scientific reports

OPEN

The impact of the European Union emissions trading system on carbon dioxide emissions: a matrix completion analysis

Francesco Biancalani[1], Giorgio Gnecco[1]✉, Rodolfo Metulini[2] & Massimo Riccaboni[1,3]

Despite the negative externalities on the environment and human health, today's economies still produce excessive carbon dioxide emissions. As a result, governments are trying to shift production and consumption to more sustainable models that reduce the environmental impact of carbon dioxide emissions. The European Union, in particular, has implemented an innovative policy to reduce carbon dioxide emissions by creating a market for emission rights, the emissions trading system. The objective of this paper is to perform a counterfactual analysis to measure the impact of the emissions trading system on the reduction of carbon dioxide emissions. For this purpose, a recently-developed statistical machine learning method called matrix completion with fixed effects estimation is used and compared to traditional econometric techniques. We apply matrix completion with fixed effects estimation to the prediction of missing counterfactual entries of a carbon dioxide emissions matrix whose elements (indexed row-wise by country and column-wise by year) represent emissions without the emissions trading system for country-year pairs. The results obtained, confirmed by robust diagnostic tests, show a significant effect of the emissions trading system on the reduction of carbon dioxide emissions: the majority of European Union countries included in our analysis reduced their total carbon dioxide emissions (associated with selected industries) by about 15.4% during the emissions trading system treatment period 2005–2020, compared to the total carbon dioxide emissions (associated with the same industries) that would have been achieved in the absence of the emissions trading system policy. Finally, several managerial/practical implications of the study are discussed, together with its possible extensions.

Keywords Matrix completion with fixed effects estimation, Counterfactual analysis, Policy impact analysis, Green economy, Pollution

Global warming is mainly a consequence of human activities and the use of fuels in an economic system. Currently, there is an extensive literature that examines various aspects related to global warming associated with CO_2 emissions[1–4]. Limiting yearly carbon dioxide (CO_2) emissions can be an effective way to reduce the effects of global warming (e.g., by slowing the rise in temperature). In fact, increases in such production lead, for example, to sea level rise (and thus a reduction in available dry land). Global warming is a hot research topic because, among other things, it causes natural disasters such as hurricanes and floods. They all can cause persistent damage to agriculture and more generally to the whole economic system. For a debate on these issues, the reader is referred to[5–10].

In recent years, some possible measures to mitigate climate change were proposed by governments, international organizations, and associations. But not all countries took significant action. One notable example of a policy to reduce CO_2 emissions is the Emissions Trading System (ETS), which was introduced by the European Union (EU) in 2005 and came into effect in various stages. The ETS sets an annual cap on CO_2 emissions for companies in certain industries. The basic idea behind this policy is that CO_2 emissions are the main cause of current global warming and that reducing CO_2 emissions can lead to stopping or slowing climate change. A significant portion of new CO_2 emissions is caused by human impacts on the environment during manufacturing, transportation, and energy production (from fossil sources) that use large quantities of hydrocarbons. Since the amount of (EU) allowances is set by the authorities and a fine per ton is imposed if emissions are exceeded, the

[1]AXES – IMT School for Advanced Studies, Lucca, Italy. [2]Department of Economics, University of Bergamo, Bergamo, Italy. [3]Istituto Universitario di Studi Superiori, Pavia, Italy. ✉email: giorgio.gnecco@imtlucca.it

EU can effectively curb CO_2 emissions. Up to 2020, this policy has come into force in different phases (the first was in 2005, the second in 2008, and the third in 2013). The EU ETS policy is consistent with the 2016 Paris Agreement and the EU Nationally Determined Contributions (NDC). The latter calls for a 55% reduction in CO_2 emissions by 2030, using 1990 levels as the basis for calculation[11]. On the other hand, some non-EU countries implemented their CO_2 reduction policies later and in a softer way.

Although some effects of the EU ETS policy were already studied (see the literature review in the "Empirical review" section), an examination of the literature reveals the following gaps: (i) few studies analyzed the impact of the EU ETS policy at the European level, while the rest of the analyses focused on specific EU countries; (ii) the results of the analyses conducted in different papers were often contradictory; (iii) few studies used a rigorous counterfactual analysis; (iv) typically, only the first phase of the EU ETS policy was analyzed (by considering quite a short period of time), not its long-term impact.

In this work, we have addressed the above gaps by using a state-of-the-art machine learning method (namely, Matrix Completion with Fixed Effects estimation, or MCFE, hereafter), which was recently shown by[12] to be a more effective method for evaluating policies than other, more traditional econometric methods used for panel data analysis. From the policy-making perspective, our study has several managerial/practical implications (also discussed in the "Conclusions, policy implications, and possible future research developments" section). First, the study highlights the need to use more reliable and general methods to estimate the impact of policies, such as EU ETS, on pollution reduction. This issue is particularly important given the current prominence of climate change and, more generally, environmental issues and the lessons we can draw from the EU experience for policymakers in developing countries. Second, a similar analysis as ours could be used to suggest the adoption of policies similar to the EU ETS one to other countries. Third, the methodology used in this study could be used as one of the tools to assess the economic, social, and environmental impact of policies, as established by recent EU legislation. In this context, a possible motivation for the application of such methodology would be a sequential (rather than simultaneous) data availability.

The work is organized as follows: the "Literature review" section provides a literature review; the "Idea of the work and its original contributions" section illustrates the idea of the work and its original contributions; the "Description of the data set" section describes the available data set, used for our study; the "Methodology" section details the methodology adopted; the "Results" section shows the results obtained by applying this methodology to a suitable pre-processed CO_2 emission matrix; the "Conclusions, policy implications, and possible future research developments" section concludes the work and sheds light on possible future developments. Some technical appendices provide additional analyses, and comparisons with other methods.

Literature review

In the following, we provide a literature review, which is divided into two parts: (1) a methodological review, mainly focused on the description of some methodologies commonly adopted in the literature for policy evaluation; (2) an empirical review, mainly focused on the state of the art on the evaluation of the EU ETS policy. These two parts are connected to the present work in the following way: our study deals with policy evaluation, in which the set of treated units is a subset of EU countries, and the set of untreated (control) units is a subset of non-EU countries; the study applies a recently developed methodology (Matrix Completion with Fixed Effects estimation, or MCFE) for policy evaluation, which overcomes most of the limitations of previous methods presented in the methodological review; MCFE is used in our study to investigate the effects of the EU ETS policy on total yearly CO_2 emissions of (treated) industries in EU countries.

Methodological review

When doing policy evaluation, it is often the case that the ideal situation of a randomized sample in which treated subjects have ex-ante the same characteristics as untreated subjects cannot be achieved. Therefore, one should use techniques that can provide good counterfactual data for the elements of the treated group.

In the case in which randomized controls are not available (which occurs in most policy impact analyses), various techniques, such as Instrumental Variables (IVs), are often used in the literature to evaluate interventions. A relevant example of the use of IVs in the context of ecological economics comes from[13], whose authors studied the effects of voluntary adoption of green programs on mayoral elections, using as IV the existence of a Covenant Territorial Coordinator. They found that participation in non-mandatory green programs at the local level was not a barrier to re-election. Another typical IV application was considered by[14], where the authors examined the impact of environmental Non-Governmental Organisations (NGOs) on air quality. In this case, IVs were represented by the number of international NGOs per capita and membership density of international NGOs.

Another technique is Regression Discontinuity Design (RDD), which can be applied when there is at least one specific threshold that separates treated and untreated units. For example, water conservation in California was examined by[15] using three discontinuity points in the timeline: June 2015, February 2016, and November 2016 (i.e., the dates of key legislative events). Similarly, a spatial regression discontinuity design was used by[16] to examine whether the Forest Stewardship Council (FSC) has changed (or not) the standard of living of indigenous people in Congo.

The use of Difference-in-Differences (DiD) (*Remark 1*: A brief introduction to the DiD method is provided in the "Supplementary Material".) may provide a good alternative approach[17], but its application requires the so-called parallel trend assumption, which is often difficult to be met. Adjustments to the control group such as Propensity Score Matching (or PSM)[18] and Mahalanobis distance matching and entropy (or Hainmueller) balancing[19] are often not conclusive in case of a large heterogeneity and small number of units available. These problems (especially the comparison between a small number of states or regions) can be partially solved with the Synthetic Control Method (SCM) (*Remark 2*: A non-technical introduction to the SCM is provided in the

"Supplementary Material".). This method was applied by[20] to analyze the impact of oil production in Basilicata (a small region in southern Italy) on socioeconomic indicators.

Empirical review

The topic of the implementation of markets for emission rights, and in particular the implementation of the one designed by the EU (ETS), was already studied in the literature from various points of view[21,22]. One of them is the empirical analysis of the effective reduction of total CO_2 emissions in the countries of the old continent. However, there is no consensus in the current literature on the impact of the EU ETS policy[23]. A relevant literature review[24] reports, among other things, on the following effects of policies such as the EU ETS: they tend to increase prices of emission-intensive goods[25], thus generating incentives to reduce such emissions (in the case of the EU ETS and other policies, such a reduction was highlighted by several studies (i.e.,[21,26–36]) reviewed in Chapter 13 of[24]; they can stimulate technological change by participants and even other actors not involved in such policies[30,37,38]; they may impact other countries in various ways, including changes in their emissions (leakage effects, see[39]), although no (or not significant) evidence of such changes was observed for the case of the EU ETS[23,27,30,40–45]. Still, the study of the impact of ETS policies is complicated by the possible simultaneous presence of other similar policies and exogenous factors (e.g., fossil fuel price changes and fluctuating economic conditions)[30,31,46], and by the fact that they apply to emissions associated with a subset of industries, possibly with different shares of emissions in different countries[47].

In more detail, focusing on the specific techniques of analysis, one very relevant paper on the environmental impact of the ETS is certainly[48], which combined a nearest-neighbor matching approach, between treated large plants and untreated small plants, with DiD and Difference-in-Means (DiM) estimators. However, in that work, whose analysis was focused on the United Kingdom from 2000 to 2012, no significant effect of the ETS on CO_2 emissions reductions was found for that country. The impact of the EU ETS on Lithuanian companies was studied by[49], analyzing data from 2003 to 2010 (*Remark 3*: Lithuania joined the EU in 2004, so 2004 was considered the pre-treatment period for all its observed firms.). In their analysis, the authors combined nearest neighbor matching with DiD and then applied kernel matching as a robustness check. They concluded that the ETS did not significantly reduce CO_2 emissions in Lithuania (in some treated years, only minor effects were achieved as the old plants of the large polluters were released). A similar methodological approach was used by[50] for the Norwegian case (*Remark 4*: Although Norway is not an EU member, it adopted the EU ETS policy in 2008.). In that work, the authors used a fixed-effects DiD approach and selected a control group through nearest-neighbor matching, specifically assuming exact industry-level matching between treated and untreated firms. However, the results obtained related to emissions were not statistically significant.

Another stream of literature showed that the EU ETS had a positive impact on reducing CO_2 emissions of selected European countries. For example, a relevant reduction in CO_2 production in Germany was found by[51], motivated by an increase in the energy efficiency of plants. Their econometric methodology used PSM to weigh treated and non-treated firms. Similarly, a significant reduction in CO_2 production in France was observed by[52]. In our opinion, these approaches may hide a problem in obtaining a fair evaluation of the policy, since the treated plants were quite large, while the ones in the control group were small. As a result, there may be economies of scale in CO_2 emissions that were not captured by the models. In other words, if the control group has different characteristics (i.e., in particular, a different order of magnitude in size) than the treated group, approaches based on (classical) matching cannot produce an adequate control group because the control group obtained cannot be entirely similar to the treated group.

A very recent article[53] found a reduction of about 10% in CO_2 emissions between 2005 and 2012 in four countries studied (i.e., France, the Netherlands, Norway, and the United Kingdom). However, in their one-to-one matching approach, it was necessary to exclude a number of companies for which it was not possible to find a good match (e.g., large electricity production companies). This might have biased the results of their analysis due to the possible exclusion of some of the most important examples of potential CO_2 emissions reductions.

Finally, a significant methodological improvement in studying the performance of the EU ETS in reducing pollution was made by[31], where researchers applied the SCM at the industry level and concluded that the presence of the EU ETS policy significantly reduced CO_2 production in the EU by 3.8% between 2008 and 2016, compared to its absence. However, SCM may fail under various circumstances, especially if the period of pre-treatment observations is not long enough[54]. The same method was also used by[55], where the scholars concentrated their analysis on estimating the effects on emissions for Australia, if Australia had adopted the EU ETS scheme. They found a statistically significant reduction in the CO_2 emissions per capita.

Idea of the work and its original contributions

Considering the limitations of the methods presented in the "Methodological review" section and reviewed in their application to the EU ETS policy analysis in the "Empirical review" section, in our analysis we have chosen to use another recently developed method coming from the Statistical Machine Learning literature (SML), namely Matrix Completion (MC), in order to verify whether the results found by[31] are confirmed or not with this novel approach. Moreover, our use of MC allows us to fill the four gaps in the literature highlighted at the end of the "Introduction" section, namely: (i) the opportunity of focusing the analysis on a larger set of countries; (ii) the necessity of using reliable estimation methods; (iii) the requirement of performing a rigorous counterfactual analysis; (iv) the need of analyzing a period covering various phases of the EU ETS policy.

The following is a non-technical introduction to MC. In essence, MC deals with the challenge of reconstructing a data matrix with missing entries (see Chapter 4 in[56]). A typical example is given by predicting missing elements in a recommender system's rating matrix, which reflects users' preferences for various items (such as movies). Imagine a partially filled matrix in which each entry represents a user's rating for a movie (if the user has seen that movie). MC aims to predict the missing entries (users' ratings for movies that they have not watched, but that have been seen by other users). In this application, the final goal could be, e.g., to suggest unseen movies that users might enjoy. To achieve this, MC leverages user and item similarities. The idea is that users who give similar ratings to movies likely share similar tastes for new ones. The resulting underlying pattern of the rating matrix, known as its low-rank structure, allows MC to make predictions for unseen entries in that matrix. For this reason, MC typically includes a regularization term in its mathematical formulation, whose goal is to enforce a low-rank structure to the reconstructed matrix. In its application to counterfactual analysis, detailed by[12] and also considered in the present work, the goal of MC is, instead, to predict counterfactual values of an outcome variable for treated units during the treatment period, starting from observations of the outcome variable for the treated units in the pre-treatment period, and the observations of that variable during the whole (pre-treatment and treatment) period for the control units.

In more detail, from an optimization point of view, the main idea of (low-rank) MC is to minimize a proper tradeoff between a suitably-defined approximation error on a set of observed entries of a matrix (training set) and a proxy for the rank of the reconstructed matrix, e.g., its nuclear norm (i.e., the summation of its singular values). More technical details are reported in the "Methodology" section. MC is a state-of-the-art quantitative method particularly suited for counterfactual analyses, as recently demonstrated by[12], where it was successfully compared with other commonly-adopted econometric methods such as DiD and SCM. Other effective applications of MC were made by[57], in which MC was exploited in the context of international trade for the reconstruction of World Input–Output Database (WIOD) subtables[58], by[59,60], in which MC was used for the analysis of economic complexity, and by[61,62], where MC was exploited for job analysis.

Non-technical overview of the original contributions of the work

According to the framework detailed above, our main research question is to investigate, by means of the application of a specific MC method (suitable for the prediction of counterfactuals), whether EU countries—through the EU ETS policy—reduced CO_2 production significantly during a particular treatment period (2005–2020), which is equivalent to assessing the effectiveness of the EU ETS policy in reducing CO_2 emissions in that period. Our goal is not limited to assessing whether or not reductions occurred but also includes quantifying (through a robust SML approach) the reduction in CO_2 emissions due to the presence of the EU ETS policy. Specifically, we perform a counterfactual analysis based on MC to estimate the (unobserved) CO_2 emissions of EU countries in the years of treatment in the absence of the EU ETS policy.

In this work, we aim to contribute to the academic debate by examining the impact of the EU ETS policy on reducing CO_2 emissions. This work can be viewed as a development of the research made by[63] and of our earlier conference article[64] on the application of MC to the prediction of CO_2 emissions, each based on two different data sets (*Remark 5*: The analysis made by[64] was further extended recently by[65], showing that the predictive accuracy of MC, applied to a matrix of CO_2 emissions, can be improved by combining it with a baseline estimate (e.g., an estimate of fixed effects). In that work, an ensemble machine-learning approach was followed, in which first the baseline estimate was generated, then MC was applied to the residual. The MC approach by[12], used in the present work, is based on a similar but more sophisticated idea, in which fixed-effects estimation and MC are performed simultaneously.). In contrast to these papers (in which only the predictive accuracy of MC was evaluated), here we perform a counterfactual analysis, based on MC. Moreover, this analysis is based on a different choice of the matrix to which MC is applied, as well as an appropriate choice of matrix elements provided as inputs to MC. We also use a different MC method that is more appropriate for estimating causal effects, perform robustness checks for the application of that MC method, and compare it with several other methods for counterfactual prediction (*Remark 6*: To improve the readability of the work, details on this comparison are reported in the "Supplementary Material".). To our knowledge, no other previous work analyzed the effects of the EU ETS policy using MC. Finally, in our analysis, we apply MC to a data set related to total CO_2 emissions coming from various EU and non-EU countries during a period of several years in the past.

With the aforementioned goal in mind, in the present paper we propose to perform a counterfactual analysis for the EU ETS policy by referring to the approach used by[12] (MC with Fixed Effects estimation, or MCFE, hereafter), based on a nuclear norm MC optimization problem, which is an extension of the optimization problem introduced by[66] (baseline MC or MCB, hereafter, i.e., MC without fixed effects estimation) and solved numerically by applying the soft-impute algorithm developed in the latter work. The MCFE method is specifically designed for panel data analysis (where the rows and columns of the matrix may refer to individuals and time points, for example). It introduces a two-way (individual and time) fixed effects component to the MCB optimization problem considered by[66], with the aim to increase the performance of matrix completion (or matrix reconstruction) (*Remark 7*: The performance of MCFE compared to MCB was recently evaluated in our conference paper[67] using a simulation study for CO_2 emission data. Therein, we found that the inclusion of individual and time fixed effects in the MC optimization problem, as well as an appropriate pre-processing of the original data achieved by applying a suitable l_1 row-normalization, increases the predictive performance of the MC method. In particular, the latter normalization filters out the possible side effect of differences in CO_2 emission levels between countries. Therefore, also in the present work, we apply l_1 normalization by row (i.e., by country) as an appropriate pre-processing of the matrix.). One important reason for its use in the present work is that the MCFE method was recently found to achieve better performance than alternative traditional methods used for causal panel data analysis (such as DiD and SCM), in several problems for which a ground truth is available [12] (*Remark 8*: As a side note, just to highlight the significance of MCFE as a state-of-the-art methodology for policy evaluation, it is worth remarking here that the work by[12] in which MCFE was developed reached a huge number of citations since its recent publication (more than 630 citations at the time of writing the present study), and that it was co-authored by a recent recipient (in 2021) of the Nobel prize in Economics, namely

Prof. Guido Imbens.). More technical motivations for which MCFE is preferable, for the study of the impact of the EU ETS policy on CO_2 emissions, to MCB, DiD, and SCM, are reported, respectively, in Supplementary Sections A, B, and C of the Supplementary Material. It is also worth observing that the MCFE method does not suffer from multicollinearity issues, since it works directly on a matrix of (partially) observed realizations of the outcome variable (*Remark 9*: Actually, the presence of multicollinearity could even support adopting a low-rank approximation of the matrix to be reconstructed, which is enforced by the presence of the nuclear-norm regularization term in the objective function of the MC optimization problem (1), which is reported later in the "Methodology" section.). In other words, its goal is not to estimate the coefficients associated with covariates in a linear regression model, but just to generate predictions of counterfactual values of the outcome variable (*Remark 10*: This remark is valid also for the extension of the MC method to the inclusion of covariates in that method, which was examined by[12]. Indeed, also in that case focus is on prediction of counterfactual values of the outcome variable, not on the estimation of the coefficients associated with the covariates.). Moreover, the MCFE method does not rely on matrix inversion, which would suffer from numerical issues in case of approximate multicollinearity (*Remark 11*: These last two remarks hold also for the MCB method but, as already discussed, the MCFE method is more suitable for policy evaluation.).

We conclude this subsection by providing, in Fig. 1, a broad overview of the conceptual framework of the work, which is detailed in the following sections. We anticipate here that, in summary, as described in the next sections, our analysis finds that the EU ETS policy reduced CO_2 emissions by about 15.4% in the European countries studied during the 2005–2020 period, ranging from almost no impact for Austria to a reduction of about 35% for Denmark.

Detailed overview of the original contributions of the work

The main original contributions of the work are related to an extensive evaluation of the MCFE application to assess the impact of the EU ETS policy on CO_2 emissions on EU countries. For a better reading, the results of such an analysis are reported in the following subsections of the "Results" section, whose titles provide a more detailed overview of the original contributions of this work: application of statistical tests about the presence of individual and time fixed effects ("Application of statistical tests about the presence of individual and time fixed effects" section); evaluation of the MCFE reconstruction accuracy in the pre-treatment period ("Evaluation of the MCFE reconstruction accuracy in the pre-treatment period" section); additional comparisons with other methods ("Additional comparisons with other methods" section); comparison between counterfactual/actual values in the treatment period ("Comparison between counterfactual/actual values in the treatment period" section); application of statistical tests about the comparison between estimated and actual data ("Application of statistical tests about the comparison between estimated and actual data" section); robustness checks ("Robustness checks" section); analysis of the effects of the EU ETS policy on CO_2 emissions ("Analysis of the effects of the EU ETS policy on CO_2 emissions" section).

Additional more technical original contributions are reported in the "Supplementary Material". They refer, respectively, to: a comparison between MC with/without fixed effects estimation (Supplementary Section A); a comparison with the DiD model (Supplementary Section B); a comparison with the SCM (Supplementary Section C); the presentation of additional tables related to the results obtained in the work (Supplementary Section D).

Description of the data set

In this paper, we use data on CO_2 emissions by country. In our analysis of the causal effects of the EU ETS policy (covering the period from 2000 to 2020, being this policy introduced in 2005), we can consider EU countries as "treated" and selected high-income non-EU countries as "untreated", since for the latter the potential treatment (before 2020) is limited compared to that of the EU countries[31]. The data set used for our analysis is extracted

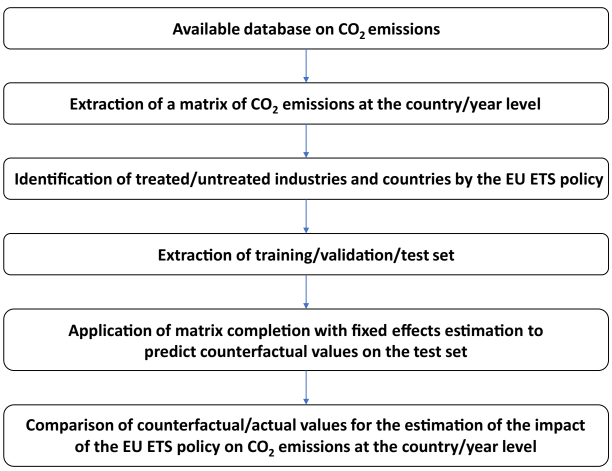

Figure 1. Conceptual framework of the work.

from the Emissions Database for Global Atmospheric Research (EDGAR), which covers all countries and can be accessed for free at https://edgar.jrc.ec.europa.eu/ (*Remark 12*: This hyperlink was accessed in July 2024.). The reader is referred to[68] for details about how missing data for specific sectors and years are handled in the construction of the EDGAR database (basically, suitable interpolation procedures are used therein). The issue of missing data is discussed also at the end of this section.

Since the ETS is mandatory for EU countries, we consider the following 17 countries as "treated" for our study: Austria (AUT), Belgium (BEL), Czech Republic (CZE), Germany (DEU), Denmark (DNK), Spain (ESP), Finland (FIN), France (FRA), Hungary (HUN), United Kingdom (GBR), Greece (GRC), Ireland (IRL), Italy (ITA), Poland (POL), Portugal (PRT), Slovakia (SVK), and Sweden (SWE). All these countries were members of the EU between 2005 (the beginning of the ETS policy) and 2020 (*Remark 13*: Brexit came into effect in 2021, so the United Kingdom was a member of the EU during the study period. Indeed, although the United Kingdom left the EU at the end of January 2020, it continued to participate in many EU institutions, during a transition period that covered 11 months. Then, starting from January 2021, the EU ETS policy in the United Kingdom was replaced by the UK ETS policy.). Among the 25 countries that were part of the EU in 2005, we do not consider in our analysis small countries having less than 5 million inhabitants at the time (i.e., Cyprus, Estonia, Latvia, Lithuania, Luxembourg, Malta, Slovenia), as their data are expected to be highly sensitive even to small shocks (since even the cessation of production of a single plant could have had a large impact on their respective total CO_2 emissions). In the case of the Netherlands, only partial data are available in the EDGAR database for the period considered for our study. Hence, also that country is excluded from the analysis. Additionally, data related to other 6 non-EU countries (as of 2005) influenced by the ETS policy, i.e., Bulgaria, Croatia, Iceland, Liechtenstein, Norway, and Romania are not included in our analysis partly because of the small-country issue mentioned above, and partly because their emissions in the period of analysis could have been significantly influenced by other ad-hoc policies beside the EU ETS. Altogether, the selected 17 EU countries represent, respectively, about 87% of the population in 2005 of the 25 EU countries at the time and of the 6 non-EU countries influenced by the ETS policy reported above, and 92% of total CO_2 emissions of the same set of countries.

As a control group (non-treated countries, without ETS), we consider countries that were not members of the EU between 2000 and 2020 and that were not highly related to the EU (e.g., as a result of very relevant special agreements) during this period (i.e., Iceland, Liechtenstein, Norway, and Switzerland are excluded) (*Remark 14*: Iceland, Liechtenstein, and Norway introduced the EU ETS in 2008 (i.e., during its phase 2), while Switzerland has several bilateral agreements with EU countries. A second reason to exclude from the control group countries such as Iceland, Liechtenstein, and Switzerland is the small-country issue just mentioned in the main text.). Our control group for this work consists of large non-EU high-income countries in the available database that are almost in the same phase of the so-called environmental Kuznets curve (*Remark 15*: Emerging economies are commonly regarded as the world's biggest polluters. Conversely, developed countries are generally considered cleaner. This perception is theoretically supported by the concept of environmental Kuznets curve (see, e.g.,[69]), according to which there is an inverted "U" relationship between Gross Domestic Product (GDP) per capita and environmental degradation, measured as CO_2 emissions per capita.) as the selected EU countries, and that are far from such countries geographically. Indeed, to obtain a valid control group, one needs to exclude as much as possible spillover effects of the EU ETS policy on its countries (due, e.g., to possible delocalization induced by the policy, which is unlikely to occur for the selected countries in the control group). Specifically, we include the following seven industrialized countries in the control group: Australia (AUS), Canada (CAN), Israel (ISR), Japan (JPN), New Zealand (NZL), South Korea (KOR), and the United States of America (USA). Although some non-EU countries also took specific measures to reduce CO_2 emissions, the impact of these measures was relatively negligible compared to the EU ETS policy at least until 2016[28]. For instance, Australia adopted a policy similar to the EU ETS one only some years later (in 2011), and ratified the Paris agreement only in November 2016[55].

In our analysis, we aggregate to the country-year level the values of CO_2 emissions originally available at the country-industry-year level in the EDGAR database. Instead of aggregating CO_2 emissions coming from all the industries, we consider only those associated with a subset of industries that, according to[70], are potentially influenced by the EU ETS policy. Specifically, for both the treated and untreated countries, we aggregate only CO_2 emissions coming from the following industries (represented here by their EDGAR codes): 1.A.1.a (Main Activity Electricity and Heat Production), 1.A.1.bc (Petroleum Refining—Manufacture of Solid Fuels and Other Energy Industries), 1.A.2 (Manufacturing Industries and Construction), 2.A.1 (Cement Production), 2.A.2 (Lime Production), 2.A.3 (Glass Production), 2.A.4 (Other Process Uses of Carbonates), 2.B (Chemical Industry), 2.C (Metal Industry) (*Remark 16*: This partial aggregation has the advantage of limiting the potential influence of other policies similar to the EU ETS on our counterfactual estimates of CO_2 emissions of treated countries.). Emission data related to these industries are available for all the countries considered in our analysis (which motivates the exclusion of the Netherlands, for which this does not hold). In this way, the resulting 24×21 CO_2 emission matrix represents the (total) amount of CO_2 emissions for each country and year, restricted to the selected industries. The portions of this matrix that are associated, respectively, with the pre-treatment period of EU countries (2000–2004) and with their treatment period (2005–2020) are reported in Tables 1 and 2. For illustrative reasons, rows corresponding to treated/untreated countries appear separately in the two tables.

The goal of this data aggregation is threefold. First, we reduce the computational burden of repeatedly applying MC(FE) by using a smaller matrix as input. Second, we simplify the analysis by focusing on the aggregate level for each country. Comparing countries, industries, and years (i.e., using three dimensions in the analysis) would make the approach to completing the matrix much more complex, possibly calling for its extension to the tensor case[71]. Third, by aggregating emissions patterns at the national level, our pre-processing provides an effective way to reduce the noise possibly due to missing disaggregated data.

Country/year	2000	2001	2002	2003	2004
Pre-treatment period (treated countries)					
AUT	34707	36675	36982	39740	41280
BEL	70990	70182	64423	66120	63616
CZE	101026	99733	96098	96161	96573
DEU	511535	515928	515322	525809	519996
DNK	33324	34789	34588	39439	33786
ESP	182709	179756	195454	194093	203825
FIN	39177	44337	46464	54446	51046
FRA	160145	155427	156114	158261	154559
GBR	305317	317019	307621	319570	314799
GRC	65497	66026	65267	66137	66197
HUN	33198	33586	32267	33085	31015
IRL	23463	24829	23705	22549	22386
ITA	250885	246360	255464	269626	277563
POL	231400	225987	217256	225924	226089
PRT	40896	39497	43004	37459	38431
SVK	28023	28698	27911	29518	28765
SWE	27286	27133	28851	29770	29408
Pre-treatment period (untreated countries)					
AUS	256160	263192	268015	264705	274725
CAN	300356	298056	300351	313143	305502
ISR	41785	41560	44628	45121	44459
JPN	794764	779417	809513	824833	818929
KOR	315217	322808	322926	326878	350700
NZL	17240	19003	18553	18625	17927
USA	3518332	3480104	3304577	3327424	3398937

Table 1. Portion of the emission matrix during the pre-treatment period of EU countries, which is obtained from the aggregation of data available in the EDGAR database. The values reported in the table are total yearly CO_2 emissions in thousands of tons, restricted to selected industries (see the "Description of the data set" section for details).

Methodology

In this paper, we opt for an innovative methodological approach in the field of policy evaluation. Clearly, with the available data set, it is not possible to conduct a randomized control trial analysis, because we have data on sovereign countries, and the ideal case of a randomized sample in which treated/untreated units have ex-ante the same characteristics is not realized. Some possible methods to perform policy evaluation in this situation (and their limitations) have been discussed in the "Methodological review" section (*Remark 17*: The application of some of the methods presented in the "Methodological review" section was excluded immediately, taking into account the peculiarities of the available data set: for instance, in our case, it was not possible to find at least one valid IV among the available variables; also RDD did not appear to be applicable because there was no sharp temporal or spatial threshold between treated and untreated countries; adjustments to the control group such as PSM and Mahalanobis distance matching and entropy balancing were not conclusive in our case, given the large heterogeneity and quite small number of countries in our sample.).

Given the peculiarities of our specific problem and data structure, in our analysis we prefer to adopt the MCFE approach for the following reasons: (i) the numerical results reported by[12] show that their proposed MCFE method for policy evaluation generally outperforms other alternative methods, such as the SCM and elastic net estimators; (ii) the MCFE approach can also be interpreted as a generalization of earlier approaches such as the SCM. Indeed, these approaches share the same objective function (based on the Fröbenius norm of a suitable projection of the difference between a latent matrix and the observed matrix), but have different constraints (which are less stringent in the case of the MCFE method).

For the application considered in this paper, the use of the MCFE method is justified by the fact that the counterfactual CO_2 emission levels for the treated countries (namely, the selected EU countries) are not known in the years of treatment when the EU ETS policy was in force. Therefore, we use the MCFE method to generate estimates of such counterfactual values and compare them to the actual CO_2 emission values, with the ultimate goal of estimating the effect of the treatment on CO_2 emission values through the EU ETS policy.

The main idea is to consider the treated values (i.e., the CO_2 emission values of EU countries in the years of treatment) as missing values and the other entries of the CO_2 emission matrix as given data. Specifically, in this paper, we apply the following formulation of the MC optimization problem, namely, the MCFE optimization problem proposed by[12]:

Country/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Treatment period (treated countries)																
AUT	41417	40742	39733	40069	35094	39861	40340	37671	36671	34461	34912	34244	35846	33279	34720	32021
BEL	61890	60810	59912	57445	52916	58905	54153	52234	51810	50732	50801	50117	49241	50198	50881	47430
CZE	95541	95299	97959	91662	86821	86597	85194	81446	77078	75693	75553	75779	75975	75507	70759	63544
DEU	512214	518967	532729	513605	467501	500953	493731	500504	503910	482378	481942	480249	464755	449741	395238	356315
DNK	30274	38249	33377	30500	29163	29192	25106	21608	23545	20042	17247	18606	16610	16532	13541	12288
ESP	215340	202179	214060	191705	162382	149795	160615	165360	138852	138820	150250	138143	151717	142705	126248	105940
FIN	39001	50620	48331	41277	39319	47368	40756	34952	36562	33877	30684	32182	30055	31619	28564	24816
FRA	162353	157648	159113	151980	139882	148646	144459	140928	133894	116696	117419	120495	124085	119375	115227	102829
GBR	314999	319266	312715	299790	260934	267828	249959	264253	249522	224321	202852	173242	160151	154349	143756	130593
GRC	67216	65638	70021	66959	60441	57247	54679	55892	51880	49085	45454	44284	44910	43557	37577	30431
HUN	29768	29937	30898	29738	24245	25029	25746	24258	22353	22226	23253	23553	28170	27433	25258	24298
IRL	23127	22749	22002	21258	18010	18286	16379	17527	16218	16598	17477	18485	17853	16791	15429	14711
ITA	275508	272742	271390	259886	215099	224502	222088	209769	179866	167416	174652	171107	171902	164078	157317	142954
POL	220887	227740	228082	218310	204236	212864	216447	209140	209250	200601	201770	202755	207730	207356	194394	180448
PRT	41905	38068	36517	35048	34350	29485	29616	29690	28366	27813	31687	30047	34284	30005	24931	22192
SVK	28592	28300	27633	26162	23433	24840	24628	24001	23576	22463	22978	23270	24187	23859	21747	20578
SWE	26913	26409	24824	24729	21907	27070	24255	23206	22220	21534	21920	22490	22582	21225	20353	20666
Treatment period (untreated countries)																
AUS	275978	281321	289438	291008	296375	295491	290440	288395	278607	267909	272227	279623	279245	274799	273692	265331
CAN	311545	308482	327438	308867	284589	296059	302167	305267	305731	307571	308974	310726	310847	309819	307753	288648
ISR	44135	46428	48183	48243	46020	50370	50329	56015	51204	48631	49158	48162	47911	43548	44775	42083
JPN	830547	819719	867927	816586	766791	825911	888195	932825	945707	908041	871998	863760	846319	816234	783470	740558
KOR	346953	362790	375267	391885	399810	441847	472250	472054	468071	461935	476183	481839	488572	495141	475400	446528
NZL	18544	18609	17305	18384	15971	15940	15100	16803	16703	17113	16506	15815	16157	15359	16988	16079
USA	3409149	3361812	3417242	3325690	3007161	3196881	3045237	2922073	2937381	2943427	2799234	2714036	2629591	2688014	2542358	2348769

Table 2. Portion of the emission matrix during the treatment period of EU countries, which is obtained from the aggregation of data available in the EDGAR database. The values reported in the table are total yearly CO_2 emissions in thousands of tons, restricted to selected industries (see the "Description of the data set" section for details).

$$\begin{aligned}&\underset{\hat{\mathbf{M}}\in\mathbb{R}^{m\times n},\hat{\mathbf{L}}\in\mathbb{R}^{m\times n},\hat{\boldsymbol{\Gamma}}\in\mathbb{R}^{m\times 1},\hat{\boldsymbol{\Delta}}\in\mathbb{R}^{n\times 1}}{\text{minimize}} \left(\frac{1}{|\Omega^{\text{tr}}|} \sum_{(i,j)\in\Omega^{\text{tr}}} \left(M_{i,j} - \hat{M}_{i,j} \right)^2 + \lambda \|\hat{\mathbf{L}}\|_* \right),\\ &\text{subject to } \hat{\mathbf{M}} = \hat{\mathbf{L}} + \hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top + \mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top,\end{aligned} \tag{1}$$

where

- Ω^{tr} is a subset of pairs of indices (i, j) corresponding to the positions of known entries of a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ (using a machine learning expression, Ω^{tr} can be called a training set of pairs of indices);
- $\mathbf{1}_n$ and $\mathbf{1}_m$ are column vectors consisting of n entries and m entries, respectively, all equal to 1;
- $\hat{\mathbf{M}}$ is the completed matrix decomposed as

$$\hat{\mathbf{M}} = \hat{\mathbf{L}} + \hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top + \mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top \tag{2}$$

  (where $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Delta}}$ must be chosen to solve the above optimization problem);
- $\lambda \geq 0$ is a regularization constant;
- $\|\hat{\mathbf{L}}\|_*$ is the nuclear norm of the matrix $\hat{\mathbf{L}}$, i.e., the summation of all its singular values.

In the above, the two terms $\hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top$ and $\mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top$ model, respectively, estimates of row-fixed effects (e.g., of unit-fixed effects) and of column-fixed effects (e.g., of time-fixed effects) in the reconstruction $\hat{\mathbf{M}}$ of $\mathbf{M}$ according to Eq. (2). The regularization constant λ controls the tradeoff between adequately fitting the known entries of the matrix $\mathbf{M}$ and achieving a small nuclear norm of the first term $\hat{\mathbf{L}}$ of its reconstruction. Here, the nuclear norm plays a similar role as the l_1-norm regularization term used in the well-known and widely used Least Absolute Shrinkage and Selection Operator (LASSO) regularization method[56]. It is worth noting that, in contrast to earlier formulations of the MC optimization problem—see, e.g.,[66]—the nuclear norm $\|\hat{\mathbf{L}}\|_*$ is used in the optimization problem (1) instead of the nuclear norm $\|\hat{\mathbf{M}}\|_*$. In other words, in the case of the MCFE method, the estimated fixed effects

$\hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top$ and $\mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top$ are not regularized. In the present context, this is an important issue because otherwise, by using the alternative regularization term $\lambda\|\hat{\mathbf{M}}\|_*$ instead of $\lambda\|\hat{\mathbf{L}}\|_*$ (i.e., by regularizing the entire reconstructed matrix $\hat{\mathbf{M}}$, which is the case of the MCB method) (*Remark 18*: Or, equivalently, by further constraining the optimization vectors $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Delta}}$ to be equal to $\mathbf{0}$.), one could obtain biased estimates that might underestimate the actual values[72]. In other words, any estimated element $\hat{M}_{i,j}$ could be a systematic underestimate of the corresponding element $M_{i,j}$ for the optimal choice of λ, making it difficult to obtain reliable estimated counterfactual values (*Remark 19*: The issue of underestimation under MC is investigated in more detail in Supplementary Section A of the Supplementary Material, where it is shown by means of one representative example that MC without fixed effects estimation (i.e., MCB) indeed underestimates the missing elements systematically, when it is applied to the specific data set considered in the present work.). As described in the literature, this is a common problem when using regularization methods. For instance, it is well-known that the LASSO regularization method may be affected by underestimation problems[73,74].

In this paper, the optimization problem (1) is solved numerically by applying the soft-impute algorithm developed by[66] and adapted by[12] to the case of the optimization problem (1) (*Remark 20*: Specifically, we apply the function "mcnnm" contained in the R package "MCPanel", which is freely available at the following hyperlink: https://rdrr.io/github/susanathey/MCPanel/man/mcnnm.html. This function allows the possibility of either estimating or not the row-fixed and column-fixed effects.). It is worth noting that the MCFE estimator used in this work was demonstrated by[12]—using two applications to smoker data and stock market data—to outperform several alternative methods such as DiD, SCM, vertical regression with elastic net regularization, and horizontal regression with elastic net regularization[12].

The soft-impute algorithm for calculating the MCFE estimator goes as follows. Let the projection operator $\mathbf{P}_{\Omega^{\rm tr}} : \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{m\times n}$ be defined as $[\mathbf{P}_{\Omega^{\rm tr}}(\mathbf{M})]_{i,j} := M_{i,j}$ if $(i,j) \in \Omega^{\rm tr}$, 0 otherwise. Similarly, let the projection operator $\mathbf{P}^{\perp}_{\Omega^{\rm tr}} : \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{m\times n}$ be defined as $[\mathbf{P}^{\perp}_{\Omega^{\rm tr}}(\mathbf{M})]_{i,j} := M_{i,j}$ if $(i,j) \notin \Omega^{\rm tr}$, 0 otherwise. For a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ with rank r, let its Singular Value Decomposition (SVD) be

$$\mathbf{S}_r(\mathbf{M}) = \mathbf{U}\mathbf{D}_r\mathbf{V}^T\,, \tag{3}$$

where $\mathbf{D}_r \in \mathbb{R}^{r\times r}$ is a diagonal matrix, which collects the r singular values $d_1,\ldots,d_r$ of $\mathbf{M}$. Then, the soft-thresholded version of the SVD of $\mathbf{M}$ reads as

$$\mathbf{S}_\lambda(\mathbf{M}) := \mathbf{U}\mathbf{D}_\lambda\mathbf{V}^T\,, \tag{4}$$

where

$$\mathbf{D}_\lambda := \mathrm{diag}[(d_1-\lambda)_+,\ldots,(d_r-\lambda)_+]\,, \tag{5}$$

and the subscript "+" stands for the non-negative part of a real number.

According to the soft-impute algorithm, we first initialize $\hat{\mathbf{L}}$ as $\hat{\mathbf{L}}^{\rm old} = \mathbf{P}_{\Omega^{\rm tr}}(\mathbf{M}) \in \mathbb{R}^{m\times n}$ and generate an increasing sequence of K values ($\lambda_1 < \ldots < \lambda_K$) for the regularization constant $\lambda \geq 0$. Let $\varepsilon > 0$ denote a selected tolerance, and $\|\cdot\|_F$ the Fröbenius norm. Then, for each $k = 1,\ldots,K$, set $\lambda = \lambda_k$ and

1. Iterate until convergence the following:
    (a) Given the current $\hat{\mathbf{L}} = \hat{\mathbf{L}}^{\rm old}$, get $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Delta}}$ by imposing the first-order optimality conditions in the optimization problem (1);
    (b) Compute $\hat{\mathbf{L}}^{\rm new} \leftarrow \mathbf{S}_{\frac{\lambda|\Omega^{\rm tr}|}{2}}\left(\mathbf{P}_{\Omega^{\rm tr}}(\mathbf{M} - \hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top - \mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top) + \mathbf{P}^{\perp}_{\Omega^{\rm tr}}(\hat{\mathbf{L}}^{\rm old})\right)$;
    (c) If $\frac{\|\hat{\mathbf{L}}^{\rm new}-\hat{\mathbf{L}}^{\rm old}\|_F^2}{\|\hat{\mathbf{L}}^{\rm old}\|_F^2} \leq \varepsilon$, go to Step 2;
    (d) Set $\hat{\mathbf{L}}^{\rm old} \leftarrow \hat{\mathbf{L}}^{\rm new}$;
2. Set $\hat{\mathbf{L}}_\lambda \leftarrow \hat{\mathbf{L}}^{\rm new}$ and $\hat{\mathbf{M}}_\lambda \leftarrow \hat{\mathbf{L}}_\lambda + \hat{\boldsymbol{\Gamma}}\mathbf{1}_n^\top + \mathbf{1}_m\hat{\boldsymbol{\Delta}}^\top$.

Since previous investigations showed that MC performs better when the elements of the matrix to which it is applied have similar magnitudes (e.g., when they are row-normalized, as in the cases considered by[65,67]), in our application the original matrix of annual CO_2 emissions is pre-processed by dividing each row (country) by the l_1-norm of that row restricted to the training set (*Remark 21*: This restriction is applied to avoid any use of the validation and test sets in the pre-processing phase.), and multiplied by the fraction of observed entries in that row. Then, MCFE is actually applied to the resulting matrix $\mathbf{M}$ (then, in a post-processing phase, a row de-normalization is performed, to go back to the original scale of the data).

In our application, where $\mathbf{M}$ is derived from the 24×21 true CO_2 emission matrix, with rows referring to countries and columns to years, the tolerance parameter ε is chosen as $\varepsilon = 10^{-30}$, in order to avoid early stopping of the algorithm. If convergence is not achieved, the soft-impute algorithm is stopped after $N^{\rm it} = 10^4$ iterations to reduce the total computational time.

In the present application, as shown in Fig. 2:

- the training set $\Omega^{\rm tr}$ corresponds to the union of the positions of all entries for the years 2000–2004 (pre-treatment period) and 75% (randomly selected) of the positions of entries belonging to non-EU countries in the years 2005–2020 (treatment period covered by the data set);

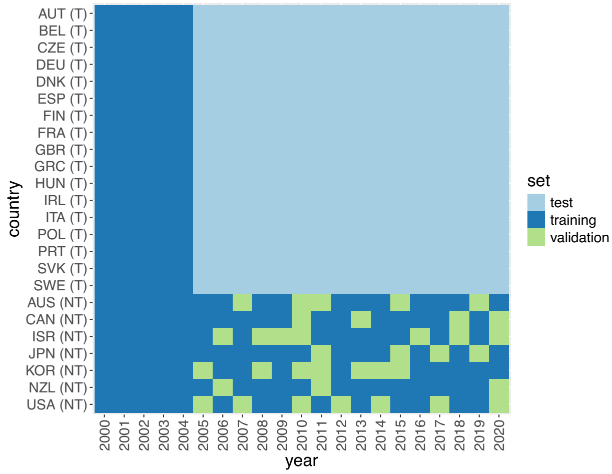

Figure 2. Partition of the considered matrix into training, validation and test sets. Countries are indicated on the *y* axis, while years are indicated on the *x* axis. T stands for "Treated" and NT stands for "Non Treated" (untreated). The training set Ω^{tr} corresponds to the union of the positions of all entries for the years 2000–2004 and 75% (randomly selected) of the positions of entries belonging to industrialized non-EU countries in the years 2005–2020. The validation set Ω^{val} corresponds to the other 25% of the positions of the entries belonging to industrialized non-EU countries in 2005–2020 that are not part of the training set. The test set Ω^{test} corresponds to the positions of the items belonging to the 17 considered EU countries in 2005–2020.

- the validation set Ω^{val} corresponds to the other 25% of the items of the entries belonging to non-EU countries in 2005–2020 that are not part of the training set;
- the test set Ω^{test} corresponds to the items belonging to EU countries in the treatment period covered by the data set (2005–2020).

It is noteworthy that while ground truth without treatment is available for the validation set Ω^{val} (which refers to untreated non-EU countries), this is not true for the test set Ω^{test} (which refers to treated EU countries). To generate confidence intervals and represent the best/worst scenarios for the estimates for each treated country, MC is applied 100 times (*Remark 22*: This number was chosen as a tradeoff between reducing machine processing time and achieving a satisfactory number of generations.), each time randomly selecting the training and validation sets as described above.

In each application of MCFE, the regularization constant λ is selected via an approach similar to that proposed by[12]. In particular, the optimization problem (1) is solved for multiple choices λ_k for λ. To explore different scales, these values are exponentially distributed as $\lambda_k = 2^{k/2-25}$, for $k = 1, \ldots, K = 100$. For each λ_k, the Root Mean Square Error (RMSE) of the matrix reconstruction on the validation set is calculated as follows:

$$RMSE_{\lambda_k}^{\text{val}} := \sqrt{\frac{1}{|\Omega^{\text{val}}|} \sum_{(i,j)\in\Omega^{\text{val}}} \left(M_{i,j} - \hat{M}_{\lambda_k,i,j}\right)^2}, \tag{6}$$

then the choice λ_k° that minimizes the $RMSE_{\lambda_k}^{\text{val}}$ for $k = 1, \ldots, 100$ is found. For each λ_k, the RMSE of the matrix reconstruction on the training set ($RMSE_{\lambda_k}^{\text{tr}}$) is defined in a similar way, as

$$RMSE_{\lambda_k}^{\text{tr}} := \sqrt{\frac{1}{|\Omega^{\text{tr}}|} \sum_{(i,j)\in\Omega^{\text{tr}}} \left(M_{i,j} - \hat{M}_{\lambda_k,i,j}\right)^2}. \tag{7}$$

In particular, the focus is on the values of $RMSE_{\lambda_k}^{\text{val}}$ and $RMSE_{\lambda_k}^{\text{tr}}$ calculated for $\lambda = \lambda_k^\circ$. Since there is no ground truth for the counterfactual values in the test set (i.e., the values without treatment), the RMSE for the test set is not calculated in this application of the MCFE method.

Results

The MCFE method described in the "Methodology" section was applied starting from the 24 × 21 country-year level CO_2 emissions matrix where selected industries are aggregated (see the "Description of the data set" section), and then pre-processed according to the methodology described in the "Methodology" section.

Application of statistical tests about the presence of individual and time fixed effects

To further motivate the adoption of the MCFE method described in the "Methodology" section (which also includes individual and time fixed effects), we performed some statistical tests for the presence of significant individual and time effects in our data. More specifically, we conducted an F-test for the null hypothesis of the absence of such effects in the context of a within regression model for panel data[75] where the 1-year lagged term is the only model covariate. We did not find a significant departure from the null hypothesis of absence of individual effects ($F =$ 0.610, p-value = 0.923). On the contrary, time-fixed effects, conditional to allowing for individual fixed effects, were found to be statistically significant (F = 2.946, p- value = 0.000). The null hypothesis of the F-test for the joint absence of time and individual effects was also rejected (F = 1.694, p-value = 0.006). In addition, it should be noted that, as mentioned by[12], the inclusion of the individual and temporal components aims to improve the quality of the imputation by penalizing only the residual component of the completed matrix in the optimization problem (1) (i.e., the one obtained by filtering out the individual and temporal components). Thus, it would have made sense to include them even in the case in which the F-test had not rejected the null hypothesis. As mentioned in the "Methodology" section, not including such components would lead to systematic underestimates.

Evaluation of the MCFE reconstruction accuracy in the pre-treatment period

It is worth noting that a necessary condition for obtaining credible counterfactual results is that the MCFE method achieves satisfactory performance in reconstructing the original matrix without any treatment. In the specific analysis, underestimation of the predicted values would not be recommended, as it could lead to incorrect conclusions about the impact of the ETS policy on different EU countries. In other words, in this context, it is crucial to avoid systematically underestimating the predicted values. It is worth mentioning that the performance of nuclear norm-based MC methods was assessed through a simulation study by[67] by applying such methods to a total CO_2 emission matrix, limited to the period 2000–2005, and generated based on the World Input–Output Database (WIOD) environmental accounts, 2016 release[76]. In that work, we found that the MCFE method proposed by[12] outperformed the MCB method developed by[66] in terms of goodness of fit (Mean Absolute Percentage Error, or MAPE, was used in that study). In particular, the MAPE of MCFE was very low even with a rather large number of unobserved entries in the matrix.

Additional comparisons with other methods

An additional comparison between MC with/without fixed effects estimation (i.e., between MCFE and MCB) is reported in Supplementary Section A of the Supplementary Material, which shows that MCFE turns out to be more suitable than MCB for assessing the impact of the EU ETS policy on CO_2 emissions. Moreover, Supplementary Sections B and C of the Supplementary Material compare MCFE with two other econometric methods commonly used to estimate counterfactuals (respectively, DiD and SCM), highlighting the advantages of MCFE for the specific case of assessing the impact of the EU ETS policy on CO_2 emissions. To ease the reading, these three comparisons are reported in the "Supplementary Material" rather than in the main text.

Comparison between counterfactual/actual values in the treatment period

The following results refer to the comparison of the counterfactual values (without treatment) with the actual values (with treatment). To summarize the results of the analysis, we report our main findings in Fig. 3. The estimated CO_2 emission values of the treated countries were obtained by applying MCFE. In other words, we used MCFE to estimate CO_2 emissions in the years of treatment for EU countries (i.e., treated countries) as if they had not received the treatment. We repeated the estimation process 100 times, each time randomly splitting the untreated portion of the matrix into training and validation sets, as described in the "Methodology" section. Then, Fig. 3 shows, for the elements of the test set related to the treated countries, the actual values of CO_2 emissions (i.e., those obtained in the case of treatment) against the appropriate statistics of the corresponding estimated values obtained by MCFE in the case of no treatment. Points are used for the medians (black), 10th percentiles (red), and 90th percentiles (blue) of the distributions of estimated values (obtained in the no-treatment case) in the 100 repetitions (one distribution for each treated country). The actual values (corresponding to the case of treatment) are shown through dark green points. It is worth noting that the actual and estimated values presented in Fig. 3 and in the successive figures of the main text have been de-normalized by row, since the application of the MCFE method was done after performing the l_1-norm row-normalization, then it was necessary to invert that process to go back to the original scale of the data. At first glance, we can see that the estimated values are, at a different extent, higher than the actual values for all treated countries in our analysis. In other words, according to our results, the ETS policy generally reduced CO_2 emissions of treated countries, particularly in the years after the second year of the second treatment phase (i.e., from 2009 onwards), as intended by the policy itself. It is worth noting that to obtain such a result, it was necessary to use the MCFE method by[12] instead of the MCB method (without fixed effects estimation) by[66], as explained in the "Methodology" section.

Application of statistical tests about the comparison between estimated and actual data

A parametric t-student test for the difference between means in independent populations (paired data t-test) was also performed to test whether the difference between actual values and estimated values was statistically significant under the hypothesis of no treatment. The test was performed for both the raw data and their natural logarithmic transformation (to more easily satisfy the normality assumption). To perform this statistical test, we considered two samples (S_1, with the actual values, and S_2, with the imputed values) with the same sample size of $n_1 = n_2 = 272$, where 272 is the product of the number of countries treated (17) and the number of years of treatment (16).

For all 100 simulations performed (with both raw and log normalized data), we rejected the null hypothesis of equal means. This simple evidence combined with Fig. 3 might suggest that the introduction of the EU ETS policy had a significant effect on reducing CO_2 emissions. At the same time, without a further check, we cannot

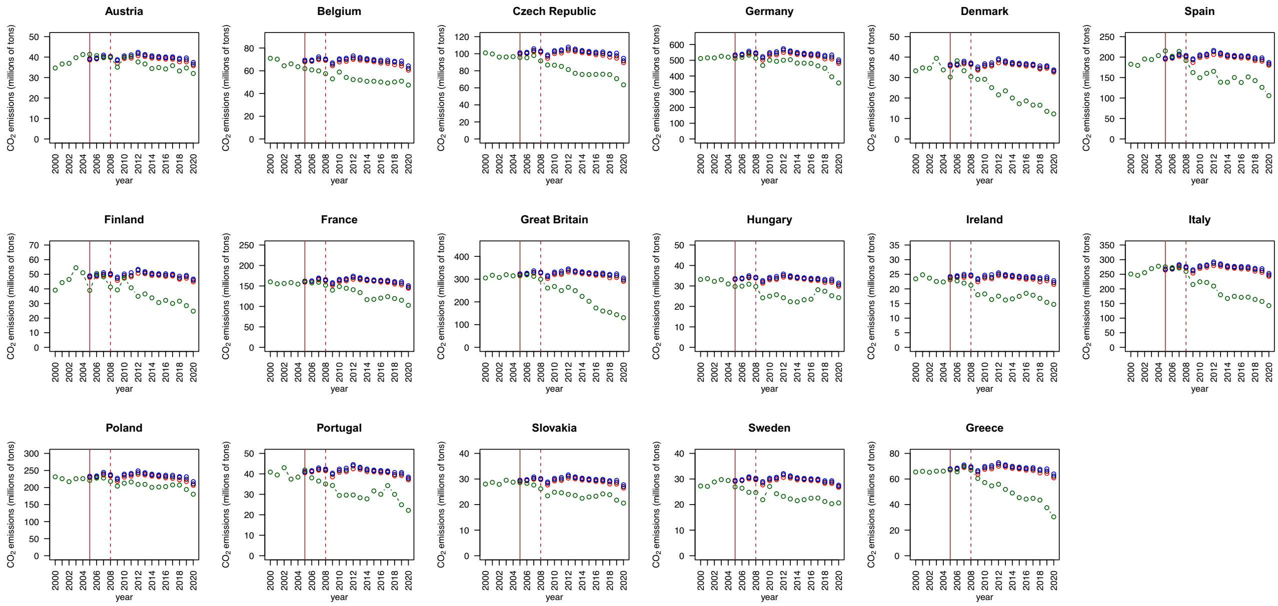

Figure 3. CO_2 emissions of treated countries. Actual values (dark green points) compared to values calculated by MCFE for the no-treatment hypothesis (test set). Medians (black points), 10th percentiles (red points), and 90th percentiles (blue points) considering the 100 MCFE random simulations. The solid vertical red line divides the period into pre-treatment and treatment periods. The dashed vertical red line represents the start of the second phase of ETS. The reader is referred to the "Comparison between counterfactual/actual values in the treatment period"–"Analysis of the effects of the EU ETS policy on CO_2 emissions" sections for an additional discussion of this figure.

rule out that this preliminary result is the consequence of using a positively biased estimator, i.e., the method used might tend to yield higher values than the true (unknown) counterfactual values.

To verify that this was not the case, we compared, as a diagnostic test, the actual values and the values estimated by the MCFE method for both the training and validation sets in the case of the untreated countries (this comparison was not possible for the test set because the counterfactual values were not available as ground truth). If the MCFE method we used were robust (i.e., if there were no significant overestimates or underestimates), then the true and estimated values for these countries would be essentially indistinguishable (especially in the case of the training set). This would be particularly important in the case of the validation set because it would rule out the overfitting of the training set.

Figures 4 and 5—referring respectively to the training set (restricted to the untreated countries after the start of treatment for treated countries) and the validation set (which, by construction, refers only to the untreated countries),—show that the differences between the actual values and the values estimated by MCFE were, as expected, quite negligible with respect to the corresponding differences obtained, in the original analysis, on the elements of the test set related to the treated countries.

Robustness checks

To further verify that our main results, related to the significant reduction of CO_2 emissions by the treatment, were not affected by a systematic overestimation, we decided to perform a counter-proof as a robustness test. To this end, we repeated the counterfactual analysis by reversing the roles of treated and untreated countries. In other words, this time we considered the EU countries as untreated and the non-EU countries as treated. So, for this second analysis, the (modified) test set was for the non-EU countries over the period 2005–2020. As can be seen in Fig. 6, the treatment effects (artificial this time) remained very strong (though reversed in sign, as expected) and, in particular, it was possible to rule out the problem of systematic overestimation of the MCFE method used, since the predicted values on the new test set looked much lower than the observed values.

Analysis of the effects of the EU ETS policy on CO_2 emissions

To return to the original analysis presented in the "Comparison between counterfactual/actual values in the treatment period" section, as can be seen in Fig. 3, when comparing the actual values of treated countries in the years of treatment with the medians of the corresponding counterfactual estimates (*Remark 23*: For simplicity, we consider here and below the medians of the estimated values instead of the estimated values themselves, since these are random variables.) obtained (in the case of no treatment) by the MCFE simulations, we can conclude that the EU ETS policy was effective in reducing CO_2 emissions. This is in line with other literature such as[31,51].

As shown in the "Supplementary Material", during the whole treatment period covered in the database (2005–2020), the majority of the EU countries included in our analysis achieved a ratio between the sum of the observed values and the sum of the medians of the estimated values (expressed as a percentage) of about 75–80%. The smallest value (65.01%) was obtained in the case of Denmark (i.e., Denmark's CO_2 emissions were

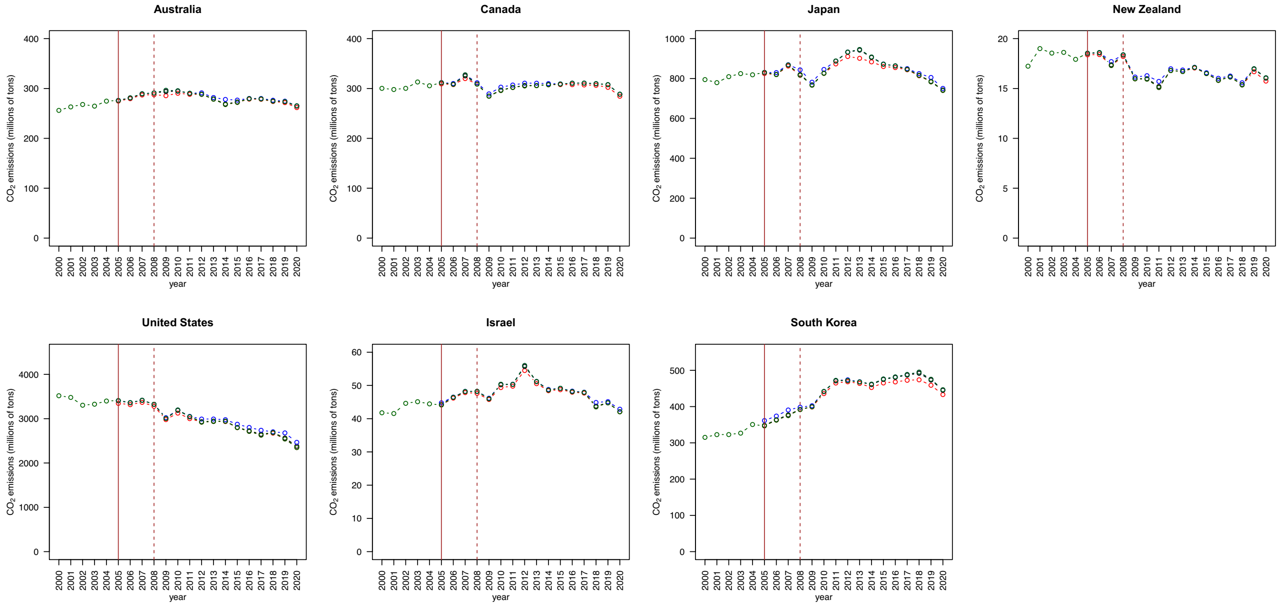

Figure 4. CO_2 emissions of untreated countries. Actual values (dark green points) compared to values calculated by MCFE for the no-treatment hypothesis (training set only, after the start of treatment of treated countries). Medians (black points), 10th percentiles (red points), and 90th percentiles (blue points) considering the 100 MCFE random simulations. The solid vertical red line divides the period into pre-treatment and treatment periods. The dashed vertical red line represents the start of the second phase of ETS. The reader is referred to the "Application of statistical tests about the comparison between estimated and actual data" section for an additional discussion of this figure.

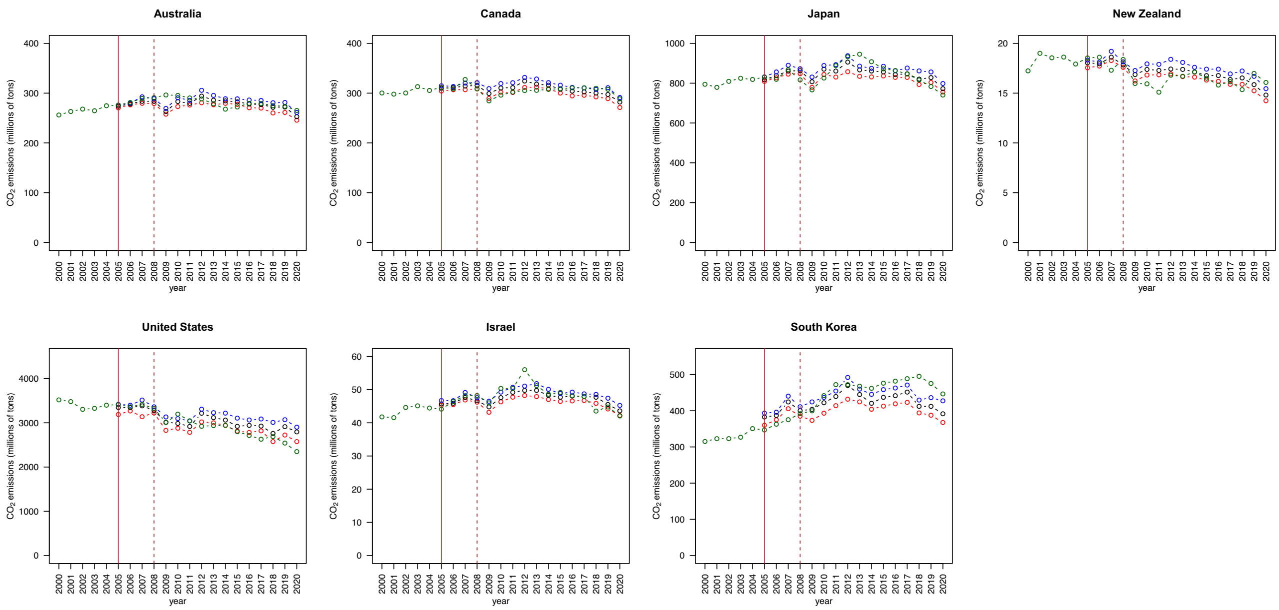

Figure 5. CO_2 emissions of untreated countries. Actual values (dark green points) compared to values calculated by MCFE for the no-treatment hypothesis (validation set only). Medians (black points), 10th percentiles (red points), and 90th percentiles (blue points) considering the 100 MCFE random simulations. The solid vertical red line divides the period into pre-treatment and treatment periods. The dashed vertical red line represents the start of the second phase of ETS. The reader is referred to the "Application of statistical tests about the comparison between estimated and actual data" section for an additional discussion of this figure.

reduced to more than 3/10 of the sum of the medians of the estimated counterfactual values associated with no treatment throughout the analysis period). The largest value (93.32%) was found in the case of Austria (this means that the amount of CO_2 emissions of Austria decreased at a minimal extent during the whole treatment period covered in the database, i.e., 2005–2020, due to the EU ETS policy). In general, however, we can conclude that the EU ETS policy did not have an irrelevant impact on CO_2 emissions in the EU. According to the results

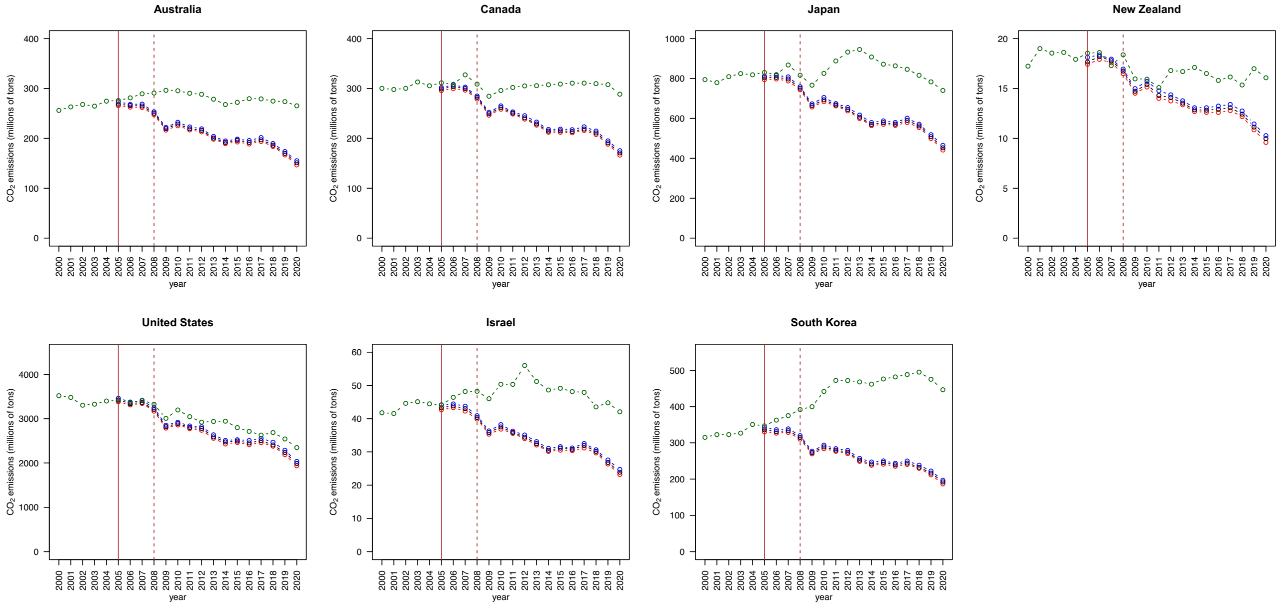

Figure 6. Inversion of treated and untreated countries in MCFE analysis. CO_2 emissions of untreated countries. Actual values (dark green points) versus values calculated by MCFE in the treatment hypothesis (modified test set). Medians (black points), 10th percentiles (red points), and 90th percentiles (blue points) considering the 100 MCFE random simulations. The solid vertical red line divides the period into pre-treatment and treatment periods. The dashed vertical red line represents the start of the second phase of ETS. The reader is referred to the "Robustness checks" section for an additional discussion of this figure.

of our analysis, the reduction was even larger than that estimated by [31]. In that paper, the authors estimated a reduction in CO_2 emissions (with respect to the case of the absence of the EU ETS policy) of about 3.8% using SCM across the EU for the period 2008–2016. According to the results of our analysis (based on the medians of the estimates), the reduction in CO_2 emissions for all EU countries included in our analysis (*Remark 24*: This reduction was calculated by comparing the sum of CO_2 emissions for each period for all EU countries included in the analysis with the sum of the medians of the estimated counterfactual CO_2 emissions for the same period and the same EU countries.) was approximately 19.0% in the same period 2008–2016 which was considered by[31]; 15.4% in the entire 2005–2020 treatment period (these results were obtained from the comparison of the sums of actual and counterfactual median values included in Fig. 7 and also in the Supplementary Section D of the Supplementary Material) (*Remark 25*: Still, we point out that the vertical distance between the various 10th and 90th percentile curves reported in Fig. 3 tends to increase when moving towards the end of the time horizon of analysis, so the results obtained at the end of the period turn out to be less reliable than, e.g., the ones related to the first year in the treatment period.).

Our results are consistent with the large increase in CO_2 emissions compared to the pre-treatment period, as shown in Fig. 5. Although both our analysis and that presented by[31] show positive effects of the EU ETS policy, some differences are observed in the magnitude of the effects achieved. This result could be explained not only by the different selection of EU countries considered in the two analyses (*Remark 26*: Supplementary Sections B and C of the Supplementary Material show that a similar magnitude of the estimated effects was obtained by replacing MCFE with either DiD or SCM (thus, providing even stronger support to the results obtained by MCFE). Nevertheless, MCFE still presented several advantages over these two other methods, as detailed in those appendices.), but also by the fact that the work[31], neglecting possible transmission effects, derived all their control and treated units within the same group of EU countries (i.e., their control and treated units were, respectively, economic sectors of EU countries directly affected by EU ETS policy and other economic sectors of the same EU countries not directly affected by the EU ETS policy). Instead, our analysis is done at a more aggregate level (i.e., EU countries are treated as a whole, limiting to industries potentially affected by that policy, while the control units are other countries outside the EU).

It is worth noting that our results change somewhat if we consider CO_2 emissions reductions only in the first phase of the EU ETS policy (2005–2007), as indicated in the "Supplementary Material". In this case, the reduction in CO_2 emissions from the EU ETS policy was much smaller. In particular, looking at the median estimates, the reduction in CO_2 emissions across the entire group of EU countries included in the analysis was about 2.5% in 2005–2007 (*Remark 27*: It is worth noting that the results obtained by our analysis, at the level of the whole set of treated countries, turned out to be quite robust with respect to variations in the data set. Indeed, we performed a similar analysis using CO_2 emissions data from a different source (the WIOD database environmental accounts, see[76]). Two smaller subsets of treated and untreated countries were selected, a shorter period of analysis was considered, and the aggregation of CO_2 emissions was made by considering all the industries. In spite of these changes, our estimated reductions of CO_2 emissions in the whole set of treated countries in the periods 2005–2007 and 2008–2016 (3.1% and 19.4%, respectively) were quite similar to the ones

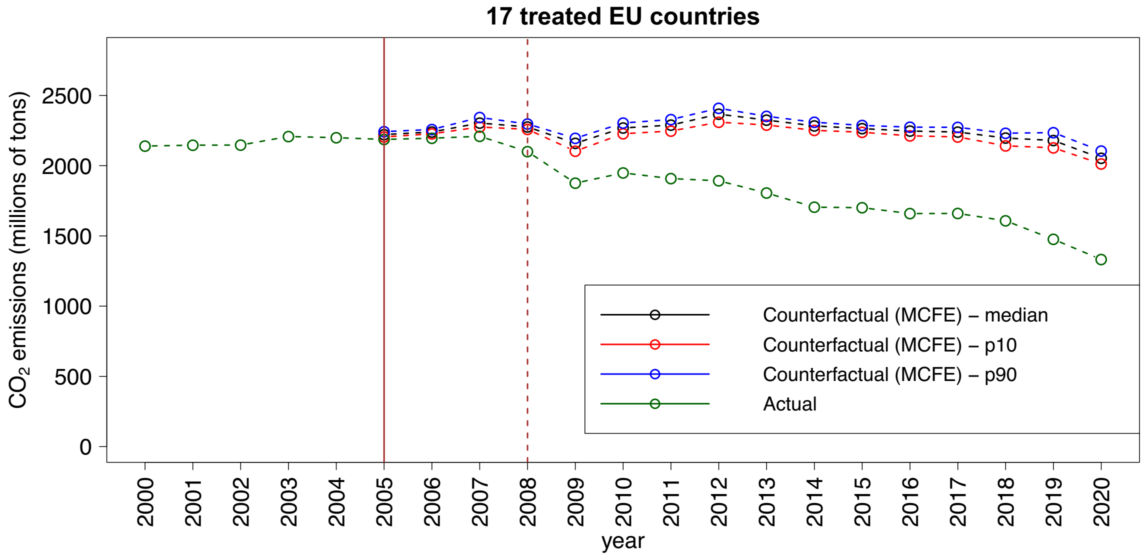

Figure 7. Sums of (observed and estimated) CO_2 emissions of the entire group of the 17 treated EU countries. Actual values (dark green points) compared to values calculated by MCFE for the hypothesis without EU ETS treatment (test set). Sum of medians across treated countries (black points), sum of 10th percentiles across treated countries (red points), and sum of 90th percentiles across treated countries (blue points) considering the 100 MCFE random simulations. The solid vertical red line divides the period into the pre-treatment and treatment periods. The dashed vertical red line represents the start of the second phase of ETS. The reader is referred to the "Analysis of the effects of the EU ETS policy on CO_2 emissions" section for an additional discussion of this figure.

obtained in the present analysis, based on the EDGAR database (i.e., 2.5% and 19.0%, respectively). This issue, combined with the extremely high coverage of CO_2 emissions by the data set considered in the present analysis, leads us to believe that the impacts estimated by our analysis are quite robust to different approaches to identify treated and control units of analysis.). Moreover, in three countries (Austria, Italy and Spain) the sums of the medians of the estimated counterfactuals were even higher than the corresponding sums of the observed values. In the first phase of the policy, the penalty for CO_2 emissions exceeding the quota was indeed small. This fact could be a possible explanation for the lower reduction in the first years of the policy.

Conclusions, policy implications, and possible future research developments

CO_2 emissions represent a growing problem closely related to pollution and climate change. Economic systems produce large amounts of CO_2 through the use of fossil energy. Therefore, governments are trying to shift production to new systems in order to reduce emissions[77]. In this context, the EU has introduced a market for emission rights, called the Emissions Trading Scheme (ETS), which was launched in 2005 and further expanded in subsequent years, as the second phase began in 2008. The impact of EU ETS on reducing CO_2 emissions is still debated in the literature.

In this paper, we present a new approach to quantify the impact of EU ETS policy on CO_2 emissions reductions. A counterfactual analysis allows us to quantify the reduction in CO_2 emissions from the ETS policy. The novelty of our work is that we apply, for the evaluation of the EU ETS policy, a state-of-the-art Statistical Machine Learning (SML) method based on Matrix Completion with Fixed Effects estimation (MCFE) for counterfactual analysis. The importance of using MCFE for this task becomes clear when one considers that conventional policy evaluation methods such as matching techniques—e.g., Propensity Score Matching (PSM), Mahalanobis and Hainmueller balancing—and the Synthetic Control Method (SCM) are not always suitable for performing true policy evaluation, since in some applications it may be nearly impossible to identify an appropriate control group for these methods, or small-sample issues may arise (*Remark 28*: See Supplementary Section C of the Supplementary Material for a more detailed discussion on these issues for the SCM case.). Applying the MCFE method to the CO_2 emissions matrix at the country-year level allowed us to quantitatively assess the impact of EU ETS on reducing emissions.

Using robust statistical tests and diagnostic controls, the effect of the EU ETS was found to be statistically significant, in line with some recent contributions. Based on our analysis, the CO_2 reduction from this policy appears to be higher than that found in the previous literature. We believe that the previous literature tends to underestimate the CO_2 reduction because it focused on the first phase of the policy and selected countries using less sophisticated methods to establish valid counterfactual estimates. We believe that overcoming such drawbacks through the adoption of MCFE is a significant result in terms of policy evaluation.

From the policy-making perspective, a first managerial/practical implication of this study concerns suggesting the adoption of policies similar to the EU ETS one to other similar countries (e.g., in the same phase of the environmental Kuznets curve) (*Remark 29*: This kind of application would be similar to the counter-proof reported in the "Robustness checks" section.), in the spirit of the classical Matrix Completion (MC) application to recommender systems, which has been outlined at the beginning of the "Non-technical overview of the original contributions of the work" section. A second managerial/practical implication of our study concerns following the same pipeline as in the "Results" section for the evaluation of the environmental impact of other policies different from the EU ETS one. As a third important managerial/practical implication, the methodology adopted in the study could be employed as one of the tools to assess the economic, social, and environmental impact of policies, as established by recent EU legislation (*Remark 30*: See https://commission.europa.eu/law/law-making-process/planning-and-proposing-law/impact-assessments_en.). In this context, MC could be used to do a more rigorous impact analysis of various policies, also with predictive aims: for instance, when one was given timely information for a group of countries but less recent information for a second group of countries, then one aimed to estimate trends for the countries in the latter group. This situation would likely arise when data related to different countries were not available simultaneously, but sequentially, because of the presence of heterogeneous data sources (*Remark 31*: A similar argument was used to motivate the recent MC application made by[57] in the context of input–output analysis.).

The results of our analysis make it possible in a rather simple way to express the reduction in environmental damage resulting from the introduction of the EU ETS policy in monetary terms. As this issue is not the focus of the present work, we plan to address it in a future extension of this research. Other possible future research developments concern overcoming some of the limitations of the current MCFE application, which are listed below along with possible ways to overcome these limitations.

First, the assumption that EU countries represent the treated group while the selected non-EU high-income countries are untreated by the EU ETS policy might oversimplify the complexity of that policy implementation and of its environmental impact. In particular, the possible presence of confounding factors—such as heterogeneous population growth and Gross Domestic Product (GDP) growth—could influence the estimates (*Remark 32*: It is also worth noting that the MC methodology adopted in our work already deals partially with these issues, because, differently from other econometric methods such as the Difference-in-Differences (DiD), it does not require the parallel trend assumption in the pre-treatment phase, implying the possibility to extrapolate time series to the treatment period even in the presence of heterogeneous trends in such time series during the pre-treatment period.) and would deserve further investigation in a future development of the present study (*Remark 33*: The possible presence of such factors is one additional reason for the exclusion of small countries from the data set used for the analysis made in the present study, in view of their expected high sensitivity to small shocks (see "Description of the data set" section).). Still remaining in the context of a MCFE analysis, the validity of the assumption above could be verified in the following way. First, one could include additional covariates in the MCFE optimization problem, either by inserting them at the beginning of each row of the matrix to be reconstructed (as is done in some applications of SCM[78]), or by applying the extension to the presence of covariates, still proposed by[12], of the MCFE optimization problem considered in this study. Then, one would compare the resulting estimates with the ones achieved in the present work, assessing if only minor changes were obtained. Such additional analyses are expected to require an extensive integration of the available data set, with the aim of including a sufficiently large number of potential confounders. An alternative way to address this issue could consist in replacing MC(FE) with either collective MC[79] or tensor completion[80] (possibly still with the insertion of covariates in the model). This could have the advantage of making it possible to consider different levels of treatment. However, both the latter extensions are expected to be much more computationally expensive than the MCFE approach adopted in this work.

Second, the MCFE application made in the present study works at quite an aggregate level, since the entries of the matrix to be reconstructed refer to CO_2 emissions at the country-year level, by restricting the aggregation to CO_2 emissions coming from selected industries. On one hand, a first possible future development could consist in a further data aggregation in each country by considering CO_2 emissions coming from all the industries, with the aim of verifying the absence of leakage effects of the EU ETS policy with respect to other industries in the same country. On the other hand, as a second possible future development, one could further disaggregate the data. For instance, after estimating the yearly impact of the ETS policy for a specific EU country (as made in the present work), one could distinguish between treated/untreated industries in that country, and perform the analysis at the industry-year level (focusing on the specific country). This would help assessing possible heterogeneous impacts of the ETS policy on different industries of the selected country. It is worth mentioning that, differently from the MCFE application considered in the present article, at this less aggregate level the average effect of the ETS policy of a given EU country on CO_2 emissions related to the activity of specific industries could be even negative, while at the level of the whole country, the effect of the policy on CO_2 emissions of the activity of all the industries in that country is expected to be positive, at least after some years since the adoption of the policy. So, identifying industries for which a negative average effect is obtained could help policymakers to investigate improvements of the policy and/or the possible presence of confounding factors that have influenced negatively the policy implementation.

Third, the EU ETS policy is a rather complex policy, which actually covers several phases of actuation. In our work, we have simplified the analysis of the treated countries by including information for them only on the fact they were treated/untreated in a specific period. Indeed, a direct application to causal inference of the matrix completion methodology developed by[12] allows one only to distinguish, for treated countries, between periods of treatment and periods of no treatment, but does not allow to include (at least in a straightforward manner) information about the occurrence of a specific phase of treatment. Further developments of this methodology would make it more feasible to extend the present MC analysis to the estimation of phase-dependent CO_2 counterfactual emission levels.

Fourth, for future research, we also consider developing a more sophisticated model that examines (in monetary terms) whether or not the decrease in output due to the price of CO_2 emissions outweighs the reduced environmental damage (this would require an appropriate definition of green gross domestic product). In addition, the whole analysis could be extended to a less aggregate level (with larger matrices) after accelerating/parallelizing the MCFE implementation, as was done recently by[60] for another application of the related baseline MC (MCB) method. This is important for the potential use of MCFE with three-dimensional data set of countries, industries, and years, possibly achieved by associating country-industry pairs with a single index, to avoid replacing MC with tensor completion. Finally, more sophisticated methods could be applied to obtain an estimate of the possible indirect impact of the EU ETS policy (e.g., related to carbon leakage to other non-EU countries) to better estimate the overall impact of the EU ETS policy.

Fifth, another possible research development concerns the investigation of potential adverse economic effects of the EU ETS policy. This would require replacing/combining the data set used for the current analysis with one related to the yearly level of economic activities of each country. Beside MCFE, other possible methods of analysis could be the already-mentioned collective MC and tensor completion.

Finally, the recent Machine Learning (ML) literature points out other methods for time-series prediction, such as multivariate Gaussian processes regression[81], and deep learning techniques[82], which could be also considered as possible alternative methods for counterfactual prediction in a future extension of this research.

Data availibility

The database used (i.e., the Emissions Database for Global Atmospheric Research, or EDGAR) can be accessed for free at https://edgar.jrc.ec.europa.eu/. Population data are taken from the World Bank's free database, available at https://data.worldbank.org/. The MCFE analysis performed in this work is based on the application of the function "mcnnm" contained in the R package "MCPanel", which is freely available at the following hyperlink: https://rdrr.io/github/susanathey/MCPanel/man/mcnnm.html. This function allows the possibility of either estimating or not the row-fixed and column-fixed effects (hence, to apply also the MCB method). The SCM application is based on the MATLAB package "Synth", which is freely available at the following hyperlink: https://web.stanford.edu/~jhain/Synth_Matlab/Synth_MATLAB.zip.

Received: 12 April 2024; Accepted: 14 August 2024
Published online: 24 August 2024

References

1. El Ouardighi, F., Kogan, K., Gnecco, G. & Sanguineti, M. Transboundary pollution control and environmental absorption efficiency. *Ann. Oper. Res.* **287**, 653–681 (2020).
2. Machado, F. L. V., Halmenschlager, V., Abdallah, P. R., da Silva Teixeira, G. & Sumaila, U. R. The relation between fishing subsidies and CO_2 emissions in the fisheries sector. *Ecol. Econ.* **185**, 107057 (2021).
3. Sun, X., Dong, Y., Wang, Y. & Ren, J. Sources of greenhouse gas emission reductions in OECD countries: Composition or technique effects. *Ecol. Econ.* **193**, 107288 (2022).
4. Zhang, Y., Wang, F. & Zhang, B. The impacts of household structure transitions on household carbon emissions in China. *Ecol. Econ.* **206**, 107734 (2023).
5. Van Aalst, M. K. The impacts of climate change on the risk of natural disasters. *Disasters* **30**, 5–18 (2006).
6. Tol, R. S. J. The economic effects of climate change. *J. Econ. Perspect.* **23**, 29–51 (2009).
7. Ackerman, F. & Munitz, C. Climate damages in the FUND model: A disaggregated analysis. *Ecol. Econ.* **77**, 219–224 (2012).
8. Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic productions. *Nature* **527**, 235–239 (2015).
9. Burke, M., Davis, W. M. & Diffenbaugh, N. S. Large potential reductions in economic damages under UN mitigation targets. *Nature* **557**, 549–53 (2018).
10. Burke, M. & Tanutama, V. *Climatic Constraints on Aggregate Economic Output* NBER working paper 25779 (2019).
11. Martin, R., Muûls, M. & Wagner, U. J. The impact of the European Union emissions trading scheme on regulated firms: What is the evidence after ten years? *Rev. Environ. Econ. Policy* **10**, 129–148 (2016).
12. Athey, S., Bayati, M., Doudchenko, N., Imbens, G. & Khosravi, K. Matrix completion methods for causal panel data models. *J. Am. Stat. Assoc.* **116**, 1716–1730 (2021).
13. Martelli, S. *et al.* Do voters support local commitments for climate change mitigation in Italy? *Ecol. Econ.* **144**, 27–35 (2018).
14. Binder, S. & Neumayer, E. Environmental pressure group strength and air pollution: An empirical analysis. *Ecol. Econ.* **55**, 527–538 (2005).
15. Soliman, A. Prescriptive drought policy and water supplier compliance. *Ecol. Econ.* **197**, 107429 (2022).
16. Doremus, J. Unintended impacts from forest certification: Evidence from indigenous Aka households in Congo. *Ecol. Econ.* **166**, 106378 (2019).
17. Koch, N. & Themann, M. Catching up and falling behind: Cross-country evidence on the impact of the EU ETS on firm productivity. *Resour. Energy Econ.* **69**, 101315 (2022).
18. Heckman, J. J., Ichimura, H. & Todd, P. E. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econ. Stud.* **64**, 605–654 (1997).
19. Hainmueller, J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20**, 25–46 (2012).
20. Pellegrini, L., Tasciotti, L. & Spartaco, A. A regional resource curse? A synthetic-control approach to oil extraction in Basilicata, Italy. *Ecol. Econ.* **185**, 107041 (2021).
21. Green, J. F. Does carbon pricing reduce emissions? A review of ex-post analyses. *Environ. Res. Lett.* **16**, 043004 (2021).
22. Teixidó, J., Verde, S. F. & Nicolli, F. The impact of the EU emissions trading system on low-carbon technological change: The empirical evidence. *Ecol. Econ.* **164**, 106347 (2019).
23. Verde, S. F. The impact of the EU emissions trading system on competitiveness and carbon leakage: The econometric evidence. *J. Econ. Surv.* **34**, 320–343 (2020).
24. IPCC. *Climate Change 2022: Mitigation of Climate Change* (Cambridge University Press, 2022). Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (eds Shukla, P. R., Skea, J., Slade, R., Al Khourdajie, A., van Diemen, R., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., Malley, J.).
25. Stavins, R. N. *The Future of U.S. Carbon-Pricing Policy* Vol. 56 (Harvard Environmental Economics Program, Cambridge, 2019).

26. Aydin, C. & Esen, O. Reducing CO_2 emissions in the EU member states: Do environmental taxes work? *J. Environ. Plan. Manag.* **61**, 2396 (2018).
27. Haites, E. *et al.* Experience with carbon taxes and greenhouse gas emissions trading systems. *Duke Environ. Law Policy Forum* **29**, 109–182 (2018).
28. Narassimhan, E., Gallagher, K. S., Koester, S. & Alejo, J. R. Carbon pricing in practice: A review of existing emissions trading systems. *Clim. Policy* **18**, 967–991 (2018).
29. Andersson, J. J. Carbon taxes and CO_2 emissions: Sweden as a case study. *Am. Econ. J. Econ. Pol.* **11**, 1–30 (2019).
30. FSR Climate. A literature-based assessment of the EU ETS. *Florence School of Regulation* 224 (European University Institute, 2019).
31. Bayer, P. & Aklin, M. The European Union emissions trading system reduced CO_2 emissions despite low prices. *Proc. Natl. Acad. Sci.* **117**, 8804–8812 (2020).
32. Diaz, G., Munoz, F. D. & Moreno, R. Equilibrium analysis of a tax on carbon emissions with pass-through restrictions and side-payment rules. *Energy J.* **41**, 93–122 (2020).
33. Arimura, T. H. & Abe, T. The impact of the Tokyo emissions trading scheme on office buildings: What factor contributed to the emission reduction? *Environ. Econ. Policy Stud.* **23**, 517–533 (2021).
34. Pretis, F. Does a carbon tax reduce CO_2 emissions? Evidence from British Columbia. *Environ. Resour. Econ.* **83**, 115–144 (2022).
35. Metcalf, G. E. & Stock, J. H. The macroeconomic impact of Europe's carbon taxes. *Am. Econ. J. Macroecon.* **15**, 265–286 (2023).
36. Rafaty, R., Dolphin, G. & Pretis, F. *Carbon pricing and the elasticity* of CO_2*emissions*. Resources for the Future Working Paper 21–23 (2023). https://media.rff.org/documents/wp_21-33_tUca3j8.pdf.
37. Calel, R. & Dechezleprêtre, A. Environmental policy and directed technological change: Evidence from the European carbon market. *Rev. Econ. Stat.* **98**, 173–191 (2016).
38. van den Bergh, J. & Savin, I. Impact of carbon pricing on low-carbon innovation and deep decarbonisation: Controversies and path forward. *Environ. Resour. Econ.* **80**, 705–715 (2021).
39. Zhang, Z. & Zhang, Z. Intermediate input linkage and carbon leakage. *Environ. Dev. Econ.* **22**, 725–746 (2017).
40. Branger, F., Quirion, P. & Chevallier, J. Carbon leakage and competitiveness of cement and steel industries under the EU ETS: Much ado about nothing. *Energy J.* **37**, 109–135 (2016).
41. Koch, N. & Basse Mama, H. Does the EU emissions trading system induce investment leakage? Evidence from German multinational firms. *Energy Econ.* **81**, 479–492 (2019).
42. aus dem Moore, N., Grokurth, P. & Themann, M. Multinational corporations and the EU emissions trading system: The specter of asset erosion and creeping deindustrialization. *J. Environ. Econ. Manag.* **94**, 1–26 (2019).
43. Borghesi, S., Franco, C. & Marin, G. Outward foreign direct investment patterns of Italian firms in the European Union's emission trading scheme. *Scand. J. Econ.* **122**, 219–256 (2020).
44. Kuusi, T., *et al. Carbon Border Adjustment Mechanisms and Their Economic Impact on Finland and the EU* (ETLA Economic Research, 2020). https://www.etla.fi/en/publications/carbon-border-adjustment- mechanisms-and-their-economic-impact-on-finland-and-the-eu/.
45. Venmans, F., Ellis, J. & Nachtigall, D. Carbon pricing and competitiveness: Are they at odds? *Clim. Policy* **20**, 1070–1091 (2020).
46. Chéze, B., Chevallier, J., Berghmans, N. & Alberola, E. On the CO_2 emissions determinants during the EU ETS phases I and II: A plant-level analysis merging the EUTL and Platts power data. *Energy J.* **41**, 153–183 (2020).
47. World Bank. *State and Trends of Carbon Pricing 2021* 87 (World Bank Group, 2021).
48. Calel, R. Adopt or innovate: Understanding technological responses to cap-and-trade. *Am. Econ. J. Econ. Pol.* **12**, 170–201 (2020).
49. Jaraitė, J. & Maria, C. D. Did the EU ETS make a difference? An empirical assessment using Lithuanian firm-level data. *Energy J.* **37**, 1–23 (2016).
50. Klemetsen, M., Rosendahl, K. E. & Jakobsen, A. L. The impacts of the EU ETS on Norwegian plants' environmental and economic performance. *Clim. Change Econ.* **11**, 2050006 (2020).
51. Petrick, S. & Wagner, U. J. *The impact of carbon trading on industry: Evidence from German manufacturing firms* Kiel Working Paper No. 1912 (2014).
52. Wagner, U. J., Muûls, M., Martin, R. & Colmer, J. The causal effect of the European Union emissions trading scheme: Evidence from French manufacturing plants. In *Proceedings of the IZA Workshop: Labor Market Effects of Environmental Policies* 33 (2014).
53. Dechezleprêtre, A., Nachtigall, D. & Venmans, F. The joint impact of the European Union emissions trading system on carbon emissions and economic performance. *J. Environ. Econ. Manag.* **118**, 102758 (2023).
54. Abadie, A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J. Econ. Lit.* **59**, 391–425 (2021).
55. Anderson, H. M., Gao, J., Turnip, G., Vahid, F. & Wei, W. Estimating the effect of an EU-ETS type scheme in Australia using a synthetic treatment approach. *Energy Econ.* **125**, 106798 (2023).
56. Hastie, T., Tibshirani, R. & Wainwright, M. *Statistical Learning with Sparsity* Vol. 143 (CRC Press, 2015).
57. Metulini, R., Gnecco, G., Biancalani, F. & Riccaboni, M. Hierarchical clustering and matrix completion for the reconstruction of world input–output tables. *Adv. Stat. Anal.* **107**, 575–620 (2023).
58. Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. & De Vries, G. J. An illustrated user guide to the world input–output database: The case of global automotive production. *Rev. Int. Econ.* **23**, 575–605 (2015).
59. Gnecco, G., Nutarelli, F. & Riccaboni, M. A machine learning approach to economic complexity based on matrix completion. *Sci. Rep.* **12**, 9639 (2022).
60. Gnecco, G., Nutarelli, F. & Riccaboni, M. Matrix completion of world trade: An analysis of interpretability through Shapley values. *World Econ.* **46**, 2707–2731 (2023).
61. Gnecco, G., Landi, S. & Riccaboni, M. Can machines learn creativity needs? An approach based on matrix completion. *Ital. Econ. J.* **9**, 1111–1151 (2023).
62. Gnecco, G., Landi, S. & Riccaboni, M. The emergence of social soft skill needs in the post COVID-19 era. *Qual. Quant.* **58**, 647–680 (2024).
63. Huang, W., Wei, D., Wang, C. & Lin, C. Matrix completion-based prediction analysis in carbon emissions. *Int. J. Embed. Syst.* **14**, 143–148 (2021).
64. Biancalani, F., Gnecco, G., Metulini, R. & Riccaboni, M. Matrix completion for the prediction of yearly country and industry-level CO_2 emissions. In *Machine Learning, Optimization, and Data Science, 8th International Workshop, LOD 2022, Certosa di Pontignano, Italy, September 19–22, 2022, Revised Selected Papers, Part I* 14–19 (2023).
65. Biancalani, F., Gnecco, G., Metulini, R. & Riccaboni, M. Prediction of annual CO_2 emissions at the country and sector levels, based on a matrix completion optimization problem. *Optim. Lett.*https://doi.org/10.1007/s11590-023-02052-2 *(2023).*
66. Mazumder, R., Hastie, T. & Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11**, 2287–2322 (2010).
67. Metulini, R., Biancalani, F., Gnecco, G. & Riccaboni, M. Assessing the performance of nuclear norm-based matrix completion methods on CO_2 emissions data. In *Book of the Short Papers—SIS 2023—Statistical Learning, Sustainability and Impact Evaluation (eds. Chelli, F. M., Ciommi, M., Ingrassia, S., Mariani, F., Recchioni, M. C.)* 776–781 (Pearson, 2023).
68. Crippa, M., *et al. GHG emissions of all world countries* Publications Office of the European Union, Luxembourg, JRC134504. https://doi.org/10.2760/953322 (2023).
69. Grossman, G. M. & Krueger, A. B. Economic growth and the environment. *Q. J. Econ.* **110**, 353–377 (1995).

70. Raude, M., Mazzarano, M. & Borghesi, S. EU emissions trading system sectoral environmental and economic indicators. EUI, RSC, Research Data, 2024, Florence School of Regulation, LIFE COASE—Collaborative Observatory for Assessment of the EU ETS [Climate]. https://hdl.handle.net/1814/76817.
71. Song, Q., Ge, H., Caverlee, J. & Hu, X. Tensor completion algorithms in big data analytics. *ACM Trans. Knowl. Discove. Data* **13**, 1–48 (2019).
72. Wang, L., Ma, S. & Han, Q. Enhanced sparse low-rank representation via nonconvex regularization for rotating machinery early fault feature extraction. *IEEE/ASME Trans. Mechatron.* **27**, 3570–3578 (2022).
73. An, B., Zhao, Z., Wang, S., Chen, S. & Chen, X. Sparsity-assisted bearing fault diagnosis using multiscale period group lasso. *ISA Trans.* **98**, 338–348 (2020).
74. Feng, Z. Z., Yang, X., Subedi, S. & McNicholas, P. D. The LASSO and sparse least squares regression methods for SNP selection in predicting quantitative traits. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **9**, 629–636 (2012).
75. Baltagi, B. H. *Econometric Analysis of Panel Data* 4th edn. (Springer, 2008).
76. Corsatea, T. D., *et al. World Input–Output Database Environmental Accounts. Update 2000–2016*. JRC Technical Report (2019). https://op.europa.eu/en/publication-detail/-/publication/df9c194b-81ba-11e9-9f05-01aa75ed71a1/language-en.
77. Sgarciu, S., Scholz, D. & Müsgens, F. How CO_2 prices accelerate decarbonisation—The case of coal-fired generation in Germany. *Energy Policy* **173**, 113375 (2023).
78. Abadie, A., Diamond, A. & Hainmueller, J. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Am. Stat. Assoc.* **105**, 493–505 (2010).
79. Alaya, M. Z. & Klopp, O. Collective matrix completion. *J. Mach. Learn. Res.* **20**, 1–43 (2019).
80. Liu, A. & Moitra, A. Tensor completion made practical. In *Proceeding of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada*, 18905–18916 (ACM, 2020).
81. Chen, Z., Fan, J. & Wang, K. Multivariate Gaussian processes: Definitions, examples and applications. *Metron* **81**, 181–91 (2023).
82. Masti, D. & Bemporad, A. Learning nonlinear state-space models using autoencoders. *Automatica* **129**(1–9), 109666 (2021).

Acknowledgements

The work was supported in part by the "Contributi Liberali 2022" project "Application of Matrix Completion Techniques to the Definition of Economic and Financial Recommender Systems", granted by the Bank of Italy, by the project "Aspetti Socioeconomici e di Sostenibilitá—Green Economy e Startup" (project code: PO232, CUP code: D65F21001600001), granted by MUR, by the PRIN 2022 project "MAHATMA" (CUP: D53D23008790006), funded by the European Union - Next Generation EU program, by the PRIN PNRR 2022 project "MOTUS" (CUP: D53D23017470001), funded by the European Union - Next Generation EU program, and by the PRIN PNRR 2022 project "SIGNUM" (CUP: F53D23010910001), funded by the European Union - Next Generation EU program.

Author contributions

The authors equally contributed to the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-70260-6.

Correspondence and requests for materials should be addressed to G.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.