inspired by such emblematic contributions as those of Judea Pearl and The Book of Why (Pearl and Mckenzie, 2018).

This new orientation will lead to a second improvement, and that a large number of methods and algorithms are included under the ML label. Some of them are more traditional and well-known from traditional statistics (e.g., ordinary least squares, logistic regression, instrumental variables), sometimes renewed with less conventional forms of application (e.g., cross validation), whereas other are new methods from data science and computer modelling (e.g., trees, forests, neural networks). This battery of new methods is growing at an extraordinary rate, fuelled by the need to solve new problems and use "unconventional" data, including big data and non-tabular data such as those derived from voice and image recognition.

A third improvement is the way of understanding, presenting, and communicating the results that these new methods make possible. Here are two examples. First, the way a rule-based method or a Regression Tree presents the results can be much more intuitive for a policy maker than a regression coefficient and suggest a richer and more complex orientation of policy strategies (Athey and Imbens, 2019). Second, the ability of non-parametric methods to observe "effect heterogeneity" (Athey and Imbens 2019). This is, not only the average effects but also effects for individuals or for groups of interest. Boix, Peiró, and Rausell (2021) and Boix, De Miguel, and Rausell (2021) show that this last property is essential in the case of the CCI, as it makes it possible to identify in which regions a policy based on increasing the weight of CCI can have potentially positive or negative effects, or in which regions these effects may be greater. In practice, all these improvements will imply a more flexible way of approaching and solving modelling problems, a tremendous expansion of the available methods, and greater precision for the design of economic policy strategies.

ML has begun to be applied recently to the resolution of economic problems. The well-known article by Hal Varian (2014), current Chief Economist at Google, positions himself on the need to start using these methods in economics. Initially, the opinion seems to have spread that ML methods find their place in the econometric toolbox as prediction tools, compared to traditional econometric models, more aimed at parameter estimation (e.g., Mullainathan and Spies, 2017). However, the use of ML methods has also quickly spread to parameter estimation problems, where they show their ability for causal inference, experimental design and optimal resource allocation policies. Susan Athey's papers (2019) on The impact of Machine Learning in Economics, and Susan Athey's with Nobel laureate in Economics Guido Imbens on

Machine Learning Methods that Economists Should Know About (Athey and Imbens, 2019) are excellent introductions. to the state of the art and applications of ML to measure causal effects and solve policy problems.

And in the field of CCI? Until now, the ML-based approach seems not to have been used for the analysis of the economic effects of CCI. For this reason, the objective of this paper is to introduce ML methods in the analysis of the economic effects of CCI, specifically the effects of CCI on the wealth of places and provide examples of their application.

We depart from the framework and data used by Boix, De Miguel & Rausell (2021) for the measurement of the impacts of the CCIs on the GDP per capita of countries, regions and municipalities. Then we use this framework to discuss and test the performance of different machine learning methods in each context, the differences that they produce on the estimated impacts, and the usability of those methods to evaluate the impact of the CCIs.

This empirical application has some peculiarities that are worth highlighting. First, we use three databases with three different scales of analysis (countries, regions, cities). It encompasses three sample sizes: a small sample (less than 100 observations), a small-medium size (almost 300 observations) and a medium sample of more than 1000 observations. Second, the measurement of effects is based on a causal framework based on a structural generator model derived from endogenous growth theory. This point is important because it presupposes that the results of the methods will not be biased as a consequence of the model (unconfoundedness), and also that the number of variables used by the final structural model will be small.

The paper is structured as follows. After the introduction, section 2 explains the ML techniques used in the empirical part. Section 3 presents the analytical model and the databases. Section 4 presents and discusses the results. Finally, section 5 draws conclusions.

References

Alpaydin E (2010): Introduction to Machine Learning (Adaptive Computation and Machine Learning Series). MIT Press.

Athey S and Imbens G (2019): “Machine learning methos that economists should know about”, Annual Review of Economics. 11:685-725.

Breiman L (2001): “Statistical Modeling: The Two Cultures”, Statistical Science Vol. 16, No. 3, pp. 199-215

Boix R, De Miguel B and Rausell P (2021): “The impact of cultural and creative industries on the wealth of countries, regions and municipalities", European Planning Studies, DOI: 10.1080/09654313.2021.1909540.

Boix R, Peiró J and Rausell P (2021): “Creative industries and productivity in the European regions. Is there a Mediterranean effect?”, Regional Science Policy & Practice, 13(5), p.1546-1564

Boix R and Rausell P (2018): “The Economic Impact of the Creative Industry in the European Union ”, in V. Santamarina-Campos, M. Segarra-Oña (eds.), Drones and the Creative Industry. Springer. p.19-36.

European Parliament (2016): Report on a coherent EU policy for cultural and creative industries(2016/2072(INI)). European Parliament, Brussels.

Mullainathan S and Spiess J (2017): “Machine Learning: An Applied Econometric Approach”, Journal of Economic Perspectives, 31(2), pp. 87-106.

Pearl J. and Mackenzie D. (2019): The book of why. Penguin Books.

Potts J and Cunningham S (2008): “Four models of the creative industries”. International Journal of Cultural Policy 14(3):233–247.

UNCTAD (2008): Creative economy Report 2010. Geneva - New York: UNCTAD.

UNESCO (2012): Measuring the economic contribution of cultural industries. UNESCO, Geneva.

Varian HR (2014): “Big Data: New Tricks for Econometrics” Journal of Economic Perspectives, American Economic Association, vol. 28(2), pages 3-28.

Keywords: *cultural and creative industries, machine learning*
JEL codes: R11, R12