Received: 3 September 2021 | Revised: 25 May 2022 | Accepted: 29 August 2022

DOI: 10.1111/jors.12618

RESEARCH ARTICLE

JOURNAL OF REGIONAL SCIENCE WILEY

A socioeconomic impact assessment of three Italian national parks

Riccardo D'Alberto[1] | Francesco Pagliacci[2] | Matteo Zavalloni[3]

[1]Department of Statistical Sciences "P. Fortunati", University of Bologna, Bologna, BO, Italy

[2]Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, PD, Italy

[3]Department of Agricultural and Food Sciences, University of Bologna, Bologna, BO, Italy

Correspondence
Francesco Pagliacci, Department of Land, Environment, Agriculture and Forestry, University of Padova, Viale Dell'Università 14, 35020, Legnaro, PD, Italy.
Email: francesco.pagliacci@unipd.it

Abstract

The expansion of protected areas (PAs) is feared to negatively affect the local economy, as every PA, albeit to different degrees, entails restriction to the economic activities. The literature on the topic has started assessing what is the socioeconomic impact of PAs, mostly focusing on the Global South. The objective of this article is the analysis of the socioeconomic impact of three Italian national parks (NPs), established in the 2000s, using a counterfactual approach based on both the outcome regression diff-in-diff and the doubly robust diff-in-diff combined with different propensity score-based and Mahalanobis distance matching procedures. We find that the three Italian NPs have a robust and statistically significant impact on average income of residents in municipalities hosting them. Conversely, there is weak evidence that population and local establishments are positively affected, and touristic local establishments and employment are negatively affected by the three NPs. All together the results indicate that the three NPs have no negative effect on the socioeconomic dynamics of the territories impacted, although additional investigations are required to shed lights to the impact mechanisms.

KEYWORDS

biodiversity, conservation policy, difference-in-differences, eco-tourism, impact evaluation

1 | INTRODUCTION

The establishment of protected areas (PAs) is a crucial strategy for the protection of biodiversity, and their expansion is highly advocated (Visconti et al., 2019). Especially effective for forest areas (Geldmann et al., 2019) and in halting the destruction of habitats (Andam et al., 2008), PAs were originally established to preserve spectacular areas with little economic relevance (Runte, 1977). However, currently covering almost 15% of the terrestrial surface (UNEP-WCMC & IUCN, 2016), they have expanded in regions more sensitive from a socioeconomic point of view than in the past (Naughton-Treves et al., 2005). As a result, conflicts among conservation and socioeconomic development goals are more likely to arise (Oldekop et al., 2016).

Indeed, as they restrict economic activities and limit or ban the extraction of natural resources, concerns over the possibility that PAs would endure poverty trap have been raised (Norton-Griffiths & Southey, 1995; Wilkie et al., 2006). Moreover, the management of PAs itself entails direct costs that conflict with other state-funded development programmes (Lindsey et al., 2018). Furthermore, despite possibly substantial eco-tourism revenues, the distribution of the benefits of PAs is suspected to be highly skewed, so that they do not reach the poorest segment of the populations affected by the restrictions in the economic activities (Wilkie et al., 2006). Recognizing these potential problems and the fact that their effectiveness depends on the acceptance by local communities (Bennett et al., 2019), the scopes of PAs are no longer limited to biodiversity conservation, but they include for example the reduction of poverty (Naughton-Treves et al., 2005).

Scientific literature has also broadened its scope beyond the analysis of the environmental effects of PAs (Jones et al., 2017; Schleicher et al., 2019). An increasing number of papers deals with the socioeconomic impact of PA establishments, even though they are still limited, and most of the studies rely on qualitative indicators (Oldekop et al., 2016). Contrary to the common wisdom, PAs are not linked to poverty trap (Mammides, 2020; Naidoo et al., 2019), and the environmental effectiveness of PAs is positively related to the socioeconomic outcome (Oldekop et al., 2016).

Most of the works dealing with the socioeconomic effects of PAs focus on the Global South (Jones et al., 2020; Oldekop et al., 2016), where the trade-offs between conservation goals and economic objectives are feared to be the most severe.[1] For example, analyzing the impact of PAs in Thailand and Costa Rica, Ferraro et al. (2011) show a win-win situation, where PAs are linked to the alleviation of both poverty and deforestation. Such a result holds particularly for lands associated to low returns from agriculture. Deepening the topic, Ferraro and Hanauer (2014) assess the heterogeneity of the impact of PAs. They find that even though there is no poverty trap associated to a PA, a sort of trade-off is still present, as the degree of the effectiveness of PAs in halting deforestation is negatively related to their capacity to alleviate poverty. Similarly, the analysis of Braber et al. (2018) indicates that PAs in Nepal have a positive effect on the reduction of poverty, but such an effect is geographically limited to the area of the parks, and it is characterized by time lags. The positive impact on poverty reduction holds in other countries, such as Thailand (Sims, 2010), or Tanzania (McNally et al., 2011).

More limited evidence exists for western countries, mostly for the USA. Sims et al. (2019) show that land protection in New England has had positive, albeit small, impact on employment levels, whereas it did impact neither population nor median income. They also analyze whether the effect depends on whether land protection is financed through private or public funds, and on the distance from the major cities. Chen et al. (2016) find a positive impact of the North West Forest plan (protecting 11 million acres in the Pacific Northwest region of the United States) on income, population and property values, but only for small communities, and not for medium ones. Other works for the same area include the one by Lewis et al. (2002) who do not find any impact on employment level. On a different line, Weiler and Seidl (2004) find that the effect of changing designation, from *monument* to *park*, yields more than 10,000 tourist a year, indicating a reputational effect of the park *brand*. Also across Europe, national

[1]Oldekop et al. (2016) conduct a global meta-analysis on the social and environmental outcome of PAs, based on 171 articles covering 160 PAs. 72% of the covered PAs are in Africa or Asia, 3% in Europe, 22% in the Americas, and the remaining in Oceania.

parks (NPs) and PAs can represent unique tourist attractions, serving as engines of economic development in otherwise weak regional economies, by means of the attraction of spending from outside the PAs (Bushell & Eagles, 2006; Mayer et al., 2010; Reinius & Fredman, 2007). Beyond monetary assessments, European PAs have been perceived to have a positive effect on wellbeing, taking into consideration dimensions such as health and social equity (Jones, Malesios, et al., 2020).

The objective of this paper is to assess the socioeconomic impact of three Italian NPs. We assess the impact of their establishments at the municipality level on: (i) per capita (taxable) income, (ii) population level, (iii) number of local establishments (i.e., firms or firm sections that operate at municipality level), (iv) employment in local establishments (i.e., the number of workers employed by local establishments operating in the municipality), (v) number of local establishments in the tourism sector, (vi) employment in local establishments in the tourism sector, (vii) number of agricultural holdings, and (viii) utilized agricultural area (UAA). The evaluation is performed by applying a counterfactual approach to panel data on Italian municipalities by means of the diff-in-diff (DID) estimator for the average treatment effect on the treated (ATT). Both the outcome regression DID (Heckman et al. 1997, 1998) and the Doubly Robust DID (Sant'Anna & Zhao, 2020) models are used. These two DID estimators are combined with five different matching approaches, used as identification strategies, based on the propensity score matching (PSM) and the Mahalanobis distance matching techniques. We restrict the analysis to the NPs that were established in the years 2000s, namely the Parco Nazionale dell'Appennino Tosco-Emiliano (2001), the Parco Nazionale dell'Alta Murgia (2004) and the Parco Nazionale dell'Appennino Lucano Val d'Agri Lagonegrese (2007).

Figure 1 maps the Italian NPs and the municipalities covered by the three NPs under consideration here. This choice is due to the availability of data on taxable income, available only since the taxable year 2000.[2]

We find evidence that NPs have a positive impact on per capita income and on population level, while we have more ambiguous results on the other socioeconomic dimensions analyzed. While some analyses exist for marine NPs (Di Franco et al., 2016), to the best of our knowledge, this is the first time that socioeconomic impact of terrestrial Italian NPs is assessed.[3] Moreover, note that most of the articles focusing on western countries address the socioeconomic impact of PAs that are less activity-restrictive than NPs, according to the IUCN categories (Dudley, 2008). Finally, this paper represents the first application of the Doubly Robust Diff-in-Diff estimator for the analysis of NPs impact. Such an estimator helps to improve the inference results since, with the panel data at hand, after conditioning on a vector of key pretreatment covariates, it remains consistent for the estimate of the ATT when either (but not necessarily both) a propensity score model or an outcome regression model is correctly specified (Sant'Anna & Zhao, 2020).

2 | METHODS

Different methods have been implemented to assess the effects of PAs, mostly referring single case studies, and focusing on the costs and the benefits associated to the establishment. The main "costs" are: the restriction of the conventional land use of agriculture, forestry or mining; the limits to the overall intensity of the economic activity (Oldekop et al., 2016) as well as to the development of infrastructure-based tourism (e.g., new hotels, ski resorts, other services...). Conversely, a new PA is also expected to deliver specific "benefits" to the local communities that host it. On the one hand, direct and tangible payment flows into PA regions (such as the state- or the region-level

[2]By considering the NPs established during the 2000s, it was possible to exploit the information on the outcome variables before (2001) and after (2011) treatment, hence setting up the baseline and follow-up years, according to the prescriptions of the counterfactual analysis framework. Given the fact that the information on taxable income at municipality level could not be retrieved before the year 2000, all the NPs established before such date could have not been analyzed in relation to such outcome variable.

[3]The only paper that addresses the socioeconomic effect of terrestrial Italian NPs, it only qualitatively compares the levels of socioeconomic indicators of municipalities within NPs with the levels of those indicators considering the entire set of Italian municipalities (Romano et al., 2021).

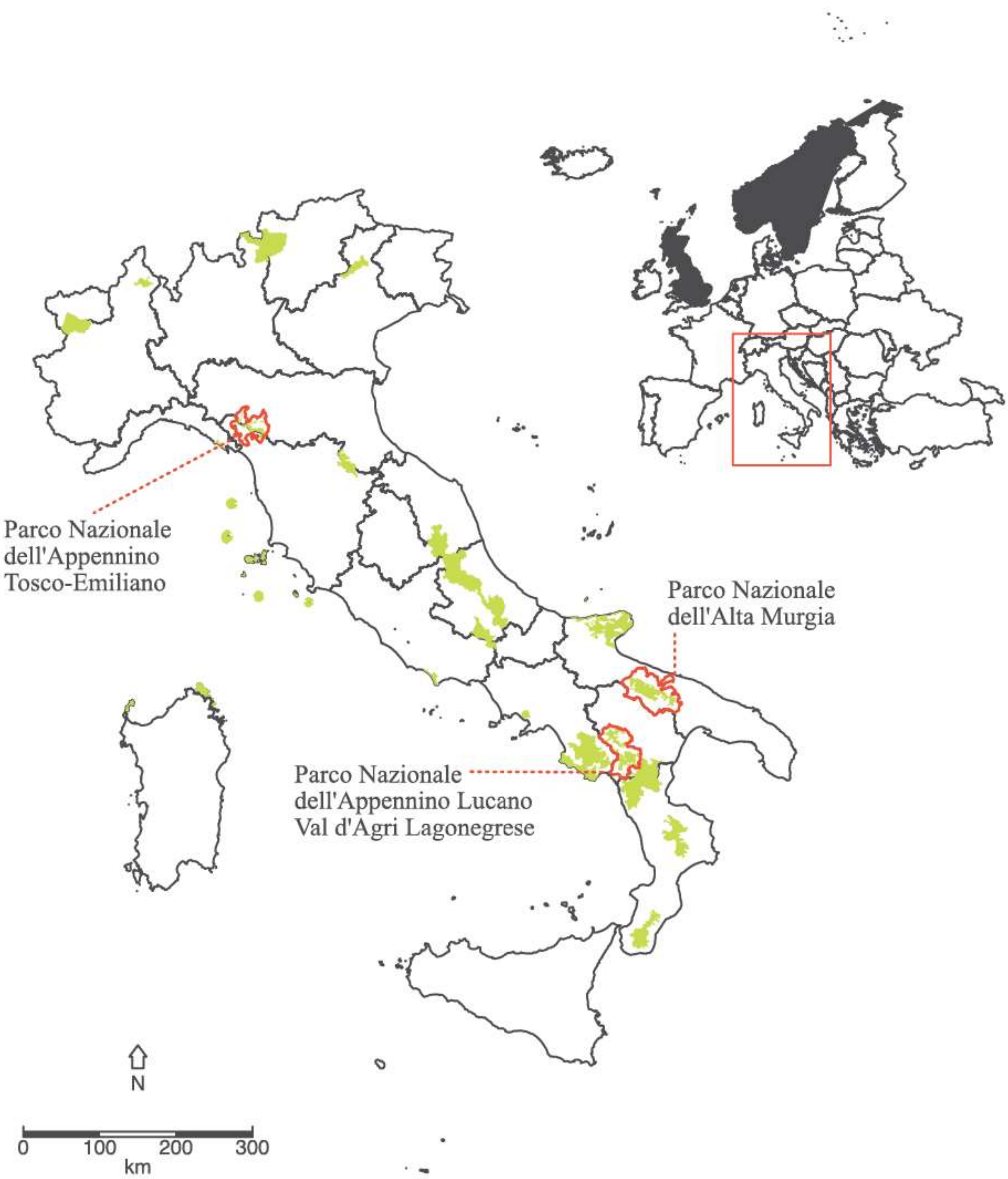

FIGURE 1 Italian national parks (NPs) and the municipalities in the three national NPs considered. NPs areas are depicted in green. Red bold borders indicate the boundaries of the municipalities under analysis for each of the three NPs. We did not indicate the Parco Nazionale del Golfo di Orosei e del Gennargentu that, while established in 1998, it was never put into effect. The map does not show the NP of the Island of Pantelleria (Sicily), a maritime park established in 2016.

funding for PA management/investments in visitor facilities) (Dixon and Sherman, 1991). On the other hand, local communities receive a flow of ecosystem services (Palomo et al., 2013; Spanò et al., 2017) that provide intangible benefits as well as monetary ones, for example, through tourism (Mayer et al., 2010).

In line with the recent literature (e.g., Andam et al., 2010; Ferraro & Hanauer, 2014; Geldmann et al., 2019; Joppa & Pfaff, 2011), this study follows the counterfactual approach to assess the impact of the establishment of a PA. The assessment of its impact is problematic because once a municipality becomes covered by it (i.e., it is "treated"), we cannot observe what would have happened in the same municipality if the area had not been protected. This is the well-known "fundamental problem of causal inference" (Holland, 1986). To overcome this issue, a matching rationale is applied. The intuition behind such a method is that, through matching, a subsample of "untreated" municipalities is selected according to the similarity (in terms of certain relevant characteristics) with the treated ones. Such a procedure mimics the design of a randomized experiment and, in principle, it is able to mitigate the treatment-selection bias, that is the issue of having treated units that systematically differ from the untreated ones (Rässler, 2002). The result is the balance of the covariates, that is, the most similar distribution of the observed covariates between the groups of treated and matched control municipalities. Given the matching outcome, the estimate of the impact of a NP considers only the difference between the treated municipalities and the (untreated) matched ones which serve as comparison group for the treated (they are also called "control" units).

More in details, first, we choose the observable covariates to be used for matching. Following the suggestions by Ho et al. (2007), these covariates are selected by means of a stepwise logistic regression procedure for the propensity score estimation. In other words, the covariates are chosen among the most statistically significant explanatory variables affecting the uptake of the treatment (i.e., the establishment of a NP) such that the "selection on observable" (Heckman & Robb, 1985) or "ignorability" condition (Rubin, 1978) is in place. Therefore, the selected covariates are used for matching since, as far as we can observe by means of the data at hand, they constitute all the statistically significant variables that are causally before the treatment, associated with it and do affect the outcome conditional on it.

Second, we match the municipalities that are part of the three Italian NPs with the most similar municipalities that are not included in any Italian NPs. This procedure is carried out by different matching methods, based on the aforementioned covariates. The outcome of the matching provides the most robust and statistically significant matching models in terms of covariates balance results.

Third, we estimate the ATT with the subset panel data at hand, that is, by applying the DID estimators to panel data on the municipalities in NPs (the treated) and their matched counterfactual observations (the matched controls).

The adopted matching procedures follow the suggestions and results provided by Ho et al. (2011, 2007). The matching on covariates through nonparametric techniques, like for example, the Mahalanobis distance, should be preferred to PSM since the former has been proved to be better justified when very large sample sizes are available after matching (King & Nielsen, 2019). Since this is not the case of the present study, we start from the most common matching approach, based on a linear probit PSM specification (Rosenbaum & Rubin 1983, 1985) and, then, we differentiate the "complexity" of the matching model up to nonparametric matching solutions based on the Mahalanobis distance (Cochran & Rubin, 1973; Rubin, 1980). The goal of such a strategy is the achievement of the highest balance between the two groups (the municipalities in NPs and the ones not in NPs) following the rule of balance maximization to avoid the "balancing test fallacy," as prescribed by Imai et al. (2008). Reaching the balance of covariates is crucial to obtain unconfounded causal comparisons (see, e.g., Rosenbaum, 2010) since when all relevant differences between treatment and comparison group units that affect outcomes are captured in the observed covariates (i.e., potential outcomes are independent of assignment to treatment, conditional on pretreatment observed covariates) matching yields a consistent estimate of the treatment impact (Lee, 2013).

After having selected the counterfactual municipalities, the matched data are used to build the DID estimator to identify the ATT. Then, the DID estimator for the ATT is applied to the matched subset of municipalities. Let be the case that we have two time periods t, a baseline ($t = 0$) and a follow-up ($t = 1$). Let Y_{it} be the outcome of interest for the ith municipality at time t. Let be $P_{it} = 1$ if the ith municipality is treated (i.e., if it hosts a NP) before time t, $P_{it} = 0$ otherwise. Since $_{Pi0} = 0$ for every i, we can simplify writing $_{Pi} = P_{i1}$. Adopting the notation of Rubin's potential outcome framework (Rubin, 2004), let be $Y_{it}(0)$ the outcome of municipality i at time t if it is not part of a NP, $Y_{it}(1)$ the outcome of the same municipality if it is part of the NP, instead. The outcome for the municipality i at time t is: $Y_{it} = P_i Y_{it}(1) + (1 - P_i) Y_{it}(0)$. Being available a set of covariates X_i, following the standard assumptions of DID methods (Abadie et al., 2004 and the references therein; Heckman et al., 1997) we assume that:

Assumption 1. $\{Y_{i0}, Y_{i1}, P_i, X_i\}$ for $i = 1, \ldots, n$ are independent and identically distributed (iid).

We are interested in estimating the ATT and hence, our parameter of interest is $\tau = E[Y_{i1}(1) - Y_{i1}(0)|P_i = 1]$. It is easy to rewrite the ATT as follows: $\tau = E[P_i = 1] - E[P_i = 1] = E[P_i = 1] - E[P_i = 1]$, where the last expectation on the right side of the equation represents the crucial entity to estimate with the data at hand.

We also assume that:

Assumption 2. $E[P_i = 1, X_i] = E[P_i = 0, X_i]$, which means that in absence of an established NP the average conditional outcome of the municipalities within a NP and the municipalities out of a NP would have been the same. In other words, this is the so-called "parallel trend assumption."

Assumption 3. For some $\xi > 0$, $P(P_i = 1) > \xi$ and $P(P_i = 1|X_i) \leq 1 - \xi$, meaning that there is at least a small portion of the municipalities that are included in a NP, and, for every value of the covariates, there is at least a small probability that the municipality is not part of a NP. This is, in other words, the so-called "overlap condition."

We estimate the ATT by means of (i) the Outcome Regression DID (ORDID) model and (ii) the Doubly Robust DID (DRDID). Under the above-mentioned assumptions we have that:

$$E[P_i = 1] = E[P_i = 1, X_i] + E[P_i = 0, X_i] - E[P_i = 0, X_i\,|P_i = 1] = E[P_i = 1] + E[P_i = 0, X_i] \\ - E[P_i = 0, X_i\,|P_i = 1].$$

Therefore, the ATT can be estimated by applying the ORDID estimator (Heckman et al., 1997) (to ease the notation, we drop the subscript i), as follows:

$$\hat{\tau}^{ORDID} = \underline{Y}_{1,1} - \left[\underline{Y}_{1,0} + \frac{1}{n_{P=1}} \sum_{i|P_i=1} (\hat{\mu}_{0,1}(X_i) - \hat{\mu}_{0,0}(X_i))\right]$$

with $\underline{Y}_{p,t} = \sum_{i|P_i=p,T_i=t} \frac{Y_{it}}{n_{p,t}}$ being the sample average outcome of municipalities in group p at time t and $\hat{\mu}_{p,t}(x)$ being the estimator of the true but unknown $m_{p,t}\ \ (x) = E[Y_t|P = p, X = x]$.

To introduce the Doubly Robust DID we have to consider the approach proposed by Abadie et al. (2004). Here, the ATT is introduced as follows: $\tau = \frac{1}{E[P]} E[\frac{P - ps(X)}{1 - ps(X)}(Y_1 - Y_0)]$, where $ps(X) = P(P = 1|X)$, i.e., the propensity score. Consequently, the estimator for the ATT proposed by Abadie et al. (2004) is such that:

$$\hat{\tau}^{ps} = \frac{1}{E_n[P]} - E_n\left[\frac{P - \hat{\pi}(X)}{1 - \hat{\pi}(X)}(Y_1 - Y_0)\right]$$

with $\hat{\pi}(x)$ being the estimator of the true but unknown $ps(x)$.

Since the consistency of the ATT estimator in the ORDID model relies on the fact that the estimators for $m_{p,t}(\cdot)$, $\hat{\mu}_{p,t}(\cdot)$ are correctly specified, whereas the consistency of the estimator proposed by Abadie needs the estimator $\hat{\pi}(\cdot)$ for $ps(\cdot)$ to be not mis-specified, Sant'Anna and Zhao (2020) developed the Doubly Robust DID, that combines them in a way that the resulting estimand is robust even if either the outcome regression model or the model for the propensity score are mis-specified. Consequently, let be $\Delta Y = Y_1 - Y_0$ and let be $\mu^{ps}_{p,\Delta}(X) = \mu^{ps}_{p,1}(X) - \mu^{ps}_{p,0}(X)$ where $\mu^{ps}_{p,t}(x)$ is the model for the true but unknown outcome regression $m^{ps}_{p,t}(x) = E\left[Y_t|P = p, X = x\right]$ with $p, t = 0{,}1$. The DRDID estimator for the ATT results to be the following one:

$$\hat{\tau}^{DRDID} = E\left[\left(w_1^{ps}(P) - w_0^{ps}(P, X, \pi)\right)\left(\Delta Y - \mu^{ps}_{0,\Delta}(X)\right)\right],$$

where, for a generic function g, we have that $w_1^{ps}(P) = \frac{P}{E(P)}$ and $w_0^{ps}(P, X; g) = \frac{g(X)(1-P)}{1-g(X)} / E\left[\frac{g(X)(1-P)}{1-g(X)}\right]$. For the sake of brevity, in relation to the DRDID estimator efficiency bounds and asymptotic properties, as well as with respect to the Monte Carlo simulations results that offer proofs of finite sample properties of the DRDID estimator, we refer to Sant'Anna and Zhao (2020).

The matching procedure and the estimation of the ATT have been performed by using the "Matching" (Sekhon, 2020), "MatchIt" (Ho et al., 2020), and "did" (Callaway & Sant'Anna, 2022) R packages, respectively, in addition to user-written R coding.

3 | DATA

3.1 | The case study

According to the official list of PAs by the Italian Ministry of Ecological Transition, in 2019 there were 25 NPs in Italy, covering 16,000 km^2, namely 5.3% of the total area of the country. In the country, the first introduction of NPs dates to the 1920s, when Parco Nazionale del Gran Paradiso, in the North-Western of Italy, was established. Since then, new NPs have been mostly set up throughout mountain areas (both the Alps and the Apennines range). Later, and since the 1990s, several marine NPs have been created as well, especially throughout Southern regions. Regarding the time dimension, the largest expansion in the overall area under protection was experienced during the 1990s, when 12 new NPs—covering an area of 6946 km^2—were established. Only four new NPs were established during the 2000s, but still encompassed more than 2300 km^2 under protection.

A vast majority of the Italian NPs surface lies in mountain areas (i.e., at an altitude which is generally greater than 600 m above sea level). Since the 1960s, these areas have suffered from steady depopulation processes (Romano et al., 2021), as a direct consequence of remoteness, poor accessibility, and lack of economic development. Also, economic wealth is below the national average.[4]

As already mentioned, the choice to consider only those NPs created in the 2000s is justified by the availability of economic data to compare the ex-ante and the ex-post situation. Indeed, the data on (taxable) per capita income—which is one of the economic dimensions under investigation in this study—are available only for the period from 2000 onwards. We focus on the following NPs: Parco Nazionale dell'Appennino Tosco-Emiliano (hereinafter, PAT—established in 2001–), Parco Nazionale dell'Alta Murgia (hereinafter, PAM—established in 2004–), and Parco Nazionale dell'Appennino Lucano Val d'Agri Lagonegrese (hereinafter, PAL—established in 2007–).[5]

The impact assessment of the introduction of a new NP and the framing of the counterfactual analysis in a territorial perspective are based on the municipality level, as it represents the level for which data with the most detailed spatial granularity are available.[6] To this regard, the treated municipalities are those that are officially included into the three parks, according to the lists provided by the National Ministry of Ecological Transition.[7] As a result, 50 Italian municipalities are considered as treated: 11 of them are covered by the PAT, 13 of them by the PAM, 26 of them by the PAL (see Figure 1). The municipalities representing the case study area are from different Italian regions (i.e., both from the Northern and the Southern part of Italy). They are located at an average altitude of 599 m above sea level, being at a distance from the coast equal to 32 km, on average. They are quite large municipalities in terms of km^2, and they are characterized by high percentages of forested land and a low population density. To this regard, they are similar to the vast majority of the municipalities in the Italian NPs. Taken together,

[4]While here we do not assess their environmental impact, Italian NPs seem to be located in area with low human pressure, as suggested at the global level by Joppa and Pfaff (2011).

[5]We exclude from the analysis the fourth NP established in the 2000s, namely the Parco Nazionale della Sila (located in Calabria), as it was created upon an older NP, that is, the Parco Nazionale della Calabria, originally established in 1968. That NP was firstly reduced in 1989, with the establishment of the Parco Nazionale dell'Aspromonte, and then definitely canceled in 2002, when the Parco Nazionale della Sila was established. Due to the existence of this former NP, it would be misleading to consider the institution of the new PA as a pure treatment within a counterfactual approach framework.

[6]Regarding the layer of the boundaries of the Italian municipalities, we have considered the 2019 layer, hence a total number of 7,926 municipalities. Note that, since 1991, the number of the Italian municipalities has decreased by 174, due to an overall merger process.

[7]https://www.mite.gov.it/pagina/elenco-dei-parchi.

these municipalities account for 5515 km^2, with a total PA equal to 1520 km^2 (30.41 km^2 per municipality, on average). Their total population is equal to 524,480 inhabitants (at the baseline year 2001), namely 10,490 inhabitants on average. Their per capita income is equal to € 10,205 on average (at the baseline year 2001).

3.2 | Variables description

The empirical analysis is grounded on a set of secondary-level variables, which have been retrieved from different open access official statistics sources. The largest part of the variables under analysis are retrieved from the Italian National Institute of Statistics (Istat) sources. In particular, from: (i) the 1991, 2001, and 2011 Census of Population and Housing; (ii) the 1991, 2001, and 2011 Census of Industry, Services and Non-profit Institutions; and iii) the 1990, 2000, and 2010 Census of Agriculture (available at http://dati.istat.it/). In addition, Istat also provides data on the geographical characteristics of the Italian municipalities (e.g., the altitude above sea level and the area of the municipality). Additional information–namely, the distance from the coast of each municipality—has been elaborated starting from the Istat data, by computing the distance between each municipality's centroid and the centroid of the closest coastal municipality. Two additional sources of data are the 1990 CORINE Land Cover for land use,[8] and the open data of the fiscal declaration data set provided by the Italian Ministry of Economy and Finance (available at http://dati.istat.it/). Collected data refer to different years. For those municipalities that over the years have been merged (see footnote 6), past data have been converted according to the 2019 layer.[9]

By referring to the aforementioned data sources, two sets of variables are extracted and considered for the analysis. The first set of variables is used for the matching of the treated municipalities with the control ones; the second set refers to the outcome variables, adopted to assess the effect of the establishment of a NP.

With regard to the matching procedure, following the prescriptions of Andam et al. (2010) and Garrido et al. (2014), preliminary data management is implemented to remove from the group of possible counterfactual observations (to be matched with the municipalities in the NPs) those that belong to any other NP. According to these criteria, the set of possible control municipalities includes 7424 municipalities. To them and to the ones covered by the three NPs under consideration (i.e., the 50 treated municipalities), the matching techniques are applied, to identify the most similar municipalities at the baseline year 2001.

The matching covariates include altitude above sea level (m), distance from the coast (km), area of the municipality (km^2), percentage of forested land at year 1990, nr. of local establishments (enterprises) located in the municipality at year 1991, nr. of workers employed in local establishments at year 1991, nr. of agricultural holdings at the year 1990, population density at year 1991 (inhabitants/km^2) and per capita income at year 2000 (Euro). The characteristics under consideration refer to the geographical and socioeconomic conditions. Both the altitude above sea level and the distance from the coast are admitted as proxies for remoteness of the municipality and its mountain degree, as the NPs under consideration are mostly located across the Apennines.[10] The percentage of forested land allows to single out those municipalities characterized by large natural and seminatural habitat. As noticed by Falcucci et al. (2007), forested areas have shown a marked increase in Italy, in the period 1960–2000,

[8]Land cover data are provided as geospatial data (available at land.copernicus.eu/pan-european/corine-land-cover). For this analysis, they have been superimposed onto the boundaries of the Italian municipalities, to get municipality-level data (see footnote 6 for the administrative layer that is adopted).

[9]Note that, among the treated municipalities, two of them underwent a process of merger in 2015 (Sillano Giuncugnano) and in 2016 (Ventasso). They are both located within the boundaries of the PAT.

[10]Following the suggestion of one anonymous reviewer, other factors (different from the distance from the coast) potentially affecting the dynamic of infrastructural connection between the municipalities in NPs and the metropolitan areas around them were also worthy to be included in the matching. Namely, with the data at disposal, we explored the possibility to include in the matching procedure two variables: the distance from the administrative capital of the province (where the municipalities in NPs are located) and the distance from the closest administrative province-level capital. Nevertheless, in the several matching models that we have explored, neither the former, nor the latter variable showed statistically significant coefficients. In addition, the different matching procedures carried out by including them (despite the fact that they were not statistically significant) did not provide a sufficient balance of the treated and nontreated groups. In certain cases, the balance even worsened with respect to the current one. Therefore, we decided to not consider these covariates for matching. For further details, the related additional material is available upon request to the authors.

especially in the mountainous and hilly areas of the country. Lastly, socioeconomic data, such as population density, per capita income, number of local establishments (and number of workers employed) as well as the number of agricultural holdings, sheds light on the similarities about the ex-ante economic conditions of the municipalities. Table 1 shows the covariates observed for the municipalities in NPs and those not in NPs.[11]

We assess the effect of the establishment of a NP from the baseline year 2001 (year 2000 for the covariates taken from the Agricultural Census) to the follow-up year 2011 (year 2010 for the covariates taken from the Agricultural Census), by considering the following outcome variables: per capita (taxable) income of the resident population (Euro), population level (nr. of inhabitants), nr. of local establishments (enterprises) per municipality, nr. of workers employed in local establishments,[12] nr. of tourism sector establishments, nr. of workers employed in local tourist establishments, nr. of agricultural holdings, Utilized Agricultural Area (UAA, in hectares—ha). Table 2 shows the descriptive statistics related to the outcome variables (at both the baseline and follow-up years), for the municipalities within the NPs and the ones outside.

4 | RESULTS

4.1 | Covariates and matching models selection

Table 3 depicts the results of the best logistic regression model (in terms of both the number of covariates included and the coefficients' significance) that is finally elected for the covariates selection. All the covariates considered in the model are statistically significant in identifying a municipality as being part of a NP and are crucial to single out those municipalities that share major similarities with the ones included in the treatment group to build the ex-post counterfactual set of observations to be compared with it. In addition to these covariates, the area of the municipality (km^2) is included in the matching approaches that are not strictly based on the propensity score estimation. This variable is not statistically significant in any model for the propensity score estimation, however, it is included in the nonparametric matching procedures that fully exploit the observable covariates' values since it is very relevant to account for the differences in the municipalities size.

These covariates are included in five matching approaches that have been selected since they are the most robust and statistically significant in terms of covariates balance. They are: the linear probit PSM[13] (M01), the logistic PSM with smoothing spline function[14] (M02), the mere Mahalanobis matching (M03), its combination with the estimated propensity score that is used for caliper-based units discarding (M04) and the full-covariates matching with controls discarding based on the common support (M05). The models M03, M04, and M05 also include the area of the municipality as one of the covariates. The models M01, M02, M04, and M05 include a variable named "score," which is the absolute difference between the propensity score distance of treated-control pairs of units, that is, a "position" measure.[15] Moreover, as it is explained in Section 1 of the Supplementary materials, the nonparametric matching models are based on the full exploitation of the covariates' values by means of the Mahalanobis distance function (eventually, in combination with the estimated propensity score). For them (M03–M05), the computation of the Mahalanobis distance is based also on the area of the municipality.

[11]Table 1 presents only the covariates selected by the stepwise logistic regression procedure whose results are described in Section 4.1. Therefore, they are the final variables selected for matching and used in the present study. In addition to them, other variables were available from the data at hand but have been discarded due to the methodological rationale described. Namely, the following covariates resulted to be nonstatistically significant: ski routes (in km), percentage of urbanized land at year 1990, nr. of local establishments at year 1981, nr. of workers employed in local establishments at year 1981, nr. of agricultural holdings at year 1982, Utilized Agricultural Area at years 1982 and 1990.

[12]According to the official statistics, the workers employed in local establishments are considered disregarding the municipalities where they live in.

[13]Table A1 in Appendix shows the results of the linear probit PSM model (M01).

[14]Table A2 in Appendix shows the results of the logistic PSM with a smoothing spline function on per capita income model (M02).

[15]We decided not to name the variable "score" as "distance," as it is usually done in the literature (Ho et al., 2020), to better distinguish it from the observable covariates referred to the municipalities (e.g., the distance from coast, the geographic distance among municipalities...).

TABLE 1 Descriptive statistics for the selected covariates (unmatched sample)

Variable	Mean (standard deviation) Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 7424)	Data source
Altitude above sea level (m)	598.840 (242.771)	340.883 (288.479)	Istat, "Principali statistiche geografiche sui comuni"
Distance from the coast (km)	32.401 (12.413)	67.182 (55.853)	Authors' elaboration on Istat data
Area of the municipality (km^2)	110.300 (95.877)	36.205 (49.060)	Istat, "Principali statistiche geografiche sui comuni"
Percentage of forested land (year 1990)	0.541 (0.295)	0.360 (0.328)	1990 CORINE Land Cover for land use
Nr. of local establishments (year 1991)	521.780 (868.246)	421.017 (2202.923)	Istat, "7th General Census of Industry and Services"
Nr. of workers employed in local establishments (year 1991)	1520.620 (2690.658)	1785.599 (12,251.040)	Istat, "7th General Census of Industry and Services"
Nr. of agricultural holdings (year 1990)	1192.140 (1634.191)	348.937 (531.762)	Istat, "4th General Census of Agriculture"
Population density (year 1991, inhabitants/km^2)	71.212 (61.904)	281.192 (641.953)	Istat, "13th General Census of Population and Housing"
Per capita income (year 2000, Euro)	10,204.688 (1733.821)	12,521.294 (2984.664)	Italian Ministry of Economy and Finance

Tables A3–A7 in Appendix show the balance results of the covariates for the different models. In other words, they depict how much similar the groups of treated and control municipalities have become due to matching. The results indicate a very good balance between the treated and untreated municipalities after matching in terms of the standardized mean difference (SMD), variance ratio and percentage balance improvement (see Appendix A for details). Indeed, for almost all the covariates, SMD values are below the 0.1 threshold (or between 0.1 and 0.2) thus indicating negligible differences between the matched municipalities. Moreover, covariates variance ratios, overall, are not far from 1, indicating close variances of the covariates' distribution within the matched pairs of units. Finally, the percentage balance improvement indicates an increase of similarity in terms of SMD, variance ratio as well as mean and maximum of the empirical cumulative density function.

In line with the literature in balancing diagnostics when counterfactual inference is addressed (see, e.g., Austin, 2009), we focus on the discussion of the "critical" SMD values. By ordering for the "level of unbalance," attention should be paid to the population density (i.e., slightly unbalanced in models M03 and M04—note that, for the latter model, the variance ratio coefficient is also far from 1–), the distance from coast in models M01 and M04 (the related variance ratio coefficients are also "worrying," as per the ones in models M02, M03, and M05), the

TABLE 2 Descriptive statistics for the outcome variables (unmatched sample)

Variable	Baseline year 2001, mean (standard deviation)		Follow-up year 2011, mean (standard deviation)	
	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 7424)	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 7424)
Per capita income (Euro)	10579.676 (1635.155)	12903.111 (3026.829)	18810.210 (1729.546)	20553.192 (3060.582)
Population level (nr. of inhabitants)	10489.601 (18753.738)	7261.932 (40856.847)	10717.901 (19676.153)	7590.914 (41264.356)
Nr. of local establishments	594.301 (1149.230)	482.378 (2971.128)	736.040 (1416.305)	618.814 (4284.523)
Nr. of workers employed in local establishments	1809.121 (3622.802)	1884.454 (12834.406)	1991.505 (3960.982)	2131.703 (16149.904)
Nr. of tourism sector establishments	30.500 (38.357)	33.300 (187.326)	41.960 (60.822)	42.958 (249.232)
Nr. of workers employed in tourism sector establishments	74.320 (106.426)	110.578 (887.194)	126.000 (206.854)	157.199 (1218.724)
Nr. of agricultural holdings	1062.800 (1596.966)	292.620 (485.447)	784.802 (1365.831)	197.319 (350.355)
Utilized Agricultural Area (UAA, ha)	5443.178 (7076.524)	1616.093 (2590.875)	5518.751 (7905.362)	1574.349 (2685.891)

Source: Italian Ministry of Economy and Finance for per capita income. 14th (2001) and 15th (2011) General Censuses of Population and Housing for the population level. 8th (2001) and 9th (2011) General Censuses of Industry and Services for the nr. of local establishments, nr. of workers employed in local establishments, nr. of tourism sector establishments. 5th (2000) and 6th (2010) General Censuses of Agriculture for the nr. of agricultural holdings and the Utilized Agricultural Area (UAA).

TABLE 3 Logistic regression results for the covariates selection

Variable	Coefficients	Standard error
Altitude above sea level	0.00324***	0.00059
Distance from the coast	0.04074**	0.01874
Percentage of forested land	1.10800*	0.66340
Nr. of local establishments	0.00302***	0.00090
Nr. of workers employed in local establishments	−0.00066***	0.00024
Nr. of agricultural holdings	0.00053***	0.00014
Population density	−0.00999***	0.00313
Per capita income	0.00024**	0.00011
Constant	−7.64000***	1.23900

Note: Observations: 7424. Pseudo R^2: 0.251.

Significance levels: *0.1; **0.05; ***0.01.

percentage of forested land in models M03 (that presents "severe" unbalance). However, these variables are not exogenous by construction (i.e., they cannot change over time, while the others could). Hence, by conditioning (also) on the exogenous-by-construction covariates, we cannot occur in violations of the identification assumption (Assumption 1) since exogeneity holds anyway (Lechner, 2010). Therefore, considerations on even "severe" imbalance of these covariates cannot be thought as alarming (as it is in the only case of percentage of forested land).

In addition, Figures 1–12 of Section 2 of the Supporting Information presents the distributions of the covariates for the unmatched and matched municipalities in NPs and not in NPs, side-by-side boxplots, as suggested by the literature in observational studies and counterfactual approach estimation (Austin, 2009). The distributions of the covariates between unmatched and matched municipalities become more similar after matching, with a great improvement in balance for all the variables considered (and the score). The only covariate that, after matching, still presents relevant differences in the distribution between the treated and the matched controls is the distance from coast.

4.2 | Impact of the three NPs

Tables A8–A12 in Appendix A show the descriptive statistics related to the eight outcome variables at the baseline year 2001 and the follow-up year 2011 for the matched treated and controls.[16] Some differences exist among the outcomes of the different matching models, that is, in terms of distribution of the outcome variables observed for the subsamples of the matched treated and controls. The outcome variables hint at the fact that the matched treated and controls change differently during time. These changes are coherent among the matched subsamples generated by the different matching models.

Table 4 presents the estimated impact that the three NPs' establishment produced on the eight outcome variables considered. Namely, the ATT estimated by means of both the Outcome Regression DID and the Doubly Robust DID is presented for each of the matching approaches adopted to select the treated and control municipalities (M01–M05). The ORDID and DRDID models show a coherent pattern for the ATT estimation with respect to all the outcome variables but for the number of local establishments. In general terms, the only estimates that are never statistically significant are those for the proxies of the agricultural activities (namely, number of agricultural holdings and the hectares of UAA).

Results hint at a statistically significant, coherent, strong, and positive impact of NPs on income (but for DRDID in M05). The three Italian NPs seem to increase the taxable income with estimates between a minimum €408.45 and a maximum €763.30 of per capita income increase. Regarding population, the results from both ORDID and DRDID in models M03-M05 show that the three NPs addressed have a positive impact on the population level in comparison with the control municipalities, even if at different levels of statistical significance.

The other results are somewhat weaker. The impact on income does not seem to couple with the creation of new businesses in the treated municipalities, but the results are only significant for a limited number of model specifications. There is evidence that for a municipality being part of a NP, it induces a negative impact on the number of establishments settled within its boundaries, as well as on the number of workers employed (and the same results are showed by the number of tourism local establishments and the number of workers they employ). This happens despite the fact that the statistical significance is not uniform between the ORDID and DRDID estimates, neither among the different matching approaches.

Finally, the effect of the establishment of a new NP on the hectares of UAA and on the number of agricultural holdings is not statistically significant, mainly due to the large standard errors of the estimated quantities of interest.

[16]Refer to the Figure A1 for the geographical distribution of the matched treated and control municipalities for the five matching models adopted.

TABLE 4 Estimated average treatment effect on the treated (ATT)

	Outcome regression DID					Doubly robust DID				
Outcome variable	M01	M02	M03	M04	M05	M01	M02	M03	M04	M05
Per capita income	571.966** (255.713) [112.538, 1031.395]	763.298*** (267.798) [222.180, 1304.417]	589.384*** (198.582) [201.388, 977.381]	422.841** (191.267) [44.480, 801.202]	408.454** (206.572) [4.376, 812.532]	568.837** (230.095) [123.693, 1013.980]	673.328*** (241.231) [123.693, 1013.980]	587.568*** (189.539) [201.971, 973.165]	451.206** (183.484) [139.041, 763.370]	205.082 (160.021) [−91.593, 501.757]
Population level	−12.147 (112.217) [−242.744, 218.450]	53.813 (114.175) [−207.417, 315.043]	257.963** (125.571) [6.658, 509.267]	142.364* (66.784) [8.796, 275.932]	242.772*** (103.209) [24.304, 461.241]	19.792 (90.185) [−173.737, 213.320]	47.776 (100.962) [−148.301, 243.852]	264.754** (121.245) [27.627, 501.881]	142.678* (70.281) [2.116, 282.926]	177.808* (101.520) [−80.782, 436.398]
Nr. of local establishments	−37.345** (17.539) [−72.961, −1.728]	−11.124 (18.878) [−52.529, 30.282]	1.452 (15.667) [−32.313, 35.217]	7.883 (10.030) [−15.744, 31.509]	2.435 (10.025) [−20.850, 25.719]	−36.253** (18.000) [−67.372, −5.134]	−10.131 (14.478) [−51.084, 30.822]	2.687 (18.102) [−27.365, 32.739]	8.896 (9.603) [−10.602, 28.393]	−9.688 (12.875) [−38.861, 19.486]
Nr. of workers employed in local establishments	−147.363** (66.307) [−276.223, −18.502]	−117.368** (60.038) [−247.552, 12.817]	−106.285 (67.808) [−251.148, 38.577]	−84.805* (50.932) [−195.499, 25.890]	−38.682 (53.747) [−156.608, 79.244]	−165.280** (66.523) [−295.665, −34.895]	−116.807* (67.582) [−249.268, 15.655]	−106.572 (69.727) [−243.238, 30.093]	−78.462 (54.142) [−184.580, 27.657]	−107.055 (75.439) [−254.915, 40.805]
Nr. of tourism sector establishments	−6.345*** (1.522) [−9.828, −2.862]	−2.698 (1.822) [−6.448, 1.052]	−1.562 (1.490) [−4.708, 1.584]	−3.066* (1.751) [−6.782, 0.651]	−3.726*** (1.470) [−7.171, −0.282]	−5.922*** (1.533) [−9.024, −2.819]	−2.555** (1.250) [−6.555, 1.445]	−0.883 (1.515) [−3.609, 1.843]	−2.548** (1.235) [−5.606, 0.511]	−5.169*** (1.756) [−9.143, −1.196]
Nr. of workers employed in tourism sector establishments	−29.450*** (9.373) [−48.947, −9.952]	−7.453 (9.004) [−25.846, 10.939]	1.114 (6.554) [−13.056, 15.283]	−9.989 (6.468) [−24.567, 4.591]	−8.038 (6.719) [−25.040, 8.964]	−27.656*** (11.652) [−46.976, −8.336]	−7.600 (7.036) [−21.175, 5.974]	2.172 (5.916) [−9.529, 13.872]	−8.195 (5.463) [−18.461, 2.071]	−13.659* (7.143) [−32.601, 5.283]

TABLE 4 (Continued)

Outcome variable	Outcome regression DID M01	M02	M03	M04	M05	Doubly robust DID M01	M02	M03	M04	M05
Nr. of agricultural holdings	−21.995 (46.614) [−123.708, 79.719]	−3.863 (41.605) [−95.300, 87.574]	−35.701 (39.927) [−119.986, 48.584]	−11.380 (40.463) [−101.330, 78.571]	−39.261 (51.195) [−142.194, 63.673]	−20.283 (46.458) [−125.677, 85.110]	3.652 (44.603) [−74.089, 81.393]	−38.623 (33.300) [−97.200, 19.954]	−8.601 (42.027) [−109.639, 92.437]	31.943 (48.506) [−55.323, 119.208]
Utilized agricultural area	−128.096 (228.891) [−619.017, 362.825]	−156.784 (256.693) [−654.420, 340.851]	152.687 (217.692) [−299.244, 604.618]	−106.697 (214.389) [−582.046, 368.652]	−39.261 (51.195) [−142.194, 63.673]	−256.296 (223.057) [−621.761, 109.169]	−155.730 (240.685) [−634.686, 323.226]	172.587 (199.615) [−276.699, 621.873]	−134.662 (197.225) [−561.214, 291.891]	−154.168 (230.853) [−571.512, 263.176]

Note: Values are (standard errors) and [95% confidence intervals].

Abbreviation: DID, diff-in-diff.

Significance levels: *0.1; **0.05; ***0.01.

5 | DISCUSSION

The results discussed in Section 4—which are quite robust among alternative model specifications– seem suggesting that the establishment of one of the three NPs under consideration here has led to a positive effect on the socioeconomic conditions of those municipalities that are part of them, at least in a comparison with what observed in the untreated municipalities.

The most robust finding is the one about the positive variation of per-capita income, after the establishment of a NP. Potential comparisons with other similar analyses suffer from the fact that other articles focus on different outcome variables and on different IUCN categories of PAs. For example, we use the *average* income as a proxy for wealth, whereas, for example, Sims et al. (2019) and Chen et al. (2016) use *median* income. Furthermore, the proxy that is adopted in this study for employment refers to the local establishments and the workers employed in the local establishments at municipality level (hence, disregarding the municipality they live in). Even with these caveats, our results further confirm that NPs are not linked to poverty traps, as the great majority of the literature finds (Norton-Griffiths & Southey, 1995; Wilkie et al., 2006).

However, given the data at disposal, we cannot assess the effect on the distribution of income, and hence we cannot determine whether the increase in the average is due to an overall increase in income, or to a shift from low-income categories to higher ones, or to an increase in the income of the richest segment of the population only.

Moreover, the finding is (apparently) in contrast with the negative impact of a new NP on the presence of economic activities (i.e., the number of local establishments and of workers employed) in the affected area. Considering the contrasting results between income and other economic indicators, several possible explanations could be admitted (but could not be tested given the currently available data).

A first explanation for this phenomenon could lie in the fact that the establishment of a new NP, while attracting raising number of tourists and their associated money flows, imposes restrictions on the creation of new infrastructures (Mayer et al., 2010; Oldekop et al., 2016), hence limiting the growth in the number of both local establishments and employment. This outcome can be considered as a sort of disequilibrium between demand and supply, eventually creating a rent for the few economic activities that were already allocated in those municipalities that host a NP. The occurrence of an increase in touristic flows after the establishment of a new PA has been long studied in the literature both in the USA (Loomis, 1999; Weiler & Seidl, 2004) and in Europe (Fredman et al., 2007). For instance, Fredman et al. (2007), by focusing on a single Swedish NP, observe a visitor increase by almost 40%, by comparing the year before and the year after the park designation. Further research, if data becomes available, should then focus on the touristic flows, and on the prices of touristic services. Moreover, as observed by Romano et al. (2021), the large number of day-trip visitors together with the fact that the number of second homes in Italian NPs is particularly large (i.e., nearly 50% of the total number of homes) could also justify the poor effect of the creation of a new NP on new creation of new (or larger) touristic establishments.

Second, one could argue an additional motivation to explain the negative effect on economic activities (nr. of establishments and employment), in contrast with the positive effect observed on per-capita income after the establishment of a new NP. This establishment—which is in part exogenous to the local community– contributes to a dramatic change in the local economy characteristics of the hosting municipalities. By introducing specific restrictions on the construction of new buildings and infrastructure, this event creates a kind of "gentrification" of the labor market in the affected communities. As a result of the higher opportunity cost of available land (hence, of the rent), low-income jobs and economic activities are pushed out of the area, which instead retains only high-income jobs. This hypothesis would be verifiable provided that we had data on the distribution of income at the municipality level.

Third, probably the positive impact on income is due to larger working opportunities also in the municipalities that are outside the NPs under consideration, but still in their neighborhood (e.g., within a given commuting distance, or within the same Labor Market Area). Future research should then focus on what happens outside the

NPs, whether the restrictions on the economic activities within the border of PAs create positive spillovers on the neighboring areas.

In addition to the aforementioned ones, further limitations apply. Referring to methodological issues, the heterogeneity of the treated municipalities under consideration here, that is, the differences that they present in terms of socioeconomic characteristics and, also, with respect to their territorial peculiarities, is of major concern for the further development of the analysis. For example, very recently, new methods arose allowing to consider in the estimation of the ATT the treatment heterogeneity due to both observable and unobservable factors (Sakaguchi, 2020). Moreover, it is envisaged an extension of the analysis to the whole set of PAs, including the other IUCN categories to see whether the effect depends on the IUCN categories (Nelson & Chomitz, 2011), for example, by means of a "multiple-treatment" approach (Callaway & Sant'Anna, 2020; Lopez & Gutman, 2017).

While our analyses represent a first step for a proper assessment of the impact of the Italian NPs, policy implications are limited by the impossibility of further disentangling the mechanisms through which they affect the local economies. The main message drawn by the results is that the fear of negative economic consequence of NP establishment seems not to be justifiable. However, the design of new NPs would benefit by further research. For example, it would be relevant to have a picture of how the economic effect changes over time. While in the medium period the impact is positive, it might not be in the short one, and if this is the case, temporary transfers could be useful to improve the acceptability of new PAs. Second, if NPs are associated to changes in the composition of the economy, policies should be put in place to prepare the human capital necessary to face the change. Moreover, if the impact of NPs goes beyond their borders, attention to the area outside their boundaries would be crucial.

6 | CONCLUSION

PAs have experienced a terrific increase in the latest years, from being confined to spectacular and remote regions to locations potentially more sensitive from a socioeconomic point of view. This expansion is feared to be linked to negative effects for the local populations, fueling a trade-off between land preservation and economy enhancement, as every PA, albeit to different degrees, entails restriction to the economic activities. These concerns are even greater in a country like Italy since, there, most of the PAs include within their boundaries, municipalities suffering from remoteness and weaker economic development. Following these worries, NPs have expanded their scope well beyond the mere environmental goals, to include, for example, poverty reduction. An increasing literature has started assessing what is the socioeconomic impact of PAs. The great bulk of the literature has focus on the Global South, where the trade-offs between economic and protectionist goals are likely to be exacerbated. Works focusing on the socioeconomic impact of PAs in high income countries are rather scarce and limited to the USA.

In this article we analyze the socioeconomic impact of three NPs in Italy, using a counterfactual approach based on both the Outcome Regression Diff-in-Diff and the Doubly Robust Diff-in-Diff combined with both the Mahalanobis distance matching (with and without replacement) and optimal matching procedures. The results indicate that the three NPs established in the 2000s are not evidently linked to poverty traps (Naughton-Treves et al., 2005). The most important result, especially for those local policymakers or stakeholders who are asked to evaluate the establishment of a new NPs is its positive impact on the average income of the people living within their boundaries. Conversely, the other effects are more ambiguous, with just weak evidence of the increase of population, small (but negative) effects on local establishments and touristic activities. A potential interpretation for this seemingly odd result is that at the same time the NPs attract a flow of tourists and money but that the restrictions imposed by the PAs rules impede to further boost the development of infrastructures and additional economic activities.

Our results call for future analyses aimed at further disentangling the mechanisms through which PAs impact on the territorial socioeconomics. More specifically, future research should highlight to what extent the establishment of NPs polarize the hosting territories in terms of economic activities within them, and whether the economic impact of NPs spillover effects in the neighboring territories and whether the economic impact depends on the biodiversity protection effectiveness.

ACKNOWLEDGEMENT

The authors thank Daniele and Stefano whose discussions inspired the research. The authors also thank the anonymous reviewers and the editor for their valuable comments and suggestions improving the paper. The usual disclaimers apply. Open Access Funding provided by Universita degli Studi di Padova within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Riccardo D'Alberto http://orcid.org/0000-0002-7227-7485
Francesco Pagliacci http://orcid.org/0000-0002-3667-7115
Matteo Zavalloni http://orcid.org/0000-0002-6291-7653

REFERENCES

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, *4*, 290–311. https://doi.org/10.1177/1536867X0400400307

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., Robalino, J. A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 16089–16094. https://doi.org/10.1073/pnas.0800437105

Andam, K. S., Ferraro, P. J., Sims, K. R. E., Healy, A., & Holland, M. B. (2010). Protected areas reduced poverty in Costa Rica and Thailand. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9996–10001. https://doi.org/10.1073/pnas.0914177107

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, *28*, 3083–3107. https://doi.org/10.1002/sim.3697

Bennett, N. J., Franco, A. D., Calò, A., Nethery, E., Niccolini, F., Milazzo, M., & Guidetti, P. (2019). Local support for conservation is associated with perceptions of good governance, social impacts, and ecological effectiveness. *Conservation Letters*, *12*, e12640. https://doi.org/10.1111/conl.12640

Braber, B., den, Evans, K. L., & Oldekop, J. A. (2018). Impact of protected areas on poverty, extreme poverty, and inequality in Nepal. *Conservation Letters*, *11*, e12576. https://doi.org/10.1111/conl.12576

Bushell, R., & Eagles, P. F. (2006). *Tourism and protected areas: Benefits beyond boundaries: The Vth IUCN World Parks Congress*. Cabi.

Callaway, B., & Sant'Anna, P. (2022). *DID: Difference in differences. R package version 2.1.2*.

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*, 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chen, Y., Lewis, D. J., & Weber, B. (2016). Conservation land amenities and regional economies: A postmatching difference-in-differences analysis of the northwest forest plan. *Journal of Regional Science*, *56*, 373–394. https://doi.org/10.1111/jors.12253

Cochran, W. G., & Rubin, D. B. (1973). *Controlling bias in observational studies: A review. Sankhyā: The Indian Journal of Statistics, Series A (1961–2002)*, *35*, 417–446.

Di Franco, A., Thiriet, P., Di Carlo, G., Dimitriadis, C., Francour, P., Gutiérrez, N. L., Jeudy de Grissac, A., Koutsoubas, D., Milazzo, M., Otero, M. delM., Piante, C., Plass-Johnson, J., Sainz-Trapaga, S., Santarossa, L., Tudela, S., & Guidetti, P. (2016). Five key attributes can increase marine protected areas performance for small-scale fisheries management. *Scientific Reports*, *6*, 38135. https://doi.org/10.1038/srep38135

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics*, *95*, 932–945. https://doi.org/10.1162/REST_a_00318

Dixon, J. A., & Sherman, P. B. (1991). Economics of protected areas. *Ambio*, *20*, 68–74.

Dudley, N. (2008). *Guidelines for applying protected area management categories*. IUCN. https://doi.org/10.2305/IUCN.CH.2008.PAPS.2.en

Falcucci, A., Maiorano, L., & Boitani, L. (2007). Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. *Landscape Ecology*, *22*, 617–631. https://doi.org/10.1007/s10980-006-9056-4

Ferraro, P. J., & Hanauer, M. M. (2014). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 4332–4337. https://doi.org/10.1073/pnas.1307712111

Ferraro, P. J., Hanauer, M. M., & Sims, K. R. E. (2011). Conditions associated with protected area success in conservation and poverty reduction. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 13913–13918. https://doi.org/10.1073/pnas.1011529108

Fredman, P., Friberg, L. H., & Emmelin, L. (2007). Increased visitation from national park designation. *Current Issues in Tourism*, *10*, 87–95. https://doi.org/10.2167/cit293.0

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health Services Research*, *49*, 1701–1720. https://doi.org/10.1111/1475-6773.12182

Geldmann, J., Manica, A., Burgess, N. D., Coad, L., & Balmford, A. (2019). A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 23209–23215. https://doi.org/10.1073/pnas.1908221116

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*, 1017–1098. https://doi.org/10.2307/2999630

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *The Review of Economic Studies*, *64*, 605–654. https://doi.org/10.2307/2971733

Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, *30*, 239–267. https://doi.org/10.1016/0304-4076(85)90139-3

Ho, D., Kosuke, I., King, G., Stuart, E., Whitworth, A., & Greifer, N. (2020). *MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. R package version 4.1.0*. https://cran.r-project.org/web/packages/MatchIt/index.html

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*, 199–236. https://doi.org/10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*, 1–28. https://doi.org/10.18637/jss.v042.i08

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. https://doi.org/10.2307/2289064

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society A*, *171*, 481–502. https://doi.org/10.1111/j.1467-985X.2007.00527.x

Jones, N., Graziano, M., & Dimitrakopoulos, P. G. (2020). Social impacts of European protected areas and policy recommendations. *Environmental Science & Policy*, *112*, 134–140. https://doi.org/10.1016/j.envsci.2020.06.004

Jones, N., Malesios, C., Kantartzis, A., & Dimitrakopoulos, P. G. (2020). The role of location and social impacts of protected areas on subjective wellbeing. *Environmental Research Letters*, *15*, 114030. https://doi.org/10.1088/1748-9326/abb96e

Jones, N., McGinlay, J., & Dimitrakopoulos, P. G. (2017). Improving social impact assessment of protected areas: A review of the literature and directions for future research. *Environmental Impact Assessment Review*, *64*, 1–7. https://doi.org/10.1016/j.eiar.2016.12.007

Joppa, L. N., & Pfaff, A. (2011). Global protected area impacts. *Proceedings of the Royal Society B: Biological Sciences*, *278*, 1633–1638. https://doi.org/10.1098/rspb.2010.1713

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*, 435–454. https://doi.org/10.1017/pan.2019.11

Lechner, M. (2010). The estimation of causal effects by difference-in-difference methods estimation of spatial panels. *FNT in Econometrics*, *4*, 165–224. https://doi.org/10.1561/0800000014

Lee, W.-S. (2013). Propensity score matching and variations on the balancing test. *Empirical Economics*, *44*, 47–80. https://doi.org/10.1007/s00181-011-0481-0

Lewis, D. J., Hunt, G. L., & Plantinga, A. J. (2002). Public conservation land and employment growth in the Northern forest region. *Land Economics*, *78*, 245–259. https://doi.org/10.2307/3147271

Lindsey, P. A., Miller, J. R. B., Petracca, L. S., Coad, L., Dickman, A. J., Fitzgerald, K. H., Flyman, M. V., Funston, P. J., Henschel, P., Kasiki, S., Knights, K., Loveridge, A. J., Macdonald, D. W., Mandisodza-Chikerema, R. L., Nazerali, S., Plumptre, A. J., Stevens, R., Zyl, H. W. V., & Hunter, L. T. B. (2018). More than $1 billion needed annually to secure Africa's protected areas with lions. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E10788–E10796. https://doi.org/10.1073/pnas.1805048115

Loomis, J. B. (1999). Do additional designations of wilderness result in increases in recreation use? *Society & Natural Resources*, *12*, 481–491. https://doi.org/10.1080/089419299279551

Lopez, M. J., & Gutman, R. (2017). Estimation of causal effects with multiple treatments: A review and new ideas. *Statistical Science*, *32*, 432–454. https://doi.org/10.1214/17-STS612

Mammides, C. (2020). Evidence from eleven countries in four continents suggests that protected areas are not associated with higher poverty rates. *Biological Conservation*, *241*, 108353. https://doi.org/10.1016/j.biocon.2019.108353

Mayer, M., Müller, M., Woltering, M., Arnegger, J., & Job, H. (2010). The economic impact of tourism in six German national parks. *Landscape and Urban Planning*, *97*, 73–82. https://doi.org/10.1016/j.landurbplan.2010.04.013

McNally, C. G., Uchida, E., & Gold, A. J. (2011). The effect of a protected area on the tradeoffs between short-run and long-run benefits from mangrove ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 13945–13950. https://doi.org/10.1073/pnas.1101825108

Naidoo, R., Gerkey, D., Hole, D., Pfaff, A., Ellis, A. M., Golden, C. D., Herrera, D., Johnson, K., Mulligan, M., Ricketts, T. H., & Fisher, B. (2019). Evaluating the impacts of protected areas on human well-being across the developing world. *Science Advances*, *5*, eaav3006. https://doi.org/10.1126/sciadv.aav3006

Naughton-Treves, L., Holland, M. B., & Brandon, K. (2005). The role of protected areas in conserving biodiversity and sustaining local livelihoods. *Annual Review of Environment and Resources*, *30*, 219–252. https://doi.org/10.1146/annurev.energy.30.050504.164507

Nelson, A., & Chomitz, K. M. (2011). Effectiveness of strict vs. multiple use protected areas in reducing tropical forest fires: A global analysis using matching methods. *PLoS One*, *6*, e22722. https://doi.org/10.1371/journal.pone.0022722

Norton-Griffiths, M., & Southey, C. (1995). The opportunity costs of biodiversity conservation in Kenya. *Ecological Economics*, *12*, 125–139. https://doi.org/10.1016/0921-8009(94)00041-S

Oldekop, J. A., Holmes, G., Harris, W. E., & Evans, K. L. (2016). A global assessment of the social and conservation outcomes of protected areas. *Conservation Biology*, *30*, 133–141. https://doi.org/10.1111/cobi.12568

Palomo, I., Martín-López, B., Potschin, M., Haines-Young, R., & Montes, C. (2013). National parks, buffer zones and surrounding lands: Mapping ecosystem service flows. *Ecosystem Services, Special Issue on Mapping and Modelling Ecosystem Services*, *4*, 104–116. https://doi.org/10.1016/j.ecoser.2012.09.001

Rässler, S. (2002). *Statistical matching, lecture notes in statistics*. Springer. https://doi.org/10.1007/978-1-4613-0053-3

Reinius, S. W., & Fredman, P. (2007). Protected areas as attractions. *Annals of Tourism Research*, *34*, 839–854. https://doi.org/10.1016/j.annals.2007.03.011

Romano, B., Zullo, F., Fiorini, L., & Marucci, A. (2021). "The park effect"? An assessment test of the territorial impacts of Italian national parks, thirty years after the framework legislation. *Land Use Policy*, *100*, 104920. https://doi.org/10.1016/j.landusepol.2020.104920

Rosenbaum, P. R. (2010). *Design of observational studies*. Springer Series in Statistics. Springer-Verlag. https://doi.org/10.1007/978-1-4419-1213-8

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. https://doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33. https://doi.org/10.2307/2683903

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. the. *Annals of Statistics*, *6*, 34–58.

Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics*, *36*, 293. https://doi.org/10.2307/2529981

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, *31*, 161–170. https://doi.org/10.1111/j.1467-9469.2004.02-123.x

Runte, A. (1977). The national park idea: Origins and Paradox of the American experience. *Forest & Conservation History*, *21*, 64–75. https://doi.org/10.2307/3983461

Sakaguchi, S. (2020). Estimation of average treatment effects using panel data when treatment effect heterogeneity depends on unobserved fixed effects. *Journal of Applied Econometrics*, *35*, 315–327. https://doi.org/10.1002/jae.2752

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*, 101–122. https://doi.org/10.1016/j.jeconom.2020.06.003

Schleicher, J., Zaehringer, J. G., Fastré, C., Vira, B., Visconti, P., & Sandbrook, C. (2019). Protecting half of the planet could directly affect over one billion people. *Nature Sustainability*, *2*, 1094–1096. https://doi.org/10.1038/s41893-019-0423-y

Sekhon, J. (2020). *Matching: Multivariate and propensity score matching with balance optimization. R package version 4.9.7*. https://cran.r-project.org/web/packages/Matching/index.html

Sims, K. R. E. (2010). Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, *60*, 94–114. https://doi.org/10.1016/j.jeem.2010.05.003
Sims, K. R. E., Thompson, J. R., Meyer, S. R., Nolte, C., & Plisinski, J. S. (2019). Assessing the local economic impacts of land protection. *Conservation Biology*, *33*, 1035–1044. https://doi.org/10.1111/cobi.13318
Spanò, M., Leronni, V., Lafortezza, R., & Gentile, F. (2017). Are ecosystem service hotspots located in protected areas? Results from a study in Southern Italy. *Environmental Science & Policy*, *73*, 52–60. https://doi.org/10.1016/j.envsci.2017.04.008
Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services & Outcomes Research Methodology*, *14*, 166–182. https://doi.org/10.1007/s10742-014-0123-z
Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*, S84–S90. https://doi.org/10.1016/j.jclinepi.2013.01.013
UNEP-WCMC & IUCN. (2016). *Protected planet Report 2016*.
Visconti, P., Butchart, S. H. M., Brooks, T. M., Langhammer, P. F., Marnewick, D., Vergara, S., Yanosky, A., & Watson, J. E. M. (2019). Protected area targets post-2020. *Science*, *364*, 239–241. https://doi.org/10.1126/science.aav6886
Weiler, S., & Seidl, A. (2004). What's in a name? Extracting econometric drivers to assess the impact of national park designation. *Journal of Regional Science*, *44*, 245–262. https://doi.org/10.1111/j.0022-4146.2004.00336.x
Wilkie, D. S., Morelli, G. A., Demmer, J., Starkey, M., Telfer, P., & Steil, M. (2006). Parks and people: Assessing the human welfare effects of establishing protected areas for biodiversity conservation. *Conservation Biology*, *20*, 247–249. https://doi.org/10.1111/j.1523-1739.2005.00291.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: D'Alberto, R., Pagliacci, F., & Zavalloni, M. (2023). A socioeconomic impact assessment of three Italian national parks. *Journal of Regional Science*, 63, 114–147. https://doi.org/10.1111/jors.12618

APPENDIX A

See Tables A1–A12 and Figure A1.

TABLE A1 Linear probit propensity score matching model (M01)

Variable	Coefficients	Standard Error
Altitude above sea level	0.00144***	0.00026
Distance from the coast	0.01909*	0.00830
Percentage of forested land	0.44320*	0.26860
Nr. of local establishments	0.00123**	0.00044
Nr. of workers employed in local establishments	−0.00028*	0.00011
Nr. of agricultural holdings	0.00029***	0.00007
Population density	−0.00429**	0.00131
Per capita income	0.00011*	0.00005
Constant	−3.79500***	0.51710

Note: Observations: 7424. Pseudo R^2: 0.262.
Significance levels: *0.1; **0.05; ***0.01.

TABLE A2 Logistic propensity score matching with smoothing spline on per capita income model (M02)

Variable	Coefficients	Standard error
Altitude above sea level	0.00283***	0.00063
Distance from the coast	0.03751*	0.01795
Percentage of forested land	1.05500	0.65920
Nr. of local establishments	0.00304**	0.00093
Nr. of workers employed in local establishments	−0.00067**	0.00024
Nr. of agricultural holdings	0.00058***	0.00015
Population density	−0.00882**	0.00312
Per capita income	0.00020**	0.00011
Constant	−7.72900***	1.22000

Note: Observations: 7424. Pseudo R^2: 0.250.
Significance levels: *0.1; **0.05; ***0.01.

For both treated and untreated municipalities, as well as for the matched controls, Tables A3–A7 depict the main statistics on the covariates used for matching. Namely, the mean and standard deviation, the difference in means between the aforementioned groups, the standardized mean difference (SMD), the SMD percent balance improvement, the variance ratio, the mean and maximum of the empirical cumulative density function (eCDF). The SMD is the difference in the means of each covariate between the groups, standardized by the standard deviation of the covariate in the treated group. Standardization prevents the mean difference from being confounded by changes in the standard deviation of the covariate, hence achieving the same scale for all covariates. Since there is high correlation between the mean or maximum absolute SMD and the level of bias in the ATT, SMDs close to zero are assumed to indicate good balance between the groups (Austin, 2009; Stuart et al., 2013). The rule of thumb suggests that SMD values 0.2 hints at concerns about covariates imbalance (see, e.g., Stuart et al., 2014 and the references therein). The variance ratio represents the ratio of the variance of a covariate in one group to that in the other, such that variance ratios close to 1 indicate good balance because they imply similar variances of the samples (Austin, 2009; Ho et al., 2020). Finally, the eCDF considers the whole covariate distribution (rather than just the mean or variance), offering supplementary insights about the overall imbalance (note that the maximum eCDF is also known as the Kolmogorov–Smirnov statistics) (Diamond & Sekhon, 2013).

TABLE A3 Balance results for M01

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (n = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in std. mean diff.]	Variance ratio eCDF mean eCDF max	% balance improvement in variance ratio eCDF mean eCDF max
Score	−1.795 (0.541)	−4.176 (3.152)	2.381 (4.405)	0.029 0.432 0.750	−1.800 (0.514)	0.005 (0.009) [99.8]	1.106 0.000 0.040	97.2 100.0 94.7
Altitude above sea level	598.840 (242.771)	340.883 (288.479)	257.957 (1.063)	0.708 0.233 0.464	641.060 (212.742)	−42.220 (−0.174) [83.6]	1.302 0.055 0.201	23.5 76.3 56.9
Distance from coast	32.401 (12.413)	67.182 (55.853)	−34.781 (−2.802)	0.049 0.212 0.474	29.047 (20.075)	3.354 (0.270) [90.4]	0.382 0.059 0.262	68.0 71.9 45.2
Percentage of forested land (at 1990)	0.541 (0.295)	0.360 (0.328)	0.181 (0.612)	0.805 0.193 0.342	0.585 (0.279)	−0.044 (−0.150) [75.5]	1.109 0.071 0.183	52.5 63.3 47.4
Nr. of local establishments (at 1991)	521.780 (868.246)	421.017 (2202.923)	100.763 (0.116)	0.155 0.059 0.155	396.140 (720.066)	125.640 (0.145) [−24.7]	1.453 0.055 0.261	79.9 6.6 −68.1
Nr. of workers employed in local establishments (at 1991)	1520.620 (2690.658)	1785.599 (12,251.040)	−264.979 (−0.099)	0.048 0.027 0.153	1173.240 (2441.401)	347.380 (0.129) [−31.1]	1.215 0.072 0.303	93.6 −165.5 −96.5
Nr. of agricultural holdings (at 1990)	1192.140 (1634.191)	348.937 (531.762)	843.203 (0.516)	9.444 0.260 0.508	992.260 (1592.616)	199.880 (0.122) [76.3]	1.053 0.098 0.445	97.7 61.9 13.3

(Continues)

TABLE A3 (Continued)

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (n = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in std. mean diff.]	Variance ratio eCDF mean eCDF max	% balance improvement in variance ratio eCDF mean eCDF max
Population density (at 1991)	71.212 (61.904)	281.192 (641.953)	−209.980 (−3.392)	0.009 0.206 0.356	61.819 (68.701)	9.393 (0.152) [95.5]	0.812 0.057 0.185	95.5 72.5 49.5
Per capita income (at 2000)	10,204.688 (1733.821)	12,521.294 (2984.664)	−2316.606 (−1.336)	0.338 0.242 0.443	10,077.791 (2001.207)	126.897 (0.073) [94.5]	0.751 0.025 0.081	73.6 89.6 82.0

Abbreviations: eCDF, empirical cumulative density function; Std. Dev., standard deviation; Std. Mean Diff., standardized mean difference.

TABLE A4 Balance results for M02

Matching variable	Mean(Std. Dev.) Treated (n = 50)	Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (n = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	% balance improvement invariance ratio eCDF mean eCDF max
Score	0.062 (0.120)	0.006 (0.020)	0.056 (0.459)	35.344 0.446 0.751	0.060 (0.109)	0.002 (0.013) [97.2]	1.221 0.000 0.024	94.4 100.0 97.3
Altitude above sea level	598.840 (242.771)	340.883 (288.479)	257.957 (1.063)	0.708 0.233 0.464	606.420 (315.172)	−7.580 (−0.031) [97.1]	0.593 0.049 0.121	−51.3 78.9 74.29
Distance from coast	32.401 (12.413)	67.182 (55.853)	−34.781 (−2.802)	0.049 0.212 0.474	31.198 (35.675)	1.203 (0.097) [96.5]	0.121 0.111 0.381	29.8 47.6 19.9
Percentage of forested land (at 1990)	0.541 (0.295)	0.360 (0.328)	0.181 (0.612)	0.805 0.193 0.342	0.589 (0.331)	−0.048 (−0.157) [74.4]	0.794 0.064 0.281	−6.3 66.9 18.2
Nr. of local establishments (at 1991)	521.780 (868.246)	421.017 (2202.923)	100.763 (0.116)	0.155 0.059 0.155	506.260 (887.477)	15.520 (0.018) [84.6]	0.957 0.051 0.221	97.6 12.9 −42.3
Nr. of workers employed in local establishments (at 1991)	1520.620 (2690.658)	1785.599 (12,251.040)	−264.979 (−0.099)	0.048 0.027 0.153	1618.980 (3132.144)	−98.360 (−0.037) [62.9]	0.738 0.052 0.222	90.0 −92.1 −44.1
Nr. of agricultural holdings (at 1990)	1192.140 (1634.191)	348.937 (531.762)	843.203 (0.516)	9.444 0.260 0.508	1039.540 (1717.488)	152.600 (0.093) [81.9]	0.905 0.110 0.360	95.6 57.5 29.1

(Continues)

TABLE A4 (Continued)

Matching variable	Mean(Std. Dev.) Treated (*n* = 50)	Unmatched controls (*n* = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (*n* = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	% balance improvement invariance ratio eCDF mean eCDF max
Population density (at 1991)	71.212 (61.904)	281.192 (641.953)	−209.980 (−3.392)	0.009 0.206 0.356	70.011 (75.398)	1.201 (0.019) [99.4]	0.674 0.045 0.222	91.6 78.2 38.3
Per capita income (at 2000)	10,204.688 (1733.821)	12,521.294 (2984.664)	−2316.606 (−1.336)	0.338 0.242 0.443	10,541.397 (1791.816)	−336.709 (−0.194') [85.5]	0.936 0.036 0.148	93.9 85.2 68.4

Abbreviations: eCDF, empirical cumulative density function; Std. Dev., standard deviation; Std. Mean Diff., standardized mean difference.

TABLE A5 Balance results for M03

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (n = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Altitude above sea level	598.840 (242.771)	340.883 (288.479)	257.957 (1.063)	0.708 0.233 0.464	575.980 (255.415)	22.860 (0.094) [91.1]	0.903 0.026 0.141	70.6 88.9 69.8
Distance from coast	32.401 (12.413)	67.182 (55.853)	−34.781 (−2.802)	0.049 0.212 0.474	32.845 (17.695)	−0.444 (−0.036) [98.7]	0.492 0.026 0.140	76.4 88.0 70.5
Area of the municipality	110.297 (95.877)	36.205 (49.060)	74.092 (0.773)	3.819 0.359 0.593	100.663 (83.012)	9.634 (0.101) [87.0]	1.321 0.020 0.100	79.3 94.4 83.1
Percentage of forested land (at 1990)	0.541 (0.295)	0.360 (0.328)	0.181 (0.612)	0.805 0.193 0.342	0.429 (0.281)	0.112 (0.380) [37.9]	1.099 0.109 0.340	56.7 43.4 0.7
Nr. of local establishments (at 1991)	521.780 (868.246)	421.017 (2202.923)	100.763 (0.116)	0.155 0.059 0.155	492.600 (764.806)	29.180 (0.034) [71.0]	1.289 0.028 0.121	86.4 52.7 22.4
Nr. of workers employed in local establishments (at 1991)	1520.620 (2690.658)	1785.599 (12,251.040)	−264.979 (−0.099)	0.048 0.027 0.153	1408.600 (2306.733)	112.020 (0.042) [57.7]	1.361 0.025 0.123	89.8 9.0 21.4
Nr. of agricultural holdings (at 1990)	1192.140 (1634.191)	348.937 (531.762)	843.203 (0.516)	9.444 0.260 0.508	1150.400 (1571.067)	41.741 (0.026) [95.0]	1.082 0.025 0.145	96.5 90.4 72.4

(Continues)

TABLE A5 (Continued)

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched controls (n = 50)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Population density (at 1991)	71.212 (61.904)	281.192 (641.953)	−209.980 (−3.392)	0.009 0.206 0.356	84.568 (86.752)	−13.356 (−0.216) [93.6]	0.509 0.029 0.144	85.6 85.8 60.7
Per capita income (at 2000)	10,204.688 (1733.821)	12,521.294 (2984.664)	−2316.606 (−1.336)	0.338 0.242 0.443	10,316.216 (1965.837)	−111.528 (−0.064) [95.2]	0.778 0.030 0.162	76.9 87.7 63.9

Abbreviations: eCDF, empirical cumulative density function; Std. Dev., standard deviation; Std. Mean Diff., standardized mean difference.

TABLE A6 Balance results for M04

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched treated (n = 48)	Mean (Std. Dev.) Matched controls (n = 48)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Score	0.056 (0.120)	0.006 (0.019)	0.050 (0.411)	35.344 0.446 0.751	0.037 (0.033)	0.036 (0.033)	0.001 (0.002) [99.6]	0.988 0.007 0.063	99.7 98.4 91.8
Altitude above sea level	598.840 (242.771)	340.883 (288.479)	257.957 (1.063)	0.708 0.233 0.464	610.917 (238.033)	605.042 (263.911)	5.875 (0.024) [97.7]	0.814 0.025 0.125	40.2 89.1 73.1
Distance from coast	32.401 (12.413)	67.182 (55.853)	−34.781 (−2.802)	0.049 0.212 0.474	32.719 (12.403)	29.595 (16.888)	3.124 (0.252) [91.0]	0.539 0.042 0.229	79.5 80.1 51.7
Area of the municipality	110.297 (95.877)	36.205 (49.060)	74.092 (0.773)	3.819 0.359 0.593	97.512 (73.497)	91.342 (65.431)	6.17 (0.064) [91.7]	1.262 0.025 0.104	82.7 93.2 82.4
Percentage of forested land (at 1990)	0.541 (0.295)	0.360 (0.328)	0.181 (0.612)	0.805 0.193 0.342	0.560 (0.284)	0.563 (0.308)	−0.003 (−0.009) [98.5]	0.850 0.067 0.188	25.1 65.2 45.2
Nr. of local establishments (at 1991)	521.780 (868.246)	421.017 (2202.923)	100.763 (0.116)	0.155 0.059 0.155	394.708 (594.302)	336.000 (650.398)	58.708 (0.068) [41.7]	0.835 0.045 0.188	90.3 22.8 −21.3
Nr. of workers employed in local establishments (at 1991)	1520.620 (2690.658)	1785.599 (12,251.040)	−264.979 (−0.099)	0.048 0.027 0.153	1122.604 (1855.688)	893.750 (1939.832)	228.854 (0.085) [13.6]	0.915 0.054 0.167	97.1 −99.5 −9.2

(Continues)

TABLE A6 (Continued)

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched treated (n = 48)	Mean (Std. Dev.) Matched controls (n = 48)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Nr. of agricultural holdings (at 1990)	1192.140 (1634.191)	348.937 (531.762)	843.203 (0.516)	9.444 0.260 0.508	999.458 (1306.054)	775.313 (964.473)	224.145 (0.137) [73.4]	1.834 0.048 0.167	73.0 81.6 67.2
Population density (at 1991)	71.212 (61.904)	281.192 (641.953)	−209.980 (−3.392)	0.009 0.206 0.356	66.727 (58.281)	53.479 (46.969)	13.248 (0.214) [93.7]	1.540 0.052 0.188	90.8 74.7 47.4
Per capita income (at 2000)	10204.688 (1733.821)	12521.294 (2984.664)	−2316.606 (−1.336)	0.338 0.242 0.443	10181.434 (1761.244)	10110.105 (1806.288)	71.329 (0.041) [96.9]	0.951 0.016 0.125	95.4 93.4 71.8

Abbreviations: eCDF, empirical cumulative density function; Std. Dev., standard deviation; Std. Mean Diff., standardized mean difference.

TABLE A7 Balance results for M05

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched treated (n = 48)	Mean (Std. Dev.) Matched controls (n = 48)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Score	0.056 (0.120)	0.006 (0.019)	0.050 (0.411)	38.549 0.436 0.758	0.037 (0.033)	0.036 (0.034)	0.001 (0.003) [99.2]	0.972 0.009 0.063	99.2 97.9 91.8
Altitude above sea level	598.840 (242.771)	340.883 (288.479)	257.957 (1.063)	0.708 0.233 0.464	610.917 (238.03)	622.813 (256.648)	−11.896 (−0.049) [95.4]	0.860 0.027 0.104	56.3 88.3 77.6
Distance from coast	32.401 (12.413)	67.182 (55.853)	−34.781 (−2.802)	0.049 0.212 0.474	32.719 (12.403)	30.255 (17.812)	2.464 (0.199) [92.9]	0.485 0.044 0.208	75.9 79.2 56.19
Area of the municipality	110.297 (95.877)	36.205 (49.060)	74.092 (0.773)	3.819 0.359 0.593	97.512 (73.497)	89.636 (61.052)	7.876 (0.082) [89.4]	1.449 0.021 0.104	72.3 94.2 82.4
Percentage of forested land (at 1990)	0.541 (0.295)	0.360 (0.328)	0.181 (0.612)	0.805 0.193 0.342	0.560 (0.284)	0.567 (0.297)	−0.007 (−0.022) [96.4]	0.917 0.073 0.188	60.1 62.2 45.2
Nr. of local establishments (at 1991)	521.780 (868.246)	421.017 (2202.923)	100.763 (0.116)	0.155 0.059 0.155	394.708 (594.302)	319.229 (614.002)	75.479 (0.087) [25.1]	0.937 0.047 0.188	96.5 19.3 −21.3
Nr. of workers employed in local establishments (at 1991)	1520.620 (2690.658)	1785.599 (12251.040)	−264.979 (−0.099)	0.048 0.027 0.153	1122.604 (1855.688)	867.813 (1913.007)	254.791 (0.095) [3.8]	0.941 0.057 0.146	98.0 −111.9 4.5

(Continues)

TABLE A7 (Continued)

Matching variable	Mean (Std. Dev.) Treated (n = 50)	Mean (Std. Dev.) Unmatched controls (n = 7424)	Difference in means (Std. Mean Diff.)	Variance ratio eCDF mean eCDF max	Mean (Std. Dev.) Matched treated (n = 48)	Mean (Std. Dev.) Matched controls (n = 48)	Difference in means (Std. Mean Diff.) [% balance improvement in Std. Mean Diff.]	Variance ratio eCDF mean eCDF max	%balance improvement in variance ratio eCDF mean eCDF max
Nr. of agricultural holdings (at 1990)	1192.140 (1634.191)	348.937 (531.762)	843.203 (0.516)	9.444 0.260 0.508	999.458 (1306.054)	866.271 (1131.701)	133.187 (0.082) [84.2]	1.332 0.040 0.188	87.2 84.5 63.1
Population density (at 1991)	71.212 (61.904)	281.192 (641.953)	−209.980 (−3.392)	0.009 0.206 0.356	66.727 (58.281)	56.923 (50.763)	9.804 (0.158) [95.3]	1.318 0.045 0.167	94.1 78.4 53.2
Per capita income (at 2000)	10204.688 (1733.821)	12521.294 (2984.664)	−2316.606 (−1.336)	0.338 0.242 0.443	10181.434 (1761.244)	10028.016 (1817.565)	153.418 (0.089) [93.4]	0.939 0.019 0.125	94.2 92.0 71.8

Abbreviations: eCDF, empirical cumulative density function; Std. Dev., standard deviation; Std. Mean Diff., standardized mean difference.

TABLE A8 Descriptive statistics for the outcome variables of the matched municipalities, model M01

Variable	Mean (standard deviation)			
	Baseline year 2001		Follow-up year 2011	
	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)
Per capita income (Euro)	10579.676 (1635.155)	10416.565 (1823.504)	18810.210 (1729.546)	18410.799 (1979.686)
Population level (nr. of inhabitants)	10489.601 (18753.738)	7923.180 (15091.870)	10717.901 (19676.153)	7922.200 (15545.016)
Nr. of local establishments	594.301 (1149.230)	423.460 (845.051)	736.040 (1416.305)	560.940 (1163.549)
Nr. of workers employed in local establishments	1809.121 (3622.802)	1235.040 (2640.166)	1991.505 (3960.982)	1492.660 (3194.124)
Nr. of tourism sector establishments	30.500 (38.357)	26.960 (45.645)	41.960 (60.822)	41.100 (73.031)
Nr. of workers employed in tourism sector establishments	74.320 (106.426)	67.980 (137.106)	126.000 (206.854)	133.200 (283.994)
Nr. of agricultural holdings	1062.800 (1596.966)	876.480 (485.447)	784.802 (1365.831)	582.660 (931.943)
Utilized agricultural area (UAA, ha)	5443.178 (7076.524)	4349.838 (5007.453)	5518.751 (7905.362)	4722.367 (5758.520)

Abbreviation: NP, national park.

TABLE A9 Descriptive statistics for the outcome variables of the matched municipalities, model M02

Variable	Mean (standard deviation)			
	Baseline year 2001		Follow-up year 2011	
	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)
Per capita income (Euro)	10579.676 (1635.155)	10968.505 (1644.454)	18810.210 (1729.546)	18738.869 (1937.169)
Population level (nr. of inhabitants)	10489.601 (18,753.738)	9251.300 (16,145.389)	10717.901 (19,676.153)	9522.200 (16,951.384)
Nr. of local establishments	594.301 (1149.230)	535.040 (976.815)	736.040 (1416.305)	709.320 (1383.496)
Nr. of workers employed in local establishments	1809.121 (3622.802)	1590.800 (3171.722)	1991.505 (3960.982)	1957.580 (3965.073)
Nr. of tourism sector establishments	30.500 (38.357)	40.240 (74.567)	41.960 (60.822)	58.600 (118.436)
Nr. of workers employed in tourism sector establishments	74.320 (106.426)	125.880 (254.586)	126.000 (206.854)	188.320 (384.256)
Nr. of agricultural holdings	1062.800 (1596.966)	820.740 (1383.650)	784.802 (1365.831)	520.780 (821.222)
Utilized agricultural area (UAA, ha)	5443.178 (7076.524)	4029.334 (4898.761)	5518.751 (7905.362)	4110.402 (5050.794)

Abbreviation: NP, national park.

TABLE A10 Descriptive statistics for the outcome variables of the matched municipalities, model M03

Variable	Mean (standard deviation)			
	Baseline year 2001		Follow-up year 2011	
	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)	Municipalities in NPs (n = 50)	Municipalities not in NPs (n = 50)
Per capita income (Euro)	10579.676 (1635.155)	10557.354 (1917.666)	18810.210 (1729.546)	18526.697 (1805.892)
Population level (nr. of inhabitants)	10489.601 (18753.738)	9898.200 (17457.190)	10717.901 (19676.153)	9959.500 (17675.770)
Nr. of local establishments	594.301 (1149.230)	539.980 (899.015)	736.040 (1416.305)	685.580 (1180.610)
Nr. of workers employed in local establishments	1809.121 (3622.802)	1528.200 (2629.844)	1991.505 (3960.982)	1851.580 (3282.678)
Nr. of tourism sector establishments	30.500 (38.357)	34.220 (49.169)	41.960 (60.822)	48.060 (75.050)
Nr. of workers employed in tourism sector establishments	74.320 (106.426)	89.780 (153.307)	126.000 (206.854)	139.420 (239.881)
Nr. of agricultural holdings	1062.800 (1596.966)	1048.240 (1571.557)	784.802 (1365.831)	745.460 (1200.140)
Utilized agricultural area (UAA, ha)	5443.178 (7076.524)	4565.314 (5476.083)	5518.751 (7905.362)	4327.682 (5660.999)

Abbreviation: NP, national park.

TABLE A11 Descriptive statistics for the outcome variables of the matched municipalities, model M04

Variable	Mean (standard deviation)			
	Baseline year 2001		Follow-up year 2011	
	Municipalities in NPs (n = 48)	Municipalities not in NPs (n = 48)	Municipalities in NPs (n = 48)	Municipalities not in NPs (n = 48)
Per capita income (Euro)	10560.780 (1659.757)	10395.289 (1739.034)	18846.213 (1752.259)	18439.352 (1672.388)
Population level (nr. of inhabitants)	7597.083 (11,937.571)	5695.208 (10,985.943)	7631.542 (12,268.141)	5561.255 (11,135.477)
Nr. of local establishments	409.917 (658.645)	326.396 (652.446)	503.625 (810.687)	397.271 (857.839)
Nr. of workers employed in local establishments	1216.500 (2171.629)	867.938 (1879.557)	1326.958 (2246.702)	1039.354 (2362.447)
Nr. of tourism sector establishments	25.083 (26.087)	21.813 (34.646)	33.000 (41.960)	30.604 (62.987)
Nr. of workers employed in tourism sector establishments	57.750 (62.962)	51.333 (95.554)	92.250 (123.453)	86.542 (204.393)
Nr. of agricultural holdings	878.104 (1229.690)	693.083 (1060.906)	614.396 (1018.870)	464.604 (743.843)
Utilized agricultural area (UAA, ha)	4595.511 (5740.511)	3767.813 (3821.539)	4368.877 (5534.311)	3792.148 (4192.274)

Abbreviation: NP, national park.

TABLE A12 Descriptive statistics for the outcome variables of the matched municipalities, model M05

Variable	Mean (standard deviation)			
	Baseline year 2001		Follow-up year 2011	
	Municipalities in NPs (n = 48)	Municipalities not in NPs (n = 48)	Municipalities in NPs (n = 48)	Municipalities not in NPs (n = 48)
Per capita income (Euro)	10560.780 (1659.757)	10331.379 (1792.363)	18846.213 (1752.259)	18365.921 (1729.568)
Population level (nr. of inhabitants)	7597.083 (11,937.571)	6137.438 (11,381.999)	7631.542 (12,268.141)	6001.979 (11,493.289)
Nr. of local establishments	409.917 (658.645)	321.229 (631.726)	503.625 (810.687)	395.604 (840.545)
Nr. of workers employed in local establishments	1216.500 (2171.629)	825.792 (1822.152)	1326.958 (2246.702)	989.396 (2317.217)
Nr. of tourism sector establishments	25.083 (26.087)	23.250 (36.565)	33.000 (41.960)	32.979 (65.113)
Nr. of workers employed in tourism sector establishments	57.750 (62.962)	53.021 (98.192)	92.250 (123.453)	89.333 (205.903)
Nr. of agricultural holdings	878.104 (1229.690)	752.625 (1165.687)	614.396 (1018.870)	502.875 (819.388)
Utilized agricultural area (UAA, ha)	4595.511 (5740.511)	3566.372 (3344.708)	4368.877 (5534.311)	3592.054 (3716.001)

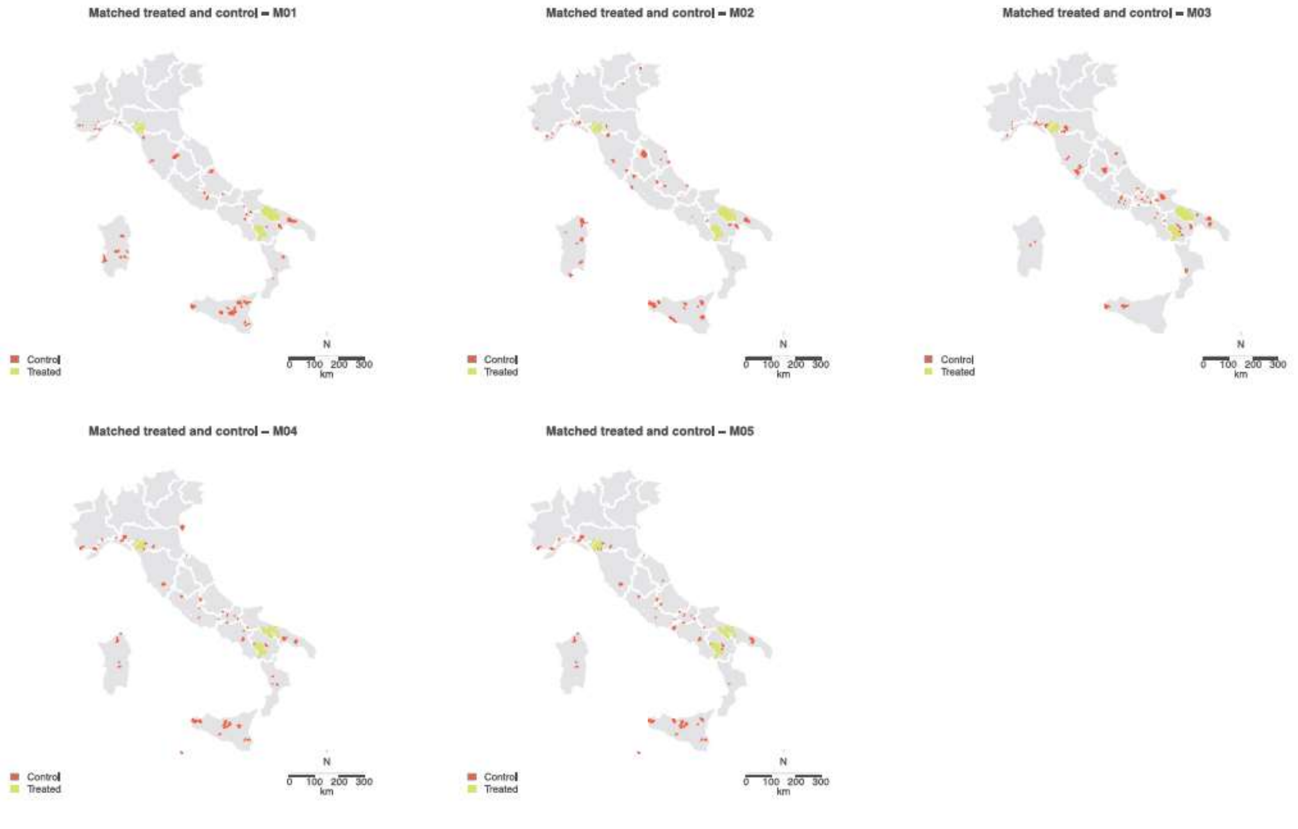

FIGURE A1 Geographical distribution of the matched treated and control municipalities for the five matching models adopted