**Estimating Cultural Tourists in Spanish Regions through Machine Learning**

Jordi Sanjuán[1], Pau Rausell-Köster[1] and Fernando Álvarez-Teresa[1]

[1]Econcult Research Unit, Department of Applied Economics, Universitat de València (Spain)

**Biographies:**

*Jordi Sanjuán* is an economist, Master in Economic Policy and Public Economics and PhD in Social Sciences. He is currently a researcher in the Department of Applied Economics at the Universitat de València and a member of the research unit in Economics of Culture and Tourism (Econcult).

*Pau Rausell-Köster* is an economist and professor of the Department of Applied Economics. Director of the Research Area in Economics of Culture and Tourism (Econcult). He has led and participated in numerous European and national competitive projects in the field of cultural economics.

*Fernando Álvarez-Teresa* is a pre-doctoral researcher in the Department of Applied Economics at the Universitat de València. Member of the research group ECONCULT and of the association for the promotion of research in creative industries and economic impact INNOPAC.

**Abstract.** Cultural tourism is receiving increasing attention and plays a significant role in shaping the global tourism landscape. However, defining cultural tourism remains ambiguous, and there are no standardized methods to measure its extent. This study introduces a novel methodology to quantify cultural tourists in Spanish regions. The proposed approach involves calculating the elasticity of tourist flows in a region concerning the dimension of the cultural and creative sectors. To achieve this, machine learning techniques, in particular Causal Forest, are applied, employing a comprehensive database that gathers information from European regions spanning the years 2008 to 2019. A counterfactual scenario is simulated, assuming the absence of cultural and creative workers in each region in order to identify the number of overnight stays attributed to cultural tourism. The results indicate that cultural tourism accounts for 18.6% of total overnight stays in Spain, though there are important differences between regions.

**Keywords:** cultural tourism; machine learning; causal forest; Spanish regions; cultural and creative sectors

## Introduction

Culture, regardless of how it is defined, undoubtedly plays a significant role in people's travel choices. Considering culture in a broad sense to encompass the interest in different lifestyles and objects from other cultures, it becomes evident that most tourists, at some point, engage with cultural products—whether in authentic or commodified forms, akin to theme parks. Additionally, virtually every tourist destination offers some degree of cultural experiences.

Almost three decades ago, Greg Richards and Carolina Bonink (1995) provided a definition of cultural tourism as “the movement of individuals to cultural attractions away from their usual place of residence with the intention of acquiring new information and experiences to satisfy their cultural needs”. As per the definition adopted by the UNWTO General Assembly during its 22nd session in 2017, Cultural Tourism refers to “a type of tourism activity where the primary motivation of the visitor is to learn, explore, experience, and engage with the tangible and intangible cultural attractions and products offered by a tourism destination”. These attractions and products encompass a unique combination of material, intellectual, spiritual, and emotional aspects of a society. This includes arts and architecture, historical and cultural heritage, culinary heritage, literature, music, creative industries, as well as the living cultures and their associated lifestyles, value systems, beliefs, and traditions.

Ultimately, cultural tourism can be seen as the act of traveling with the purpose of engaging in cultural experiences to fulfil one's cultural needs. However, the ever-expanding notion of tourists' cultural consumption has made the definition of cultural tourism increasingly elusive (Noonan & Rizzo, 2017). If we take a broad interpretation of cultural tourism, it might encompass almost any form of tourism since every tourist activity involves some level of interaction with the local culture. Alternatively, we can adopt a narrower approach that includes only those tourist activities intentionally seeking interactions with cultural information and where cultural needs serve as the essential motivation for travel. This narrower approach focuses on specific “types” of culture, which can be categorized into two main aspects. The first aspect revolves around “objects”, encompassing buildings, museum artefacts, paintings, crafts, and sculptures found in galleries. The second aspect revolves around “performances”, such as theatre plays, music, dance performances in theatres or concert halls, or observing the creation of arts and crafts (Hughes, 2002). In any case, it is widely accepted that the study of the cultural tourist's experience remains relatively limited in the literature (Seyfi et al., 2020).

In some moment around the 80’s, cultural tourism was recognized as specific niche of the “mass tourism” (Richards, 2020). Since then, an increasing fragmentation into various niche segments, including creative tourism, festival tourism, gastronomic tourism, and literary tourism, prompts the question of whether there is still an aggregate concept of “the cultural tourist”.

Nowadays, with the threat of tourism-phobia on the rise, interest in cultural tourism is growing, because it implies a narrative that contains a certain moral superiority. Since the inception of tourism, the Grand Tour – as an archetype of cultural tourism- has been associated with elevating the soul and intimately linked to immersing individuals in classical culture, representing a quest for knowledge and self-improvement (Paquette & Wright, 2021). By adopting the ways of the ancient travelers and embracing cultural experiences along the way, individuals embark on a transformative path of personal growth. In this sense, culture is not seen as a complementary supply but the core motivation of the travel.

But from another perspective it seems that the increasing cultural content of travel is not a consequence of a growing interest in culture but rather of what Richards (2021) highlights as the *Cultural Turn of Tourism Industry*, which transformed tourism in a cultural practice. This means that including cultural products in tourism packages is a supply-side strategy to enrich and complement travel demand. This implies that in mature markets and in very competitive contexts culture may enrich the attributes of the tourism product, but demand does not really increase its willingness to pay for the cultural add-on.

Whether from a supply-side perspective or from the demand-side motivations, it is clear that becoming *a cultural tourist* necessitates access to cultural resources for consumption, which, in turn, relies on a certain level of cultural capital or competence. Additionally, cultural tourism holds personal significance, such as the opportunity to acquire knowledge or reaffirm one's identity. In contemporary cultural tourism, diverse practices embody these aspects, often converging at specific sites, destinations, or periods of time.

In any case, the implications of having more or fewer cultural tourists affect the characteristics of destinations and the quality and quantity of impacts generated on them. There are many studies that analyze the differential effects that cultural tourism generates compared to other types of tourism from economic perspectives (Noonan & Rizzo, 2017), social value creation (Cannas, 2018), reputational and branding (Kostopoulou et al., 2022), quality of life (B. McKercher & Ho, 2012), cultural (Basri et al., 2021; Guccio et al., 2017, 2018) and even sustainability (Durovic & Lovrentjev, 2014) impacts. Hence, considering the varying impacts that cultural tourism can have, the pivotal issue at hand is determining the extent and significance of its presence.

Defining a cultural tourist is undoubtedly a challenging task, and even more so is accurately determining the number of these tourists amidst the vast ocean of tourism (Bonet, 2013). The ATLAS Cultural Tourism Project's initial attempts to quantify the number of cultural tourists revealed a smaller percentage of people who self-identified as such (5–10%) compared to broader measures that encompassed all visitors to cultural attractions (40%) (Richards & Munsters, 2010). The focus of this effort was primarily on gauging the market's size, and relatively simple methods were employed, leading to impressive figures that sparked considerable enthusiasm about the potential of this undertaking (Bob McKercher, 2020).

**Cultural tourism: a new definition**

In statistical terms, four fundamental methods exist for quantifying cultural tourists, all approached from the lens of tourism demand. The initial method relies on self-identification, where a cultural tourist is one who self-identifies through questionnaires. However, given the multifaceted nature of most trips, a second approach involves querying tourists about their primary motivations in questionnaires. For instance, the Frontur-Egatur survey by the Spanish Statistical Institute (INE) encompasses 16 options, including cultural tourism, alongside others with a broad cultural scope, such as religious or gastronomic tourism

Both approaches hinge on subjective self-perception, susceptible to the bias of the "cultural" concept. A more objective method involves observing whether tourists undergo cultural experiences during their trips (Richards, 2018), yet defining cultural experiences poses challenges due to the broad nature of culture. Determining the level of cultural experiences required for one to be considered a cultural tourist remains an open question. Some studies link sociodemographic characteristics, such as age, income, and education, to the condition of being a cultural tourist (Richards & van der Ark, 2013), but the concept of "cultural tourists" evolves over time. Despite longstanding studies emphasizing culture's role in regional tourist attraction, the extent of its impact remains unclear (Ritchie & Zins, 1978).

Our paper introduces an unconventional methodology providing enhanced accuracy and reliability. We employ a counterfactual approach to identify and quantify "cultural tourists" based on the supply side. If cultural and creative activities vanished –heritage, performing arts festivals, handcrafts, local guides, museums– what portion of tourism would disappear? This approach sheds light on the indispensable role of cultural and creative sectors in attracting tourists.

We chose Spain to test this methodology firstly because Spain is the most competitive tourist destination in the world and has been producing tourism statistics of a high level of excellence for decades. Moreover, with the detailed knowledge of the Spanish tourism reality, it is relatively easy to interpret the coherence of the results at regional level since there are island and coastal regions where the core product is the sun and beach, while there are other regions where it can be identified that cultural and creative elements explain a large part of the tourism flows.

To obtain a more accurate quantitative measurement of cultural tourism, we need to refine the definition used in the preceding paragraphs. This involves considering cultural tourism as encompassing *the number of tourists who would cease to visit a territory if it lacked any cultural resources.* While it may be challenging to envision a territory completely devoid of cultural resources, our operational approach suggests that we should imagine a scenario where there are no individuals professionally involved in producing market cultural goods and services, effectively reducing the number of people working in the cultural sectors to zero. Building upon this approach, we have formulated a novel methodological proposition for assessing the quantitative magnitude of cultural tourism in the various Spanish regions.

**New methodologies to extract cultural tourism from overall tourism**

If we were to find a mechanism to estimate the elasticity of visitors to cultural activities, we could make the assumption that these activities are reduced to zero and check what number of tourists continue to visit the territory. With the simple operation of subtracting the current visitors from those who would remain in the territory if all cultural activities were to disappear, we would obtain the number of cultural tourists.

Despite the relatively small overall difference between the proportions of tourists and overnight stays, we have chosen to emphasize the calculation of overnight stays. This decision is because overnight stays provide a more accurate representation of the tourism impact on a region, avoiding distortions caused by varying average stays resulting from different cultural resources. There are not many studies, but in general, the evidence suggests that those tourists who prioritize culture tend to reduce their length of stay in a statistically significant way (Aguilar & Díaz, 2019; de Menezes et al., 2008). For instance, visits to urban cultural destinations are usually shorter and concentrated in specific times and locations, while cultural visits to rural areas with natural and cultural resources tend to be longer and spread across the territory.

***Model***

To develop this strategy, we set up a model in which the total number of nights spent by tourists during a year in a region is determined by:

- the cultural offer and dynamism of the region, the main motivation for cultural tourism;
- the total population of the region, which influences its capacity to receive tourism;
- the tourism competitiveness of the country where the region is located, based on multiple factors; and
- tourism services and facilities in the region, measured in bed places.

In addition, travel decisions are often made some time in advance, and are based on factors that have information transaction costs, which are not instantaneous but are formed over time (e.g., reputation of the destination). Therefore, we consider a time lag of one year between each of the variables involved in the model and the outcome in terms of tourism overnight stays. This also allows us to circumvent or smooth out possible distortions created by endogeneity and double causality; i.e., the pre-existing cultural offer may attract tourists, but the tourist flow could also constitute a source of demand for the cultural offer to develop further. To verify the existence of this dual link, Granger's (non-)causality test for panel data (Dumitrescu & Hurlin, 2012) is applied, which confirms that there is a causal relationship, with a time lag of one year, of both CCS affecting overnight stays ($\tilde{Z} = 2.22$, p-value = 0.026) and tourist overnights affecting CCS ($\tilde{Z} = 3.36$, p-value = 0.001). This results in the following equation:

$$Overnights_{i,t} = f\left(CCS_{i,t-1}, Population_{i,t-1}, TTCI_{i,t-1}, Bedplaces_{i,t-1}\right) + \varepsilon_{i,t} \qquad (1)$$

where, for a given region *i* in year *t*, *Overnights* represents the total number of nights spent by tourists, *CCS* represents the share of employment in Cultural and Creative Sectors, *Population* represents the total population, *TTCI* represents the Travel & Tourism Competitiveness Index of the country where the region is located, *Bedplaces* represents the total number of bed-places in tourist accommodation, and $\varepsilon$ represents the random error attributable to factors not included in the model.

It is crucial to highlight that the model's results, obtained through the employed techniques, indicate causality. Hence, any changes in the model's variables lead to corresponding causal variations in the dependent variable. To illustrate, if we were to manipulate the cultural resources of a territory, specifically the number of individuals employed in the cultural and creative sectors, in a controlled environment, we would observe corresponding variations in the number of overnight stays of tourists in that same territory.

### ***Data***

For the estimations, a regional database has been compiled covering the European OECD countries, resulting in a total of 209 regions. Data have been collected on an annual basis from 2008 to 2019, forming panel data. The 12-year time series, by introducing a time lag of one year between the dependent variable and the independent variables, remains at 11 periods. The following indicators are used to capture the variables identified in the model in data.

The dependent variable, *overnights*, is the total annual overnight stays in all tourist accommodation establishments in a region. Data are from Eurostat.

As for the treatment variable, *CCS* is defined as the percentage of workers in these sectors over total employment in the region. It also comes from Eurostat through a specific extraction from the Labor Force Survey. For the classification of sectors considered within the CCS, we start from the proposal of the project for the European Commission “Measuring CCS in the EU” (Vilares et al., 2022), although with some small adaptations because the disaggregation of the data only allows to reach the 3-digit detail in the NACE codes of economic activities. Thus, the activities considered are: Printing and reproduction of recorded media (NACE 18); Manufacture of jewelry, bijouterie and related articles (NACE 32.1); Manufacture of musical instruments (NACE 32.2); Publishing activities (NACE 58); Motion picture, video and television program production (NACE 59); Programming and broadcasting activities (NACE 60); Advertising (NACE 73.1); Specialized design activities (NACE 74.1); Photographic activities (NACE 74.2); Translation and interpretation activities (NACE 74.3); Creative, arts and entertainment activities (NACE 90); and Libraries, archives, museums and other cultural activities (NACE 91). These sectors include the activities that produce and communicate heritage services, which allow the production of festivals, handcrafted souvenirs purchased by tourists, audiovisual services that serve as communication and branding of tourist destinations, etc.

Some other covariates are introduced into the model, namely:

- *Population*: Population of the region on 1 January, the source being Eurostat.
- *TTCI*: Travel & Tourism Competitiveness Index. This index comprises up to 90 indicators covering enabling environment, travel and tourism policy and enabling conditions, infrastructure, and natural and cultural resources. The data is at country level, on a scale of 1 to 7, and the same value is applied to all regions in each country. The index is compiled every two years, in odd-numbered years, so in even-numbered years the average value of the two adjacent years is assigned. The source is the World Economic Forum (WEF) (Uppink Calderwood & Soshkin, 2019).
- *Bedplaces*: total number of bed-places in all tourist accommodation establishments, the source being Eurostat.

Table 2 summarizes the descriptive statistics of the variables used in the model.

Table 2. Descriptive statistics.

| | Avg. | Std. Dev. | Max. | Min. | N |
|---|---|---|---|---|---|
| *Overnights*$_t$ | 13,249,321 | 16,884,272 | 123,882,181 | 126,378 | 2,299 |
| *CCS*$_{t-1}$ | 2.49 | 1.32 | 9.35 | 0.33 | 2,299 |
| *Population*$_{t-1}$ | 2,319,091 | 2,448,383 | 17,996,621 | 27,153 | 2,299 |
| *TTCI*$_{t-1}$ | 4.94 | 0.42 | 5.68 | 3.84 | 2,299 |
| *Bedplaces*$_{t-1}$ | 143,531 | 152,843 | 894,605 | 764 | 2,299 |

Source: Own elaboration from Eurostat and WEF.

### *Method: Causal Forest*

In recent years there have been important developments in the field of machine learning for causal inference (Athey & Imbens, 2019; Jacob, 2021). This research uses causal forest technique, a rather novel ML algorithm developed by Wager & Athey (2018), for estimation. It is a further development of the random forest algorithm that allows for the identification of heterogeneous treatment effects in observational studies.

The primary goal of the Causal Forest algorithm is to estimate the causal effect of a treatment variable (in our case, CCS) on an outcome variable (tourist overnights) while accounting for potential heterogeneity in treatment effects across different units or observations (regions).

Causal Forest provides both global and local estimates of the causal effect of the treatment variable under unconfounding (this is, if the causal model is correctly specified). In the presence of heterogeneous effects among regions (as expected), this allows to identify which regions of the sample are likely to benefit most by the individual treatment effect depending on the covariates (Athey & Imbens, 2019). Causal Forests have been shown to perform well in situations where traditional methods may struggle due to heterogeneity in treatment effects. They provide a flexible and powerful approach for estimating causal effects in observational studies while addressing issues related to confounding and other biases.

It is important to note that only the estimate for the effect of the treatment variable is obtained, not for the rest of features, as these do not have a direct causal interpretation. Causal forests pose no problems when using variables with very different scales and variances, so we will not apply algorithms or any other mathematical transformation to the variables. In addition, the algorithm does not need to assume any rigid and restrictive assumption, and is able to capture complex interactions that occur between variables.

The main advantages of this technique are therefore that 1) the results are straightforward and transparent, leading to causal interpretation if the model is well-specified, 2) the flexibility of the model allows for a fairly accurate fit, and 3) it provides local estimates that take into account heterogeneous effects, which is our main objective. Consequently, we can obtain differentiated and reliable estimates for each region based on its particular features, which we could not do with other more standard approaches such as OLS. Thus, Causal Forest achieves an optimal balance in the trade-off between accuracy and interpretability, as well as being based on solid theoretical foundations in causal logic (Wager & Athey, 2018). Despite the novelty of this technique, there have already been very recent applications to the study of the regional impacts of CCS on other economic and social indicators (Boix-Domènech et al., 2022; Sanjuán, 2023).

### Results

The overall results of the model are presented in Table 3. It is found that the effect of CCS on tourist overnight stays is statistically significant and positive. For the set of regions analyzed, on average, an additional percentage point of CCS employment increases the number of total tourist overnight stays in the region in the following year by 1.26 million.

Table 3. Main results of the model.

| | **Estimate** | **Std. Error** | **t value** | **Pr(>\|t\|)** | |
|---|---|---|---|---|---|
| Average treatment effect | 1,262,725 | 125,331 | 10.08 | 0.000 | *** |
| *Calibration tests:* | | | | | |
| Mean forest prediction | 0.87 | 0.04 | 22.21 | 0.000 | *** |
| Differential forest prediction | 1.03 | 0.07 | 14.34 | 0.000 | *** |

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The calibration tests are satisfactory. In addition to being statistically significant, the coefficients of both should be reasonably close to 1 (Wager & Athey, 2018). Values close to 1 for ‘mean forest prediction’ indicate that the estimate of the average treatment effect is accurate, while values close to 1 for ‘differential forest prediction’ indicate that the heterogeneity of the effects is well calibrated. Both conditions are met.

***Results for the Spanish regions***

In addition to the average marginal effect, the Causal Forest technique reports a differentiated marginal effect for each region on the basis of the heterogeneity of the effects. From these regional estimates, our approach is to take each region's marginal effect (the average value from 2009 to 2019) that one percentage point of CCS employment would cause and multiply it by the weight of CCS in the region. That is, if, for a region, an additional percentage point of CCS employment would lead to 1 million more overnight stays, and the region has 2% of employment in CCS, we can assume that this 2% CCS employment is responsible for 2 million overnight stays that would not exist if there were no CCS employment. If we put this amount in relation to the total number of tourist overnight stays, we obtain the percentage of overnight stays attributable to CCS. The results indicate that cultural tourism accounts for 18.6% of total overnight stays in Spain, above the traditional method which accounted for 10.1% percent of overnight stays.

The primary notable finding is that, with the exception of Cantabria, all Spanish regions exhibit positive elasticities. This indicates that the cultural activity generated by professionals in the cultural sectors, relative to the total number of workers, has a substantial impact on the number of tourist overnight stays. The range of elasticities varies from 0.65 to almost 0, with Cantabria being the exception at -0.02. This negative value suggests that if all cultural resources were to vanish in Cantabria, the number of visitors might actually increase. This could be interpreted as suggesting that in the absence of cultural tourists, there might be a substitution effect whereby other types of tourists could potentially prolong their stays, leading to an increase in the overall number of overnight stays. To check that there have been no errors or inconsistencies, we have repeated the exercise using the number of visitors as the dependent variable, verifying that the elasticity in Cantabria is in this case positive, albeit small.

The findings demonstrate a high level of consistency with the reality of tourism in the Spanish regions. Galicia ranks first in terms of the average number of overnight stays, which can be attributed to the significance of cultural and religious motivations associated with the Pilgrim's Way to Santiago. In this region, cultural tourism constitutes two-thirds of all tourism.

Following closely is Madrid, one of Spain's major urban destinations (alongside Barcelona), where cultural tourism accounts for almost half (47%) of the total overnight stays. Catalonia and Andalusia, two large coastal regions, also see cultural tourism contributing to a quarter of their overnight stays, a proportion similar to Castile and Leon. Despite having a more modest number of tourists, Castile and Leon boasts a considerable number of heritage resources, compensating for the absence of sun and beach tourism.

Table 4. Results: Percentage of overnight stays explained by the existence of CCS in Spanish Regions.

| NUTS | Region | Overnight stays 2019 | Overnight stays explained by the existence of CCS (%) |
|---|---|---:|---:|
| ES11 | Galicia | 10,938,537 | 65% |
| ES30 | Madrid | 28,975,266 | 47% |
| ES42 | Castile-La Mancha | 5,394,480 | 33% |
| ES21 | Basque Country | 8,142,554 | 32% |
| ES51 | Catalonia | 84,140,872 | 26% |
| ES61 | Andalusia | 72,044,756 | 25% |
| ES41 | Castile-Leon | 11,737,934 | 24% |
| ES64 | Melilla | 146,310 | 21% |
| ES23 | La Rioja | 1,639,201 | 18% |
| ES12 | Principality of Asturias | 5,778,325 | 17% |
| ES24 | Aragon | 8,302,457 | 17% |
| ES63 | Ceuta | 167,989 | 15% |
| ES62 | Region of Murcia | 5,410,417 | 15% |
| ES22 | Navarre | 3,277,283 | 15% |
| ES52 | Valencian Country | 50,063,663 | 13% |
| ES53 | Balearic Islands | 68,376,034 | 13% |
| ES43 | Extremadura | 3,581,316 | 12% |
| ES70 | Canary Islands | 96,113,149 | 1% |
| ES13 | Cantabria | 5,583,008 | -2% |

Source: Own elaboration and Eurostat.

Conversely, the lower end of the table includes prominent island destinations like the Canary Islands and the Balearic Islands, as well as regions highly specialized in coastal tourism such as the Valencian Country and the Region of Murcia on the Mediterranean coast. It is noteworthy that in the Canary Islands, only 1% of its nearly 100 million overnight stays can be attributed to cultural tourism. Looking at the regions as a whole, cultural tourism appears to contribute approximately 18.6% of all overnight stays. This aligns well with the data presented in the introductory paragraphs of the article and underscores the significant impact of culture on the overall landscape of tourism.

The results therefore show high levels of consistency and are in line with our knowledge of regional tourism in Spain, opening a fruitful way for the estimation of cultural tourism in other regions of Europe.

## Discussion and conclusions

There is little contention regarding the importance of cultural tourism as a major component of overall tourism flows, and numerous authors concur that it is a steadily expanding and worldwide phenomenon. Despite this consensus, official statistics often fail to adequately capture its full extent, making the object of analysis rather elusive (Noonan & Rizzo, 2017). The provided figures vary, ranging from data collected from individuals who claim culture as their primary reason for travel to encompassing all tourists who engage in cultural visits, which could potentially represent up to 40% of overall tourism.

And we have to add that the cultural tourism market is not uniform and fixed; instead, similar to the broader tourism market, it comprises tourists with progressively diverse characteristics, needs, and expectations (Pulido-Fernández & Sánchez-Rivero, 2010).

This article introduces innovative and consistent methodologies that help refine and specify these approaches, without relying on demand-side characteristics considerations. Instead, the definition of cultural tourists is derived by exclusion, referring to those visitors who would not choose to visit the region if there were no cultural market resources available in the territory. Thus, this definition of cultural tourists is conditioned by the attributes of the tourist supply present in the territory.

Recent papers have already incorporated machine learning techniques to predict tourism flows and demand in Spain (Claveria et al., 2016; Perles-Ribes et al., 2020); but this is, to our knowledge, the first to apply them to the measurement of cultural tourism. The application of a machine learning algorithm designed for causal inference tasks gives us the opportunity to make this proposal. Causal Forest utilizes the strengths of ensemble learning from Random Forest, which is widely used for supervised learning tasks, such as classification and regression, and integrates it with causal inference techniques to estimate treatment effects in observational data. Thus, it allows to comprehend cause-and-effect connections between variables when it is not feasible challenging to conduct randomized controlled trials.

The underlying logic of this methodological approach is straightforward. The aim is to calculate the elasticity of tourist flows concerning the cultural content of a particular region and then utilize this elasticity in a counterfactual manner, simulating a scenario where the cultural resources cease to exist. This allows us to observe the potential impact on tourist flows if the cultural resources were to disappear.

However, this macro-level approach requires some abstractions, so it is not exempt from limitations. On the one hand, employment in CCS is a highly aggregated variable, and may include some sectors that have little to do with cultural tourism, such as advertising or television. On the other hand, regions can have strong internal heterogeneity, including multiple specializations in urban, rural or coastal tourism.

All in all, the results obtained exhibit high coherence in analyzing the tourism landscape of the Spanish regions and shed light on cultural tourism figures that surpass those recorded by conventional statistics, which rely on asking tourists about the primary reason for their trip. However, these figures avoid the overestimation that might occur if all individuals engaging in any cultural activity were classified as cultural tourists. Based on our findings in Spain, cultural tourists contribute to 18.6% of the total overnight stays. However, regional variations are considerable, ranging from 65% in Galicia to merely 1% in a prominent destination like the Canary Islands.

The accuracy and consistency of these results, with optimal goodness-of-fit measures, indicate a promising approach for estimating cultural tourism in numerous OECD regions. Furthermore, the relative ease of the methodology implementation and the availability of data for most tourism territories suggest that it has the potential to become an accepted standard in assessing cultural tourism.

**References**


Aguilar, M. I., & Díaz, B. (2019). Length of stay of international tourists in Spain: A parametric survival analysis. *Annals of Tourism Research*, *79*(August), 102768. https://doi.org/10.1016/j.annals.2019.102768

Athey, S., & Imbens, G. (2019). Machine Learning Methods Economists Should Know About. *Annual Review of Economics*, *11*, 685–725. https://doi.org/10.1146/annurev-economics080217-053433

Basri, H., Harriami, D. R., Rahadi, I., Sugandi, Y. B. W., Ramli, M., Zaitun, & Pauzi, M. H. (2021). Cultural Tourism: The Socio-Cultural Impact on Moyo Festival, Sumbawa-Indonesia. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, *12*(8), 162–176.

Boix-Domènech, R., Sanjuán, J., & Rausell-Köster, P. (2022). Machine Learning to Measure the

Effect of Cultural and Creative Industries on Wealth: an application to countries, regions and cities. *XLVII International Conference on Regional Science*. https://reunionesdeestudiosregionales.org/granada2022/actas-del-congreso/
Bonet, L. (2013). Heritage tourism. In I. Rizzo & A. Mignosa (Eds.), *Handbook on the economics of cultural heritage* (pp. 286–401). Cheltenham: Edward Elgar.
Cannas, R. (2018). Diverse Economies of Collective Value Co-creation: The Open Monuments Event. *Tourism Planning and Development*, *15*(5), 535–550. https://doi.org/10.1080/21568316.2018.1505651
Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, *23*(6), 428–431. https://doi.org/10.1080/13504851.2015.1078441
de Menezes, A. G., Moniz, A., & Vieira, J. C. (2008). The determinants of length of stay of tourists in the Azores. *Tourism Economics*, *14*(1), 205–222. https://doi.org/10.5367/000000008783554866
Dumitrescu, E.-I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling*, *29*(4), 1450–1460. https://doi.org/10.1016/j.econmod.2012.02.014
Durovic, M., & Lovrentjev, S. (2014). Indicators of sustainability in cultural tourism. *The Macrotheme Review*, *3*(7), 180–189.
Guccio, C., Lisi, D., Martorana, M., & Mignosa, A. (2017). On the role of cultural participation in tourism destination performance: an assessment using robust conditional efficiency approach. *Journal of Cultural Economics*, *41*(2), 129–154. https://doi.org/10.1007/s10824-017-9295-z
Guccio, C., Lisi, D., Mignosa, A., & Rizzo, I. (2018). Does cultural heritage monetary value have an impact on visits? An assessment using official Italian data. *Tourism Economics*, *24*(3), 297–318. https://doi.org/10.1177/1354816618758729
Hughes, H. L. (2002). Culture and tourism: A framework for further analysis. *Managing Leisure*, *7*(3), 164–175. https://doi.org/10.1080/13606710220000013701
Jacob, D. (2021). CATE Meets ML - Conditional Average Treatment Effect and Machine Learning. *Digital Finance*, *3*(2), 99–148. https://doi.org/10.1007/s42521-021-00033-7
Kostopoulou, S., Sofianou, P. K., & Tsiokanos, K. (2022). Silk Road Architectural Heritage and Polycentric Tourism Development. *Advances in Science, Technology and Innovation*, 33–58. https://doi.org/10.1007/978-3-030-95564-9_3
McKercher, B., & Ho, P. (2012). Cultural Tourism and the Enhancement of Quality-of-Life. In Uysal, M., R. Perdue, & M. Sirgy (Eds.), *Handbook of Tourism and Quality-of-Life Research: Enhancing the Lives of Tourists and Residents of Host Communities* (pp. 105–132). Springer. https://doi.org/10.1007/978-94-007-2288-0
McKercher, Bob. (2020). Cultural tourism market: a perspective paper. *Tourism Review*, *75*(1), 126–129. https://doi.org/10.1108/TR-03-2019-0096
Noonan, D. S., & Rizzo, I. (2017). Economics of cultural tourism: issues and perspectives. *Journal of Cultural Economics*, *41*(2), 95–107. https://doi.org/10.1007/s10824-017-9300-6
Paquette, J., & Wright, C. (2021). The Morality of Cultural Itineraries and Tourism: The Ethics of the Grand Tour. In J. Paquette, A. Lacassaggne, & C. Alcantara (Eds.), *Cultural Roads and Itineraries. Concepts and Cases* (pp. 129–151). Palgrave Macmillan. https://doi.org/10.1007/978-981-16-3533-5
Perles-Ribes, J. F., Ramón-Domínguez, A. B., Moreno-Izquierdo, L., & Such-Devesa, M. J. (2020). Machine learning techniques as a tool for predicting overtourism: The case of Spain. *International Journal of Tourism Research*, *22*, 825–838. https://doi.org/10.1002/jtr.2383
Pulido-Fernández, J. I., & Sánchez-Rivero, M. (2010). Attitudes of the cultural tourist: A latent segmentation approach. *Journal of Cultural Economics*, *34*(2), 111–129. https://doi.org/10.1007/s10824-010-9115-1
Richards, G. (2020). Culture and tourism: natural partners or reluctant bedfellows? A perspective paper. *Tourism Review*, *75*(1), 232–234. https://doi.org/10.1108/TR-04-2019-0139
Richards, G. (2021). *Rethinking cultural tourism*. Edward Elgar Publishing.
Richards, G., & Bonink, C. (1995). Marketing cultural tourism in Europe. *Journal of Vacation Marketing*, *1*(2), 172–180. https://doi.org/10.1177/135676679500100205
Richards, G., & Munsters, W. (2010). *Cultural tourism research methods.* (G. Richards & W. Munsters (eds.)). Cabi.

Richards, G., & van der Ark, L. A. (2013). Dimensions of cultural consumption among tourists: Multiple correspondence analysis. *Tourism Management*, *37*, 71–76. https://doi.org/10.1016/j.tourman.2013.01.007

Ritchie, J. R. B., & Zins, M. (1978). Culture as determinant of the attractiveness of a tourism region. *Annals of Tourism Research*, *5*(2), 252–267. https://doi.org/https://doi.org/10.1016/0160-7383(78)90223-2

Sanjuán, J. (2023). *Cultural and Creative Industries and the Well-Being of Regions* [Universitat de València]. https://hdl.handle.net/10550/90744

Seyfi, S., Hall, C. M., & Rasoolimanesh, S. M. (2020). Exploring memorable cultural tourism experiences. *Journal of Heritage Tourism*, *15*(3), 341–357. https://doi.org/10.1080/1743873X.2019.1639717

Uppink Calderwood, L., & Soshkin, M. (2019). *The Travel and Tourism Competitiveness Report 2019: Travel & Tourism at a Tipping Point*. World Economic Forum. https://es.weforum.org/reports/travel-and-tourism-competitiveness-report-2019

Vilares, M., Rausell, P., Martínez, C., Boix, R., Sanjuán, J., Álvarez, F., Cabedo, A., Cordero, J., Alves, B., Colaço, D., Fonseca, I., Borrego, P., Justo, S., Bulian, G., & Turon i Loren, J. (2022). *Measuring the Cultural and Creative Sectors in the EU*. Publications Office of the European Union. https://www.measuring-ccs.eu/the-measuring-ccs-consortium-publishes-the-final-report/

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