

19-21 de Octubre 2022 | Granada

INTERNATIONAL CONFERENCE ON REGIONAL SCIENCE

Challenges, policies and governance of the territories in the post-covid era

Desafíos, políticas y gobernanza de los territorios en la era post-covid

XLVII REUNIÓN DE ESTUDIOS REGIONALES

XIV CONGRESO AACR



RESUMEN AMPLIADO

Título: Industrias culturales y creativas y bienestar regional: una evaluación de sus efectos mediante el aprendizaje automático causal.

Autores y e-mail de todos ellos: Jordi Sanjuán-Belda (jordi.sanjuan@uv.es)¹, Rafael Boix-Domènech (rafael.boix@uv.es)², Pau Rausell-Köster (pau.rausell@uv.es)¹.

Departamento: ¹ Economía Aplicada, ² Estructura Económica (Economía Aplicada II)

Universidad: Universitat de València

Área Temática: S06 – *Inteligencia artificial, industrias creativas y robótica.*

Resumen: (*mínimo 1500 palabras*)

1. Introducción

Los desequilibrios económicos, sociales y ambientales que han salido a relucir a partir de la crisis financiera de 2008, la pandemia de la Covid-19 o la acuciante emergencia climática ponen de relieve la necesidad de redirigir las economías regionales hacia nuevos modelos productivos sostenibles y centrados en el bienestar de la población. Para ello, resulta imprescindible obtener evidencia cuantitativa sobre los impactos que la estructura productiva regional puede generar sobre distintos ámbitos del bienestar.

En particular, las Industrias Culturales y Creativas (ICC) han suscitado un creciente interés en las últimas décadas y han sido señaladas, tanto desde la esfera académica como desde las instituciones públicas, como un posible vector de generación de crecimiento económico (p. ej. Boix-Domènech, De-Miguel-Molina, et al., 2021; Boix-Domènech & Soler-i-Marco, 2017; De-Miguel-Molina et al., 2012; Marco-Serrano et al., 2014) y de bienestar, si bien este último aspecto solo ha sido abordado empíricamente de forma indirecta, centrándose en la participación cultural individual y no en el papel de las ICC en la economía regional (p. ej. Blessi et al., 2016; Fancourt & Finn, 2019; Grossi et al., 2011, 2012; Taylor et al., 2015).

Las nuevas técnicas que se han desarrollado en el ámbito de la inteligencia artificial y el aprendizaje automático (o *machine learning*) ofrecen a su vez enormes potencialidades y permiten abordar con mayor rigurosidad problemas de la complejidad del que aquí se discute. El trabajo presentado es, de hecho, una de las primeras aplicaciones de métodos avanzados de *machine learning* al estudio de los impactos regionales de las ICC. Tan solo se han producido, recientemente, algunas primeras incursiones para la medición de sus efectos en el PIBpc (Boix-Domènech, De-Miguel-Molina, et al., 2021; Boix-Domènech, Rausell-Köster, et al., 2021), pero muchas de sus

potencialidades están todavía por explotar. En este trabajo se amplía dicho análisis al conjunto de dimensiones de bienestar que componen el *Better Life Index* elaborado por la OCDE, en su versión de ámbito regional. Esto es: vivienda, ingresos, empleo, comunidad, educación, medio ambiente, compromiso cívico, salud, satisfacción con la vida, seguridad y acceso a servicios.

La presente investigación tiene como objetivo identificar y cuantificar los impactos de las ICC sobre cada uno de dichos componentes del bienestar, y se realiza sobre el conjunto de regiones (en su mayoría a nivel NUTS 2 y, en algunos casos, NUTS 1) de los países europeos de la OCDE. Se trata del primer estudio, al menos del que tengamos conocimiento, acerca de los efectos de las ICC a escala regional sobre un conjunto amplio de componentes del bienestar.

2. Metodología

2.1. Datos

Los datos utilizados corresponden a las regiones de todos los países europeos que son miembros de la OCDE. Aunque el propósito inicial era utilizar el nivel NUTS 2, en algunas regiones se ha tenido que optar por el nivel NUTS 1 debido a cambios en las fronteras a lo largo de la serie temporal o, en algunos países, por cuestiones de disponibilidad de los datos de bienestar. El resultado es una muestra de 209 regiones (176 regiones NUTS 2 y 33 regiones NUTS 1) de un total de 26 países.

Se ha elaborado una serie temporal de 2008 a 2019 para este conjunto de regiones a partir de diversas fuentes. En particular, los datos de empleo provienen de una explotación específica de la *Labour Force Survey*. La lista de actividades económicas, siguiendo la Clasificación Nacional de Actividades Económicas (CNAE) 2009, que incluimos dentro de las ICC se define a partir de la clasificación de actividades culturales que utiliza el INE en su cuenta satélite de Cultura, junto con otras aplicadas en varios estudios académicos (Boix-Domènech et al., 2015; Marco-Serrano et al., 2014).

Por otra parte, los indicadores de bienestar provienen del *Better Life Index*, elaborado por la OCDE, que define 11 dimensiones del bienestar incluyendo aspectos relacionados con las condiciones materiales, la calidad de vida y la sostenibilidad del bienestar futuro. Se utilizan, naturalmente, los indicadores disponibles en la versión regional. Todos ellos parten de cuatro principios:

Además, cumplen una serie de características técnicas idóneas para nuestra investigación: resultan fáciles de interpretar; son comúnmente aceptados y utilizados como medidas de bienestar por la comunidad académica y estadística; son susceptibles de ser alterados con intervenciones públicas; están basados, en la mayoría de los casos, en datos oficiales que se actualizan periódicamente; y pueden ser comparados, en un marco bastante armonizado, entre los países de la OCDE (Durand, 2015). Se ha recurrido a las fuentes originales de los distintos indicadores para completar la serie temporal, además de aplicar un proceso de *in-filling* para estimar los datos vacíos.

Finalmente, la base de datos se completa con una serie de variables que requieren los modelos causales y que actuarán como variables de control, mediadoras o instrumentales (p. ej. densidad poblacional). Estas provienen de fuentes muy diversas, si bien se ha procurado recurrir a fuentes oficiales como Eurostat siempre que fuera posible.

2.2. Técnicas

El reciente desarrollo de métodos de aprendizaje automático (*Machine Learning*) aplicables a grandes bases de datos ofrece nuevos instrumentos para tratar con problemas complejos como el que aquí se aborda. Haremos uso de técnicas avanzadas de *machine learning* y de minería de datos preparadas para el análisis causal (Athey et al., 2019), pero evitando “cajas negras”. Esto es, utilizamos técnicas que nos permitan:

(1) obtener resultados causales, interpretables y fundamentados en la lógica económica, ya que se persigue entender los procesos e identificar las cadenas de transmisión a través de las cuales se producen los impactos; (2) obtener estimaciones de estos impactos a nivel global (conjunto de regiones) y local (para cada región en particular); y (3) que sean de aplicabilidad directa para la política económica.

En la fase actual de la investigación, se aplicarán *Random Forest* (Breiman, 2001; Strobl et al., 2008), así como *Causal Forest*, un desarrollo posterior de los primeros que permite extraer interpretaciones causales a partir de una variable de tratamiento (Athey et al., 2019; Athey & Imbens, 2016; Wager & Athey, 2018).

Es importante resaltar que, aunque inicialmente los métodos de *machine learning* tenían el inconveniente de su interpretación causal, su reciente integración con los marcos causales de Rubin y Pearl (Imbens & Rubin, 2015; Pearl, 2019; Pearl & Mackenzie, 2018) ha engendrado una nueva generación que explota la estructura interna de los algoritmos generando resultados causales mediante contrafactuales, y generalmente supera ya a la de los métodos tradicionales. En particular, nos apoyaremos en herramientas tales como *partial dependence plots* (PDP), *individual conditional explanation* (ICE) o *local interpretable model-agnostic explanations* (LIME) (véase Goldstein et al., 2015; Ribeiro et al., 2016). Estos instrumentos permiten obtener efectos marginales a partir de los cuales aproximar pseudo-elasticidades, así como extraer conclusiones diferenciadas para diferentes regiones.

En próximas fases de la investigación, se utilizará también la técnica de estimación lineal no paramétrica *local-linear least-squares* (LLLS) para determinar relaciones de tipo causal (Henderson & Parmeter, 2015) y obtener, en su caso, estimaciones locales para cada región del análisis. En caso de relaciones causales cruzadas, se aplicarán también modelos de ecuaciones estructurales (SEM). Finalmente, dada la naturaleza dinámica de los datos, cabe también la posibilidad de usar modelos de sistemas dinámicos para integrar la retroalimentación de los sistemas con el estudio de las políticas económicas, sociales y culturales.

2.3. Modelos causales

Para estudiar la relación entre las ICC y las diferentes dimensiones del bienestar, se definen los modelos causales a través de diagramas causales guiados por el criterio de ajuste de puerta trasera de Pearl y Mckenzie (2018). Esto permite identificar los llamados caminos causales evitando factores de confusión.

Se parte, tomando como base la dimensión de ingresos, del modelo de crecimiento semiendógeno de Romer-Jones (Jones, 1995, 2001), adaptado y ampliado por Boix et al. (2021) incluyendo las ICC. Las ICC actúan como promotoras de la innovación, en tanto que su fuerza de trabajo contribuye en mayor medida a la generación de ideas para el resto de la economía a través de la creatividad.

A partir de este primer modelo, se despliegan el resto relacionando los indicadores de las distintas dimensiones del bienestar. Para identificar los determinantes de cada una de ellas, se acude a la literatura académica y se adoptan modelos ya configurados y validados, en ocasiones adaptados a la naturaleza de este estudio. Por ejemplo, para la esperanza de vida, se toma como referencia el modelo de Bilas et al. (2014); para la participación electoral, el de Sundström y Stockemer (2015); para el acceso a conexión de banda ancha, el de Flamm & Chaudhuri (2007) y el de Reddick et al. (2020), etc. En otros casos, como en la vivienda (número de habitaciones por persona) o la comunidad (porcentaje de la población que cree poder confiar en alguien en caso de necesidad), no había estudios previos de características similares o asimilables, por lo que se han elaborado modelos propios a partir del razonamiento teórico.

3. Resultados preliminares

Esta investigación pretende aportar nuevas evidencias y mejorar la comprensión respecto a los impactos de las ICC sobre el bienestar de las regiones y cada una de las dimensiones que lo forman. Se trata del primer estudio que aborda los efectos de las ICC sobre el bienestar para una muestra amplia de regiones y un conjunto de indicadores de bienestar. Proporcionará, por tanto, evidencia generalizada sobre la dirección e intensidad de estos efectos sobre cada una de las facetas del bienestar estudiadas.

Asimismo, los resultados extraídos de los modelos causales y predictivos pueden constituir una fuente de información útil para las instituciones públicas a la hora de adoptar políticas económicas.

En el momento de enviar el presente resumen, todavía no se han podido desarrollar y testar todos los modelos, pero los resultados se desglosarán con mayor detalle durante la celebración del Congreso en caso de aceptación de la propuesta de comunicación. Con todo, sí que se han realizado algunas primeras pruebas con *random forest* en algunas de las dimensiones del bienestar, las cuales arrojan resultados preliminares potencialmente interesantes pero que todavía necesitan ser refinados y validados.

Las primeras estimaciones confirman, por un lado, los hallazgos previos sobre el impacto positivo de las ICC en el PIBpc regional (Boix-Domènech, De-Miguel-Molina, et al., 2021; Boix-Domènech & Soler-i-Marco, 2017; De-Miguel-Molina et al., 2012; Marco-Serrano et al., 2014). Además, su efecto marginal es creciente para mayores niveles de empleo en ICC. Ocurre lo mismo en el caso de la participación electoral (indicador del *compromiso cívico*), para la cual la presencia de ICC resulta ser un determinante nada desdeñable. Un efecto positivo más modesto se identifica en otras dimensiones como la salud (esperanza de vida), el acceso a servicios (porcentaje de hogares con conexión de banda ancha) o la comunidad (porcentaje de la población que cree poder confiar en alguien en caso de necesidad). Sin embargo, también se aprecian efectos negativos, si bien leves, en otras dimensiones. En particular, en la vivienda (número de habitaciones por persona).

Referencias

- Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Arthur M. Sackler Colloquium on Drawing Causal Inference from Big Data*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), 1179–1203. <https://doi.org/10.1214/18-AOS1709>
- Bilas, V., Franc, S., & Bošnjak, M. (2014). Determinant factors of life expectancy at birth in the European union countries. *Collegium Antropologicum*, 38(1), 1–9.
- Blessi, G. T., Grossi, E., Sacco, P. L., Pieretti, G., & Ferilli, G. (2016). The contribution of cultural participation to urban well-being. A comparative study in Bolzano/Bozen and Siracusa, Italy. *Cities*, 50, 216–226. <https://doi.org/10.1016/j.cities.2015.10.009>
- Boix-Domènech, R., De-Miguel-Molina, B., & Rausell-Köster, P. (2021). The impact of cultural and creative industries on the wealth of countries, regions and municipalities. *European Planning Studies*. <https://doi.org/10.1080/09654313.2021.1909540>
- Boix-Domènech, R., Hervás-Oliver, J. L., & De-Miguel-Molina, B. (2015). Micro-geographies of creative industries clusters in Europe: From hot spots to assemblages. *Papers in Regional Science*, 94(4), 753–772. <https://doi.org/10.1111/pirs.12094>
- Boix-Domènech, R., Rausell-Köster, P., & Sanjuán-Belda, J. (2021). Assessing the impact of cultural and creative industries on the wealth of the regions through causal Machine Learning. *International Conference on Regional Science*.

- Boix-Domènech, R., & Soler-i-Marco, V. (2017). Creative service industries and regional productivity. *Papers in Regional Science*, 96(2), 261–279. <https://doi.org/10.1111/pirs.12187>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- De-Miguel-Molina, B., Hervás-Oliver, J. L., Boix-Domènech, R., & De-Miguel-Molina, M. (2012). The Importance of Creative Industry Agglomerations in Explaining the Wealth of European Regions. *European Planning Studies*, 20(8), 1263–1280. <https://doi.org/10.1080/09654313.2012.680579>
- Durand, M. (2015). The OECD better life initiative: How's life? And the measurement of well-being. *Review of Income and Wealth*, 61(1), 4–17. <https://doi.org/10.1111/roiw.12156>
- Fancourt, D., & Finn, S. (2019). *What is the evidence on the role of the arts in improving health and well-being? A scoping review*.
- Flamm, K., & Chaudhuri, A. (2007). An analysis of the determinants of broadband access. *Telecommunications Policy*, 31(6–7), 312–326. <https://doi.org/10.1016/j.telpol.2007.05.006>
- Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44–65. <https://doi.org/10.1080/10618600.2014.907095>
- Grossi, E., Blessi, G. T., Sacco, P. L., & Buscema, M. (2012). The Interaction Between Culture, Health and Psychological Well-Being: Data Mining from the Italian Culture and Well-Being Project. *Journal of Happiness Studies*, 13(1), 129–148. <https://doi.org/10.1007/s10902-011-9254-x>
- Grossi, E., Sacco, P. L., Blessi, G. T., & Cerutti, R. (2011). The Impact of Culture on the Individual Subjective Well-Being of the Italian Population: An Exploratory Study. *Applied Research in Quality of Life*, 6(4), 387–410. <https://doi.org/10.1007/s11482-010-9135-1>
- Henderson, D. J., & Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*. Cambridge University Press.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences: an introduction*. Cambridge University Press.
- Jones, C. I. (1995). R & D-Based Models of Economic Growth Charles I. Jones. *Journal of Political Economy*, 103(4), 759–784.
- Jones, C. I. (2001). *Introduction to Economic Growth*. WW Norton & Company.
- Marco-Serrano, F., Rausell-Köster, P., & Abeledo-Sanchis, R. (2014). Economic development and the creative industries: a tale of causality. *Creative Industries Journal*, 7(2), 81–91. <https://doi.org/10.1080/17510694.2014.958383>
- Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54–60. <https://doi.org/10.1145/3241036>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic Books.
- Reddick, C. G., Enriquez, R., Harris, R. J., & Sharma, B. (2020). Determinants of broadband access and affordability: An analysis of a community survey on the digital divide. *Cities*, 106(September), 102904. <https://doi.org/10.1016/j.cities.2020.102904>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional

- variable importance for random forests. *BMC Bioinformatics*, 9(1), 307.
<https://doi.org/10.1186/1471-2105-9-307>
- Sundström, A., & Stockemer, D. (2015). Regional variation in voter turnout in Europe: The impact of corruption perceptions. *Electoral Studies*, 40, 158–169.
<https://doi.org/10.1016/j.electstud.2015.08.006>
- Taylor, P., Davies, L., Wells, P., Gilbertson, J., Tayleur, W., Christy, E., Cooley, E., Taylor, A., Jones, R., & Dumas, V. (2015). *A review of the Social Impacts of Culture and Sport* (Issue March). <http://shura.shu.ac.uk/id/eprint/9596%0A>
- Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>

Financiación

Este proyecto ha recibido financiación del programa de investigación e innovación de la Unión Europea Horizonte 2020 bajo el acuerdo de subvención nº 870935.

Esta investigación ha sido financiada por el Ministerio de Universidades a través del programa de Formación del Profesorado Universitario.

Palabras Clave: *industrias culturales y creativas, bienestar, machine learning, economía regional*

Clasificación JEL: I31, R11, Z11