



# Unilateral climate policies can substantially reduce national carbon pollution

Alice Lépissier<sup>1</sup> · Matto Mildenberger<sup>2</sup>

Received: 31 August 2020 / Accepted: 26 April 2021 / Published online: 03 June 2021  
© The Author(s) 2021

## Abstract

Following the failure of climate governance regimes that sought to impose legally binding treaty-based obligations, the Paris Agreement relies on voluntary actions by individual countries. Yet, there is no guarantee that unilateral policies will lead to a decrease in carbon emissions. Critics worry that voluntary climate measures will be weak and ineffective, and insights from political economy imply that regulatory loopholes are likely to benefit carbon-intensive sectors. Here, we empirically evaluate whether unilateral action can still reduce carbon pollution by estimating the causal effect of the UK's 2001 Climate Change Programme (CCP) on the country's carbon emissions. Existing efforts to evaluate the overall impact of climate policies on national carbon emissions rely on Business-As-Usual (BAU) scenarios to project what carbon emissions would have been without a climate policy. We instead use synthetic control methods to undertake an *ex post* national-level assessment of the UK's CCP without relying on parametric BAU assumptions and demonstrate the potential of synthetic control methods for climate policy impact evaluation. Despite setting lax carbon targets and making substantial concessions to producers, we show that, in 2005, the UK's CO<sub>2</sub> emissions per capita were 9.8% lower relative to what they would have been if the CCP had not been passed. Our findings offer empirical confirmation that unilateral climate policies can still reduce carbon emissions, even in the absence of a binding global climate agreement and in the presence of regulatory capture by industry.

**Keywords** Climate policy · Carbon pricing · Carbon emissions · Causal inference · Synthetic control · Non-parametric

## 1 Introduction

In recent years, policies to reduce greenhouse gas emissions (GHG) have been deployed at a rapid pace across the world. While scholars have extensively debated the theoretical

---

✉ Alice Lépissier  
alepissier@bren.ucsb.edu

<sup>1</sup> Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA, USA

<sup>2</sup> Department of Political Science, University of California Santa Barbara, Santa Barbara, CA, USA

merits of different types of policy instruments, we still do not know enough about the extent to which these policies work in the context of real-world, practical implementations. Identifying the specific impacts of climate policies on environmental outcomes is a difficult task. GHG emissions pervade industrial economies as the by-product of transportation, energy, and manufacturing processes. As a result, nearly every significant economic trend shifts carbon pollution patterns (Schleich et al. 2001; Peters et al. 2012). Moreover, the adoption of carbon pricing policy is endogenous: countries might introduce climate mitigation policies as their emissions are already falling (Downs et al. 2007). Consequently, most efforts to identify the effect of specific climate reforms on carbon pollution levels are either *ex ante* economic simulations or *ex post* sectoral impact analyses. These models excel at simulating how policy instruments will affect different sectors of the economy, identifying economic trade-offs, and exploring sector-specific policy effects. However, existing approaches struggle to evaluate the net causal effect of national policies because they compare realized outcomes to business as usual scenarios rather than counterfactual outcomes in the absence of the specific policy.

In this article, our contribution is to offer a national-level estimate of climate policy effectiveness without requiring assumptions about the pattern and shape of emissions trajectories, using the synthetic control method (SCM) (see also parallel SCM analysis by Bayer and Aklin (2020) on the European Union's Emissions Trading Scheme). SCM was developed to provide an empirically calibrated way of selecting comparison groups for policy impact analyses. It has since become a staple technique in policy impact analysis in fields such as comparative politics, economics, and criminology (Billmeier and Nannicini 2013; Costalli et al. 2017; Heersink and Peterson 2016; Robbins et al. 2017; Sills et al. 2015). The method offers a transparent and principled means of choosing comparison units that is blind to post-intervention outcomes; this means that researchers develop counterfactual scenarios without knowing how comparison group choice will shape their results.

Here, we evaluate the 2001 UK Climate Change Programme, a complex reform that included a carbon tax on large-scale energy users, industry-negotiated exemptions from the tax for meeting reduction targets, and a voluntary emissions trading scheme. The CCP was established in November 2000 to meet the Kyoto Protocol's EU-wide target of reducing emissions 8% by 2008–2012 compared to 1990 levels, including the country's more ambitious unilateral target of a 20% reduction by 2010, again compared to a 1990 baseline. The UK's CCP was one of the first comprehensive climate reform packages passed globally, in advance of action by most other OECD countries.

We leverage the synthetic control method to compare British emissions post-CCP to what would have happened if the policy had not been passed, rather than a stylized Business-As-Usual (BAU) or other benchmark scenario. We find evidence of substantial emissions reductions as a result of the policy: the UK's CO<sub>2</sub> emissions per capita were 9.8% lower relative to what they would have been if the CCP had not been passed.

SCM's ability to measure the causal effect of a complex, national climate policy contributes to debates over the potential efficacy of the current climate regime. Conventional accounts of global climate policy-making emphasize countries' weak incentives to act on climate change alone. Yet, we find that an early unilateral climate policy in the UK meaningfully reduced carbon pollution. The CCP was also effective despite the policy's hybrid nature (a combination of carbon pricing with negotiated industry agreements) and its substantial concessions to domestic polluters. Our findings thus provide evidence that even imperfect policy instruments can result in consequential reductions in national emissions.

## 1.1 Approaches to climate policy evaluations

Efforts to identify the effect of specific climate reforms on carbon pollution levels are typically *ex ante* economic simulations (Böhringer et al. 2005; Burniaux et al. 1992; Bruvoll and Larsen 2004; Svendsen et al. 2001; Agnolucci 2009; Hu et al. 2015) or *ex post* sectoral impact analyses (Ang et al. 2016; Martin et al. 2011; AEA Technology 2003). *Ex ante* approaches use Computable General Equilibrium (CGE) or Integrated Assessment models (IAM) to simulate the impact of a policy on a country's economy and environment. These models contain complex systems of equations that are stylistic representations of the relationships between different factors of production and agents in an economy, and (in the case of IAMs) the physical climate system. They are calibrated using historical data to reproduce the equilibrium state of an economy for a benchmark year. In general, CGE and IAM models can then compare a policy intervention against alternative reference scenarios that are chosen by the modeler, which often include scenarios of the form "climate stalemate" or total inaction, "Business-As-Usual" (BAU), or "optimal" scenarios where policies are implemented with welfare maximization (see for example Nordhaus 2013).

Such models are useful to understand how a policy instrument is expected to affect different sectors of the economy, to identify potentially important trade-offs, and to derive comparative statics. However, theoretical predictions on how a carbon policy is expected to perform cannot take into account institutional and political barriers that emerge during policy enactment and implementation. These models also reflect complex assumptions on functional forms and parameter values that lead to highly divergent predicted outcomes between different models (Pindyck 2017). For example, to generate BAU scenarios, modelers need to make assumptions about the growth rate of GDP, population, energy consumption elasticities (Böhringer et al. 2003), and (in the case of IAMs) environmental responses to these factors. Consequently, these models impose (often hidden) parametric assumptions on the hypothesized future emissions trajectories, leading some to criticize these approaches as akin to a "black box" (Böhringer et al. 2003; Pindyck 2017) where model runs are not always grounded in empirical or theoretical realities. Moreover, *ex ante* models are calibrated using historical benchmark data (Böhringer et al. 2003) which often rely on outdated economic snapshots. For example, the model used to generate a BAU scenario to compare the effectiveness of UK climate policy in the early 2000s was calibrated using input-output tables from 1995 (see Ekins and Etheridge 2006).

While CGE or IAM models offer clear advantages when conducting *ex ante* simulations about the general equilibrium effects of an exogenous policy treatment in comparison to a stylized reference scenario, the Business-As-Usual (BAU) scenarios they produce are not always appropriate to conduct *ex post* policy impact evaluations because those benchmarks are not clear counterfactuals for the policy outcome. In particular, the BAU assumption of no action whatsoever on climate is rarely the appropriate counterfactual to causally evaluate the effect of a climate policy. Rather, the counterfactual should be the potential outcome of carbon emissions in the absence of *that specific climate policy*.

Recognizing the weakness of these assumption-intensive counterfactuals, other analyses focus on *ex post* sectoral-level impacts rather than a policy's net capacity to decrease overall CO<sub>2</sub> emissions. The BAU scenarios in these cases are often rudimentary forward projections. For example, the consulting firm tasked by the UK government's Department for the Environment, Food and Rural Affairs (DEFRA) to estimate the results of the UK's climate change policies presents performance results as energy savings compared to what energy would have been used if sectors had produced the same throughput but at the energy-efficiency of a reference year (AEA Technology 2003, p. 13). Other studies use micro-level

data or case studies to estimate the impact of the CCP on businesses (Ang et al. 2016; Martin et al. 2011). However, the net national impact of a policy is the most important measurement with respect to climate change risk mitigation (Allen et al. 2009), and these methods don't allow for *ex post* assessment of this critical feature.

By contrast, synthetic control methods (SCM) allow for causal identification of the net national impact of a policy, offering a different form of *ex post* policy impact evaluation that supplements existing approaches. In general, the SCM has been referred to as “arguably the most important innovation in the policy evaluation literature in the last 15 years” (Athey and Imbens 2017, p. 10). While it is not possible to enumerate all of the possible drivers of CO<sub>2</sub> emissions in a given country and to specify how they interact, the synthetic counterfactual approach uses a diverse sample of countries to capture all of these latent trends in a way that does not require out-of-sample extrapolation (Abadie and Gardeazabal 2003, 2010, 2011, 2015).

This approach to causal identification of policy impacts is grounded within the potential outcomes framework (Holland 1986; Rubin 1974). Synthetic control methods borrow some elements from matching and difference-in-difference strategies. Matching is often used as part of selection-on-observables strategies, and aims to identify causal treatment effects by making the distributions of covariates that may impact an outcome as similar as possible between the treated and the control units. If the goal is to estimate the causal impact of some treatment *T* on some outcome *Y*, matching on some covariates *X* that also impact *Y* may help attenuate bias. However, in the presence of unobserved confounders *Z*, matching will not identify the causal effect of treatment. Difference-in-difference strategies exploit panel data to identify causal effects, and control for time-invariant confounders across treatment and control groups. In addition, they assume that time-varying confounders do not vary across treatment and control groups, often referred to as the “parallel trends” assumption. By contrast, SCM does not require us to make this assumption, and can accommodate time-varying unobserved confounders. The problem can also be restated as one of estimating a latent factor model, where a linear combination of time-varying trends (e.g., demand for energy) and time-fixed confounders drive a country's per capita emissions. The goal then becomes to capture the same combination of those confounders in the donor pool, in order to replicate the same factors driving the treated country's emissions. These confounders are then “differenced out” when we compare the emissions trajectories of the treated country and its synthetic control (Hazlett and Xu 2018; Xu 2017). We further explicate the synthetic control method in the Methods section.

## 1.2 The 2001 UK Climate Change Programme

Our empirical focus is an evaluation of the UK's 2001 Climate Change Programme (CCP), one of the first major reform packages passed by any OECD country. The CCP was established in November 2000 to meet the Kyoto Protocol's EU-wide target of reducing emissions 8% by 2008–2012 compared to 1990 levels, including the country's more ambitious unilateral target of a 20% reduction by 2010, again compared to a 1990 baseline.

The CCP included three interlocking policy instruments: first, a Climate Change Levy (CCL) on large-scale energy users (including the public sector); second, sector-wide Climate Change Agreements (CCA) negotiated between industry and government that discounted CCL rates if sectors hit pre-negotiated emissions reduction targets; and third, a voluntary unilateral emissions trading scheme (ETS).

The first of these components was the Climate Change Levy (CCL) which came into effect in April 2001. The CCL taxed the energy intensity of different fuel sources. It was

passed alongside a 0.3% reduction in employer National Insurance Contributions (NICs) and new renewable energy-oriented R&D funds. The CCL was not a pure carbon tax. While it did exempt most forms of renewable energy, it still included carbon-free nuclear energy. The CCL was levied on non-domestic consumers only, including the business and the public sectors.

The policy offered substantial producer flexibility through its second interlocking policy instrument, industry-level Climate Change Agreements (CCAs). CCAs exempted businesses from up to 80% of the levy if they agreed on voluntary carbon pollution reduction benchmarks.<sup>1</sup> By 2002, 44 sectoral associations had signed CCAs, including aluminum and steel (Bailey and Rupp 2005). Performance under these agreements was assessed at the sector level, but it was possible for individuals to continue under the program even if their broader sector failed to meet its target. Under the CCAs, industry could choose their own base years, which ranged from 1990 through 1999, and could set targets in different accounting “currencies” (i.e., relative energy: GJ primary energy per unit ton of production; relative carbon: tons of carbon per unit ton of production; absolute energy: GJ; absolute carbon: tons of carbon).

Additional producer-oriented flexibility was introduced with the April 2002 UK Emissions Trading Scheme (ETS), a voluntary program that allowed participants to trade emissions reduction permits relative to an absolute target baseline (the average of a participant’s 1998–2000 emissions); CCA signatories could then buy and trade these permit as insurance against failure to meet CCL carbon pollution reduction benchmarks. Conversely, sectors who over-complied with their CCA targets were able to sell their excess permits on the UK’s Emissions Trading Scheme.

In our online [Supplementary Information](#) (SI), we describe each of these policy components in more detail as part of a narrative history of UK climate policy-making from the 1980s through to 2015. We also detail the political controversy that accompanied the introduction of the CCP.

The CCP is particularly well-suited to synthetic control analysis. The UK was one of the first European countries to implement a comprehensive national climate reform package, and was the first country to unilaterally enact a domestic emissions trading scheme. With the exception of Northern European countries that enacted modest carbon tax systems in the early 1990s, most OECD countries had only implemented voluntary climate reforms up until 2005, when the EU emissions trading scheme began. This creates a window from 2001 through 2005 where domestic UK action largely stands alone against its peers. This allows us to construct of a credible counterfactual for the UK while avoiding possible policy diffusion effects from other countries.

## 2 Materials and methods

### 2.1 Causal identification using synthetic control methods

We use the Synthetic Control Method to generate a “synthetic UK” as a weighted average of other OECD, upper middle-, and high-income countries in our sample, or “donor pool.” Countries in the donor pool are selected through an algorithm so that the pre-CCP emissions

---

<sup>1</sup>If private sector actors failed to meet these negotiated benchmarks, they would be forced back into the CCL system for at least 2 years.

trajectories of the UK and of the synthetic UK match each other as closely as possible. We then evaluate the causal effect of the UK's Climate Change Programme by comparing the trajectory of emissions in the "synthetic UK" with the observed post-treatment emissions in the UK.

More formally, assume a sample of  $J + 1$  countries where  $j = 1$  corresponds to the treated UK, and  $J = \{2, \dots, J + 1\}$  is our donor pool. The intervention (i.e., the passage of the CCP) occurs at  $T_0 + 1$  and so the pre-intervention time periods are indexed by  $t = 1, 2, \dots, T_0$  and the post-intervention time periods are indexed by  $t = T_0 + 1, T_0 + 2, \dots, T$ . Let  $Y_{1t}^C$  represent the potential outcome under control for the UK, where  $j = 1$  indexes the UK. These are the potential CO<sub>2</sub> emissions in the UK if the CCP had *not* been passed. Let  $Y_{1t}^T$  represent the potential outcome under treatment; which are the potential CO<sub>2</sub> emissions in the UK if the CCP *had* been passed. The causal impact of the CCP is the difference between the two, and so our estimand of interest is  $\alpha_{1t} = Y_{1t}^T - Y_{1t}^C$ . However,  $Y_{1t}^C$  is unobserved.

Consider the following  $J \times 1$  vector  $\mathbf{W} = (w_2, \dots, w_{J+1})^T$  which contains the weights that reflect how much the  $j$ th candidate in the donor pool contributes to the synthetic counterfactual for the UK's emissions trajectory. These weights are restricted to be non-negative and sum to 1, that is,  $w_j \geq 0$  for  $j = 2, \dots, J + 1$  and  $\sum_{j=2}^{J+1} w_j = 1$ . This restriction on the weights is imposed in order to avoid extrapolating when constructing the synthetic counterfactual (Abadie et al. 2010, 2015).

Let  $\mathbf{X}_1$  be a  $K \times 1$  vector of the pre-treatment values of the  $K$  predictor variables of CO<sub>2</sub> emissions in the UK. The  $K \times J$  matrix  $\mathbf{X}_0$  contains the corresponding values of the pre-treatment values of explanatory variables for the  $J$  control countries. In our case, the  $K = 11$  attributes correspond to pre-treatment values of the outcome variable chopped up into discrete segments corresponding to CO<sub>2</sub> per capita emissions in each pre-treatment time period, respectively. Using a specification which includes all pre-treatment lags of the outcome variable has been recommended as the benchmark specification, unless researchers have strong theoretical priors on how other covariates affect the outcome (Ferman et al. 2020).

The pre-intervention characteristics of the synthetic UK will be given by  $\mathbf{X}_1^* = \mathbf{X}_0 \mathbf{W}^*$ . The optimal  $\mathbf{W}^*$  should thus be chosen so as to minimize the distance  $\|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|$ , in order to construct a synthetic counterfactual that best approximates the treated unit with respect to pre-treatment outcome values. In practice, the SCM implementation seeks a  $\mathbf{W}^*$  that solves  $\arg \min_{\mathbf{W}^*} \sqrt{(\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})^T \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})}$ .  $\mathbf{V}$  is a  $K \times K$  positive semi-definite, diagonal matrix of weights applied to the  $K$  variables that predict CO<sub>2</sub> emissions. Therefore, the loss function is a scalar. The implementation of the SCM by its authors (Abadie and Gardeazabal 2003) allows for the choosing of a custom  $\mathbf{V}$  weight matrix. This can be a fruitful approach if we possess a priori knowledge on the relative predictive power of different explanatory variables. However, in the absence of strong priors, we follow Abadie and Gardeazabal (2003) and Abadie et al. (2011) and adopt a data-driven approach whereby the matrix  $\mathbf{V}$  is the one that minimizes the mean square prediction error (MSPE) of the pre-treatment outcome variable, i.e., such that the average squared discrepancies between the pre-treatment CO<sub>2</sub> emissions of the UK and of the synthetic UK are minimized. A numerical optimization algorithm is used to solve for these optimal weights.<sup>2</sup>

<sup>2</sup>We use the *Synth* package in R with the default optimization methods of Nelder-Mead and BFGS.

Finally, the observed emissions (pre- and post-treatment) of the UK are collected in a  $T \times 1$  matrix  $\mathbf{Y}_1$ . The CO<sub>2</sub> emissions of the countries in the donor pool are recorded in a  $T \times J$  matrix  $\mathbf{Y}_0$ . The emissions of the synthetic UK are simulated as  $\mathbf{Y}_1^* = \mathbf{Y}_0 \mathbf{W}^*$ . The estimated treatment effect is thus given by  $\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_{jt}^* Y_{jt}$ .

Causal identification is achieved using SCM under less restrictive conditions than difference-in-difference strategies. First, there can be no treatment spillover to other countries in the donor pool. Although the authors of the SCM approach do not explicitly refer to this assumption as such, this assumption is the stable unit treatment values assumption, or SUTVA, which states that “[t]he potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes” (Imbens and Rubin 2015, p. 10). Second, to avoid interpolation bias, variables used to form the weights must be within the same support of the data for the treated unit and countries in the donor pool (Abadie et al. 2010, 2015). In other words, the variables used to form the weights must have values for the donor pool countries that are similar to those of the UK. This is because interpolation biases may be severe if the procedure interpolates across different regions with very different characteristics (Abadie et al. 2010).

In general, the UK during the early CCP era satisfies these conditions. The UK is the only country to be treated by the CCP in 2001, and is the only country in the sample that passed major climate legislation until the European Union launched its emissions trading scheme (EU ETS) in 2005. Our dependent variable is operationalized as CO<sub>2</sub> emissions per capita, which ensures that the outcome variable across regions is broadly on the same order of magnitude and thus avoids interpolation bias. Moreover, alternative specifications provided in SI Section G also achieve a restriction of the data to a common support for all countries in the sample by employing a rescaled dependent variable (e.g., relative to a 1990 and a 2000 baseline, respectively). Running the synthetic control estimator on absolute CO<sub>2</sub> emissions levels is not appropriate given the variance in emissions levels across countries.

## 2.2 Data sources and sample selection

To implement the synthetic control method, we use data on CO<sub>2</sub> emissions and CO<sub>2</sub> emissions per capita from the World Bank’s World Development Indicator (WDI) database, extracting indicators “EN.ATM.CO2E.KT” (CO<sub>2</sub> emissions in kilotons) and “EN.ATM.CO2E.PC” (CO<sub>2</sub> emissions per capita in metric tons), respectively. The CO<sub>2</sub> emissions measured are those stemming from the burning of fossil fuels and the manufacture of cement. We impute some missing data for Germany, Kuwait, and Liechtenstein using alternate data sources. This procedure is described in the online SI Section A.

We define our donor pool as the 51 countries which were either OECD members or classified by the World Bank as upper middle-income or high-income countries at the time of treatment in 2001, that had a population greater than 250,000, and that did not have a carbon pricing policy in place. The World Bank classifies countries into income categories according to GNI per capita in US\$. In fiscal year 2001, the World Bank classified high-income (HIC) countries as those with GNI per capita above 9265 US\$, and upper middle-income (UMC) countries as those with GNI per capita in the 2996 US\$ to 9265 US\$ range. In 2001, there were 47 high-income countries, 38 upper middle-income countries, and 30 OECD countries. Our donor pool is the union of those sets, minus countries for which data is missing or countries that were deemed “treated” in 2001, and minus countries with a very small population.



We determine whether countries in the sample were “treated” by building on the World Bank’s *State and Trends of Carbon Pricing 2019* report (World Bank 2019), albeit with some modifications. Even though the World Bank report notes that Poland had passed a carbon tax in 1990, we do not consider it “treated” until 2005 (the start of the EU ETS) because the Polish tax was so small in scope and incidence that it cannot be considered a materially important carbon pricing policy. Indeed, the Polish carbon tax of 1990 was less than 1 US\$ per ton CO<sub>2</sub>e and covered only 4% of the jurisdiction’s emissions (World Bank 2019).

Moreover, we consider the Netherlands to be “treated” in 2001, even though the World Bank report does not consider the Netherlands as having a carbon tax. However, the Netherlands introduced a tax on energy in 1996, which complemented a tax on fuel that came into force in 1992. Tax rates were set as a function of CO<sub>2</sub> per energy content, and were estimated to be around NLG 30 per metric ton of CO<sub>2</sub> (Hoerner and Bosquet 2001, p. 20).

The countries that were “treated” in 2001 were thus the following: Denmark (carbon pricing policy first passed in 1992), Estonia (2000), Finland (1990), Netherlands (1992), Norway (1991), Slovenia (1996), and Sweden (1992). These countries are excluded from the donor pool.

### 2.3 Specifications

In the main specification we report below, we construct this synthetic UK from a donor pool of countries that were either OECD, upper middle-, or high-income countries in 2001. We exclude small countries with a population less than 250,000 in 2001 since these may have different fundamental drivers of CO<sub>2</sub> emissions than the UK. Not all countries in this donor pool contribute equally to this synthetic control. In our main specification, 8 countries make up the effective sample (see Fig. S1 in the SI) accounting for 88% of the weights, with the other countries having weights of less than 1%. In the SI’s Fig. S2, we also display the CO<sub>2</sub> per capita emissions of the donor countries in the effective sample. In this specification, which generates the strongest pre-treatment fit and performs best according to diagnostics reported in the Findings section and in SI Section G, the counterfactual trend is estimated using a blend of 19% Poland, 19% Libya, 18% Bahamas, 16% Belgium, 6% Trinidad and Tobago, 5% Uruguay, 4% Luxembourg, and 1% Brunei. Here, the pre-treatment MSPE achieved with that donor pool was  $1.24 \times 10^{-4}$ . Figure S1 in the online SI displays the weights applied to each country in the donor pool.

The fact that surprising countries, such as the Bahamas and Libya are part of the top donors, while an intuitively similar country like France is at the bottom should not be cause for concern. Rather, it suggests that there were latent, unobserved forces driving British emissions, and that a weighted combination of these forces was found in the top donor countries. Specifically, the synthetic control approach estimates a latent factor model with a linear combination of time-varying and time-invariant confounds. Some combination of the unobserved factors responsible for driving British emissions was also present in donor countries, which are then re-weighted to create a credible control for the UK.

Instead, an advantage of this effective donor pool is that it rules out spatial spillover effects.<sup>3</sup> One of the assumptions required for causal identification is that the treatment affected the treated unit only and did not spillover to other control units (the SUTVA assumption). Since the UK’s untreated neighbors such as France and Germany are not part

<sup>3</sup>We thank an anonymous reviewer for making this point to us.



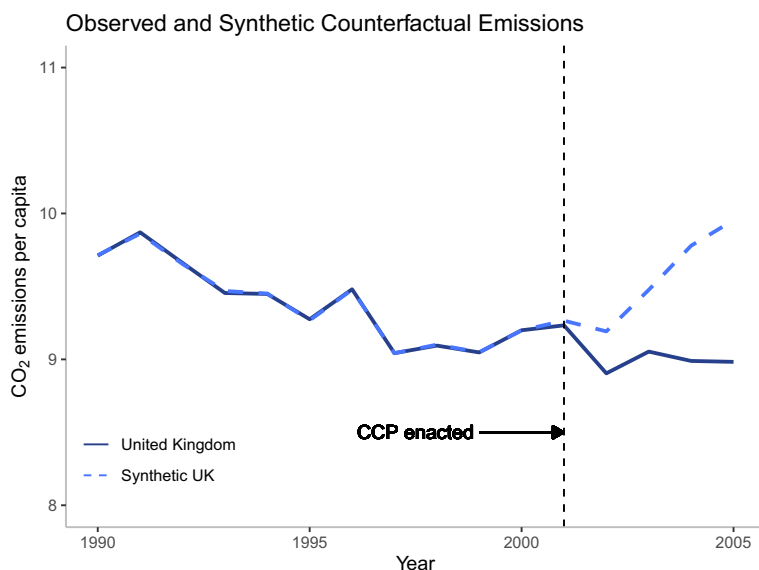
of the effective sample of countries used to generate the synthetic control, our results are not at risk of over-estimating the treatment effect of the CCP due to a violation of the SUTVA assumption.

As a robustness check, we also evaluate specifications generated by progressively smaller donor pools, again applying population filters: (1) on countries that were either OECD members or high income countries in 2001; and (2) on countries that were OECD members in 2001. The pre-treatment MSPE increases (indicating a poorer fit between the UK and the synthetic UK) as the donor pool decreases: from  $5.24 \times 10^{-4}$  (donor pool consisting of 2001 OECD and HIC countries) to  $2.13 \times 10^{-3}$  (donor pool consisting of 2001 OECD members). However, despite these specifications being slightly weaker from a SCM perspective, they still generate similar estimates of the effect of the UK policy (see section G in the SI). In this way, while we choose our specification in a principled way based on synthetic control method best practices, our results hold even for a range of donor pools that rely only on countries with substantively similar political and economic systems.

Generally, there are a multitude of observed and unobserved factors, both dynamic and constant in time, that drive British emissions in ways that are hard to specify a priori. Attempting to specify a functional form that would accurately reproduce the emissions trajectory of the UK is a difficult task. The advantage of the SCM is that it enables us to sidestep the need to enumerate all of the structural drivers of CO<sub>2</sub> emissions. By contrast, we employ a non-parametric approach where we find the combination of (latent) drivers in donor countries that serve as an appropriate control by numerically minimizing the distance between the pre-treatment trends of the UK and the control.

The predictor variables used to construct a synthetic UK are the pre-treatment values of per capita CO<sub>2</sub> emissions from 1990 to 2000, with no other covariates. Other covariates might be useful to improve the match between the UK's pre-CCP emissions and its synthetic counterpart. In Section G.2 of the SI, we show this was not the case, and therefore we report our estimates using pre-intervention values of the dependent variable only. Kaul et al. (2018) show theoretically that using all pre-treatment values of the outcome variable as separate predictors in the SCM algorithm leads to an optimization procedure that renders all other covariates irrelevant. We verify empirically that this is the case: specification 2 in our SI uses 4 covariates as predictors (GDP per capita, renewable energy consumption, fossil fuel energy consumption, and energy use per capita), in addition to the pre-treatment values of per capita CO<sub>2</sub> emissions. The weights on the 4 covariates when constructing the synthetic UK are all 0.

We construct our synthetic UK on the basis of the lagged values of CO<sub>2</sub> emissions per capita alone for three reasons. First, doing so leads to an optimal pre-treatment fit between the UK and its synthetic control. Since the goal of SCM is to create a credible counterfactual for the treated unit in the absence of treatment, a guiding heuristic is to choose the specification that minimizes the distance in potential outcomes pre-treatment. Second, this research design choice minimizes the risk of specification searching on the part of researchers. Ferman et al. (2018, 2020) suggest that despite the advantage of the transparency of the SCM, researchers have some latitude to engage in specification-searching. By restricting our choice set to specifications that only include pre-treatment values of the outcome variable, we tie our hands at the outset. Third, we do not have strong theoretical priors on the types of covariates that would capture most of the drivers of British CO<sub>2</sub> emissions. While we may account for observable characteristics that correlate with the outcome, such as income per capita, this is by no means a guarantee that we would account for the *unobservable* characteristics that determine the pattern of emissions. Ferman et al. (2020) address this problem and recommend that in the case where researchers do not have



**Fig. 1** Observed and synthetic counterfactual per capita emissions for the UK. The solid line represents actual emissions trajectory. The dashed line represents the emissions trajectory of a synthetic UK, in the absence of the country's Climate Change Programme. Treatment occurred in 2001

strong theoretical priors on the covariates to use, a specification which uses all pre-treatment lags of the outcome variable should be used and reported as the benchmark specification. Nevertheless, as a robustness check, we also estimate the treatment effect using alternative specifications, which we report below and in further detail in online SI Section G.

### 3 Findings

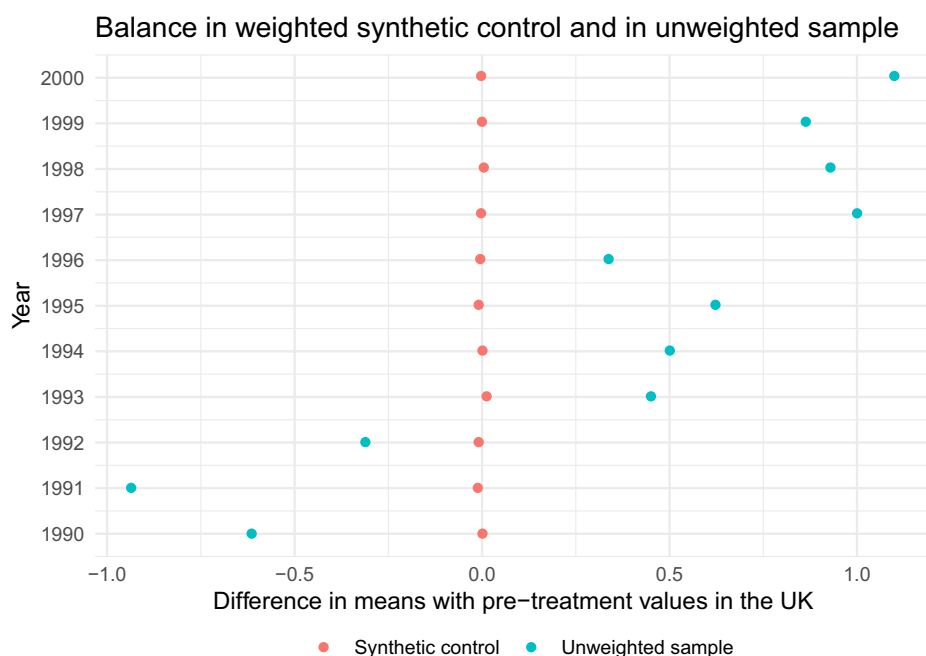
#### 3.1 Treatment effect of the CCP

We first construct a synthetic UK as a weighted average of the pre-treatment characteristics of countries in the donor pool, where weights are chosen so as to minimize the distance between the UK and its synthetic counterpart. The solid line in Fig. 1 displays the observed CO<sub>2</sub> emissions per capita path of the UK: the emissions trajectory remained relatively flat post-treatment. The dashed line represents the UK's emissions trajectory had the country *not* passed its 2001 reform, as estimated by SCM.

From 1990 to 2001, the difference in means between the pre-treatment CO<sub>2</sub> emissions of the UK and of the synthetic UK is statistically indistinguishable from 0 ( $p = 0.981$ ).<sup>4</sup>

Figure 2 displays the difference between these pre-treatment CO<sub>2</sub> emissions in the UK and the weighted means and unweighted means, respectively. It indicates that the synthetic control achieves pre-treatment balance with the treated unit.

<sup>4</sup>During the pre-treatment period, average emissions per capita in the entire sample were 6.3% lower than those of the UK's until 1992, and from 1993 onward they were 7.8% higher than those of the UK's. An unweighted sample is thus not an appropriate counterfactual for the UK.



**Fig. 2** Difference in means in pre-treatment values observed in the UK and those estimated by the synthetic control (in orange) which is a weighted sample of the donor pool comprised of OECD, high-, and upper middle-income countries. Blue points represent the difference in means in pre-treatment values observed in the UK and those observed in the same donor pool sample, but unweighted

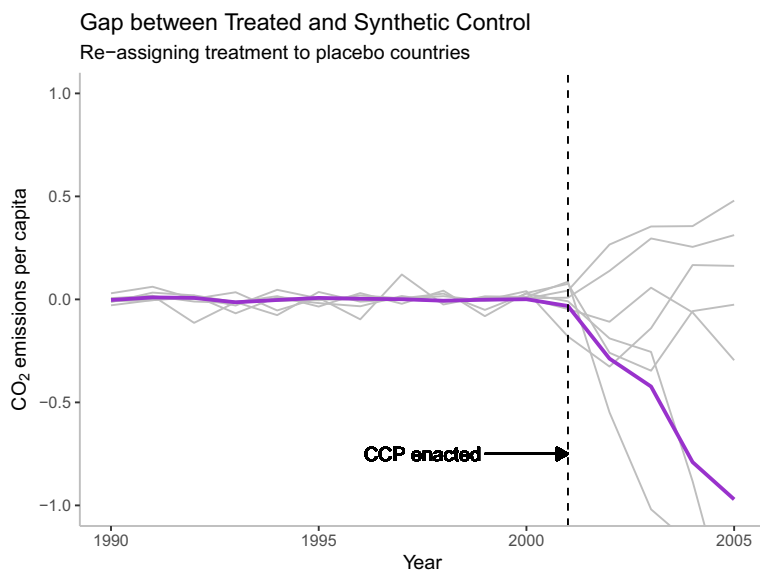
However, after the 2001 passage of the Climate Change Programme, synthetic counterfactual emissions and observed CO<sub>2</sub> emissions start to diverge. The causal impact of the CCP can then be estimated as the difference in per capita emissions between the UK and the synthetic UK in the post-treatment period. By 2005, four years after the policy's passage, we estimate a treatment effect of -9.8% emissions per person in 2005. This is equivalent to a reduction of 148 Mt CO<sub>2</sub> during the period 2002–2005, an average annual reduction of 0.6 tons of CO<sub>2</sub> per capita. We do not estimate the causal impact of the CCP after 2005, since this corresponds to the launch of the EU-wide emissions trading scheme. After 2005, many countries in the donor sample are “treated” with comprehensive climate reform, and no longer act as appropriate donor countries.

We discuss the logic of our donor pool in the Methods section. However, it is important to (1) verify that our results are not dependent on the inclusion of certain countries in the donor pool, and (2) to re-run the synthetic control estimator on a donor pool of countries that have similar political and economic institutions as the UK. First, we run a “leave-one-out” robustness check that is detailed in the section below. We show that the findings are not dependent on the inclusion of any single country in the donor pool. Second, we also run the specification on a donor pool composed of 22 OECD countries that share institutional similarities with the UK. The top donors in this case are France (0.353), Japan (0.329), Belgium (0.123), Germany (0.099), Luxembourg (0.066), and Italy (0.018). The treatment effect attenuates slightly from -9.8% per capita emissions in 2005 to -5.3% per capita emissions, but retains statistical significance ( $p < 0.05$ ). More details on this robustness check are provided as SI Section G.5.

### 3.2 Statistical inference

After estimating the treatment effect of the CCP on British emissions, we then ask whether our results are statistically significant, rather than the product of chance. Since SCM does not assume a data-generating process, nor do we estimate a specific functional form, we accomplish this through the use of falsification or placebo tests, rather than through parametric hypothesis testing. Placebo tests are commonly used in the literature to test whether an outcome or a unit that we know to be unaffected by treatment responds to a placebo treatment, in which case any positive treatment effect on the treated might be spurious (Bertrand et al. 2004; Abadie et al. 2010). To conduct our placebo analysis, we iteratively re-assign treatment to all countries in the donor pool. Since we know these countries were not treated, we should expect to see null treatment effects, other than by chance. The estimated treatment effect is given by the difference between the placebo unit and its synthetic control in post-treatment periods. This allows us to create a null distribution of gaps in post-treatment emissions trajectories for all countries in the sample. If the results in the UK are not driven by chance, we should expect the gaps in the post-CCP emissions trajectories in the UK to lie in the tails of that null distribution. This procedure is similar to testing Fisher's sharp null hypothesis, which tests a null hypothesis of no effect whatsoever (Imbens and Rubin 2015).

However, it may be the case that the pre-treatment fit between a placebo unit and its synthetic control is poor. In this case, this particular placebo test is uninformative, since synthetic control estimators hinge on finding weights that minimize the distance in pre-treatment emissions trajectories. When the fit is poor, it is unlikely that the resulting synthetic counterfactual provides a credible control for the treated unit (placebo or otherwise). We thus exclude placebo countries with a pre-treatment MSPE greater than 30 times the pre-treatment MSPE of the UK in Fig. 3. However, the choice of cut-off for the



**Fig. 3** Gaps in emissions per capita between the treated unit and its synthetic counterpart. The thick purple line represents the gaps for the UK. The grey lines represent the distribution of placebo treatment effects. Countries with a pre-treatment MSPE greater than 30 times that of the UK have been excluded (see Methods for details)

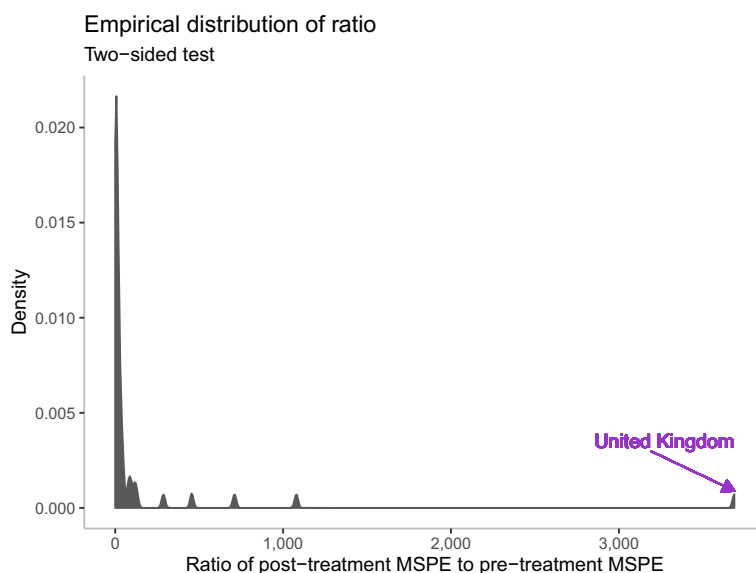
treatment MSPE is rather arbitrary. We also provide figures in SI Section D of the gaps between the treated unit and its synthetic control with cut-offs for excluding placebo runs that have a pre-treatment MSPE greater than 50 and 100 times that of the UK's for illustration.

Figure 3 displays the results of iteratively re-assigning treatment to countries in the donor pool (minus the UK). The purple line displays the gaps between the emissions in the UK and in the synthetic UK. The grey lines represent the gaps in emissions between each placebo unit and its synthetic counterpart. Only placebos with high-quality pre-treatment counterfactuals are informative to evaluate whether the treatment effect of the CCP is robust to a falsification test. Thus, Fig. 3 only includes placebos whose pre-treatment MSPE is not more than 30 times greater than that of the UK's. The causal effect of the CCP in the UK lies at the edge of this null distribution. In other words, we would be unlikely to see a treatment effect as large as we see for the UK by chance alone.

Since we know that none of the placebo countries had a climate policy, we should expect null treatment effects on each of these placebo treatments, as only the UK was treated with the Climate Change Programme in 2001. The donor pool includes countries that were Annex I parties to the United Nations Framework Convention on Climate Change (UNFCCC) in 1992. To the extent that Annex I membership might constitute a shadow treatment on these countries, this will bias against finding an effect; and our estimates can thus be seen as a lower-bound on the treatment effect of the CCP. After iteratively assigning a placebo treatment to countries in the donor pool, we then calculate the gaps in emissions between the placebo units and their synthetic controls. We should expect to see little to no variation in these post-2001, other than by chance.

It may be the case that the synthetic control algorithm on a placebo unit failed to achieve a good pre-treatment fit, in which case this placebo run would be uninformative. We account for this by calculating the mean squared prediction error (MSPE), which is the average of the squared gaps between the per capita CO<sub>2</sub> emissions in the treated unit and its synthetic control. If the fit achieved by the synthetic control algorithm was good, then we should expect a low pre-treatment MSPE; and conversely, if the fit was poor, the pre-treatment MSPE for any given country would be larger. If a country (placebo or the UK) has a large MSPE post-treatment, this is suggestive of a large treatment effect. We compute the ratio of the post- to pre-treatment MSPE for the UK and each placebo country in the sample, as recommended by Abadie et al. (2010, 2011, 2015). By dividing the post-treatment gaps with the pre-treatment gaps, the statistic downweights the ill-fitting synthetic controls. This effectively penalizes the treatment effect when the fit achieved by the synthetic control algorithm is poor. The ratio of post- to pre-treatment MSPE for all countries in the donor pool is the statistic that we use to create a non-parametric null distribution.

We can then look at the empirical distribution of this statistic to ascertain whether the ratio of post- to pre-treatment MSPE in the UK falls in the tails of this distribution, which would indicate that the results in the UK are unlikely to be driven by chance. When we re-assign treatment to all countries in the sample, we find that the UK has the largest ratio statistic. If we were to pick a country at random under uniform sampling from the entire sample, the probability of obtaining a ratio statistic as large as the UK's is  $1/51 \approx 0.02$ . In other words, the probability of obtaining a treatment effect as large as the UK's would be 0.02, which is conventionally seen as statistically significant for parametric analyses. Figure 4 displays the empirical distribution of this ratio statistic: this is our null distribution. The UK's ratio statistic is approximately 3687, and it falls in the right tail of that distribution, which suggests that we can reject the null hypothesis that the CCP had no effect in favor of the alternative hypothesis that the CCP had an effect on emissions per capita.



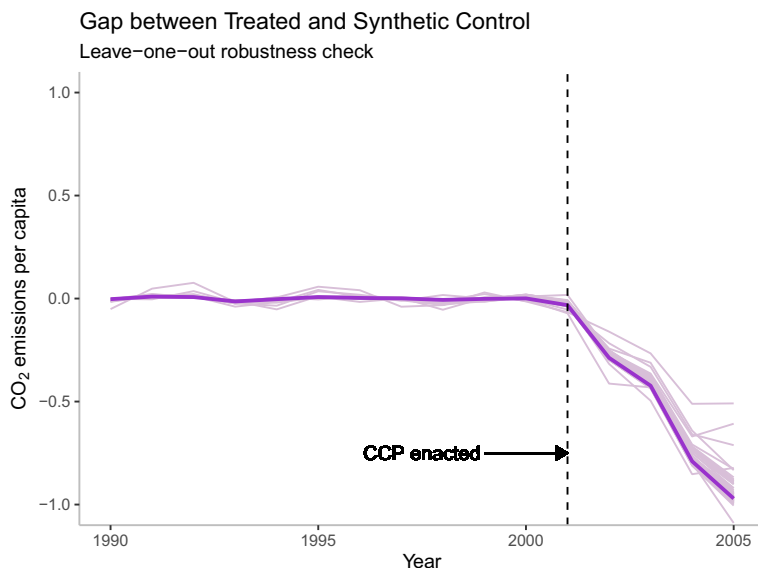
**Fig. 4** Null distribution for a two-sided test. The density represents the empirical distribution of the ratio statistic (computed as the ratio of post- to pre-treatment mean square prediction error) for all countries in the sample

### 3.3 Robustness checks

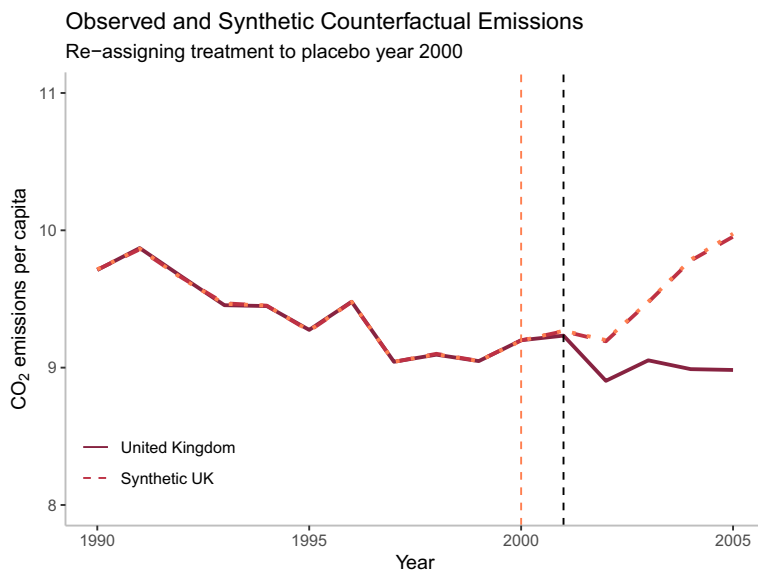
Finally, we conduct additional checks to verify that are results are robust. These include “leave-one-out” robustness checks where we iteratively drop a single country from the donor pool to ensure that our results are not an artifact of individual donor countries, placebo “in time” tests where we re-assign treatment to earlier years, and a series of alternative specifications for synthetic control construction.

First, we might ask whether the weights in the synthetic UK are driven by certain countries in particular. To test this, we conduct a “leave-one-out” robustness check where we iteratively drop a single country at a time from the donor pool used to construct the synthetic UK. This allows us to check that the emissions trajectory of the synthetic UK is not driven by a single country, and that achieving balance between the pre-CCP emissions trajectories of the UK and its synthetic control does not depend on the inclusion of a single country. As shown by Fig. 5, our results remain robust to the omission of single countries from the donor pool.

Second, we run placebo “in time” tests, where we re-assign treatment to previous years. Since we know that treatment occurred in 2001, and not earlier, we should not expect to find a large divergence between the UK and its synthetic control in those placebo years, other than by chance. Figure 6 displays the results of this test for the year immediately preceding the passage of the CCP. The emissions trajectory of the synthetic control for the placebo year 2000 do not start diverging from those of the UK until after 2001 and not earlier, which further reinforces the impression that there indeed was a structural break in emissions after the treatment. Additional placebo tests for other years prior to treatment can be found in SI Section E.



**Fig. 5** Gaps in per capita emissions between the UK and the synthetic UK. The thick purple line represents the gaps when the synthetic UK is constructed using all countries in the donor pool (51 countries). Each thin purple line represents the gaps when one country is dropped from the donor pool



**Fig. 6** Observed and synthetic counterfactual emissions for a placebo run where treatment occurs in 2000. The synthetic control's emissions trajectory for the placebo year 2000 is in the dashed orange line



Third, we run the synthetic control procedure on a variety of alternative specifications and samples. We considered two alternative ways to operationalize the outcome variable: CO<sub>2</sub> emissions rescaled to a 1990 baseline, and CO<sub>2</sub> emissions rescaled to a 2000 baseline. Both dependent variables are rescaled to ensure that they are within the common support of the data. The first outcome variable is rescaled to the baseline used in the formulation of the Kyoto targets, and can help us visualize at a glance the extent to which the UK met its targets. The second dependent variable can then help us understand the immediate impact of the CCP at  $t + 1$ .

We also consider three samples for the donor pool: (A) countries that were either OECD, high-, or upper middle-income countries in 2001; (B) countries that were either OECD or high income countries in 2001; and (C) countries that were OECD members in 2001. In all of these samples we exclude Northern European countries that we consider to have been treated by 2001.

We report the results run on donor pool (A) as our preferred model, but the results run on donor pools (B) and (C) are also statistically significant. However, the smaller donor pool sample means that achieving a good pre-treatment fit between the UK and its synthetic counterpart is dependent on the inclusion of a single country, Luxembourg. This is not a problem when we use the larger donor pool (A): if we drop Luxembourg, the treatment effect is comparable (-8.5%) and is statistically significant ( $p = 0.02$ ).

Finally, we also run the synthetic control method on the main donor pool sample (A) using a specification that includes covariates (specification 2), and one that increases the pre-treatment optimization period to 1980 (specification 3). The treatment effect of the CCP is substantively large and statistically significant in both of those cases too.

Table 1 summarizes all the specifications that have been run as a robustness check on our results. The detailed results for our alternative specifications can be found in the online SI Section G. As our main finding, we choose to report a specification where the outcome variable is CO<sub>2</sub> emissions per capita, rather than emissions rescaled to a baseline, since per capita emissions are a meaningful and readily interpretable measure of climate abatement. Within the specifications that have per capita CO<sub>2</sub> emissions as their outcome variable (specifications 1–5), we choose the specification that achieves the best pre-treatment fit (i.e., the lowest pre-treatment MSPE), which occurs when the donor pool comprises countries that were either OECD, high-, or upper middle-income countries in 2001.

Best practice in SCM analysis is to report several specifications as a robustness check (Ferman et al. 2018, 2020). Ferman et al. (2020) discuss how to approach generating a valid hypothesis test that encompasses all the different specifications. On the one hand, a decision rule that rejects the null hypothesis of no effect only if all the specifications individually reject the null would be unduly conservative, though it should be noted that our results would pass that test (at a 10% significance level). On the other hand, a decision rule that rejects the null if at least one specification has rejected the null would inflate the rate of false positives. They thus suggest to generate a new test statistic, inspired by work by Imbens and Rubin (2015): for each unit  $j$  and across all specifications  $s$ , compute the ratio of post- to pre-treatment MSPE, and compute p-values using the same statistical inference procedure as before.

We compute such a test statistic across specifications that share the same donor pool. For all 3 donor pools, these omnibus p-values are highly statistically significant—pool (A):  $p = 0.0385$ ; pool (B):  $p = 0.0303$ ; pool (C):  $p = 0.0435$ . This indicates that our findings are not the result of a single spurious specification; we can thus reasonably conclude that the CCP had a significant and negative effect on British per capita CO<sub>2</sub> emissions.

**Table 1** Summary of alternative specifications

Spec	Outcome variable	Donor pool	Covariates	Optimization	Pre-treatment MSPE	p-value
1	CO <sub>2</sub> per capita	(A) OECD, high & upper middle income <i>n</i> = 51	No	1990–2001	$1.24 \times 10^{-4}$	0.020
2	CO <sub>2</sub> per capita	(A) OECD, high & upper middle income <i>n</i> = 37	Yes	1990–2001	$2.07 \times 10^{-3}$	0.053
3	CO <sub>2</sub> per capita	(A) OECD, high & upper middle income <i>n</i> = 51	No	1980–2001	$7.86 \times 10^{-3}$	0.058
4	CO <sub>2</sub> per capita	(B) OECD & high income <i>n</i> = 32	No	1990–2001	$5.24 \times 10^{-4}$	0.030
5	CO <sub>2</sub> per capita	(C) OECD <i>n</i> = 22	No	1990–2001	$2.13 \times 10^{-3}$	0.043
6	1990 baseline	(A) OECD, high & upper middle income <i>n</i> = 51	No	1990–2001	$6.42 \times 10^{-6}$	0.038
7	1990 baseline	(B) OECD & high income <i>n</i> = 32	No	1990–2001	$1.07 \times 10^{-5}$	0.030
8	1990 baseline	(C) OECD <i>n</i> = 22	No	1990–2001	$3.12 \times 10^{-5}$	0.043
9	2000 baseline	(A) OECD, high & upper middle income <i>n</i> = 51	No	1990–2001	$8.69 \times 10^{-6}$	0.038
10	2000 baseline	(B) OECD & high income <i>n</i> = 32	No	1990–2001	$1.42 \times 10^{-5}$	0.030
11	2000 baseline	(C) OECD <i>n</i> = 22	No	1990–2001	$3.75 \times 10^{-5}$	0.043

Specification 1 is reported as the main finding. Details on alternative specifications including robustness checks are reported in the SI Section G. The p-values reported here are two-sided

## 4 Discussion

Collectively, national climate policies remain insufficient to mitigate the catastrophic risks of climate change (Peters et al. 2015). However, we show that a unilateral climate policy in the UK meaningfully reduced carbon pollution, even in the absence of a legally binding global climate treaty. Conventional accounts of global climate policy-making emphasize

countries' weak incentives to act on climate change alone. Yet, we show that the UK reduced its per capita carbon pollution by 9.8% in the face of free-riding disincentives to act.

The CCP included a mix of several policy instruments: a type of carbon tax (the Climate Change Levy collected from industry and the public sector), negotiated industry agreements (the so-called Climate Change Agreements), and a domestic emissions trading scheme (ETS). These policies individually had several shortcomings which cast doubt on the CCP's ability to achieve substantial emissions reductions. In particular, empirical evidence suggests that the Climate Change Agreement (CCA) targets negotiated with industry were too lax at the outset (Ekins and Etheridge 2006), which would have resulted in "hot air" on the emissions trading scheme (ETS) market. The CCL was not a pure carbon tax and carbon-free nuclear energy was not exempt from it. The Climate Change Agreements were negotiated with industrial polluters and made substantial concessions to producers. Sectors who overcomplied on their CCA targets could sell those surplus emissions as allowances on the UK's domestic ETS, and conversely sectors could meet their CCA targets by purchasing permits on the market. These provisions introduced additional flexibility for business managers who could decide on the least-cost way to meet their CCA targets.

While the CCA targets themselves were lax, the CCA sectors outperformed their 2002 targets. Ekins and Etheridge (2006) argue that this was due to an "awareness effect": there were many cost-effective opportunities to improve energy efficiency that had previously not been recognized by industrial business managers. The excise rates of the CCL were high enough to be considered a credible threat and succeeded in bringing industrial actors to the table to negotiate the voluntary CCA targets, and it was this process which allowed the private sector to realize that there were low-hanging fruit energy efficiency gains to be made (Ekins and Etheridge 2006). Many of those energy improvements were made on financial grounds alone, and the fact that the targets were not stringent was counterbalanced by the process of learning from industrial managers about how energy efficiency could improve their bottom line. These findings provide suggestive evidence that a combination of imperfect policy instruments can result in meaningful emissions mitigation.

SI Section H provides an additional narrative of the mobilization against the CCP by both labor and industry groups which succeeded in watering down the stringency of the policy and resulted in important concessions to polluters. Still, despite regulatory capture by industry, and even if it was voluntary and unilateral, the CCP was nevertheless able to abate 148 Mt of CO<sub>2</sub> over 4 years, or around 37 Mt of CO<sub>2</sub> per annum. The IPCC estimates that mitigation pathways that keep warming within 1.5 °C would cap emissions in 2030 to 25–30 Gt CO<sub>2</sub>e per year (Rogelj et al. 2018). Our results suggest that the UK was able to mitigate emissions on the order of magnitude of 0.5% of the *global* annual carbon budget remaining in 2030.

Finally, even though evaluating the overall impact of a given climate policy on national-level carbon emissions is crucial for the development of climate budgets, existing efforts are stifled by the reliance on unrealistic BAU scenarios. BAU scenarios used for causal impact evaluations need to be developed with the explicit aim of being counterfactual. CGE and IAM models are useful for *ex ante* simulations of the general equilibrium effects of an exogenous policy on the economy and on the environment. However, the BAU scenarios that are used by these models as comparisons are not necessarily appropriate for an *ex post* policy impact evaluation. This is because the correct counterfactual to estimate the impact of a climate policy is a scenario where the policy had *not* been passed, and not a baseline of no action or other stylized vignette. However, it is difficult to enumerate all the possible drivers of that counterfactual emissions trajectory, and furthermore to specify how they interact with each other. We demonstrate the advantage of using a non-parametric approach which

obviates the need to specify a functional form for all of the (observed and unobserved) drivers of emissions. The synthetic control estimator captures the specific combination of underlying dynamic and static structural drivers of British emissions in the control units and reweights them accordingly to create a credible synthetic control.

Alongside parallel work by Bayer and Aklin (2020), our findings show the promise of synthetic control methods as a tool for *ex post* climate policy impact analysis that can provide net national estimates of CO<sub>2</sub> abatement without relying on simplistic forward projections of emissions. More accurate climate policy evaluations can in turn inform the analysis of national and global carbon budgets, which form the basis of actionable goals for climate stabilization.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10584-021-03111-2>.

**Acknowledgements** We thank Alex Franks, Chad Hazlett, Paasha Mahdavi, Kyle Meng, Vítor Possebom, Matthew Potoski, Sam Rowan, and Vincent Thivierge, in addition to participants of the Environmental Politics Workshop at UC Santa Barbara, of the 2019 Duck Family Workshop on Environmental Politics and Governance at University of Washington, and of the 5th Annual Conference on Environmental Politics and Governance for helpful comments on earlier drafts of this manuscript.

**Author contribution** AL and MM conceptualized the study. AL collected the data, designed the data analysis, analyzed the data, and generated the results. AL and MM wrote the paper.

**Availability of data and material** The datasets generated and analyzed during the current study are available in the Zenodo repository, <http://doi.org/10.5281/zenodo.4566804>.

**Code availability** The code files to replicate the results of the study are available in the GitHub repository, <https://github.com/walice/synth>.

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abadie A, Gardeazabal J (2003) The economic costs of conflict: a case study of the Basque Country the economic costs of conflict. *Am Econ Rev* 93(1):113–132
- Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 105(490):493–505
- Abadie A, Diamond A, Hainmueller J (2011) Synth: an R package for synthetic control methods in comparative case studies. *J Stat Softw* 1(13):1–17
- Abadie A, Diamond A, Hainmueller J (2015) Comparative politics and the synthetic control method. *Am J Polit Sci* 59(2):495–510

- AEA Technology FES (2003) Climate change agreements: results of the first target period assessment (version 1.2, revised July 2004). Harwell
- Agnolucci P (2009) The effect of the German and British environmental taxation reforms: a simple assessment. *Energy Policy* 37(8):3043–3051
- Allen MR, Frame DJ, Huntingford C, Jones CD, Lowe JA, Meinshausen M, Meinshausen N (2009) Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* 458(7242):1163–1166
- Ang CP, Toper B, Gambhir A (2016) Financial impacts of UK's energy and climate change policies on commercial and industrial businesses. *Energy Policy* 91:273–286
- Athey S, Imbens GW (2017) The state of applied econometrics: causality and policy evaluation. *J Econ Perspect* 31(2):3–32
- Bailey I, Rupp S (2005) Geography and climate policy: a comparative assessment of new environmental policy instruments in the UK and Germany. *Geoforum* 36(3):387–401
- Bayer P, Aklin M (2020) The European Union Emissions Trading System reduced CO2 emissions despite low prices. In: *Proceedings of the National Academy of Sciences*, pp 1–9
- Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Q J Econ* 119(1):249–275
- Billmeier A, Nannicini T (2013) Assessing economic liberalization episodes: a synthetic control approach. *Rev Econ Stat* 95(3):983–1001
- Böhringer C, Rutherford TF, Wiegard W (2003) Computable general equilibrium analysis: opening a black box, pp 1–39
- Böhringer C, Hoffman T, Lange A, Löschel A, Moslener U (2005) Assessing emission regulation in Europe: an interactive simulation approach. *Energy J* 26(4):1–21
- Bruvoll A, Larsen BM (2004) Greenhouse gas emissions in Norway: do carbon taxes work? *Energy Policy* 32(4):493–505
- Burniaux J-M, Nicoletti G, Oliveira-Martins J (1992) GREEN: a global model for quantifying costs of policies to curb CO2 emissions. *OECD Econ Stud* (19):50–91
- Costalli S, Moretti L, Pischedda C (2017) The economic costs of civil war: synthetic counterfactual evidence and the effects of ethnic fractionalization. *J Peace Res* 54(1):80–98
- Downs GW, Locke DM, Barsoom PN (2007) Is the good news about compliance good news about cooperation? *International Law and International Relations*, pp 92–112
- Ekins P, Etheridge B (2006) The environmental and economic impacts of the UK climate change agreements. *Energy Policy* 34:2071–2086
- Ferman B, Pinto C, Possebom V (2018) Cherry picking with synthetic controls. *MPRA Paper* (85138):1–59
- Ferman B, Pinto C, Possebom V (2020) Cherry picking with synthetic controls. *J Policy Anal Manag* 39(2):510–532
- Hazlett C, Xu Y (2018) Trajectory balancing: a general reweighting approach to causal inference with time-series cross-sectional data. *SSRN* 1–38
- Heersink B, Peterson BD (2016) Measuring the vice-presidential home state advantage with synthetic controls. *Am Polit Res* 44(4):734–763
- Hoerner JA, Bosquet B (2001) Environmental tax reform: the European experience. Washington, DC
- Holland PW (1986) Statistics and causal inference. *J Am Stat Assoc* 81(396):945–960
- Hu J, Crijns-Graus W, Lam L, Gilbert A (2015) Ex-ante evaluation of EU ETS during 2013–2030: EU-internal abatement. *Energy Policy* 77:152–163
- Imbens GW, Rubin DB (2015) Causal inference for statistics, social and biomedical sciences: an introduction. Cambridge University Press, Cambridge
- Kaul A, Klößner S, Pfeifer G, Schieler M (2018) Synthetic control methods: never use all Pre-intervention outcomes together with covariates. *MPRA Paper* (83790):1–20
- Martin R, de Preux LB, Wagner UJ (2011) The impacts of climate change levy on business: evidence from microdata. *NBER Working Paper* (17446):1–62
- Nordhaus W (2013) Integrated economic and climate modeling. In: Dixon PB, Jorgenson DW (eds) *Handbook of computable general equilibrium modeling*, chap 16. Elsevier, pp 1069–1131
- Peters GP, Marland G, Le Quéré C, Boden T, Canadell JG, Raupach MR (2012) Rapid growth in CO2 emissions after the 2008–2009 global financial crisis. *Nat Clim Change* 2(1):2–4
- Peters GP, Andrew RM, Solomon S, Friedlingstein P (2015) Measuring a fair and ambitious climate agreement using cumulative emissions. *Environ Res Lett* 10(105004):1–9
- Pindyck RS (2017) The use and misuse of models for climate policy. *Rev Environ Econ Policy* 11(1):100–114
- Robbins MW, Saunders J, Kilmer B (2017) A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention. *J Am Stat Assoc* 112(517):109–126

- Rogelj J, Shindell D, Jiang K, Fifita S, Forster P, Ginzburg V, Handa C, Kheshgi H, Kobayashi S, Kriegler E, Mundaca L, Séférian R, Vilarinho M (2018) Mitigation pathways compatible with 1.5 °C in the context of sustainable development. In: Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, chapter 2, in press
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66(5):688–701
- Schleich J, Eichhammer W, Boede U, Gagelmann F, Jochem E, Schlomann B, Ziesing H-J (2001) Greenhouse gas reductions in Germany—lucky strike or hard work? *Clim Policy* 1(3):363–380
- Sills EO, Herrera D, Kirkpatrick AJ, Brandão A, Dickson R, Hall S, Pattanayak S, Shoch D, Vedoveto M, Young L, Pfaff A (2015) Estimating the impacts of local policy innovation: the synthetic control method applied to tropical deforestation. *PLOS ONE* 10(7):1–15
- Svendsen GT, Daugbjerg C, Hjøllund L, Pedersen AB (2001) Consumers, industrialists and the political economy of green taxation: CO<sub>2</sub> taxation in OECD. *Energy Policy* 29(6):489–497
- World Bank (2019) State and trends of carbon pricing 2019. Washington, DC
- Xu Y (2017) Generalized synthetic control method: causal inference with interactive fixed effects models. *Polit Anal* 25(1):57–76

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