The signalling effect of eco-labels in modern coastal tourism

Cerqua, A.

This is an Accepted Manuscript of an article published by Taylor & Francis in the Journal of Sustainable Tourism, doi:10.1080/09669582.2016.1257014 Published online: 25 Nov 2016

The final definitive version is available online:

https://dx.doi.org/10.1080/09669582.2016.1257014

The signalling effect of eco-labels in modern coastal tourism

Abstract: As the demand for environmentally sustainable tourism grows, eco-labels are becoming increasingly popular as a signal of environmental quality. However, the existence of a causal link between awarding a seaside eco-label and the increase in tourism flows is still under discussion in the literature. In this article, we gauge the signalling impact of a specific eco-label, the Blue Flag award, using detailed data on tourism flows to seaside Italian destinations during the period 2008-2012. We adopt a recent econometric modelling strategy - the synthetic control method - in shaping estimation results and testing the sensitivity and robustness of our results. We find that being awarded the Blue Flag increases the flow of domestic tourists for up to three seasons after assignment. However, we find no effect for the flow of international tourists. Investigating the mechanisms driving the results, we find that the award of a Blue Flag only positively affects the flow of domestic tourists when it is used as a driver of organisation, coordination and integrated management of the tourism supply.

Keywords: Blue Flag; Seaside eco-label; Tourism flows; Destination competitiveness; Synthetic control method

Introduction

Over the last decades, “sustainable tourism” has become one of the keywords for tourism officials (e.g. Cucculelli & Goffi, 2016), driven by a rising demand for “greener holidays” (e.g. Esparon, Stoeckl, Farr & Larson, 2015). Indeed, tourists’ increasing awareness of environmental issues, coupled with the constant development of mass beach tourism and with the intensified competition among tourism destinations (Friedman, 2006), have created the need for high-quality and sustainable beaches capable of differentiating tourist offers in a highly competitive market. However, tourists cannot easily validate the environmental sustainability of a given tourism destination. As a result, eco-labels have emerged as a popular way for tourism destinations to signal the quality of their environmental amenities (Fraguell, Martí, Pintó & Coenders, 2016). Eco-labels can help attract tourists who prefer high quality natural amenities or are concerned by the environmental impact of their travels.

One of the widely recognised eco-labels in the tourism industry is the Blue Flag programme, awarded to seaside destinations which meet the high-standard criteria set by the Foundation for Environmental Education (FEE) in the four categories of: water quality, environmental management, environmental education and safety (Blue Flag, 2015). The Blue Flag programme aims to encourage local authorities and other tourism stakeholders to constantly enhance the environmental quality of their destinations (Creo & Fraboni, 2011). For tourists, the Blue Flag is an international recognition and a symbol of good quality beaches and marinas (Cagilaba & Rennie, 2005), and it stands out among the increasing number of seaside eco-labels (Nelson & Botterill, 2002; Blackman, Naranjo, Robalino, Alpízar & Rivera, 2014). The programme is run in 49 countries (www.blueflag.org, 2016) and is a widely recognised eco-label for the public, decision-makers and tour operators.

Despite the good features of the Blue Flag and the other eco-labels, the existence of a causal link between awarding a seaside eco-label and increasing tourism flows is still under discussion in the literature. While some papers report evidence supporting this causal link, other scholars find no relationship between the two. For instance, Capacci, Scorcu & Vici (2015) use panel data models to analyse province-level data and find that the Blue Flag award positively affects future international tourism flows, while McKenna, Williams & Cooper (2011) qualitatively examine questionnaire responses finding that Blue Flags have only a marginal effect on tourists' choices.

We contribute to the literature in three ways. First, on the substantive side, we evaluate the signalling impact of the Blue Flag award with respect to tourism flows using data from official sources for both domestic and international tourists. Besides, with the exception of Capacci et al. (2015), the literature so far has analysed the signalling impact of seaside eco-labels on tourism flows using qualitative approaches; therefore, we fill the clear need for a sound data-driven evaluation. Second, this is the first article to disentangle the signalling impact on domestic and inbound tourism flows and to investigate the heterogeneity of these two groups of tourists. On the methodological side, we use data at the Italian tourism areas (*circoscrizioni turistiche*); they are the smallest territorial units (one municipality or a set of municipalities) the Italian Statistics Bureau (ISTAT) provides tourism flows data for, namely the number of arrivals and length of stay. This allows us to pinpoint the signalling effect of the Blue Flag award with a smaller margin of error than other approaches used in the past (see Capacci et al., 2015). Third, we adopt a recent econometric modelling strategy - the synthetic control method (SCM), first proposed by Abadie & Gardeazabal (2003) - in shaping estimation results and test the sensitivity and robustness of our results. Differently from other counterfactual methods, the SCM allows gauging the effect of the Blue Flag award on tourism flows looking at one awarded locality at a time. In fact, the SCM can estimate the time-varying heterogeneous effect of the Blue Flag award, while a standard diff-in-diffs or fixed effects panel estimation can only provide an estimate for the time-invariant average treatment effect (Chung, Lee & Osang, 2016).

The geographical focus of the article lies in Italy, more specifically in all Italian seaside destinations. Italy stands out among European countries with its 5,507 bathing waters, 25.6% of all bathing waters in Europe (European Environment Agency, 2015). Over the last few years, the Great Recession contributed to a sensible drop in the number of domestic arrivals at Italian tourist accommodation establishments (UniCredit, 2016). On the other hand, Italy experienced an increase in the number of international arrivals but at a slower pace than other European countries, mostly due to the emergence of new tourism destinations and to the change in tourism demand trends. The intensified competition, particularly from other Mediterranean countries, had a particular negative impact on the Italian seaside tourism segment because of a reduced “value for money” in relative terms of most coastal destinations and the lack of policies able to satisfy the more and more demanding modern tourist (Ferrero, 2015). In fact, tourism is a highly changeable industry and as such tourism officials must be prepared to adjust their marketing strategies to changes in the economy, in transportation and in tourist preferences.

Background

In a resource-based industry such as tourism, where the balance between economic and environmental goals is challenging (e.g. Hall, Gössling & Scott, 2015), it is difficult for people to figure out whether tourism destinations are environmentally friendly, even if this is what local organisations claim. Eco-labels can validate these claims from an objective point of view (Graci & Dodds, 2015) and might therefore persuade tourists to visit awarded destinations. Seaside eco-labelling schemes are commonly voluntary in nature and destinations have to meet certain established criteria in order to be awarded the eco-label (Cagilaba & Rennie, 2005). Over the last decades environmental certification schemes have flourished (Fairweather, Maslin & Simmons, 2005) and nowadays there are over 100 tourism eco-labels worldwide (Gössling & Buckley, 2016).

The development of eco-labels in the tourism industry began in France in 1985 with the creation of the Blue Flag programme for beaches and marinas. This programme has been operating in Europe since 1987 by the FEE (Font, 2002) and it meets all the criteria for an independent, trustworthy and objective environmental seal (Mihalič, 2000). Over the years it has become a highly recognised eco-label (Eijgelaar, Nawijn, Barten, Okuhn & Dijkstra, 2016) with the dual aim of preventing environmental damage and attracting tourism (Lucrezi & Saayman, 2015). It challenges local authorities and beach operators to achieve high standards in the following four categories: i) environmental education and information (e.g., litter removal and community participation); ii) water quality (e.g., bathing water quality is monitored by means of weekly or fortnightly water sampling); iii) environmental management; iv) safety and services (Blue Flag, 2015). The main goals of the Blue Flag programme are to improve the understanding of the coastal environment and to promote the incorporation of environmental issues in the decision-making processes of local authorities and their partners (FEE, 2007). Some Blue Flag criteria are imperatives, like the water quality criteria or litter bins in adequate numbers (including recycling bins), while others are merely guidelines, such as sustainable public transportation. A location that does not comply with one or more of the imperative criteria cannot be awarded a Blue Flag. Each year the criteria are reviewed and a number of new guideline criteria may become imperative (Cagilaba & Rennie, 2005). Blue Flags are awarded based on previous year’s activity and must be re-earned every year. Receiving the award in consecutive years shows how beaches are able to live up to the criteria. The programme was launched in Italy in 1987 when it assigned 37 Blue Flags. In recent years the number of assigned Blue Flags has steeply increased; for instance, in 2012, 246 Italian beaches, located in 131 municipalities or 69 tourism areas (TAs), were awarded the Blue Flag banner.

There is a growing literature addressing different aspects of seaside eco-labels. A first strand of literature focuses on the relationship between eco-labels criteria and the impact on sustainability. Despite years of research, there is still little consensus among scholars on the effectiveness of eco-labels in improving the environmental conditions (see, among others, Boevers, 2008; Esparon, Gyuris & Stoeckl, 2014; Zielinski & Botero, 2015; Pencarelli, Splendiani & Fraboni, 2016). Concerning the Blue Flag programme some criticism has been raised as it does not address all relevant aspects that are encompassed in beach ecosystem functions (Lucrezi, Saayman & Van der Merwe, 2015). Also the foundation of the Blue Flag award for bathing water quality has been questioned (Schernewski & Sterr, 2002). Analysing the behaviour of Italian local authorities in joining the programme, Pencarelli et al. (2016) find that the dynamics have more to do with the return in terms of destination image than with environmental concerns.

Pushing the sustainability critiques to the limit, Buckley (2002) claims that the main function of tourism eco-labels is as a market mechanism. As argued in Cagilaba & Rennie (2005), eco-labels are often perceived to be a promotional tool in the successful development of coastal tourism in a particular region and help to bring much needed revenue to local economies. Also, they are frequently used by local authorities and tourism promoters as an incentive to involve all parties concerned with environmental management, water quality and education activities (Nelson, Morgan, Williams & Wood, 2000).

As highlighted in Karlsson and Dolnicar (2016), despite this promotional push, in the literature there is still no agreement on whether eco-labels increase tourist demand for a destination.[1] While McKenna et al. (2011) find that Blue Flags play an insignificant role in the understanding of tourists' motivation to visit beaches in Ireland, Wales, Turkey and the USA, some analysts suggest that Blue Flag awards can stimulate investments and increase prices in hotels (Blackman et al., 2014; Rigall-I-Torrent, Fluvia, Ballester, Salo, Ariza & Espinet, 2011). Zielinski & Botero (2015) evaluate the effectiveness of 9 beach certification schemes in Latin America and the Caribbean, based on indicators of sustainable development and integrated coastal management. The authors find that the Blue Flag scheme differentiates from the other schemes as it is the most effective in terms of institutional indicators, rather than biophysical ones. Cucculelli and Goffi (2016) analyse the impact of the Blue Flag and other quality certifications on the competitiveness of several small Italian destinations. They demonstrate that factors directly referring to sustainability have a positive impact on all the competitiveness indicators. With a purpose

[1] See Capacci et al. (2015) for a systematic review of existing studies focusing on the changes in tourism demand and efficiency due to signalling.

similar to ours, Capacci et al. (2015) assess the effectiveness of the Blue Flag certification in attracting foreign tourists to Italian coasts. Using dynamic panel data models and province-level data (obtained by aggregating individual yearly survey data of foreign tourists in Italy over the period 2000-2012), they find that the Blue Flag awarded to a province during a specific year has a negligible effect on the number of foreign tourists in that year, but has a positive effect on inbound flows during the following year.

Theoretical framework

The tourism industry can be considered an "experience product" market characterized by product-quality information asymmetry (Nelson, 1974). Adapting the signalling theory developed for the job market by Spence (1973) to the tourism industry (see Smith & Font, 2014, for an adaptation to volunteer tourism), we argue that potential tourists suffer from asymmetric information as they are not fully aware of the "quality" of the potential seaside destinations (e.g., environment protection, sustainability, natural amenities). Local tourism administrators of high quality destinations own this information and they must decide whether or not they want to disclose it. In case they think there is a return from disclosing this information, they want to do it and eco-labels might be seen as an instrument to do so. In this scenario, the Blue Flag is supposed to reduce information asymmetries for potential tourists as they would be informed that a certain seaside destination meets a number of criteria concerning water quality, environmental management, environmental education and safety.

The Blue Flag award might provide a competitive advantage to tourism destinations given that it satisfies the two crucial requirements of an efficacious signal. First, it satisfies signal observability, which refers to the extent that potential tourists are able to acknowledge the signal. Indeed, the Blue Flag campaign is promoted via leaflets, press releases, on-site notice boards and display of award flags at qualifying beaches and is by far the most widely publicised in the media (particularly television and newspapers), which may account for its higher awareness level among beach users (Cagilaba & Rennie, 2005).[2] Second, it satisfies signal cost, which involves the fact that some seaside destinations are in a better position than others to absorb the associated costs. Requesting the Blue Flag award and meeting all the mandatory requirements is time consuming and costly (Sasidharan, Sirakaya & Kerstetter, 2002; Blackman et al., 2014; Fraboni, 2015). Through a questionnaire intended for the local tourism officials of all Italian municipalities which received the Blue Flag in 2012, Fraboni (2015) finds that most municipalities spent at least 3 months preparing all the compulsory documentation[3] and had an average monetary expenditure of over €5,000, despite the absence of fees. Accordingly, the cost of being awarded a Blue Flag is very high for seaside destinations

[2] However, a few studies (e.g. Nelson et al., 2000; McKenna et al., 2011) find that beach users have a low level of awareness of beach award schemes, and limited knowledge of their exact meanings.
[3] Conversely, a reduction in the time expenditure is possible as about a quarter of the municipalities took just 1 month to prepare the required documentation.

lacking in a number of the criteria imposed by this eco-label, whereas it is much less costly for high-quality destinations meeting almost all or all the imperative criteria.

Methods and data

Empirical approach

Whenever a beach is awarded a Blue Flag banner, it would be important to gauge the signalling effect of such an achievement on tourism flows (number of tourist arrivals) at the beach's municipality level. To do so, we would like to apply the principle behind the counterfactual impact evaluation, i.e. contrasting the actual tourism flows of the awarded municipality to the tourism flows of the same municipality were they without the Blue Flag. However, this is not feasible because, in a specific year, we cannot simultaneously observe the tourism flows of a seaside destination both in the case of receiving a Blue Flag and in the case of not receiving a Blue Flag. Holland (1986) refers to this as the fundamental problem of causal inference. This limits the researcher to estimate a counterfactual scenario capable of mimicking what would have been the tourism flows of the municipality in case the same beach was not awarded the eco-label. This is best achieved by looking at a subset of municipalities that are on average the same as the awarded one, except for the absence of the award. Although signalling theory predicts that local tourism administrators rationally decide whether it is convenient for them to receive a Blue Flag, this award might depend on idiosyncrasies at the municipality level with respect to administrative capabilities, tourism policies, timing in accommodating a new mandatory requirement, atmospheric conditions affecting sea quality and so on. This means that only a subset of municipalities is awarded the Blue Flag in a certain year, although a large number of municipalities have very similar environmental and sustainability features. The difference between the actual outcome and the counterfactual will give the treatment (signalling) effect on the awarded municipality. Repeating the same process for each awarded municipality we will get an estimate of the average treatment effect on the treated (ATT).

For the estimation process, tourism flows data on each coastal municipality would be ideal. Conversely, in the main analysis, we are going to compare tourism flow outcomes at the TA level (see Data section below), as TAs represent the smallest territorial units for which ISTAT releases data about the capacity and occupancy of tourist accommodation establishments. TAs are created aggregating adjacent municipalities which share a common typology of tourism. In the remaining of the article, we employ the policy evaluation terminology defining each TA with a newly awarded beach as treated TA.

Building a comparison group of non-recipient TAs with similar characteristics (i.e., similar number and typology of tourists, similar environmental amenities, and similar accommodation offer) to the treated TAs is a crucial requirement of any quasi-experimental econometric method. In this article we employ the SCM. This method allows choosing the best counterfactual scenario for each treated TA through a weighted average of TAs that did not receive or lose a Blue Flag award in the relevant years, with weights chosen so that the weighted average is similar to the treated TA in terms of lagged outcomes and covariates (Athey & Imbens, 2016). Besides, unlike most of the estimators used in the literature, the SCM accounts for the presence of time-varying unobservable confounders. This feature improves on panel models such as fixed effects or diff-in-diffs, which can only account for time-invariant unobservable confounders (Billmeier & Nannicini, 2013).

The counterfactual scenario will be derived from a weighted combination of non-recipient TAs sharing similar time-invariant characteristics and a similar ex-ante tourism history (i.e., before the treated TAs received the Blue Flag award). To this end, we first collect data on tourism flows for each seaside TA. Then, the SCM is used to choose a control group of TAs, for which the weighted-combined tourism flows mimic the ex-ante tourism flows of Blue Flag award recipients. An important feature of the SCM is that it forces the researcher to demonstrate the affinity between the treated unit and its synthetic counterpart (Abadie, Diamond & Hainmueller, 2010). This is important as our crucial assumption is that we are able to control for all possible determinants of the Blue Flag award and consequently we can consider the award as good as randomly assigned. This is why in case we find an average increase in tourism flows for awarded municipalities we can interpret it as reliable evidence of positive signalling.

In the following, we show how the SCM works for a generic treated TA. Let the index $j = (0,1,..., J)$ denotes Italian seaside TAs. While $j = 0$ indicates the TA receiving the Blue Flag award in a specific year, $j = (1,..., J)$ refers to each of the other J TAs that can potentially be included in the control group as they meet the selection criteria for the donor pool reported in the following section. Define X_0 as a $(k \times 1)$ vector with elements equal to the number of tourist arrivals at the treated TA in each year during the pre-assignment period plus additional covariates predictive of the number of tourist arrivals. Similarly, define X_1 as the $(k \times J)$ vector containing the same variables for each of the J TAs in the control group.

The synthetic control approach identifies a convex combination of the J TAs in the control group that best approximates the pre-assignment data vector for the treated TA. Define the $(J \times 1)$ weight vector $W = (w_1, w_2,..., w_J)$ such that all weights are non-negative and sum to one, that is $w_j \geq 0$ for $j = (1,..., J)$ and $\sum_{j=1}^{J} w_j = 1$. The product $X_1 W$ then gives us a weighted average of the pre-assignment vectors for all TAs omitting the treated TA, with the difference between the treated TA and this average given by $(X_0 - X_1 W)$. The SCM chooses a value for W such that

$$W^* = \arg\min_W \, (X_0 - X_1 W)' A (X_0 - X_1 W) \tag{1}$$

where A is a $(k \times k)$ diagonal positive-definitive vector with diagonal elements providing the relative weights for the contribution of the square of the elements in the vector $(X_0 - X_1 W)$ to the objective function being minimised.

Once W^* is selected, it is possible to tabulate both the pre-assignment path and the post-assignment values for the number of tourist arrivals in the synthetic control unit by calculating the corresponding weighted average for each year using the TAs with positive weights. Therefore, one can assess the Blue Flag award impact by simply comparing this counterfactual to the actual path observed.

Data

In the following we describe all data sources, putting within brackets the time periods analysed in the synthetic control analysis (considering t as the year in which the treated TA received the extra Blue Flag). The "occupancy in collective accommodation establishments for type of accommodation, country of residence of guests and tourism area" data are retrieved from ISTAT and are derived from a census survey which represents the main source of information on internal (domestic and inbound) tourism available in Italy. Data are disaggregated into hotels and other accommodations but we only use their aggregated value. This gives us the yearly number of tourist arrivals in each TA, our main outcome variable.[4] In addition to past realizations of the outcome (averaged from t-8 to t-5 and from t-4 to t-2), several variables are included in $\mathbf{X}_1$ and $\mathbf{X}_0$ as predictors of tourism flows. We also retrieve from ISTAT the average length of stay in a TA (averaged from t-4 to t-1), the percentage of tourists staying in an accommodation different from a hotel (averaged from t-4 to t-1) and the overall share of domestic tourists (averaged from t-4 to t-1). The last two variables are included in order to compare TAs offering a similar range of accommodations and being able to attract both domestic and international tourists.

Then, we control for the number of beds available in hotels and B&Bs for each TA (averaged from t-4 to t-1) aggregating the data at the municipality level from the ISTAT's annual census survey "capacity of tourist accommodation establishments". The rationale for this variable is that we aim to compare each treated TA to a weighted combination of non-treated TAs having a similar number of beds available. We exploit the same source of data also to compute the proportion of beds in 4 and 5 stars hotels (averaged from t-4 to t-1). This variable gives an indication of the "exclusivity" of the destination and spending capacity of the "typical" tourist.

We also control for the average satisfaction of foreign tourists concerning the environmental quality (over the three years preceding the additional Blue Flag

[4] We have chosen the number of arrivals as our main outcome variable because we consider it as the most suitable variable for testing whether tourists are truly persuaded by an eco-label to visit a specific destination. However, besides that, we also use the length of stay as an alternative dependent variable.

certification). This variable comes from the yearly survey conducted by the Bank of Italy named "International Tourism in Italy". The survey data are derived from a random sample of foreign travellers, who are leaving Italy, detailed at the municipality level and then aggregated up to the TA level.

Using information from the Italian FEE, we reconstruct the number of Blue Flags awarded for each seaside TA (averaged from t-4 to t-2).[5] We exploit the location of the TAs and the information about the Blue Flag award to create the variable "number of TAs located within a 50km ray having been awarded at least 1 Blue Flag in the period under analysis". In the same spirit we add two variables on the accessibility of the coastal destinations creating the variables "number of harbours in a 50 km ray" and "number of airports in a 50 km ray".

In addition, we have EU data on the quality of the water at the municipality level and we aggregate them up to the TA level. The data come from the EU Bathing Waters Directive (http://www.eea.europa.eu/data-and-maps/data/bathing-water-directive-status-of-bathing-water-7), which requires EU Member States to identify popular bathing places in fresh and coastal waters and monitor them for indicators of microbiological pollution (and other substances) throughout the bathing season. These data tell us whether a TA had one or more banned beach in the three years preceding the additional Blue Flag certification and allow TAs with a similar bathing water quality history to be compared.

We exploit data on the presence of at least one venue capable of hosting 500+ people to create a proxy for congress tourism, as some seaside locations might attract tourists not only for their beaches but also for the availability of venues suitable to large-scale national and international congresses. We have also access to the detailed financial statements of each municipality from 2004 to 2013 thanks to the "Open bilanci" project financed by the European Regional Development Fund. We exploit this information to control for the per capita expenditure on tourism and environment for each TA (averaged from t-4 to t-2).

Finally, we control for the number of municipalities within a TA, the length of the coast and the population in 2011.

Sample

In 2012, Italy counted 544 TAs. However, as we aim to estimate the impact of a Blue Flag on tourism flows, we limit our analysis to the 164 having access to the sea. Differently from Capacci et al. (2015), our tourism flows data do not suffer from the small sample issues deriving from the use of survey data. Nevertheless, our data do not allow disentangling seaside inbound tourism from total inbound tourism and so our variables

[5] We do not control for the dependent variable (number of tourist arrivals) at time t-1 as we express the dependent variable as the ratio to their corresponding t-1 value and so all the treated and control TAs take on the value 1 at time t-1. Additionally, we do not control for the number of Blue Flags at time t-1 as it is pre-determined given our selection criteria (see Sample section) for the treatment and the control group.

are subject to measurement error. To curb this issue we carry out a selection of TAs dropping 9 TAs for which we have clear evidence that the sea is not the main tourist attraction (e.g., Rome, Naples, Venice). Another 10 TAs are dropped because they are made up by a large number of municipalities (>15) and the total number of beds of the municipalities having access to the sea is less than 50% of the overall number of beds.

After the removal of such observations we are left with 145 seaside TAs.[6] As our main interest lies in picking up the signalling impact of the Blue Flag award, we impose a few restrictions to what we consider treated observations and what we consider control observations. We use annual TA-level balanced panel data for the period 2000-2014 and, in order to have a reasonable number of pre-treatment and post-treatment years, we define as treated those TAs which received an "extra" Blue Flag award[7] in the years from 2008 to 2012. Additionally, we restrict the treated sample using the following criteria:

(1) The treated TA did not receive an additional Blue Flag the year before. This criterion ensures that the impact of the Blue Flag is not driven by a previous change in the number of Blue Flags.

(2) The Blue Flag award has been assigned to a municipality having at least 15% of the sleeping accommodations of the treated TA and the treated TA is made up of no more than 20 municipalities. This way we limit the confounding factor of having data at the TA level instead of at the municipality level.

Applying these criteria we are left with 20 "treated" TAs (7 in 2008, 3 in 2009, 4 in 2010, 5 in 2011, and 1 in 2012). It is worth noting that all the treated TAs kept the additional Blue Flag at least for the following year. This allows us to discard the possibility that we are evaluating the impact of a "sporadic" award. Table 1 reports all treated TAs with their characteristics before and after treatment, while Figure 1 shows a summary map of the treated TAs and the TAs which make up the donor pool, highlighting those receiving a positive weight in the main analysis.

Insert Table 1

Insert Figure 1

[6] The majority of them (82) are made up of a single municipality, while 27 TAs are made up of a number of municipalities ranging between 2 and 10, 10 TAs have between 11 and 20 municipalities, 12 TAs have between 21 and 50 municipalities, and lastly 14 TAs are made up of more than 50 municipalities.

[7] We use the term "extra" to emphasise that TAs made up of more than 1 municipality might increase their number of Blue Flags in a certain year even if before they had already achieved 1 or more awards (for example going from 1 Blue Flags in 2008 to 2 Blue Flags in 2009). We have also considered extending the analysis to include the TAs experiencing the loss of a Blue Flag. Nevertheless, given that only 2 seaside destinations (Francavilla al Mare in 2010 and Rimini in 2012) satisfy all the selection criteria, the findings are not generalizable. The results are, however, available upon request.

In order to select a synthetic control group that includes TAs as similar as possible to the treated one, the original donor pool of each treated TA is reduced according to the following set of criteria:[8] 1) for each treated TA, the donor pool is made up of all non-treated TAs which had the same number of Blue Flags from t-1 to t+1. This way we exclude from our donor pool TAs which could not properly resemble the counterfactual scenario as they suffered a setback in or gained from a change in the number of Blue Flag awards (the aggregated estimates using only this criterion for the donor pool are reported in Figure A1 of Appendix A); 2) the donor pool only includes non-treated TAs located within a 200-km ray from the treated TA (the aggregated estimates using the above criteria for the donor pool are reported in Figure A2 of Appendix A). This will allow comparing only TAs located not too far away from each other, as this is supposed to pick up more general trends in tourism flows (e.g., the steady increase of tourism flows in Apulia); 3) the donor pool excludes TAs having a number of municipalities over 20 as they might not accurately resemble the counterfactual scenario of the treated TAs. Table A1 of Appendix A summarises the assumptions behind the selection of the treated TAs and of the synthetic counterparts.

Results

We present the findings in a set of graphs that compare the average tourism flow performance of the treated TAs, $\overline{Y}_{BF} = \sum_{i=1}^{20} Y_{BF,i} / 20$, with that of the synthetic control group, $\overline{Y}_{SC} = \sum_{i=1}^{20} Y_{SC,i} / 20$, where $Y_{SC,i} = \sum w_{ij} Y_{ij}$ for $j = 1, ..., J$ (see Chan, Frey, Gallus & Torgler, 2014). The timeline is adjusted so that year t-1 is the year before the Blue Flag award (indicated with a vertical dotted line in the following figures) for all treated TAs. In addition, following Munasib & Rickman (2015), we express the dependent variable as a ratio of its own pre-treatment value to avoid issues associated with differences in sizes of tourism areas. We employ the *synth, nested allopt* command in Stata developed by Hainmueller, Abadie, and Diamond (*synth package for Stata is available at https://web.stanford.edu/~jhain/synthpage.html*).

Figure 2 plots the trends in the number of tourist arrivals for the average treated TA (solid line) and the average synthetic control group (dashed line). As the chart suggests, the dependent variable in the synthetic control group very closely tracks the trajectory of the same variable in the average treated TA for the entire pre-treatment period. Combined with the high degree of balance on all predictors reported in Table 2, this suggests that the synthetic control group provides a sensible approximation to the tourism flows that

[8] We do not actually know whether the TAs included in the donor pool applied for an "extra" Blue Flag banner and did not get it, as this information is confidential. However, the number of applying TAs in the donor pool should be limited as usually about 90% of the applicants do obtain the Blue Flag award. This high proportion follows from the presence of imperative criteria in the assignment process of the Blue Flag: whether these criteria are satisfied can be indeed anticipated by local tourism administrations at the moment of applying.

would had been experienced by each treated TA in the 3 years after the treatment in the absence of the Blue Flag award.[9] The overall number of arrivals increases not only in the season of the extra award but also in the following two seasons reaching a peak of +2.91 percentage points in the second year after the award assignment. We can say that this increase is driven by the positive impact of the Blue Flag on domestic tourism flows. On the other hand, the signalling effect appears nil for international tourists. However, using the average length of stay as an alternative dependent variable, we find that after two years from the award, the new tourists attracted by the award might stay for a shorter period (4.78 days versus an average counterfactual scenario of 5.04 days).

Insert Table 2

Insert Figure 2

Looking at the first two Columns of Table 2, we see that the average treated and synthetic destinations are very similar in terms of all covariates; nevertheless, we observe a small difference in terms of number of beds and previous number of Blue Flags. Comparing Column 1 with Columns 3 and 4 of the same table proves that the use of a control group without any weighting would have resulted in large differences in the pre-treatment values of some covariates.

Insert Figure 3

In Figure 3 we present the tourism flow performance of every treated TA together with the corresponding synthetic control group. The pre-treatment trends of treated TAs are generally very well captured even if there is a bit of heterogeneity in the quality of the fit. This figure shows that tourism flows in Apt di Caorle (2008), Località marine Livorno (2008), Località marine Versilia (2009), Savona (2009), Ragusa (2009), Ortona (2010), Fasano (2011), and Isole Eolie (2011) diverge upward from their synthetic counterparts after the extra Blue Flag award. Estimated effects are particularly large in the latter four TAs. Conversely, no noticeable positive divergence is observable in most of the other TAs and there is even a reduction in the inflow of tourists in La Maddalena - Palau (2008), Località marine Ascoli Piceno (2008), Ancona (2010), Riviera delle Palme (2010), and Oristano (2011).

It is difficult to conceive that the Blue Flag award might have had a negative influence on the number of arrivals; therefore the decreasing tourism flow outcomes of these

[9] For TAs which received the treatment before 2012 we could have reported more than 3 years of post-treatment tourism flows. However, we limit the results to 3 post-treatment years for two reasons: i) in the aggregation we would have had a changing sample starting from the 4th year; ii) considering the peculiarities of the Blue Flag assignment process and our sample selection criteria it would be hard to argue that we are convincingly identifying the signalling impact of the award after more than 3 years from the extra Blue Flag.

localities are probably due to other issues, such as higher travel expenses (this applies particularly to La Maddalena). Nevertheless, these results prove that for some destinations the Blue Flag award was of a limited value in counterbalancing flat or even decreasing tourism trends.

Given the nature of the SCM, we must evaluate the significance of the estimates posing the question of whether or not the results could be driven entirely by chance. The most widely used test for this is what is referred to as a placebo test (Athey & Imbens, 2016). For each treated TA, this approach consists in virtually reassigning the treatment to non-treated TAs included in the final donor pool. Therefore, it is assumed that each non-treated TA had received the "extra" eco-label at the same time as the treated TA. Placebo effects are computed by comparing the divergence of actual tourism flows from its synthetic control for the treated TA with that for the other TAs included in the final donor pool. The distribution of these placebo effects gives an idea about whether the relative change in the number of tourist arrivals experienced by the treated TA after the Blue Flag assignment is different in relation to those experienced by other comparable (but non-treated) TAs. Figure 4 depicts the results of the 20 placebo tests, i.e. one placebo test for each treated TA. The placebo effects for each of the donor TAs are displayed with thin grey lines, while the corresponding effect for the treated TA is displayed with the thick black line. From Figure 4 we infer that only Savona (2009) and Fasano (2011) stand out from the placebo effects. This means that at the disaggregated level only their estimated impact can be considered statistically significant.

We exploit the SCM also to check whether the Blue Flag award allows hotels, restaurants, and other tourism operators to raise prices. Any increase in prices due to the Blue Flag might indeed encourage some tourists to visit alternative tourism destinations. To test this hypothesis we used Bank of Italy survey data on actual expenditure and satisfaction towards the level of prices of international tourists. Comparing treated and non-treated destinations we find that this is not the case, at least for international tourists. Receiving a Blue Flag does not seem to result in higher prices for local hotels, restaurants and tour operators. This is in juxtaposition with Rigall et al.' (2011) findings on hotel prices in Spain.

Appendix A reports the results obtained using a less stringent set of criteria for selecting the donor pool (see Sample section). The results of the analysis carried out using only the "stable number of Blue Flags from t-1 to t+1" criterion are reported in Figure A1 and are very similar to those shown in the main analysis; although the fit of international tourists significantly worsened in the years immediately preceding the award. Conversely, adding the "distance within a 200km ray" criterion we get the estimates reported in Figure A2. They suggest a slightly larger impact of the Blue Flag award, in particular for international tourists. These robustness checks confirm that the results presented above are not particularly sensitive to changes in the composition of the comparison group, yet they seem to suggest a possible positive impact also for international tourists.

Insert Figure 4

A possible concern of the analysis so far is that the donor pool is made up in part of coastal destinations achieving the Blue Flag award (see Sample section). This might bring about an underestimation of the actual signalling impact of the award. To mitigate these concerns we re-estimate the impact of the Blue Flag award on the 5 first-time treated destinations selecting the donor pool with the additional criterion of "only seaside destinations without a Blue Flag at the time of the treatment". However, the results are basically identical to those shown in Figure 3.

We additionally test the robustness of the results by checking for the presence of other seaside eco-labels in Italy. Indeed, as the Blue Flag is not the only seaside eco-label present on Italian territory, it is possible that the assignment of other eco-labels might confound our findings. This is why we collected data on another well-known seaside eco-label: the Blue Sails. Since 1999 Legambiente (the main Italian environmental organization) and Touring Club Italiano publish the "Guida Blu", a tourist guide about quality holidays on the Italian shores of seas and lakes. Every coastal municipality is assessed on the basis of the quality of the water, the coastline, the garbage disposal procedures (e.g., the implementation of proper sewerage schemes), the artistic heritage, and the gastronomy. Additionally, the overall environment and eco-sustainability are evaluated. Differently from the Blue Flags, Blue Sails are not assigned on voluntary basis. Each coastal municipality can receive from 0 to 5 Blue Sails. Still, adding Guida Blu variables to the set of covariates has a negligible impact on the estimates.

Why are the impacts heterogeneous?

When we look at the disaggregated picture of each single treated destination, the single most striking feature is the high heterogeneity of the results, while the average impact is slightly positive. To further investigate this heterogeneity we submitted an online questionnaire to the local tourism administrators of each treated municipalities. The questions concerned: i) the process which led to the Blue Flag award; ii) the causes behind the increase/drop in the number of tourists after the receipt of the award; iii) the overall opinion on the Blue Flag programme; iv) the signalling power of the Blue Flag programme; v) the link between the Blue Flag and sustainability; and vi) possible improvements to the programme. The answers we got from the local tourism operators (the response rate was 65%) are very consistent and show that, although the Blue Flag is useful in persuading a number of tourists to visit a certain destination, the success of some coastal destinations on domestic and foreign markets in recent years is largely due to other factors. In particular, they mentioned the work of local tourist operators and their facilities, the promotion made in consultation with public authorities, the careful management of the territory (e.g., little traffic, pedestrian areas and green spaces), the quality and safety of the bathing water, and the strict environmental policy. Indeed, only destinations capable of initiating a process of organisation, coordination and integrated management of the tourism supply fully exploit the possibility given by certified eco-labels (Lorenzini, Calzati & Giudici, 2011).

Besides, some administrators agree that the Blue Flag programme plays a crucial role in convincing policymakers that tourism management needs careful planning. In their experience, the Blue Flag created an opportunity to get a closer look at their territory and work collectively with other tourism stakeholders in improving tourist services as well as environmental protection. The role of the Blue Flag programme is particularly relevant as local tourism stakeholders often lack of a common understanding of the sustainable tourism development concept, leading to confusion in its implementation (e.g. Choi & Sirakaya, 2006). The support of tourism stakeholders is essential for the development and long-term sustainability of tourism (Dabphet, Scott & Ruhanen, 2012). For instance, this process has pushed the municipality of Piombino to introduce a planning document where the main tourism players committed themselves to deliver high environmental and tourism standards on a yearly basis.

Conclusions

In this paper we have investigated the impact of the Blue Flag award, the most widely recognised seaside eco-label for the public, decision-makers and tour operators. This programme gauges factors related to the quality of the environment as well as the quality of the services provided. Although eco-labels stress the crucial role played by sustainability, they must also provide eco-certified operators with significant private economic benefits. The economic return from receiving the award is needed particularly to compensate tourism operators for the pecuniary and non-pecuniary costs sustained in meeting certification environmental performance standards (Blackman et al., 2014). This means that increasing tourism generated revenues is a “hidden” aim of any tourism eco-label and this is usually achieved by increasing the number of tourists and/or attracting visitors who spend more money in the destination. The positive signalling impact of the award might be expected due to the certification of high standards in environmental and service features and to its promotional boost. In fact, each year newspapers, magazines, television, radio and social networks give an account of the Blue Flag award (especially highlighting new entries), enhancing the marketing campaigns of awarded municipalities with respect to domestic and inbound tourism.

The article has estimated the signalling impact of the award for Italian destinations which received an extra Blue Flag during the period 2008-2012 using a recent quasi-experimental method, the SCM. To our knowledge, this is the first case of SCM employed in a tourism context and is the first empirical attempt to disentangle the signalling impact on domestic and inbound tourism flows as well as to investigate the heterogeneity of these effects. We have found a positive average effect of the award on the number of domestic tourists, whereas the effect on the number of international tourists is negligible. This first set of results might be explained applying the consumer search behaviour framework (Moutinho, 1987) to our context. Visiting an Italian seaside destination might call for a limited problem solving behaviour in the case of a domestic tourist, while we might expect that an international tourist might adopt a more thorough search of information (extended problem solving behaviour). When extensive problem solving is needed, e.g. when planning a first-time holiday to a foreign country, a search would primarily emphasize external sources because the trip is unfamiliar. It might also be expected that the potential tourist undertaking extensive problem solving might rely on at least one decisive source (Fodness & Murray, 1999), such as travel agents, knowledgeable friends or consumer-generated review sites,[10] as well as on a wide variety of contributory sources, such as commercial guidebooks, and travel magazines, in order to reduce the perceived risk of an unfamiliar trip. On the other hand, when a limited decision making strategy is needed, a short search of both internal (past experiences) and external sources (consumer-generated review sites and other tourism related websites) might be expected. Given the widespread use of eco-labels in tourism, both search behaviours would likely uncover whether the potential destination has recently been awarded an eco-label; nevertheless, domestic tourists would put a much larger weight on it as they focus on a few pieces of information when making their decisions. This might explain why we find a positive signalling effect only for domestic tourists.

We have also investigated the heterogeneity of the results reporting separately the tourism flows of each treated municipality and their synthetic counterpart. From the analysis, it does not emerge a clear pattern as some of the awarded TAs experienced a growth in the number of arrivals while others experienced a drop in the number of arrivals. In order to shed some light on these heterogeneous results, we have augmented our empirical analysis by collecting qualitative information through an online questionnaire sent to the local tourism administrators of each treated municipality. Combining qualitative and quantitative data has led us to detect a second important result: it appears that the Blue Flag award boosts tourism flows only when combined with a clear sustainability policy. The Blue Flag programme creates an opportunity for local tourism administrators to get a closer look at their territory and work collectively with other tourism stakeholders in improving tourist services as well as environmental protection. This finding is also in line with what found by Mihalič (2000) using the Calgary tourism competitiveness model.

Overall, our findings imply that the Blue Flag "signal" by itself is not a sufficient instrument to encourage tourists to choose a particular seaside destination. Evidence suggests that seaside tourists value natural amenities, safe waters and an overall healthy environment more than other tourism-related services and take into account many pieces of information before settling for a specific destination. The main lesson to be drawn from

[10] Consumer-generated review sites (see the review in Bronner & de Hogg, 2016) are a vehicle for limiting the asymmetric information between the tourist and the quality of the seaside destinations (known by local tourism authorities and the other local tourism stakeholders) as they allow tourists to gain information about the real quality of a vast number of seaside destinations. As a result, consumer-generated review sites are likely to reduce the signalling effect of any eco-label.

our findings is that local tourism operators should not consider the Blue Flag award as the panacea for future tourism flows unless it is part of a more general tourism planning of the area.

Some limitations to this research remain. First and foremost, the data used for the study are not ideal. Although they refer to small territorial units (TAs), the availability of tourism flow data at the municipality level would have allowed analysing the signalling impact of a larger number of Blue Flag awards and identifying each impact with increased precision. Second, our empirical approach does not allow evaluating the long-term impact of an eco-label. Third, there is possibly a positive selection bias in the quality of the local tourism administrators who replied to the online questionnaire. This potential bias could have been mitigated (with a higher response rate) if we had submitted the questionnaire at the time of receiving the award instead of a few years later.

The limitations of our study call for an increased effort of national statistical institutes to collect tourism flow data as disaggregated as possible and also collect other tourism-related variables (e.g., the average price of a hotel room or the average price of a restaurant dinner) which provide valuable information for evaluating the impact of tourism policies. Future research should not only focus on the eco-label impact on tourism flows, but also evaluate whether restaurants, hotels and other local businesses use eco-labels to raise prices. Indeed, particularly for seaside destinations close to full capacity, eco-labels might be used as a tool for attracting tourists with a higher spending capacity rather than for increasing tourist numbers.

Lastly, the creation of a large cross-country panel dataset would allow testing for the external validity of our findings. It would also allow testing whether the increasing dissemination of consumer-generated review sites and the possibility to "virtually" visit destinations through freely accessible photos, videos and web mapping services could weaken the signalling power of eco-labels through time. This effect might be expected as access to technology tends to reduce asymmetric information among potential tourists. Such a result would imply that continuous technology improvements (e.g. Bronner & de Hogg, 2016) may progressively tail-off the signalling role of eco-labels in the coming years.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effects of California's tobacco control programs. *Journal of the American Statistical Association, 105*(490), 493–505.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review, 93*(1), 112–132.

Athey, S., & Imbens, G.W. (2016). The state of applied econometrics - causality and policy evaluation. Ithaca, NY: Cornell University library. arXiv:1607.00699

Billmeier, A., & Nannicini, T., (2013). Assessing economic liberalization episodes: A synthetic control approach. *The Review of Economics and Statistics, 95*(3), 983–1001.

Blackman, A., Naranjo, M.A., Robalino, J., Alpízar, F., & Rivera, J. (2014). Does tourism eco-certification pay? Costa Rica's Blue Flag program. *World Development, 58,* 41–52.

Boevers, J. (2008). Assessing the utility of beach eco-labels for use by local management. *Coastal Management, 36*(5), 524–531.

Blue Flag (2015). Blue Flag beach criteria and explanatory notes 2015. Retrieved 05/09/2015, from http://www.blueflag.org/materiale/publication-downloads/beach-criteria-and-explanatory-notes-2015.pdf-3

Bronner, F., & de Hogg, R. (2016). Travel websites: Changing visits, evaluations and posts. *Annals of Tourism Research, 57,* 94–112.

Buckley, R. (2002). Tourism ecolabels. *Annals of Tourism Research, 29*(1), 183–208.

Cagilaba, V., & Rennie, H. (2005). Literature review of beach awards and rating systems (Report No. 2005/24). Hamilton: Environmental Waikato.

Capacci, S., Scorcu, A.E., & Vici, L. (2015). Seaside tourism and eco-labels: The economic impact of Blue Flags. *Tourism Management, 47,* 88–96.

Chan, H.F., Frey S.F., Gallus, J., & Torgler, B. (2014). Academic honors and performance. *Labour Economics, 31,* 188–204.

Choi, H.C., & Sirakaya, E. (2006). Sustainability indicators for managing community tourism. *Tourism Management, 27*(6), 1274–1289.

Chung, S., Lee, J., & Osang, T. (2016). Did China tire safeguard save U.S. workers? *European Economic Review, 85,* 22–38.

Creo, C., & Fraboni, C. (2011). Awards for the sustainable management of coastal tourism destinations: The example of the Blue Flag program. *Journal of Coastal Research, 61,* 378–381.

Cucculelli, M., & Goffi, G. (2016). Does sustainability enhance tourism destination competitiveness? Evidence from Italian Destinations of Excellence. *Journal of Cleaner Production, 111,* 370–382.

Dabphet, S., Scott, N., & Ruhanen, L. (2012). Applying diffusion theory to destination stakeholder understanding of sustainable tourism development: A case from Thailand. *Journal of Sustainable Tourism, 20*(8), 1107–1124.

Eijgelaar, E., Nawijn, J., Barten, C., Okuhn L., & Dijkstra, L. (2016). Consumer attitudes and preferences on holiday carbon footprint information in the Netherlands. *Journal of Sustainable Tourism, 24*(3), 398–411.

Esparon, M., Gyuris, E., & Stoeckl, N. (2014). Does ECO certification deliver benefits? An empirical investigation of visitors' perceptions of the importance of ECO certification's attributes and of operators' performance. *Journal of Sustainable Tourism, 22*(1), 148–169.

Esparon, M., Stoeckl, N., Farr, M., & Larson, S., (2015). The significance of environmental values for destination competitiveness and sustainable tourism strategy making: Insights from Australia's Great Barrier Reef World Heritage Area. *Journal of Sustainable Tourism, 23*(5), 706–725.

European Environment Agency (2015). BWD report for the bathing season 2014 - Italy.

Fairweather, J. R., Maslin, C., & Simmons, D. G. (2005). Environmental values and response to ecolabels among international visitors to New Zealand. *Journal of Sustainable Tourism, 13*(1), 82–98.

FEE (2007). Awards for improving the coastal environment: The example of the Blue Flag. Report financed by the European Union.

Ferrero, G., (2015). Prefazione. In Pencarelli, T. (Ed.), *Comunicare le destinazioni balneari, il ruolo delle Bandiere Blu* (pp. 9–12). Milano: Franco Angeli.

Fodness, D., & Murray, B. (1999). A model of tourist information search behavior. *Journal of Travel Research, 37*, 220–230.

Font, X. (2002). Environmental certification in tourism and hospitality: Progress, process and prospects. *Tourism Management, 23,* 197–205.

Fraboni, C., (2015). Risultati dell'analisi empirica sul ruolo della Bandiera Blu nella comunicazione e valorizzazione delle destinazioni turistiche. In Pencarelli, T. (Ed.), *Comunicare le destinazioni balneari, il ruolo delle Bandiere Blu* (pp. 116–202). Milano: Franco Angeli.

Fraguell, R. M., Martí, C., Pintó J., & Coenders, G. (2016). After over 25 years of accrediting beaches, has Blue Flag contributed to sustainable management? *Journal of Sustainable Tourism, 24*(6), 882–903.

Friedman, T.L. (2006). *The world is flat. The globalized world in the twenty-first century,* 2nd ed. London: Penguin.

Graci, S., & Dodds, R. (2015). Certification and labelling. In Hall, C.M., Gössling , S., & Scott, D. (Eds.), *The Routledge Handbook of Tourism and Sustainability* (pp. 200–208). Abingdon: Routledge.

Gössling, S., & Buckley, R. (2016). Carbon labels in tourism: Persuasive communication? *Journal of Cleaner Production, 111*, 358–369.

Hall, C.M., Gössling, S., & Scott, D. (2015). The evolution of sustainable development and sustainable tourism. In Hall, C.M., Gössling , S., & Scott, D. (Eds.), *The Routledge Handbook of Tourism and Sustainability* (pp. 15–35). Abingdon: Routledge.

Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

Karlsson, L., & Dolnicar, S. (2016). Does eco certification sell tourism services? Evidence from a quasi-experimental observation study in Iceland. *Journal of Sustainable Tourism, 24*(5), 694–714.

Lorenzini, E., Calzati, V., & Giudici, P. (2011). Territorial brands for tourism development: A statistical analysis on the Marche Region. *Annals of Tourism Research, 38*, 540–560.

Lucrezi, S., & Saayman, M. (2015). Beachgoers' demands vs. Blue Flag aims in South Africa. *Journal of Coastal Research, 31*(6), 1478–1488.

Lucrezi, S., Saayman, M., & Van der Merwe, P. (2015). Managing beaches and beachgoers: Lessons from and for the Blue Flag award. *Tourism Management, 48*, 211–230.

McKenna, J., Williams, A.T., & Cooper, J.A.G. (2011). Blue Flag or Red Herring: Do beach awards encourage the public to visit beaches? *Tourism Management, 32*(3), 576–588.

Mihalič, T. (2000). Environmental management of a tourist destination: A factor of tourism competitiveness. *Tourism Management, 21*(1), 65–78.

Moutinho, L. (1987). Consumer behavior in tourism. *European Journal of Marketing, 21*, 5–44.

Munasib, A., & Rickman, D.S. (2015). Regional economic impacts of the shale gas and tight oil boom: A synthetic control analysis. *Regional Science and Urban Economics, 50*, 1–17.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy, 82*(4), 729–754.

Nelson, C., Morgan, R., Williams, A.T., & Wood, J. (2000). Beach awards and management. *Ocean & Coastal Management, 43*, 87–98.

Nelson, C., & Botterill, D. (2002). Evaluating the contribution of beach quality awards to the local tourism industry in Wales: The Green Coast Award. *Ocean & Coastal Management*, *45*(2–3), 157–170.

Pencarelli, T., Splendiani, S., & Fraboni, C. (2016). Enhancement of the “Blue Flag” Eco-label in Italy: An empirical analysis. *Anatolia, 27*(1), 28–37.

Rigall-I-Torrent, R., Fluvia, M., Ballester, R., Salo, A., Ariza, E., & Espinet, J.M. (2011). The effects of beach characteristics and location with respect to hotel prices. *Tourism Management, 32*, 1150–1158.

Sasidharan, V., Sirakaya, E., & Kerstetter, D. (2002). Developing countries and tourism ecolabels. *Tourism Management, 23,* 161–174.

Schernewski, G., & Sterr, H. (2002). Tourism and environmental quality of the German Baltic coast: Conflict or chance? In Schernewski, G., & Schiewer, U. (Eds.), *Baltic coastal ecosystems, structure, function and coastal zone management* (pp. 217–232). Berlin: Springer.

Smith, V.L., & Font, X. (2014). Volunteer tourism, greenwashing and understanding responsible marketing using market signalling theory. *Journal of Sustainable Tourism, 22*(6), 942–963.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*, 355–374.

UniCredit (2016). “UniCredit 4 Tourism” report provided by the UniCredit bank group in collaboration with the Italian Touring Club.

Zielinski, S., & Botero, C. (2015). Are eco-labels sustainable? Beach certification schemes in Latin America and the Caribbean. *Journal of Sustainable Tourism, 23*(10), *1550–1572.*

Appendix A

Figure A1 reports the results obtained selecting for each treated TA a donor pool made up of all non-treated TAs which had a stable number of Blue Flags from t-1 to t+1. Figure A2 displays the results obtained further restricting the donor pool to the TAs located within a 200 km ray from the treated TA.

Insert Figure A1

Insert Figure A2

Table A1 summarises the assumptions behind the selection of the treated TAs and of their synthetic counterparts.

Insert Table A1

Table 1. Tourism areas used for quantitative case studies

Table 2. Balancing properties of the main analysis

Table A1. Sample selection process

Figure 1. Summary map of the treated TAs and the TAs which make up the donor pool

Note: The non-treated TAs with positive weights make up the control group of 1 or more treated TAs.

Figure 2. Aggregate estimates of the main analysis

Figure 3. Number of arrivals (expressed as a ratio of the pre-treatment value) for each single treated TA and the respective synthetic control

Figure 4. Number of arrivals (expressed as a ratio of the pre-treatment value) for each single treated TA and the respective synthetic control - Placebo effects

Figure A1. Aggregate estimates obtained selecting for each treated TA a donor pool made up of all non-treated TAs which had a stable number of Blue Flags from t-1 to t+1

Figure A2. Aggregate estimates obtained restricting the donor pool to the TAs located within a 200 km ray from the treated TA

Table 1. Tourism areas used for quantitative case studies.

TAs name	TAs code	Treatment Year	Region	First Blue Flag assignment year	Average number of Blue Flags in the 5 years before treatment	Average number of Blue Flags after treatment	Average number of Italian tourists in the 5 years before treatment	Average number of Italian tourists after treatment	Average number of Foreigner tourists in the 5 years before treatment	Average number of Foreigner tourists after treatment
Apt di Caorle	275004	2008	Veneto	1990	1.2	2	430,288	477,864	698,696	859,239
Francavilla al mare	693002	2008	Abruzzi	2003	0.4	0.625	40,305	36,762	6,841	6,643
La Maddalena - Palau	1043004	2008	Sardinia	1993	0	1.5	99,200	67,935	34,129	32,405
Località marine Ascoli Piceno	445004	2008	Marche	1997	3.6	5.375	279,541	283,338	47,302	46,701
Località marine Livorno	495002	2008	Toscana	1990	3.8	6	358,670	425,583	170,344	224,592
Località marine Macerata	435004	2008	Marche	2003	2	2.75	109,624	120,293	20,311	22,755
Silvi	673005	2008	Abruzzi	2002	0.2	0.875	40,710	39,492	6,855	7,690
Località marine Versilia	465006	2009	Toscana	1990	3.8	4	327,405	324,368	197,259	226,330
Ragusa	882001	2009	Sicily	2009	0	0.86	76,092	78,870	33,998	43,297
Savona	96001	2009	Liguria	2002	0.8	1	41,983	52,859	26,395	36,465
Altri comuni Ragusa	884002	2010	Sicily	2002	1	1.83	76,691	57,716	14,474	16,511
Ancona	426001	2010	Marche	2000	0	1.00	112,480	96,344	34,904	35,029
Ortona	693003	2010	Abruzzi	2010	0	0.67	14,612	17,036	2,633	2,750
Riviera delle Palme - Località marine	95002	2010	Liguria	1987	6.4	9.67	746,388	691,823	200,836	214,882
Fasano	743002	2011	Apulia	2011	0	1	64,176	70,641	16,781	25,776
Isole Eolie	833001	2011	Sicily	2011	0	1	66,680	65,134	38,835	46,479
Località marine Rimini	995002	2011	Emilia-Romagna	1987	2.6	3	1,144,257	1,241,776	234,950	243,919
Oristano	952001	2011	Sardinia	2011	0	1	27,610	23,985	12,748	13,545
Otranto	753004	2011	Apulia	1990	0.2	1	88,250	110,237	13,135	18,132
Capri - Anacapri	633002	2012	Campania	1993	0	1	77,430	69,400	87,478	116,397

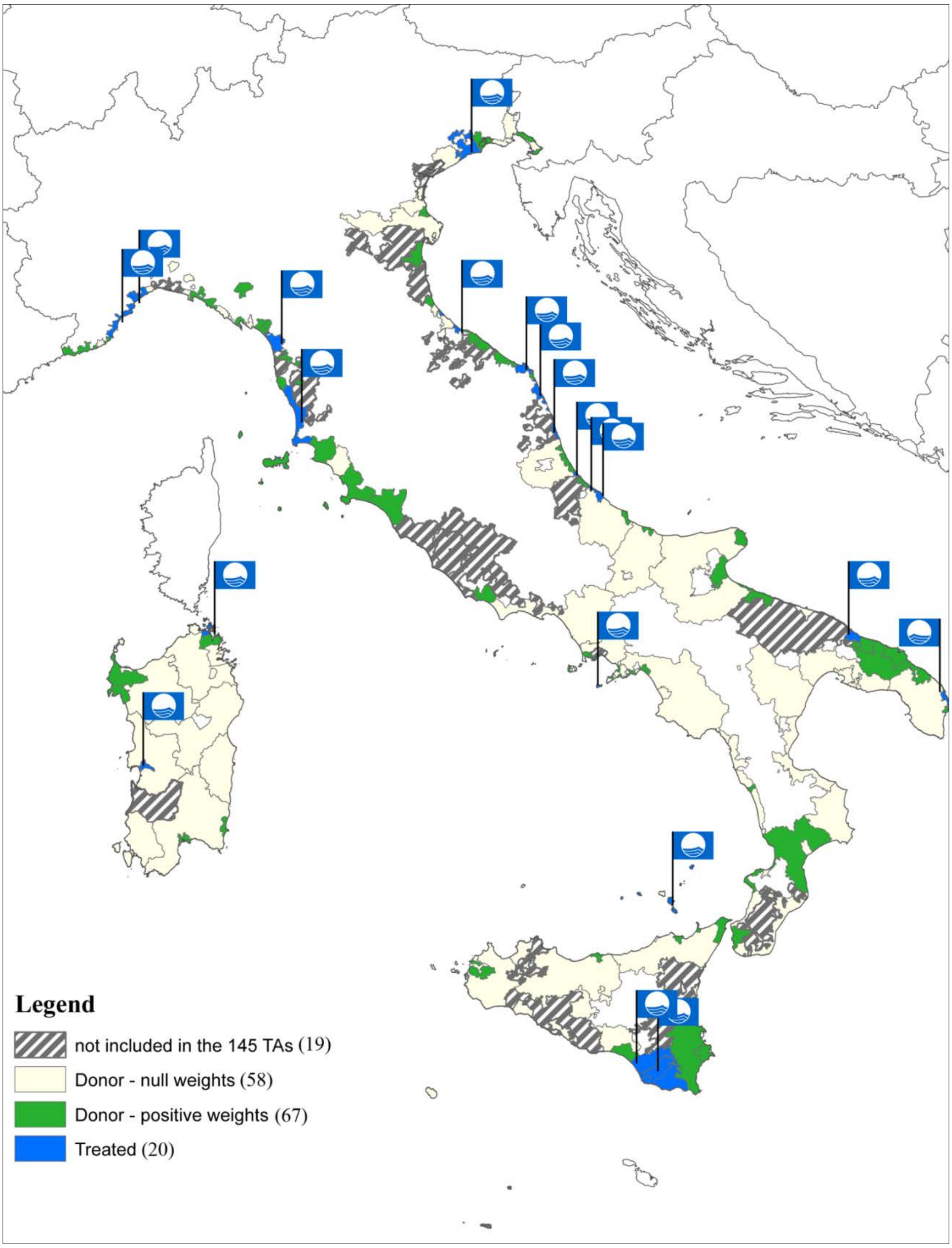
Table 2. Balancing properties of the main analysis

	Period of reference	Average of the treated TAs (1)	Average of the Synthetic controls (2)	All donors (excluding those with >20 municipalities) (3)	All donors (4)
Number of arrivals (expressed as ratio of the t-1 value)	*Between t-8 and t-5*	0.950	0.947	0.983	0.965
Number of arrivals (expressed as ratio of the t-1 value)	*Between t-4 and t-2*	0.991	0.997	1.030	1.023
Proportion of international tourists	*Between t-4 and t-1*	0.734	0.746	0.714	0.728
Proportion of tourists not staying in a hotel	*Between t-4 and t-1*	0.236	0.213	0.199	0.205
Average length of stay	*Between t-4 and t-1*	5.119	5.148	4.746	4.810
Number of beds available	*Between t-4 and t-1*	24,371	19,167	12,167	13,760
Proportion of beds available in 4 or 5 star hotels	*Between t-4 and t-1*	0.327	0.296	0.339	0.342
Number of Blue Flags awarded	*Between t-4 and t-2*	1.317	0.873	0.384	0.459
Number of bans (EU Bathing Waters Directive)	*Between t-3 and t-1*	0.383	0.359	0.330	0.381
Environmental satisfaction of international tourists (Likert-type scale from 1 to 10)	*Between t-3 and t-1*	8.729	8.822	8.754	8.765
Per capita expenditure on tourism	*Between t-4 and t-2*	€ 30.73	€ 29.88	€ 23.23	€ 22.48
Per capita expenditure on environment and territory	*Between t-4 and t-2*	€ 333.02	€ 335.72	€ 310.20	€ 306.00
Number of municipalities	*2011*	3.750	3.634	2.889	14.648
Kilometres of coastline	*2011*	50,831	46,644	36,001	55,846
Population	*2011*	68,257	70,463	67,800	125,156
Dummy venue for congress tourism	*2011*	0.100	0.217	0.232	0.248
Number of harbours in a 50 km ray	*2011*	4.800	4.578	5.172	5.416
Number of airports in a 50 km ray	*2011*	1.150	1.080	1.010	1.128
Number of TAs in a 50 km ray having achieved at least 1 Blue Flag	*Between 2008 and 2011*	3.450	3.598	2.818	2.880

Table A1. Sample selection process

	Selection criterion	Motivation
Treatment group	We define as treated those TAs which received an "extra" Blue Flag award in the years from 2008 to 2012.	Although we have an annual TA-level balanced panel data for the period 2000-2014, we limit our "treatment" definition to the period 2008-2012 in order to have a reasonable number of pre-treatment and post-treatment years.
	The treated TA did not receive an additional Blue Flag the year before.	This criterion ensures that the impact of the Blue Flag is not driven by a previous change in the number of Blue Flags.
	The Blue Flag award has been assigned to a municipality having at least 15% of the sleeping accommodations of the treated TA and the treated TA is made up of no more than 20 municipalities.	This way we limit the confounding factor of having data at the TA level instead of at the municipality level. Indeed, in case the Blue Flag has been awarded to a municipality with a very limited number of sleeping accommodations with respect to the overall number of sleeping accommodations of the TA, it would be hard to claim that each change in tourism flows could properly be attributed to the eco-label.
Control group	Sample limited to the 164 TAs having access to the sea.	The analysis concerns only seaside destinations.
	We drop 9 TAs for which there is clear evidence that the sea is not the main tourist attraction.	As data do not allow disentangling seaside tourism from total tourism, we remove the areas where tourism flows due to seaside tourism are limited with respect to the overall tourism flows.
	10 TAs are dropped because they are made up by a large number of municipalities (>15 municipalities) and the total number of beds of the municipalities having access to the sea is less than 50% of the overall number of beds.	This criterion has been added because of the inability to disentangle seaside tourism from total tourism and because of the lack of data at the municipality level. Indeed, the main tourism destination of these TAs is not the sea and they would not represent a convincing counterfactual in our empirical analysis.
	For each treated TA, the donor pool is made up of all non-treated TAs which had a stable number of Blue Flags from t-1 to t+1.	This way we exclude from our donor pool TAs which could not properly resemble the counterfactual scenario as they took advantage or disadvantage from a change in the number Blue Flag awards.
	We then additionally restrict the donor pool to the TAs located within a 200 km ray from the treated TA.	This allows comparing only TAs located not too far away between each other's. This criterion attempts to pick up more general trends in tourism flows (e.g., the steady increase of tourism flows in Apulia).
	Finally, in the main analysis we also discard the TAs having a number of municipalities over 20.	Given the selection criteria for the treatment group, these TAs might not accurately resemble the counterfactual scenario of the treated TAs.

Figure 1. Summary map of the treated TAs and the TAs which make up the donor pool.



Note: The non-treated TAs with positive weights make up the control group of 1 or more treated TAs.

Figure 2. Aggregate estimates of the main analysis

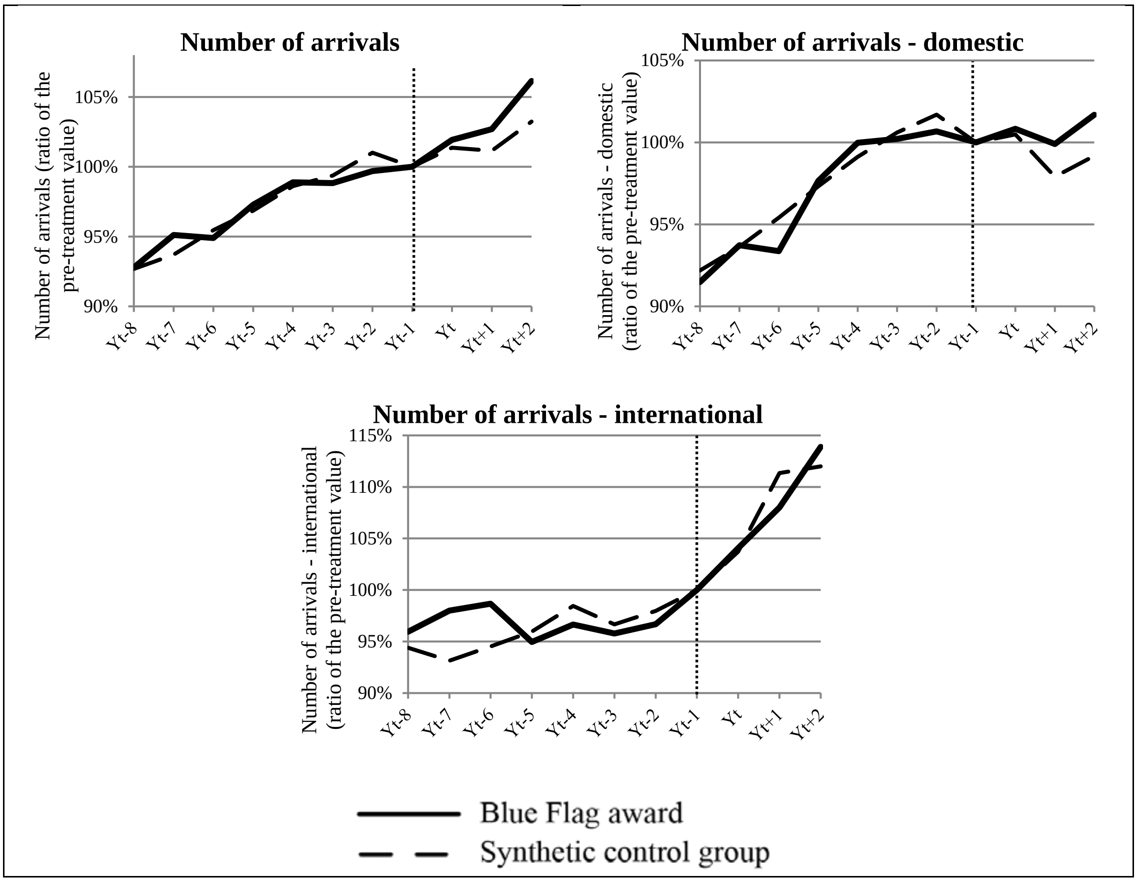


Figure 3. Number of arrivals (expressed as a ratio of the pre-treatment value) for each single treated TA and the respective synthetic control

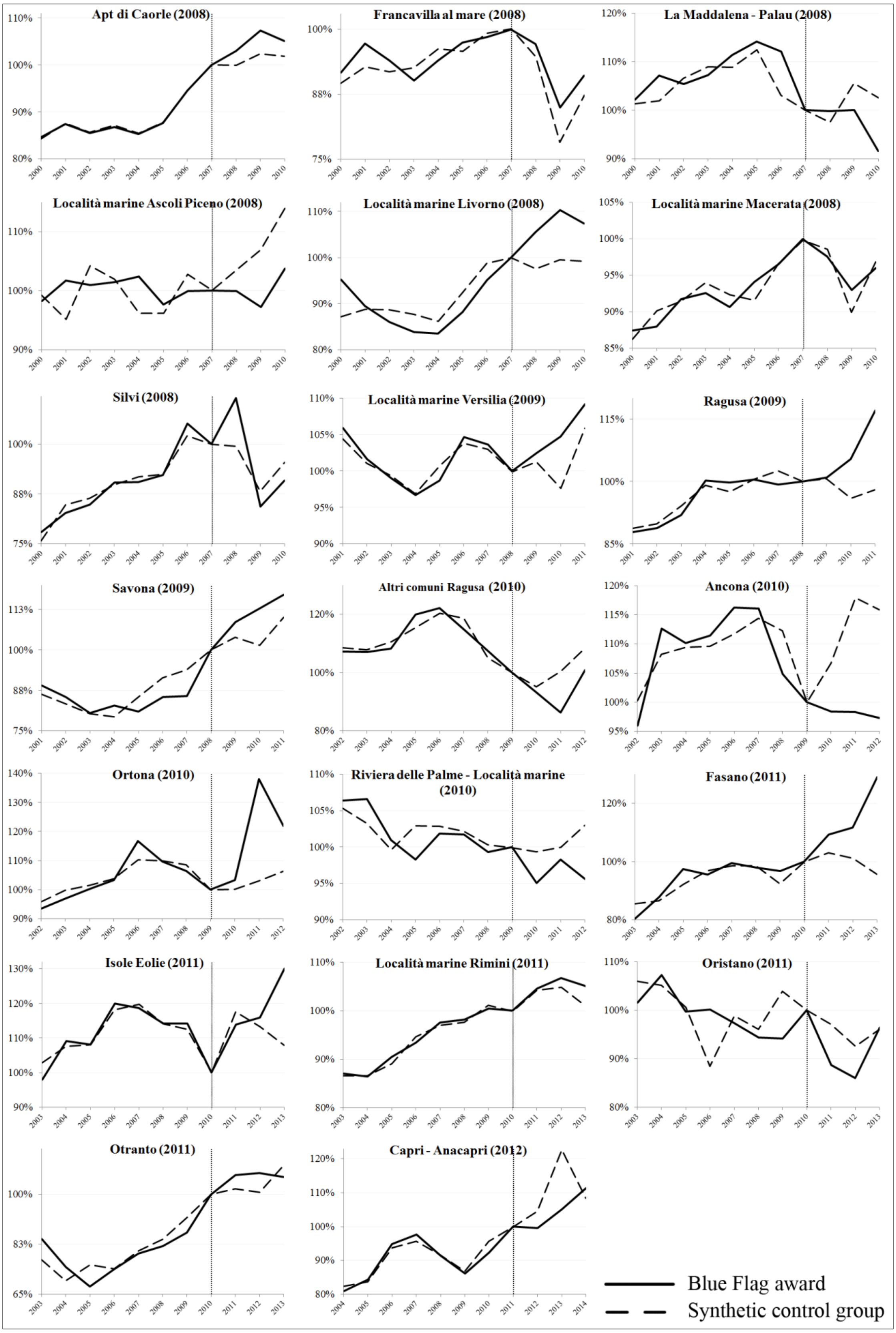


Figure 4. Number of arrivals (expressed as a ratio of the pre-treatment value) for each single treated TA and the respective synthetic control - Placebo effects

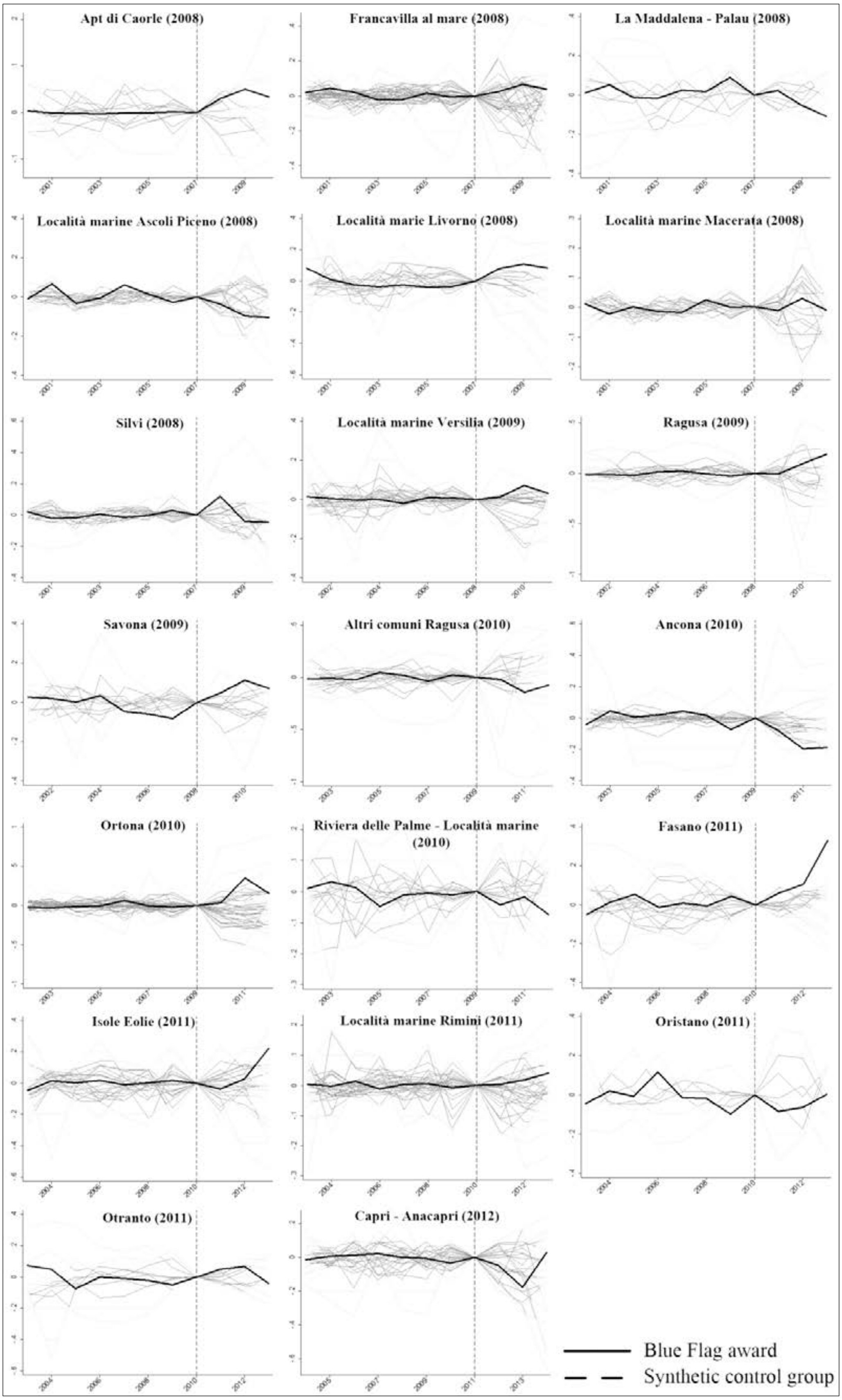


Figure A1. Aggregate estimates obtained selecting for each treated TA a donor pool made up of all non-treated TAs which had a stable number of Blue Flags from t-1 to t+1

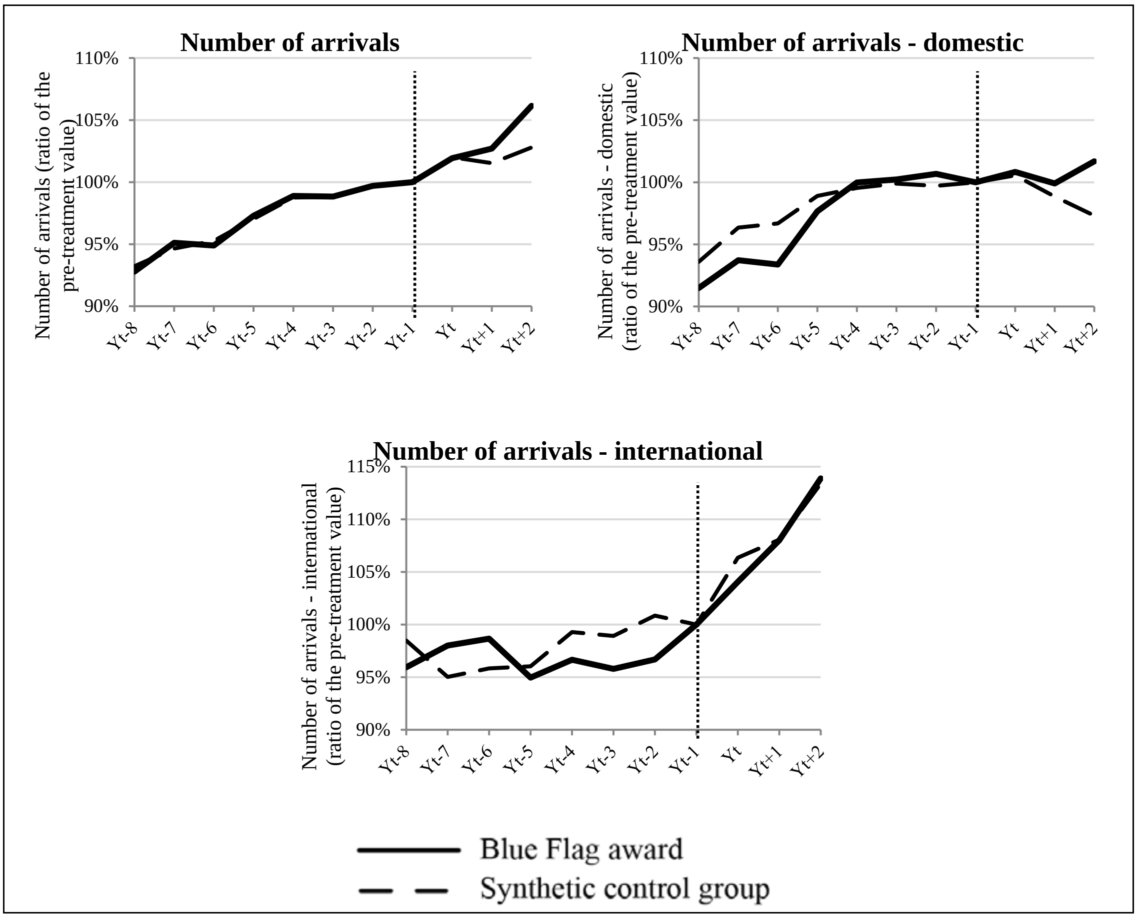


Figure A2. Aggregate estimates obtained restricting the donor pool to the TAs located within a 200 km ray from the treated TA