

Analyzing the Effectiveness of International Environmental Policies: The Case of the Kyoto Protocol*

Christian Almer[†] and Ralph Winkler[‡]

[†] Department of Economics and
Institute for Policy Research, University of Bath
3 East, Bath BA2 7AY, UK; c.almer@bath.ac.uk

[‡] (Corresponding author): Department of Economics and
Oeschger Centre for Climate Change Research, University of Bern
Schanzeneckstrasse 1, CH-3001 Bern, Switzerland; mail@ralph-winkler.de
tel: +41 31 631 4508, fax: +41 31 631 3783

Forthcoming in: *Journal of Environmental Economics and Management*

Abstract: We study the effectiveness of emission targets under the Kyoto Protocol with respect to reducing CO₂ emissions. Using country-level and US state-level panel data and employing the synthetic control method, we find very little evidence for an emission reduction effect for the major emitters among the Annex B countries with binding emission targets. More generally, we also show that evaluating the effectiveness of international environmental policies at the country level comes with a number of empirical challenges that may invalidate findings based on more traditional panel data approaches.

Keywords: Climate Policy, International Environmental Agreements, Kyoto Protocol, Synthetic Control Method

JEL-Classification: K33, Q54

* We would like to thank Alberto Abadie, Sylvain Chabé-Ferret, Philip Cooper, Maximilian von Ehrlich, Timo Goeschl, Michael Greenstone, Steve Stillman, Ulrich Wagner, two anonymous reviewers and seminar/conference participants at Bath, Chicago, Lucerne and Toulouse for invaluable comments on an earlier draft of the paper.

1 Introduction

In the Kyoto Protocol (KP), initially adopted on 11 December 1997, 37 countries (and the European Union), the so called Annex B countries, which roughly resembled the industrialized world in the 1990s, committed to reduce the emissions of six greenhouse gases (GHGs) by 5.2% on average over the period between 2008 to 2012 compared to 1990 levels by accepting individual emission targets that had to be met by the end of 2012. Ever since its emergence the KP has been heavily criticized.¹ Also the economic profession found little praise for the KP: While the theoretical literature on climate treaties produced disillusioning findings regarding the likelihood of a global agreement (see, e.g., Finus 2008 for a review of this literature), Barrett (1998) argued that the KP hardly deters non-participation and non-compliance. Copeland and Taylor (2005) criticize that its design neglects important lessons from trade theory, while some authors animadvert the level of the emission targets (e.g., Tol 2000) or discuss the challenges of the flexibility mechanisms (Zhang and Wang 2011).

Despite this widespread criticism, the KP and its approach are supported by a recent and increasing body of empirical literature on the effectiveness of the KP. Aichele and Felbermayr (2012), Aichele and Felbermayr (2013) and Grunewald and Martinez-Zarzoso (2016) consistently find that countries with binding emission targets under the KP have lower CO₂ emissions than they would have had in the absence of these targets. In fact, they estimate a statistically significant average CO₂ reduction effect of 7 to 10%.

In this paper, we test for the existence of observable emission reductions in 15 major Western Annex B countries with binding GHG emission targets under the KP. To this end, we consider a country's ratification of a binding emission target under the KP as a "treatment" and investigate its effect. To do so it is crucial to identify the counterfactual GHG emissions that each treated country would have had emitted in the absence of treatment. The resulting difference, the treatment effect, is a measure of the effectiveness of the KP in the sense that it elicits the success of a particular country to reduce GHG emissions given that the KP entered into force.²

In eliciting this treatment effect we are confronted with four major obstacles. First, to elicit

¹ In December 1997 *The Economist* (1997) already prognosticated that the US will never be able to ratify the KP, as it would never be approved by the US Senate. Prins and Rayner (2007) criticize its inflexible top-down architecture, which had been borrowed from past international treaties regulating chlorofluorocarbons, sulphur emissions and nuclear weapons, and "was always the wrong tool for the nature of the job."

² Note that this treatment effect cannot elicit any effect that the KP (or its enactment) might have had on the emissions of all countries (e.g., due to an increase in public awareness), treated or non-treated. In this sense the treatment effect may underestimate the effectiveness of the KP. To estimate this "total" effect, we would need to observe a twin Earth that is identical to our Earth in all aspects short of the enactment of the KP. Obviously, there is no way to construct such an ideal counterfactual.

the treatment effect it is crucial that there are other countries that did not receive the treatment. With the respect to the KP, the “non-Annex B countries”, i.e. all countries that are not Annex B countries, had no binding GHG emissions targets and, thus, are the natural candidates for the non-treated control group. If all countries were equal in all relevant aspects apart from receiving the treatment or not, the treatment effect would be given by the difference in GHG emissions of treated and non-treated countries. Unfortunately, there is a clear selection bias with respect to the treatment: Annex B countries roughly cover the industrialized world in the 1990s. As a consequence, non-Annex B countries differ significantly with respect to relevant country characteristics. Most importantly, this selection into treatment leads to a violation of the common trend assumption (see Section 2).

Second, the timing of the treatment is not obvious. One might argue that the date of ratification is the treatment event, as only from that point onwards a country adopted the emission target into national legislation and, thereby, certifies that it considers the emission target imposed by the KP as binding. However, there are two other plausible dates for the time of treatment: (i) We consider December 1997, when the KP was adopted, as the earliest date at which these emission reduction targets may have influenced the countries that finally ratified the KP, as the emission targets were already known then. (ii) After the US withdrew its support in 2001 it was not clear whether the KP will actually enter into force, because of the two hurdles the KP had to take.³ Due to this uncertainty, even countries that ratified the KP may have been reluctant to take costly measures to reduce GHG emissions before the KP entered into force.

Third, non-Annex B countries face an additional problem that may invalidate them as suitable controls. According to the rules of the KP, Annex B countries may achieve their emission targets either via domestic emission reductions or through the use of one of the three flexibility mechanisms: emissions trading (EM), joint implementation (JI) and clean development mechanism (CDM). Although the KP clearly states that the use of flexibility mechanisms should only be supplemental to domestic emission reduction efforts, they may blur the distinction between treated and non-treated countries. In particular the CDM encourages Annex B countries to cooperate with non-Annex B countries to reduce GHG emissions of non-Annex B countries. These emission reductions in non-Annex B countries can then be credited to the reduction target of the supporting Annex B countries. Thus, also the emissions of non-Annex B countries may be influenced by the KP, even though they do not face any direct reduction obligations.

Fourth, emissions in non-Annex B countries and Annex B countries that did not ratify

³ According to the rules of the KP, it only enters into force if at least 55 countries ratify it and ratifying countries account for at least 55% of 1990 GHG emissions from Annex B countries. The KP took both hurdles with Russia’s ratification in November 2004 and became effective in early 2005.

the KP may be influenced by an additional channel. Policies enacted in Annex B countries to comply with binding emission targets may not only cause GHG emissions reductions in these countries but may also be accompanied by GHG emission increases in countries without binding emission targets (often called carbon leakage). For example, the production of energy intensive goods may relocate to countries without binding emission targets and laxer environmental regulation.

Both CDM and carbon leakage may violate the stable unit treatment value assumption (SUTVA). However, they have effects in opposite directions. While with CDM emissions of treated countries increase and the emissions of non-treated countries decrease (compared to the situation without CDM), it is the other way round with carbon leakage. Thus, the violation of the SUTVA due to CDM (carbon leakage) is likely to underestimate (overestimate) any treatment effect the KP had on the reduction of GHG emissions of Annex B countries with binding emission targets. As a consequence, the differences in GHG emissions between treated and non-treated countries are not solely due to the adoption of binding emission targets. They are rather a combination of GHG emission reductions in treated countries because of policies enacted to comply with the binding emission target (the treatment effect we like to elicit) and GHG emission increases (reductions) in treated and GHG emission reductions (increases) in non-treated countries because of CDM (carbon leakage).

To address these four challenges, we employ the synthetic control method (SCM) developed by Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015), which constructs counterfactuals for each treated country by a weighted average of non-treated countries such that the actual country and its synthetic counterpart coincide as much as possible with respect to emissions before the treatment and in all other relevant economic characteristics that are unaffected by it. In particular, the SCM allows us to address the violation of the common trend assumption by re-balancing treated units and controls with respect to the pre-treatment development of GHG emissions and other important predictors of selection into the treatment. In addition, we use two different control groups to estimate counterfactual GHG emissions for treated countries: (i) non-Annex B countries and (ii) US state-level data. The latter has the advantage that it suffers less from a selection bias and is also immune to SUTVA violations due to CDM. However, data on some covariates that are available for non-Annex B countries is not available for US state-level data.

To address the issue of the timing of treatment we consider the date of ratification as the treatment event in our main specification and run two robustness checks with 1997, the date of adoption of the KP, and 2004, the date of the KP entering into force as the time of treatment. We also run a robustness check to control for CDM and other flexible

mechanisms, but there is little we can do to control for carbon leakage.⁴ However, as the violation of the SUTVA due to carbon leakage leads to overestimating the real treatment effect of the enactment of binding emission targets, it would only be a concern in case we found significant negative treatment effects.

Yet, we find no statistically significant and persistent treatment effect for any of the Annex B countries under investigation. This holds no matter which date we consider as the treatment event, whether we use non-Annex B countries or US states as the control group, which dependent variable we investigate (CO₂ emissions, CO₂ intensity, CO₂ per capita or GHG emissions), whether we account for flexible mechanisms under the KP or whether we run a standard difference-in-differences approach (while controlling for the opposing trends). However, more drastic interventions like the collapse of the former Soviet Union did have a significant effect on economic output and, as a consequence, on the CO₂ emissions of Eastern European countries. We also find that – overall – the use of US states as the control group is preferable, as the emission paths and other economic characteristics are very similar to those of Annex B countries. Finally, we argue that the applied econometrician faces very similar challenges when analyzing other types of international environmental policies. As a consequence, the application of the SCM, which allows us to address all these challenges simultaneously, may be preferable to more traditional panel data approaches.

2 The Synthetic Control Method

In this paper we aim to assess the effectiveness of the KP with respect to its primary target, i.e. the reduction of GHG emissions in Annex B countries with binding emission targets. In the following, we explain our empirical strategy to elicit the treatment effect of being committed to a specific emission target under the KP. The main challenge for estimating such an effect is a missing data problem (Rubin 1976), as we cannot observe a particular country having both a binding emission target and no emission target at the same time. Thus, it is crucial that we can observe other countries or regions that did not receive the treatment. In our case these can be countries/regions without any binding GHG emission targets under the KP or countries/regions that would have had binding emission targets under the KP but did not ratify it. In fact, we use two different control groups to estimate counterfactual GHG emissions for treated countries: (i) non-Annex B countries plus the US and (ii) US state-level data.⁵ We consider the US and US states as non-treated with respect

⁴ However, see Aichele and Felbermayr (2013) for an assessment of the changes in trade related GHG emissions due to the KP.

⁵ For the remainder of the paper we often refer to the former control group as “non-Annex B countries”. Obviously, this is not literally correct but simpler and shorter than “non-Annex B countries plus the US”.

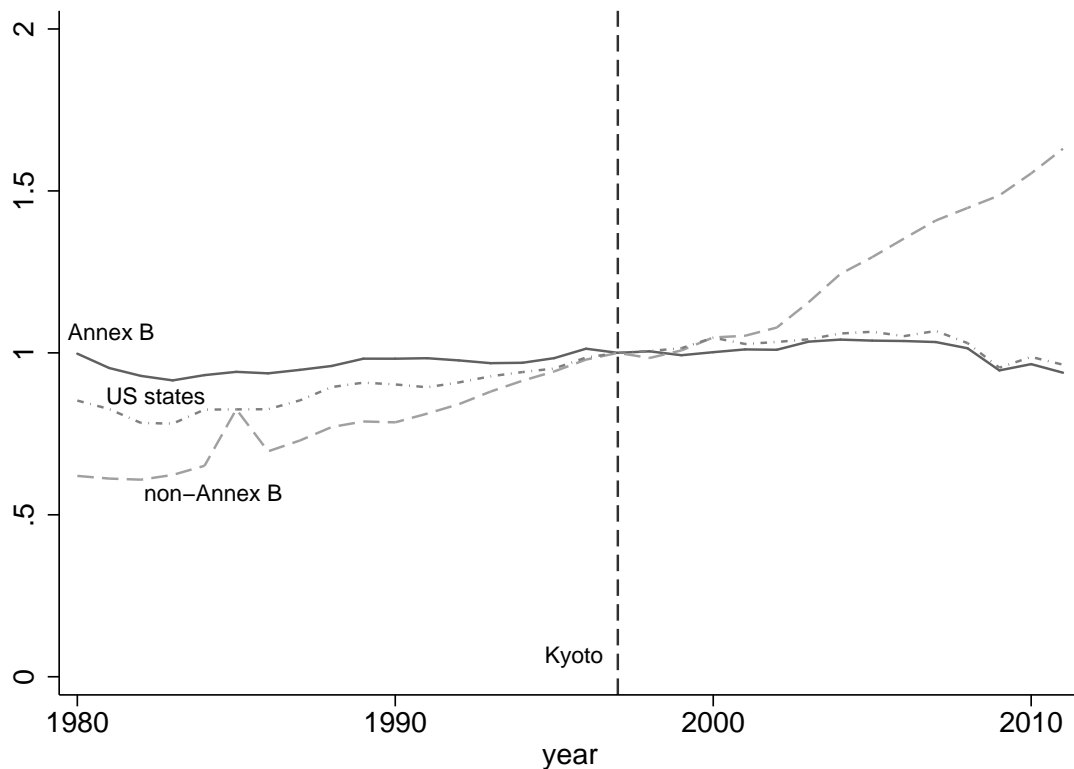
to the KP, as the US did not ratify and, thus, never had any binding obligations under the KP.

However, countries with and without binding emission targets may differ systematically with respect to both their emission paths and other important country characteristics. In particular, the common trend assumption, the key identifying assumption in settings where we observe the outcome of interest for treated units and controls prior and after the intervention, may be violated. It says that – in the absence of the treatment – treated units and controls must share a common trend.⁶ Figure 1 shows the aggregated (average) CO₂ emissions of Annex B and non-Annex B countries and US states relative to their 1997 emissions between 1980 and 2011. We observe that CO₂ emissions were relatively stable for Annex B countries, while they increased considerably for non-Annex B countries in particular after 2000. Also US states show a significantly different trend of CO₂ emissions before 1997 compared to Annex B countries. While the different development in the post-treatment period, i.e. after 1997, may be the effect of the treatment, the differences in the pre-treatment period give rise to serious concerns. Although it is impossible to directly test for a common trend without imposing strong assumptions about the treatment effect, the drastic and statistically significant differences in pre-treatment trends shown in Figure 1 provide strong support for a violation of the common trend assumption in case of the KP. Apart from the differences in trends, non-Annex B countries also differ significantly with respect to other relevant country characteristics (see Section 3). Therefore, average GHG emissions of countries with targets cannot be simply compared to average emissions of countries that have none, i.e. the assignment of emission targets cannot be treated as random (Rubin 1976, 1978, 2005).

There are several potential strategies to solve this problem (Imbens and Wooldridge 2009). In cases as ours, where different groups are either exposed or not exposed to some kind of treatment over a certain time period, the most often applied method is a differences-in-differences (DiD) approach (Bertrand et al. 2004) or some extension of it. Yet, we employ the synthetic control method developed by Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015), which constructs counterfactuals for each treated country (i.e., Annex B country that ratified the KP and, thus, is subject to GHG emission targets) by a weighted average of non-treated regions (i.e., regions without binding emission targets under the KP) such that the actual country and its synthetic counterpart coincide as much as possible with respect to emissions before the treatment and in all relevant economic characteristics that are unaffected by it. The SCM exhibits three key advantages over classical DiD estimation that renders it particularly suitable for the present research question.

⁶ Differences in the absolute value of the outcome of interest and (to some extent) covariates are less of a problem, as they can be absorbed by the use of fixed-effects.

Figure 1: Development of average CO₂ emissions for Annex B countries, non-Annex B countries and US states



Note: Graph shows aggregated CO₂ emissions relative to 1997 emissions for all Annex B countries, non-Annex B countries for which data is available for the period 1980–2011 and US states. For Annex B countries this excludes many Eastern European countries (e.g., Russia and the Ukraine), as data start in 1990 or later. Including Eastern European countries would lead to a significant dip starting in the late 1980s/early 1990s due to the collapse of the Former Soviet Union, as shown in Grunewald and Martinez-Zarzoso (2016).

First, the SCM is able to re-balance treated units and controls with respect to both (i) the pre-treatment development of the GHG emissions themselves and (ii) important predictors of selection into the treatment. Second, the SCM allows to estimate the counterfactual emissions path for every single country and every year following the adoption of the KP. Thus, we do not only get an average effect for all countries and all time periods under investigation, but we are also able to identify country-year-specific developments and characteristics. Third, the standard DiD approach faces an additional challenge when analyzing counterfactual outcomes. On the one hand, a country’s level of GHG emissions depends on several socio-economic factors that should be controlled for in a regression analysis to avoid an omitted variable bias. On the other hand, almost all of these factors may also be influenced by the treatment and therefore constitute bad controls according to the definition of

Angrist and Pischke (2008).⁷ In contrast to DiD, the SCM is flexible in the sense that one can control for these factors using only pre-treatment information.

Suppose that there are $J + 1$ countries where $j = 1$ denotes the treated country – which, in our case, corresponds to a binding emission target under the KP – and $j = 2, \dots, J + 1$ are all untreated countries or US states in the donor pool. In addition, let T_0 be the time of treatment.⁸ For the treated country we have data about the actual emission path (Y_{1t}), but we are ignorant about the counterfactual emissions which would have occurred if this country would not have been subject to the treatment (Y_{1t}^N for $t > T_0$). Thus, we have to find an estimate for Y_{1t}^N to obtain an estimate for the treatment effect $\alpha_{1t} = Y_{1t} - Y_{1t}^N$.

Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015) propose to make use of the observed characteristics of the countries in the control group or donor pool. The underlying idea is to find weights $W = (\omega_2, \dots, \omega_{J+1})'$, with $\omega_j \geq 0$ for $j = 2, \dots, J + 1$ and $\sum_{j=2}^{J+1} \omega_j = 1$, such that the weighted average of all countries in the donor pool resembles the treated country with respect to GHG emissions in the pre-intervention period and all other relevant aspects (Z). Formally, we seek W such that:

$$\sum_{j=2}^{J+1} \omega_j^* Y_{jt} = Y_{1t} \quad \text{for all } t < T_0 \quad \text{and} \quad \sum_{j=2}^{J+1} \omega_j^* Z_j = Z_1 . \quad (1)$$

Then $\sum_{j=2}^{J+1} \omega_j^* Y_{jt}$ for $t \geq T_0$ is an estimate for the unobserved counterfactual emissions path Y_{1t}^N inducing an estimate for the treatment effect:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} \omega_j^* Y_{jt} , \quad t \geq T_0 . \quad (2)$$

In general, a vector W such that equations (1) hold may not exist (in particular, if there are structural differences between treated countries and controls). However, one can choose the weights such as to

$$\min_W (X_1 - X_0 W)' V (X_1 - X_0 W) , \quad (3)$$

where X_1 denotes a $(k \times 1)$ vector of pre-intervention characteristics of the treated country, which may include the pre-intervention emission path, and X_0 denotes a $(k \times J)$ matrix of the same variables for the J countries in the donor pool (Abadie and Gardeazabal 2003,

⁷ For example, investments to reduce GHG emissions may have an impact on GDP per capita, CO₂ intensity, electricity production, etc.

⁸ Throughout the paper, we consider the year of the treatment event as the treatment year if the event took place in the third or fourth quarter and we consider the year before the event as the treatment year if the event took place in the first or second quarter.

Abadie et al. 2010, 2015). The symmetric and positive definite matrix V weights the relative importance of the various characteristics included in X . Obviously, the optimal weights W depend on the weighting matrix V . We follow Abadie et al. (2010) in choosing V by using a regression based method and equal weights.⁹

As the SCM does not provide classical standard errors to infer statistical significance, Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015) suggest to run placebo or permutation tests. The underlying idea is to estimate counterfactual emission paths for regional entities in the donor pool. In an ideal world we would find no treatment effects for all countries in the donor pool and all post-treatment years, as the countries in the donor pool did not receive any treatment. However, in practice we will always find placebo treatment effects to at least some extent. As a consequence, we only consider the actual treatment effect to be statistically significant if it is significantly larger than the placebo treatment effects.

Our inference approach rests on a combination of two refinements of the classical placebos studies that have been proposed by Abadie et al. (2010) and Abadie et al. (2015). First, we run a placebo treatment on all countries in the donor pool and select the top 19 countries/US states in terms of pre-treatment root mean squared prediction error (RMSPE) to avoid rejecting the significance of treatment effects on the basis of outliers within the placebos studies. Second, we calculate the ratio between the treatment effect (the root mean squared treatment effect) and the pre-treatment RMSPE for all 20 countries (the top 19 countries from the placebo treatment plus the treated country) for each year of the post-treatment period. For each year of the post-treatment period, we then calculate a probability that resembles the relative frequency that a randomly chosen country out of the 20 countries in the placebo test has a RMSPE ratio that is at least as large than that of the treated country. We use a combination of the two refinements as proposed by Abadie et al. (2010) and Abadie et al. (2015), as placebos with large pre-treatment RMSPE, i.e. a poor fit, are likely to experience large placebo treatment effects in the post-treatment period and, therefore, lead not only to (i) large treatment/placebo effects but also (ii) large RMSPE ratios. Finally, we consider the treatment effect of the treated country for a particular post-treatment year to be significant if none of the 19 countries with placebo treatments shows a larger RMSPE ratio than the actual treated unit, i.e. the relative frequency of finding such an effect by randomly drawing one out of the 20 countries is $1/20 = 0.05$.

⁹ For further discussion on the synthetic control method including several extensions, see Abadie et al. (2010) and Abadie et al. (2015).

3 Data

We analyze the effect of being committed to an emission target under the KP. Out of the 37 Annex B countries, the US did not ratify the KP. As a consequence, we consider it as not treated. In addition, we excluded all Eastern European countries from the empirical analysis due to limited data availability and the peculiarities of the Former Soviet Union (see also Section 5.4). We also excluded Denmark, Greece, Iceland, Ireland, Monaco, New Zealand, Liechtenstein, Luxembourg and Switzerland due to limited data availability or highly volatile GHG emission paths which are common for small countries and difficult to match.¹⁰ Thus, the remaining 15 treated countries under investigation are: Australia, Austria, Belgium, Canada, Finland, France, Germany, Great Britain, Italy, Japan, Netherlands, Norway, Portugal, Spain, and Sweden. With respect to GHG emissions, these countries are responsible for approximately 50% of total 1990 GHG emissions of the countries with binding emission targets under the KP. In order to reduce the imbalance in country characteristics between Annex B and non-Annex B countries as much as possible, we restrict the country-level control group to countries being classified as high income and upper middle income countries by the World Bank.

Throughout our analysis we use CO₂ emissions as a proxy for GHG emissions (except for one robustness check in Section 5). The reason is that GHG emissions data is only available from 1990 onwards, while data for CO₂ emissions run back to 1980 and earlier.¹¹ Moreover, CO₂ emissions are by far the most important GHG gas and there is a strong positive correlation between CO₂ and other GHG emissions.

As detailed in Section 2, we use the SCM to create the specific counterfactual country for each treated country via a convex combination of all units in the donor pool. To increase the comparability of countries we normalize the outcome of interest (CO₂ emissions) with the year of treatment T_0 as the base year (Cavallo et al. 2013). Our set of predictors when using non-Annex B countries in the donor pool includes: Several years of normalized (treatment year=1) CO₂ emissions¹²; the averages of two 5-year periods prior to the treatment for all predictors (e.g., 1987–1991 and 1992–1996 for 1997 as treatment), i.e. GDP per capita, GDP growth, human capital index, life expectancy, agricultural, industry and services value added, an indicator for the level of democracy versus autocracy, an indicator for the level of political rights and population growth; and the averages for two sub-periods of the post-treatment periods (e.g., 2000–2005 and 2006–2011) for life expectancy, human capital index

¹⁰ In addition, the GHG emissions of these countries play a very minor role for aggregate GHG emissions.

¹¹ In addition, the GHG emissions data, stemming from CAIT Climate Data Explorer of the World Resources Institute, exhibit much smaller year to year volatility than the CO₂ emissions data from the World Development Indicators provided by the World Bank.

¹² More specifically, we use 1981, 1983, 1985,

and population growth. In doing so, we assume that the latter three variables will not be affected by the treatment and, therefore, do not constitute bad controls in the sense of Angrist and Pischke (2008). Our choice of predictors is guided by three desiderata: (i) Predictors must be available for all treated countries and should be available for as many countries in the donor pool as possible. (ii) Predictors should account for the underlying causes of variations in GHG emissions across countries such as economic power and industry structure. (iii) Finally, indicators should particularly target the selection bias for treatment, i.e. developed versus developing countries. As a consequence, we include indicators such as life expectancy, human capital and political rights, as these indicators are strongly correlated with economic development.¹³

Unfortunately, several of the above listed variables are not available for US states.¹⁴ Thus, we rely on several years of normalized (treatment year=1) CO₂ emissions, GDP per capita, GDP growth, and population growth when using US states as the donor pool. As we do expect differences in the classification and coding between country level and US state-level data, in particular for GDP, we use the changes rather than the levels (as for non-Annex B countries) with respect to the treatment year values.¹⁵ More specifically, we use normalized per capita GDP, GDP growth and population growth for, e.g., 1987–1991 and 1992–1996 and, in addition, population growth data for 2000–2005 and 2006–2010 (1997 as treatment year).

The data used in the empirical analysis stem from several different sources. Data on country-level CO₂ emissions, value added for agriculture, industry and services, GDP growth, life expectancy and population growth are taken from the World Development Indicators published by the World Bank. GDP per capita and the human capital index originate from the Penn World Tables published by the University of Groningen. The two political economy indicators on democracy versus autocracy (poilty2) and political rights are taken from the Polity IV project and from data published by the Freedom House. Additional information on the KP (list of Annex B countries with targets) stems from the United Nations Framework Convention on Climate Change (UNFCCC). US state-level information on CO₂ emissions

¹³ The categorization of Annex B countries strongly resembles the industrialized world in the 1990s. Historically, industrialized countries have been predominantly characterized by their sectoral composition and their levels of GDP per capita. More recently, however, the literature draws on a more diverse set of measures including human capital, life expectancy and political economy indicators, which – in sum – are seen to adequately distinguish industrialized countries from developing countries and countries in transition (e.g., UNDP 1990).

¹⁴ However, in contrast to non-Annex B countries the differences between Annex B countries with binding emissions targets and US states are much smaller for measures such as life expectancy, human capital index, etc. (US averages over the period from 1987–1997 for life expectancy: 75.44 years, human capital index: 3.45, Serv. value added: 72.3, Pol. rights: 1, Polity2: 10). As a consequence, it is also less important to control for them in our analysis.

¹⁵ We expect these differences to be constant over time and therefore focus on the changes in order to eliminate any inconsistencies.

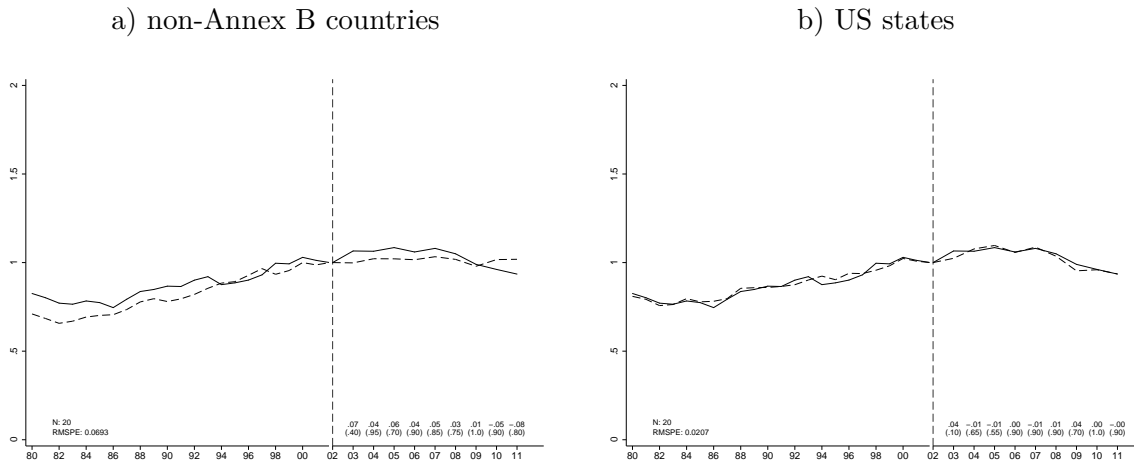
Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N	P-values
Annex B countries (under investigation)						
CO ₂ emissions	326'184.26	319'850.17	29'247.99	1'205'610.59	165	
GDP per capita	23'439.94	5'485.69	10'401.74	43'569.25	165	
GDP growth	2.15	2.09	-6.51	7.60	165	
Human Capital Index	2.77	0.29	2.19	3.3	165	
Life expectancy	76.85	1.31	73.67	80.42	165	
Agri. value added*	3.52	1.82	1.22	13.93	165	
Ind. value added*	31.05	3.1	23.56	38.41	165	
Serv. valued added*	65.42	3.21	58.46	73.23	165	
Population growth	0.48	0.39	-0.22	1.78	165	
Pol. rights	1.01	0.11	1	2	161	
Polity2	9.93	0.25	9	10	161	
GHG emissions	407.25	369.28	43.66	1300.15	120	
High income and upper middle income Non-Annex B countries⁺						
CO ₂ emissions	147'497.62	656'750.94	11.00	5.42e+06	822	0.00
GDP per capita	10'857.28	10'709.68	335.49	6'7883.89	694	0.00
GDP growth	2.13	7.50	-42.62	90.88	813	0.07
Human Capital Index	2.36	0.37	1.68	3.52	511	0.00
Life expectancy	69.83	5.99	40.97	80.13	823	0.00
Agri. value added*	11.24	10.21	0.00	67.38	687	0.00
Ind. value added*	33.20	12.64	8.02	74.67	678	0.13
Serv. valued added*	55.49	14.21	8.24	90.63	679	0.00
Population growth	1.81	1.70	-6.49	16.51	994	0.00
Pol. rights	3.61	2.22	1.00	7.00	729	0.00
Polity2	1.28	7.31	-10.00	10.00	531	0.00
GHG emissions	225.63	874.69	0.12	6592.37	542	0.01
US states						
CO ₂ emissions	100'982.17	103'210.82	4'200.00	679'600.00	561	0.00
GDP per capita	27'514.78	9'703.95	15'468	87'544	561	0.03
GDP growth	1.98	2.57	-12.90	11.90	561	0.00
Population growth	0.01	0.01	-0.04	0.07	561	0.00
GHG emissions	122.72	122.27	4.66	805.07	408	0.00

Note: Descriptive statistics for each (sub-)sample and the 11 years (1987–1997) prior to the treatment, i.e. the adoption of the KP. The p-values in the last column are based on t-tests for the equality of means between Annex B countries and the two potential control groups. GHG data are in Mt of CO₂ equivalent.

*: in % of GDP, +: not sufficient data available for some non-Annex countries, i.e. they are not part of the donor pool.

Figure 2: Synthetic Control for Canada based on



Note: Actual and synthetic CO₂ emission path for Canada. The solid line stands for the actual Canadian CO₂ emissions whereas the dashed line represents the synthetic Canada, i.e. the counterfactual emission path, based on non-Annex B country data (left) and US state-level data (right). N is the number of countries in the donor pool used for the placebo studies and we display the figures for the pre-treatment RMSPE. The numbers underneath the lines indicate the treatment effects in percent and the corresponding probabilities in parenthesis (see Section 2).

is taken from the US Energy Information Administration (EIA), real GDP per capita and GDP growth stem from the Bureau of Economic Analysis (BEA) and population data is based on the Population Estimates Program (PEP) of the United States Census Bureau. Data on GHG emissions in CO₂ equivalent emissions on country and US state level stems from the CAIT Climate Data Explorer of the World Resources Institute. Table 1 shows the summary statistics for all data used in the empirical analysis. We observe a predictor imbalance for Annex B countries, non-Annex B countries and US states for the 11 years prior to the adoption of the KP, i.e. 1987–1997. In fact, there is a statistically significant difference for most predictors between Annex B and non-Annex B countries and also between Annex B countries and US states. Yet, the differences between Annex B countries and US states are much smaller in magnitude.

4 Results

In the following, we report the results for the two different specifications, where either selected non-Annex B countries or US states act as non-treated control entities from which we construct the counterfactual CO₂ emission paths of the 15 Annex B countries under investigation. We consider ratification as the treatment event.

As a representative example, Figure 2 shows the normalized CO₂ emissions path of Canada

(solid line) and its synthetic counterfactuals (dashed line) based on non-Annex B countries (left) and US states (right). We observe that for both specifications the synthetic controls for Canada matches actual CO₂ emissions very well in the pre-treatment period (up to 2002). This is also evident from the small RMSPE values of 0.0693 and 0.0207. Yet, the counterfactual based on US state-level data matches pre treatment CO₂ emissions slightly better than the counterfactual based on non-Annex B countries, an observation that generally holds for all 15 countries under observation. Also the general pattern of counterfactual emissions in the post treatment period (after 2002), i.e. the CO₂ emissions Canada had if it were not subject to a binding emission target under the KP, is similar for both specifications and very close to actual CO₂ emissions. Again, US state-level data predicts counterfactual CO₂ emissions that slightly better match actual emissions of Canada in the post treatment period. The treatment effect ranges from -8% to +7% for non-Annex B countries and -1% to +4% for US state-level data. Note that a positive (negative) number indicates actual emissions are higher (lower) than the emissions of the synthetic control. Thus, a treatment effect in the sense that countries with a binding emission target under the KP experience lower actual emissions than predicted by their counterfactuals would result in negative numbers.

As an indicator of significance, the values in parenthesis give the probabilities that if one randomly draws one country out of the 20 countries in the placebo study (the top 19 non-treated countries with respect to low RMSPE in the pre-treatment period plus the treated country) one would draw a country that exhibits at least as high a deviation from their actual normalized CO₂ emissions than the treated country.

Tables 2 and 3 summarize the results for all 15 Annex B countries under investigation and for the two model specifications (using non-Annex B countries and US states in the donor pool). The numbers indicate the deviation of the actual normalized CO₂ emissions from its synthetic counterfactuals, i.e. the numbers for Canada are identical to the numbers shown in Figure 2. Likewise, numbers in parenthesis show probabilities as discussed in Section 2 and, thus, reflect the statistical significance of the effect.¹⁶

For the specification using non-Annex B countries to construct counterfactual CO₂ emissions (Table 2), we find negative and positive treatment effects ranging from -38% (Portugal in 2011) to +34% (Norway in 2010). Yet, none of these treatment effects is statistically significant. When looking at the specification using US state-level data to construct counterfactual CO₂ emissions (Table 3), the general pattern is that treatment effects are in a similar range as for counterfactual emissions based on non-Annex B countries, ranging from -17% (Belgium in 2007) to +45% (Norway in 2010), but are slightly shifted towards positive

¹⁶ Analogous graphs to Figure 2 of actual and counterfactual emissions paths and the corresponding placebo studies to infer statistical significance for all 15 countries under investigation for the two main specifications and all robustness checks discussed in this paper are available in the Online Appendix.

Table 2: Estimates for Treatment Effects based on non-Annex B Countries

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.04 (.50)	.11 (.40)	.00 (.95)	-.01 (.90)
	RMSPE: 0.0330									
Austria	-.01 (.90)	.08 (.35)	.04 (.75)	.08 (.65)	.05 (.90)	-.02 (1.0)	.00 (1.0)	-.02 (.95)	-.04 (.80)	-.05 (.95)
	RMSPE: 0.0681									
Belgium	-.07 (.35)	.00 (1.0)	-.06 (.75)	-.10 (.60)	-.10 (.75)	-.15 (.65)	-.10 (.85)	.01 (1.0)	-.05 (.80)	-.12 (.65)
	RMSPE: 0.0751									
Canada		.07 (.40)	.04 (.95)	.06 (.70)	.04 (.90)	.05 (.85)	.03 (.75)	.01 (1.0)	-.05 (.90)	-.08 (.80)
	RMSPE: 0.0693									
Finland	.07 (.35)	.21 (.15)	.13 (.50)	-.08 (.65)	.12 (.60)	.04 (.90)	-.04 (.95)	-.02 (.95)	.01 (.95)	-.10 (.65)
	RMSPE: 0.0864									
France	-.02 (.70)	.01 (.95)	-.02 (.85)	-.06 (.65)	-.08 (.75)	-.13 (.65)	-.09 (.85)	-.02 (.95)	-.17 (.45)	-.19 (.50)
	RMSPE: 0.0982									
Germany	-.03 (.65)	-.00 (1.0)	-.06 (.75)	-.11 (.55)	-.10 (.75)	-.16 (.60)	-.10 (.80)	-.00 (1.0)	-.18 (.45)	-.17 (.50)
	RMSPE: 0.0966									
Great Britain	-.05 (.55)	-.02 (.80)	-.05 (.75)	-.06 (.65)	-.05 (.90)	-.10 (.80)	-.07 (.90)	-.07 (.90)	-.13 (.60)	-.20 (.50)
	RMSPE: 0.1338									
Italy	-.01 (.90)	.02 (.80)	.00 (.90)	.00 (.95)	.00 (1.0)	-.03 (.95)	-.04 (.95)	-.09 (.80)	-.13 (.65)	-.13 (.60)
	RMSPE: 0.0535									
Japan	-.00 (.90)	.01 (.95)	.00 (.95)	-.01 (.90)	-.01 (1.0)	-.02 (1.0)	-.04 (.95)	-.08 (.80)	-.09 (.70)	-.08 (.80)
	RMSPE: 0.0747									
Netherlands	.02 (.70)	.04 (.70)	.02 (.85)	-.02 (.85)	-.03 (.95)	-.03 (1.0)	.03 (.95)	.11 (.75)	.08 (.70)	.03 (.95)
	RMSPE: 0.0567									
Norway	-.10 (.15)	.01 (.95)	-.02 (.85)	-.04 (.70)	.01 (.95)	.02 (1.0)	.17 (.50)	.15 (.65)	.34 (.30)	.08 (.85)
	RMSPE: 0.0882									
Portugal	.05 (.50)	-.07 (.55)	-.08 (.65)	-.08 (.65)	-.20 (.25)	-.20 (.40)	-.24 (.30)	-.24 (.35)	-.35 (.25)	-.38 (.25)
	RMSPE: 0.0745									
Spain	.06 (.35)	.07 (.55)	.09 (.65)	.11 (.55)	.08 (.75)	.09 (.85)	.01 (.95)	-.06 (.90)	-.19 (.45)	-.16 (.50)
	RMSPE: 0.0583									
Sweden	.12 (.15)	.10 (.30)	.04 (.80)	-.05 (.65)	-.07 (.80)	-.14 (.65)	-.04 (.95)	.04 (.95)	-.03 (.80)	.00 (1.0)
	RMSPE: 0.0981									

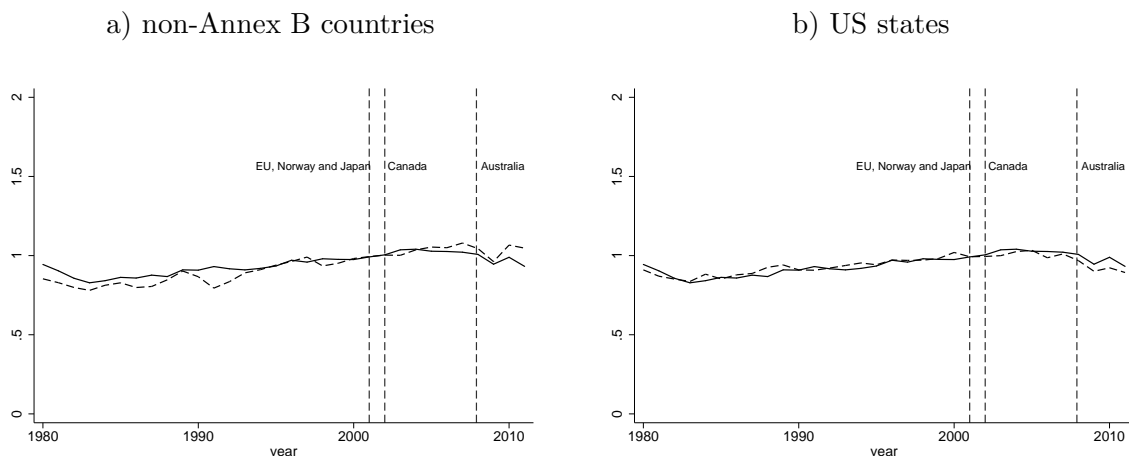
Note: The Table contains treatment effect estimates for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Table 3: Estimates for Treatment Effects based on US States

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.07 (.05)	.17 (.05)	.07 (.20)	.09 (.20)
	RMSPE: 0.0220									
Austria	.04 (.30)	.12 (.05)	.11 (.10)	.14 (.05)	.12 (.10)	.03 (.75)	.04 (.70)	.04 (.75)	.11 (.30)	.10 (.25)
	RMSPE: 0.0427									
Belgium	-.09 (.05)	-.04 (.30)	-.08 (.10)	-.11 (.10)	-.12 (.10)	-.17 (.05)	-.13 (.10)	-.05 (.55)	-.05 (.55)	-.13 (.15)
	RMSPE: 0.0537									
Canada		.04 (.10)	-.01 (.65)	-.01 (.55)	.00 (.90)	-.01 (.90)	.01 (.90)	.04 (.70)	.00 (1.0)	-.00 (.90)
	RMSPE: 0.0207									
Finland	.08 (.10)	.19 (.05)	.17 (.05)	-.06 (.30)	.20 (.05)	.13 (.05)	.04 (.65)	.04 (.75)	.21 (.05)	.11 (.25)
	RMSPE: 0.0670									
France	-.03 (.30)	-.00 (.85)	.00 (1.0)	-.00 (.85)	.06 (.25)	.01 (.85)	.06 (.50)	.09 (.50)	.06 (.50)	.04 (.60)
	RMSPE: 0.0724									
Germany	-.05 (.15)	.01 (.85)	-.02 (.45)	-.02 (.55)	.15 (.10)	.08 (.30)	.17 (.10)	.09 (.50)	.08 (.45)	.12 (.25)
	RMSPE: 0.0454									
Great Britain	-.04 (.25)	-.05 (.30)	-.03 (.20)	-.06 (.35)	.00 (1.0)	-.05 (.40)	-.02 (.75)	.04 (.80)	.05 (.60)	-.03 (.65)
	RMSPE: 0.0298									
Italy	.02 (.70)	.05 (.30)	.02 (.40)	.02 (.55)	.03 (.50)	.00 (.95)	.02 (.80)	-.01 (.85)	-.03 (.65)	-.04 (.60)
	RMSPE: 0.0238									
Japan	.01 (.95)	.03 (.40)	.01 (.85)	-.01 (.75)	-.03 (.55)	-.04 (.50)	-.03 (.75)	-.07 (.50)	-.06 (.45)	-.04 (.60)
	RMSPE: 0.0194									
Netherlands	.00 (.95)	.01 (.65)	.00 (1.0)	-.03 (.55)	-.07 (.25)	-.06 (.35)	.00 (1.0)	.06 (.55)	.09 (.35)	.03 (.65)
	RMSPE: 0.0373									
Norway	-.10 (.05)	.02 (.60)	-.00 (.90)	-.01 (.80)	.04 (.40)	.04 (.70)	.20 (.05)	.26 (.05)	.45 (.05)	.21 (.05)
	RMSPE: 0.0607									
Portugal	.15 (.05)	-.00 (1.0)	-.06 (.15)	-.07 (.15)	.02 (.85)	.04 (.70)	.02 (.75)	.04 (.65)	.00 (.95)	.06 (.60)
	RMSPE: 0.0803									
Spain	.07 (.10)	.07 (.15)	.07 (.10)	.11 (.10)	.10 (.10)	.11 (.05)	.03 (.75)	-.03 (.80)	-.11 (.10)	-.09 (.30)
	RMSPE: 0.0297									
Sweden	.10 (.05)	.10 (.05)	.08 (.10)	.03 (.55)	.16 (.10)	.09 (.25)	.20 (.05)	.08 (.55)	.25 (.05)	.32 (.05)
	RMSPE: 0.0814									

Note: The Table contains treatment effect estimates for each Annex B country under investigation using US states to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Figure 3: Average actual and synthetic CO₂ emissions for synthetic controls based on



Note: The solid line represents the (average) normalized (treatment year=1) CO₂ emission of the Annex B countries under investigation. The dashed line indicates the average normalized emissions of the synthetic controls for the treated countries based on non-Annex B countries (left) and US states (right).

treatment effects, in particular for the three Scandinavian countries Finland, Norway and Sweden. Treatment effects are also more often statistically significant. In fact, we do now find three statistically significant negative effects: for Belgium in 2002 (-9%) and in 2007 (-17%) and for Norway in 2002 (-10%). In addition, we find positive and significant effects for Australia (2 years), Austria (2 years), Finland (5 years), Norway (4 years), Portugal (1 year), Spain (1 year) and Sweden (5 years).

In summary, we do not find any evidence that a binding emission target under the KP induced a significant and persistent emission reduction effect for any of the investigated Annex B countries with binding emission targets under the KP: For non-Annex B countries in the donor pool we do not find any statistically significant treatment effects. For counterfactual CO₂ emissions paths based on US states we do find significant negative effects for two countries in one, respectively two years (out of the ten year post treatment period). However, we also find some significant positive treatment effects, in particular for Finland, Norway and Sweden. This finding is further illustrated in Figure 3 showing the average normalized CO₂ emissions of the Annex B countries under investigation (solid line) and the average counterfactual CO₂ emissions path (dashed line) based on non-Annex B countries (left) and US states (right). We observe that the average counterfactual CO₂ emissions path fits the average actual emissions path quite well not only in the pre-treatment but also in the post treatment period.

When comparing Figures 3 and 1 it is also evident that the synthetic control method suc-

Table 4: Predictor Imbalance

Variable	Diff. original	Diff. matched	rel. Decrease	P-values
Non-Annex B countries				
<i>pre-treatment</i>				
GDP per capita	7358.8	-2266.9	0.6919	0.27
GDP growth	-0.0349	-0.3070	-7.8018	0.15
Agri. value added	-6.5614	-0.3813	0.9419	0.42
Ind. value added	-3.0797	1.7101	0.4447	0.07
Serv. valued added	9.7223	-1.3421	0.8620	0.23
Pol. rights	-2.5251	-0.4555	0.8196	0.00
Polity2	7.9997	1.1811	0.8524	0.00
Human Capital Index	0.3928	-0.3668	0.0662	0.00
Life expectancy	7.0624	3.3051	0.5320	0.00
Population growth	-1.1031	-0.7855	0.2880	0.00
<i>post-treatment</i>				
Human Capital Index	0.3946	-0.3094	0.2159	0.00
Life expectancy	7.1489	3.6752	0.4859	0.00
Population growth	-0.8297	-0.3029	0.6349	0.02
US states				
<i>pre-treatment</i>				
GDP per capita	-0.0049	-0.0065	-0.3184	0.85
GDP growth	-1.2045	-0.2353	0.8047	0.16
Population growth	2.4642	0.2712	0.8900	0.60
<i>post-treatment</i>				
Population growth	4.9809	1.8151	0.6356	0.10

Note: Average differences (across countries/states and years) between the 15 investigated treated countries and the two control groups (non-Annex B countries and US States) before matching (Diff. original) and after matching (Diff. matched). Column (3) shows the relative decrease in the predictor imbalance and column (4) the p-value for the after matching predictor imbalance to be statistically significant.

cessfully balanced the pre-treatment trends, i.e. the opposing trends between Annex B and non-Annex B countries, as shown in Figure 1, vanished entirely in Figure 3. Table 4 shows the imbalance before and after matching for all other predictors. While the predictor imbalance after matching is statistically insignificant for all predictors when using US States as the control group this is not the case for non-Annex B countries in the donor pool. In fact, we observe that the SCM successfully matches all economic indicators on a 5% significance level. However, all other indicators are still significantly different (although it reduces the imbalance to a large extent). This shows that the selection bias into treatment, i.e. treated and non-treated countries differ consistently and significantly in important aspects, can only be mitigated but not completely overcome, as there are irreconcilable structural differences between Annex B and non-Annex B countries. In this respect US State level data seems to be preferable to non-Annex B countries in constructing counterfactual CO₂ emissions paths.

5 Robustness Checks

To scrutinize the robustness of our main result that the ratification of the Kyoto Protocol had no significant effect on the CO₂ emissions of 15 major Annex B countries, we run a series of alternative specifications.

5.1 Timing of treatment

In our main specification we considered the ratification of the KP as the time of treatment. However, as the emission targets were already known in 1997, it might be that the adoption of the KP already induced changes in the emission paths of treated countries. In fact, we consider 1997 as the earliest date at which the KP could have imposed a treatment effect. As a consequence, we run an additional specification, where we assume 1997 as the year of treatment. This robustness check could also be interpreted as a placebo in time analysis (Abadie and Gardeazabal 2003, Abadie et al. 2010, 2015). Results are shown in Tables 9 and 10 in the Appendix and are very similar to our main specification. We do not find any significant treatment effects using non-Annex B countries in the donor pool and hardly find any significant negative treatment effects when using US states in the donor pool (Finland in the year 2000 being the only exception). Again, we do find some positive treatment effects, in particular for Norway and Spain.

Although an Annex B country might have known its emission reduction target under the KP already in 1997, a country might not have taken any action to reduce its GHG emissions prior to the date the KP entered into force, even if the country has ratified the KP before. Of course, this does not invalidate the analysis when choosing ratification or even adoption as the time of treatment. If the synthetic counterfactual matches the country under consideration well, all we should see is that counterfactual emissions start to deviate from actual emissions not at the considered time of treatment but at some later time at which emission reduction efforts started. Yet, taking too early a treatment date comes at the disadvantage that fewer pre-treatment observations are available to construct the synthetic counterfactuals which might result in a poorer match between actual country and synthetic counterfactual. As a consequence, we run an additional specification where we consider 2004 as the time of treatment, because subsequent to Russia's ratification in November 2004 the KP entered into force in February 2005. Results are shown in Table 11 in the Appendix and are consistent with the results of our two main specifications.

5.2 Alternative dependent variables

For the 15 Annex B countries investigated we do find very little evidence for a persistent treatment effect, i.e. a permanent reduction in CO₂ emissions compared to their synthetic counterfactuals no matter what date we consider as the time of treatment. One reason for this might be that CO₂ emission levels are strongly correlated with economic performance and, thus, are vulnerable to business cycle fluctuations. In other words, favorable global economic conditions for Annex B countries (at least up to 2007) could be responsible for the lack of a significant effect on emission levels although Annex B countries might have invested in cleaner production technologies. We test this hypothesis by running alternative specifications with CO₂ intensity (CO₂ emissions relative to GDP) and CO₂ emissions per capita as dependent variables, again using non-Annex B country and US state-level data to construct counterfactuals for the 15 Annex B countries under consideration.¹⁷ Tables 12–14 in the Appendix show the results of these alternative model specifications. Our findings are consistent with our main specifications.

In addition, we run a specification using the emissions of six different GHGs¹⁸ measured in CO₂ equivalent emissions instead of CO₂ emissions as the dependent variable. In contrast to CO₂ emissions data, emissions data for other GHGs is only available since 1990. Results are reported in Tables 15 and 16 in the Appendix. Results are highly consistent with our main specifications. Yet, we observe much fewer positive treatment effects when using US states to construct counterfactual GHG emissions paths. In particular, the Scandinavian countries Finland, Norway and Sweden which accounted for most significantly positive treatment effects in our main specification do not show any significant treatment effects.

5.3 Accounting for Flexibility Mechanisms

In addition to domestic emission reductions, Annex B countries may achieve their emission targets under the rules of the KP by three flexibility mechanisms (see also Section 1). Emissions trading (ET) allows Annex B countries to directly trade their initial endowment of allowed emissions. The general idea of joint implementation (JI) and the clean development mechanism (CDM) is that a donor country invests in an emission reduction project in a host country. The donor receives in return additional emission allowances in the amount by which emissions have been reduced in the host country due to the project. Under JI the host country is another Annex B country while it is a non-Annex B country under CDM. ET and JI are redistributions of emission allowances between Annex B countries and do not

¹⁷ Due to a discontinuity in US state-level data for GDP between 1997 and 1998 we are not able to use US states in the donor pool for CO₂ intensity.

¹⁸ The six GHGs are: CO₂, CH₄, N₂O, Hydrofluorocarbons, Perfluorocarbons, and Sulfur Hexafluoride.

affect the total cap of all Annex B countries. The CDM creates additional allowances for Annex B countries. As a consequence, the KP demands that allowances created from CDMs are *additional*, i.e. the emissions reductions in the host country would not have occurred without the investment of the donor country.

The flexible mechanisms of the KP create two potential obstacles for addressing the question whether and to what extent the adoption of the KP had an influence on the CO₂ emissions of Annex B countries. First, the availability of the flexible mechanisms may crowd out incentives of Annex B countries to invest into domestic emissions reductions, in particular if emissions allowances purchased through the flexible mechanisms are cheaper than corresponding domestic emission reductions. Second, the CDM reduces emissions in non-Annex B countries. Thus, the CDM blurs the distinction between “treated” and “non-treated” countries. Of course, the second issue is non-existent when using US states as control group.

To address the first issue, we run an alternative specification, where for each year in the post treatment period, we replace actual CO₂ emissions by CO₂ emissions, for which the country is accountable according to the rules of the KP. To this end, we calculate for all countries under investigation and for all years in the post treatment period the net amount of emissions allowances purchased via the three flexible mechanism and deduct it from the actual CO₂ emissions.¹⁹ Thus, in this specification we test whether not actual but accountable CO₂ emissions of countries are significantly different from the CO₂ emissions of their synthetic counterfactuals. To address the second issue, we take advantage of the fact that only few non-Annex B countries host the vast majority of CDM projects. In fact, China, India, South Korea, Brazil and Mexico are accountable for more than 85% of all certified emission reduction units under the CDM.²⁰ As a consequence, we discard these five countries from the donor pool in the specification where we construct synthetic CO₂ emissions by using non-Annex B countries.

The results are shown in Tables 17 and 18 in the Appendix. While we do not find any negative significant treatment effects for using non-Annex B countries in the donor pool, we do find four additional significant negative treatment effects in the specification with US states as control group: Austria in the year 2009 and Great Britain, The Netherlands and Spain in the year 2011. These four negative treatment effects stem from the fact that net emission acquisitions are highly volatile. For example, Great Britain had a net acquisition of emission credits in the year 2011 of 249'462 kt CO₂ eq., which amounts to 7.31% of Great Britain's

¹⁹ We convert GHG emission credits into CO₂ emission credits by calculating the relative share of CO₂ emissions of all GHG emissions for each country and year and multiplying net acquisitions of credits by this share. Data for GHG emissions and trade in emission credits from the different flexible mechanisms are reported in the UNFCCC GHG inventories and registries available at http://unfccc.int/national_reports.

²⁰ Numbers are taken from the official CDM website <http://cdm.unfccc.int/Statistics/Public/CDMinsights>.

Table 5: Net Acquisition of Emissions Credits for the first KP Commitment Period

Country	Initial Allowance	Net Acquisition (tot.)	Net Acquisition (rel.)
Australia	2'957'579	-97'007	-3.280%
Austria	343'866	73'120	21.264%
Belgium	673'996	-28'523	-4.232%
Canada	2'791'793	331	0.012%
Finland	355'018	7'113	2.003%
France	2'819'627	-97'212	-3.448%
Germany	4'868'097	67'593	1.388%
Great Britain	3'412'081	7'418	0.217%
Italy	2'416'278	67'404	2.790%
Japan	5'928'258	638'201	10.765%
Netherlands	1'001'262	30'660	3.062%
Norway	250'577	35'196	14.046%
Portugal	381'938	25'514	6.680%
Spain	1'666'196	162'184	9.734%
Sweden	375'189	-7'134	-1.901%

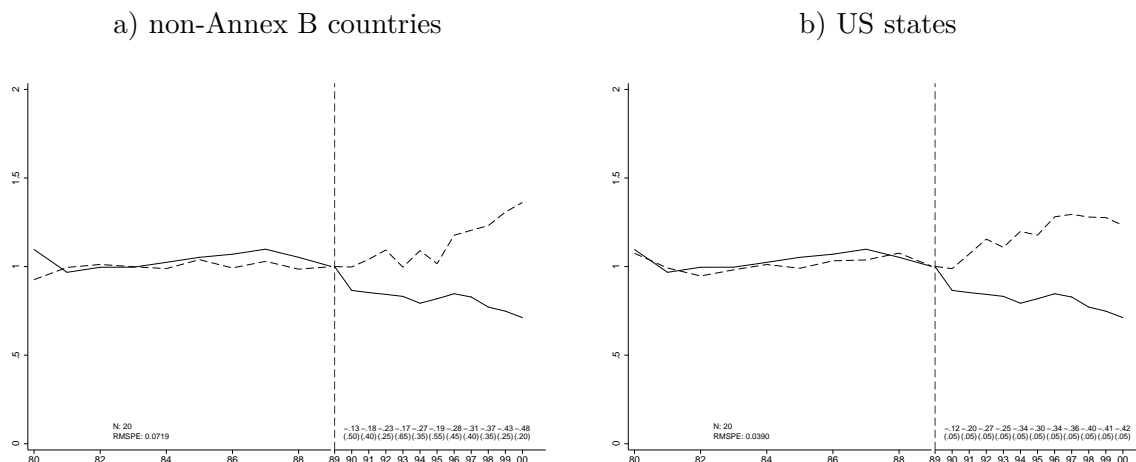
Note: The table lists initial greenhouse gas emissions endowments of Annex B countries (Assigned Amount Units) in kt CO₂eq. and the net acquisition of emission credits from all flexible mechanisms in kt CO₂eq. and relative to the initial emissions endowments.

total greenhouse gas emissions allowance under the KP for the first commitment period from 2008–2012. Yet, as can be seen from Table 5, Great Britain’s total net acquisition for the whole first commitment period only amounted to 7’418 kt CO₂ eq. (0.22% of total GHG emissions allowance under the KP). In fact, only three countries purchased greenhouse gas emission credits exceeding 10% of the total initial emission allowance under the KP (Austria, Japan and Norway) and two additional countries in excess of 5% (Portugal and Spain). Thus, even taking emission credits from the flexible mechanisms into account we do not find any significant and robust negative treatment effect for any of the 15 investigated Annex B countries.

5.4 Omission of Eastern European countries

We excluded Eastern European countries from the analysis for two reasons. First, we have insufficient data for most of these countries to construct proper synthetic counterfactuals. Second, after the collapse of the Soviet Union in 1989 the whole former “Eastern Block” experienced a severe economic downturn. Economic downturns are accompanied with lower production and, as a consequence, lower energy use and reduced CO₂ emissions. In fact, the collapse of the Former Soviet Union (FSU) had a significant impact on these countries’ CO₂ emissions that is likely to blur any additional emission reductions due to binding emission

Figure 4: Poland’s actual and synthetic CO₂ emissions for synthetic controls based on



Note: Actual normalized (1989=1) CO₂ emissions (solid line) and synthetic control (dashed line) for Poland considering the collapse of the Soviet Union in 1989 as the treatment using non-Annex B country data (left) and US state-level data to construct synthetic counterfactuals. The country weights for the synthetic control are Barbados (0.138), Fiji (0.08), Korea (0.031), Sri Lanka (0.209), Trinidad and Tobago (0.143), USA (0.193), and Vietnam (0.172) for non-Annex B countries and Illinois (0.565) and Rhode Island (0.435) for US states in the donor pool.

targets under the KP.²¹ To show this, we investigate the effect of the treatment “collapse of the FSU” on the CO₂ emissions of Poland.²² Again, we construct a synthetic counterfactual Poland (again using non-Annex B countries and US states) and compare the deviations of normalized CO₂ emissions. Figure 4 shows the results. We observe that starting from the time of the treatment in 1989 actual normalized CO₂ emissions in Poland increasingly fall short of the normalized CO₂ emissions of its synthetic counterfactual. We observe no significant treatment effect for synthetic controls based on non-Annex B countries and a highly significant treatment effect throughout the whole post treatment period for synthetic controls based on US states.

5.5 Comparison to DiD approach

Finally, we compare our SCM results to results achieved by a DiD approach (see Table 6). The most basic DiD specification includes all countries, i.e. all Annex B and non-Annex B countries with sufficient data availability, and only dummies for Annex B countries and

²¹ Despite their economic recovery, Eastern European countries exhibited emission levels far below their emission targets at the time of adoption of the KP in 1997. As these countries were not expected to reach emission levels at or even above their Kyoto targets in the near future, they had little economic incentives to reduce emissions (hot air).

²² Poland being one of the few Eastern European countries with sufficient data on CO₂ emissions prior to 1990.

Table 6: Results for Difference-in-Differences Approach

Panel A										
	Non-Annex B countries						US states			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ATT	-0.554 (0.000)	-0.490 (0.000)	-0.460 (0.000)	-0.239 (0.000)	-0.316 (0.000)	0.133 (0.436)	-0.0879 (0.028)	-0.0339 (0.465)	-0.0199 (0.632)	0.105 (0.202)
<i>N</i>	5875	5875	5323	2131	3239	2380	2715	2163	1891	1412

Panel B					
	Non-Annex B countries		US states		
	(1)	(2)	(3)	(4)	
ATT	-0.152 (0.196)	-0.0916 (0.450)	0.00381 (0.925)	0.0281 (0.554)	
<i>N</i>	167	156	77	66	

Notes: *p*-values in parenthesis. We employ clustered standard errors in Panel A and robust (Abadie and Imbens 2006, 2008, 2011) standard errors in Panel B. Column 1 of Panel A shows the estimate for the most basic DiD specification only including a dummy for Annex B countries and for the period after the treatment took place (ratification in 2002) and their interaction. Instead, column 2 uses standard country and year fixed-effects in addition to the treatment dummy. Column 3 is identical to column 2 but we restricted the treated sample to the major polluters also covered in the SCM analysis. Column 4 is identical to column 3 with all predictors used in the SCM analysis as covariates. In Column 5 we added only those variables used in our SCM analysis that can be treated as being unaffected by the treatment, i.e. life expectancy, population growth and the human capital index (otherwise identical to column 3). Finally, in column 6 we restricted the control group to have the same pre-treatment range of average CO₂ emission trends, i.e. average year-to-year emission changes of the control group are above the minimum of the 15 Annex B countries and below the maximum (otherwise identical to column 4). Estimates in columns 7–10 are based on the dataset that uses US states as the control group. Columns 7, 8, 9 and 10 are the equivalent of columns 2, 3, 5, and 6 of Panel A when using US state level data. In columns 1 and 2 of Panel B we show propensity score matching based DiD results for the full sample (1) and for the major polluters (2). Columns 3 and 4 of Panel B are the equivalent of columns 1 and 2 when using US state level data.

the post treatment period (assuming ratification in 2002 as the treatment event) and their interaction. The results of this interaction term on log CO₂ emissions are reported in column (1) in Panel A of Table (6). We find a statistically highly significant negative treatment effect of -55.4%. In column (2) we also include country and year fixed-effects, resulting in a highly significant and slightly smaller negative treatment effect of -49.0%. In column (3) we restrict attention to the 15 treated countries that we analyze with the SCM. Again, we find a highly significant and large negative treatment effect of -46.0%. In column (4) we add all the predictors that we use in the SCM analysis as covariates, rendering a highly significant negative treatment effect of -23.9%. As some of these covariates may constitute bad controls (see also Section 2), we only include life expectancy, population growth, the human capital

index, the freedom house and the polity2 index in column (5), as we consider them safe from being influenced by the treatment itself. We find a highly significant negative treatment effect of -31.6%. Finally, in column (6) we account for the violation of the common trend hypothesis by restricting the control group to have the same pre-treatment range of average CO₂ emission trends as the 15 treated countries (otherwise the specification is identical to column (5)) and find a non-significant positive treatment effect of 13.3%.

In columns (7)–(10) of Panel A we report DiD results when using US states as the control group. Column (7) corresponds to column (2), i.e. apart from dummies for treated countries and the post treatment period, we only include country (respectively state) and year fixed-effects. We find a significant (on the 5% level) treatment effect of -8.8%. Restricting the analysis to the 15 countries analyzed with the SCM results in a non-significant negative treatment effect of -3.4%. Controlling, in addition, for all covariates that are not even potentially influenced by the treatment, results in a non-significant negative treatment effect of -2.0%. Finally, restricting the control group to the same pre-treatment range of average CO₂ emission trends as the 15 treated countries to account for the violation of the common trend hypothesis yields an insignificant positive treatment effect of 10.5%.

Using average pre treatment CO₂ emissions as the only predictor, we show in Panel B of Table 6 propensity score matching based DiD results for all Annex B countries (columns (1) and (3)) and for the 15 countries analyzed with SCM (columns (2) and (4)), when using non-Annex B countries (columns (1) and (2)) and US states (columns (3) and (4)) as the control group. We find non-significant negative treatments effects for non-Annex B and non-significant positive treatment effects for US states as the control group.

In summary, we find that the DiD approach may easily generate large and significant treatment effects. However, these are artifacts of the violated common trend assumption. Taking measures to control for the different pre treatment CO₂ emission trends renders the treatment effects insignificant.

6 Discussion and Conclusion

The KP has been widely criticized by the public press and the scientific community alike. In particular, issues concerning equity, efficiency and cost-effectiveness have been raised. In this paper, we asked in how far the KP lived up to its primary goal, the reduction of domestic GHG emissions in the industrialized world. To answer this question, we analyzed the development of CO₂ emissions for major GHG emitters with binding emission targets under the KP by employing the synthetic control method. We find very little evidence for a significant emission reduction effect for all 15 investigated countries, i.e. countries with

binding emissions targets did not emit less CO₂ over the period from 1998–2011 than they would have had in the absence of the GHG emission targets.

Yet, we do find some evidence that some Annex B countries performed even worse than comparable countries from the two donor pools. On average, the CO₂ emissions of the 15 countries are rather above than below their synthetic controls when using US states as the control group (see Figure 3). This is particularly puzzling, as the violation of the SUTVA (see Section 1) should rather bias our results in the direction of negative significance. However, these positive treatment effects are restricted to very few countries. In fact, the three Scandinavian countries Finland, Norway and Sweden are responsible for the vast majority of significant positive treatment effects in our main specifications (14 out of 20) and in almost all robustness checks. In addition, these countries (like other small countries) show highly volatile CO₂ emission paths which makes the construction of counterfactual emission paths difficult.²³ This conjecture is supported by the robustness check in which we use GHG instead of CO₂ emissions. Our GHG emission data show much lower year to year variations than our CO₂ data, in particular for the three Scandinavian countries. In this specification these countries do not exhibit any positive treatment effects.

Our results show the importance of addressing a number of empirical challenges when estimating the effect of international environmental policies in general, and the KP in particular. The key to identifying the “true” treatment effect is the availability and the selection of appropriate controls. In order to do so, our empirical strategy differs from previous studies in three important aspects:

First, we employ the synthetic control method (Abadie and Gardeazabal 2003, Abadie et al. 2010, 2015), which enables us to construct counterfactual emissions paths for all 15 countries under investigation that reasonably match the observed emissions in the pre-treatment period. This is possible because the treatment effects can be estimated for each country individually, which allows to individually identify the counterfactual synthetic country to the idiosyncrasies of each treated country, such as different country characteristics, emission paths, targets and ratification dates.

Second, using non-Annex B country data to construct counterfactual CO₂ emission paths for the investigated Annex B countries may be problematic because of (i) irreconcilable structural differences between Annex B and non-Annex B countries including opposing pre-treatment emission paths and (ii) a bias resulting from the use of the flexibility mechanisms blurring the distinction between treated and non-treated countries. To circumvent these problems we run a second specification using US state-level data to construct synthetic counterfactuals. Although US state-level data comes at the disadvantage that data on some

²³ Obviously, this not just a problem for the SCM but for any econometric method.

covariates is not available, it seems to be preferable to using non-Annex B country data on the ground that (i) the imbalance in predictors is much smaller and, as a result, (ii) pre-treatment matching is considerably better.

Third, we discard all Eastern European countries from the analysis. We consider this justified, as the former “Eastern Block” experienced a severe economic downturn after the collapse of the Former Soviet Union at the beginning of the 1990s. Even after their economic recovery, former CO₂ emission levels were not reached and, at the time of the adoption of the KP, it was evident that many Eastern European countries were unlikely to reach or even exceed their GHG emission targets (also known as “hot air”). As a consequence, these countries had no incentive to reduce GHG emissions in the first place. Nevertheless, the collapse of the FSU had a significant negative impact on the CO₂ emissions of the Eastern European countries. Incorporating these countries into the treated sample when eliciting the average treatment effect of the treated may bias the result in favor of a negative significant effect if one is not able to control for the peculiarities of those countries in particular with respect to the development of CO₂ emission in the early 90s.

We believe that the empirical challenges faced in the present paper apply to many international environmental policies that can only be evaluated at the country level. Countries tend to be highly heterogeneous with significant differences in their (socio-)economic and political characteristics. As a result, there are numerous issues that need to be addressed, as discussed in Sections 1 and 2. Accounting for all of these challenges simultaneously can be difficult when employing standard panel data analysis. The synthetic control method or extensions of matching based DiD methods may be better suited to address these challenges than standard panel data approaches.

Finally, our results have important political implications. In December 2015 the UNFCCC managed to reach a consensus on the “Paris Agreement” that outlines the main pillars of the successor of the KP which is to become effective by 2020. In the Paris Agreement reduction targets are called “nationally determined contributions”. Like the KP, the Paris Agreement is lacking (so far) any enforcement mechanism for non-compliance. According to our results, the KP had no verifiable effect on the CO₂ emissions of ratifying Annex B countries. As a consequence, we are pessimistic that the Paris Agreement, which rests on similar principles, will have any discernible effect on the reduction of global GHG emissions.²⁴ In fact, bilateral or smaller multilateral emission reduction agreements, where commitment is in the best

²⁴ Note, however, that the present analysis focuses on national emission reductions from the date of ratification of the KP to 2011 which approximately covers the first commitment period (2008–2012). That is, we cannot rule out that there may be emission reduction effects after 2011 nor can we rule out sub-national or sectoral effects, as we focus on the question whether Kyoto emission targets led to a reduction in national GHG emissions of Annex B countries.

interest of all participating countries, may be more effective in reducing GHG emissions (e.g., Carbone et al. 2009).

References

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association* 105, 493–505.
- Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science* 59, 495–510.
- Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review* 93, 113–32.
- Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* 74, 235–67.
- Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica* 76, 1537–57.
- Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics* 29, 1–11.
- Aichele, R. and G. Felbermayr (2012). Kyoto and the carbon footprint of nations. *Journal of Environmental Economics and Management* 63, 336–54.
- Aichele, R. and G. Felbermayr (2013). The effect of the Kyoto Protocol on carbon emissions. *Journal of Policy Analysis and Management* 32, 731–57.
- Angrist, J. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Barrett, S. (1998). Political economy of the Kyoto Protocol. *Oxford Review of Economic Policy* 14, 20–39.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–75.
- Carbone, J. C., C. Helm, and T. F. Rutherford (2009). The case for international emission trade in the absence of cooperative climate policy. *Journal of Environmental Economics and Management* 58, 266–80.
- Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics* 95, 1549–61.
- Copeland, B. R. and M. S. Taylor (2005). Free trade and global warming: a trade theory view of the Kyoto Protocol. *Journal of Environmental Economics and Management* 49, 205–34.

- Finus, M. (2008). Game theoretic research on the design of international environmental agreements: Insights, critical remarks, and future challenges. *International Review of Environmental and Resource Economics* 2, 29–67.
- Grunewald, N. and I. Martinez-Zarzoso (2016). Did the Kyoto Protocol fail? An evaluation of the effect of the Kyoto Protocol on CO₂ emissions. *Environment and Development Economics* 21, 1–22.
- Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 5–86.
- Prins, G. and S. Rayner (2007). Time to ditch Kyoto. *Nature* 449, 973–5.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63, 581–92.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6, 34–58.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association* 100, 322–31.
- The Economist (1997). Global warming. Rubbing sleep from their eyes. *The Economist (US edition)*. December 11th 1997.
- Tol, R. S. J. (2000). Kyoto, efficiency, and cost-effectiveness: applications of FUND. *Energy Journal* 21, 131–56.
- UNDP (1990). *Human Development Report 1990*. New York, Oxford: Oxford University Press. Published for the United Nations Development Programme.
- Zhang, J. and C. Wang (2011). Co-benefits and additionality of the clean development mechanism: An empirical analysis. *Journal of Environmental Economics and Management* 62, 140–54.

A Appendix

A.1 Ratification of the KP as the treatment event

Table 7: Weights for using non-Annex B countries in the donor pool

Australia:	Chile (0.114), Ecuador (0.040), Malaysia (0.056), Panama (0.077), Singapore (0.002), USA (0.711)
Austria:	Gabon (0.031), South Korea (0.197), Thailand (0.032), USA (0.740)
Belgium:	Gabon (0.052), USA (0.948)
Canada:	China (0.017), South Korea (0.207), Singapore (0.029) USA (0.747)
Finland:	China (0.021), Gabon (0.057), South Korea (0.220), USA (0.702)
France:	Colombia (0.153), Gabon (0.113), Mexico (0.045), Mauritius (0.106), USA (0.584)
Germany:	Brazil (0.057), Gabon (0.131), Mexico (0.112), USA (0.701)
Great Britain:	Gabon (0.047), South Korea (0.144), USA (0.809),
Italy:	China (0.014), Gabon (0.012), South Korea (0.126), USA (0.848)
Japan:	Gabon (0.019), South Korea (0.335), USA (0.645)
Netherlands:	Gabon (0.051), South Korea (0.045), USA (0.904)
Norway:	Gabon (0.008), Thailand (0.137), USA (0.855)
Portugal:	Brazil (0.199), Mauritius (0.394), Thailand (0.203), USA (0.205)
Spain:	Fiji (0.073), South Korea (0.148), Mauritius (0.202), USA (0.577)
Sweden:	Gabon (0.158), Mexico (0.058), Mauritius (0.018), USA (0.766)

Note: Country Weights for Results in Table 2. The countries in the donor pool, i.e. countries that may receive positive weights, include all non-Annex B countries with sufficient data except for countries classified as low income and lower-middle income according to the World Bank. More specifically, the donor pool/control group consists of (ISO 3 country codes): BRA, BWA, CHL, CHN, COL, DOM, ECU, FJI, GAB, JOR, KOR, MEX, MUS, MYS, PAN, SAU, SGP, THA, TUN, TUR, USA, VEN, and ZAF.

Table 8: Weights for using US states in the Donor Pool

Australia:	Alaska (0.007), Arizona (0.133), Georgia (0.217), Hawaii (0.042), Louisiana (0.177), Nebraska (0.020), New Hampshire (0.014), North Carolina (0.036), South Carolina (0.0190), Utah (0.335), Wyoming (0.001)
Austria:	California (0.149), Delaware (0.250), Mississippi (0.069), Ohio (0.012), Oregon (0.161), Rhode Island (0.094), South Dakota (0.220), Wyoming (0.044)
Belgium:	California (0.030), Hawaii (0.034), Illinois (0.055), Louisiana (0.278), Ohio (0.602)
Canada:	Alaska (0.003), Arizona (0.131), California (0.021), Delaware (0.107), Hawaii (0.120), Illinois (0.201), Louisiana (0.048), Mississippi (0.0190), Nevada (0.165), Oregon (0.080), South Carolina (0.104), Wyoming (0.001)
Finland:	California (0.199), Massachusetts (0.280), New Hampshire (0.103), Ohio (0.260), Rhode Island (0.158)
France:	Delaware (0.077), District of Columbia (0.372), Illinois (0.108), Ohio (0.443)
Germany:	District of Columbia (0.829), Ohio (0.171)
Great Britain:	Delaware (0.292), District of Columbia (0.035), Massachusetts (0.225), Ohio (0.450)
Italy:	California (0.054), Connecticut (0.274), Delaware (0.127), Hawaii (0.026), Idaho (0.299), Louisiana (0.124), Mississippi (0.002), South Dakota (0.093)
Japan:	Hawaii (0.089), Idaho (0.272), Louisiana (0.323), Montana (0.007), New York (0.014), North Dakota (0.068), Oklahoma (0.002), Oregon (0.029), Rhode Island (0.054), South Dakota (0.075), Utah (0.046), Wyoming (0.021)
Netherlands:	California (0.218), Hawaii (0.092), Louisiana (0.314), North Dakota (0.002), Ohio (0.372), Wyoming (0.001)
Norway:	Arizona (0.218), District of Columbia (0.034), Maryland (0.139), Mississippi (0.268), New Hampshire (0.053), West Virginia (0.288)
Portugal:	Nevada (1.0)
Spain:	Arizona (0.492), Idaho (0.049), Mississippi (0.058), Missouri (0.253), Nevada (0.070), Oregon (0.046), South Dakota (0.032)
Sweden:	District of Columbia (0.768), Ohio (0.232)

Note: US state weights for results in Table 3. All US states are included in the donor pool, i.e. may receive positive weights.

A.2 Adoption of the KP as the treatment event

Table 9: Treatment in 1997: Estimates for Treatment Effects based on non-Annex B countries

Country	Year													
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia	.01 (.90)	-.08 (.50)	-.11 (.30)	-.13 (.45)	-.09 (.40)	-.09 (.50)	-.10 (.50)	-.08 (.60)	-.07 (.65)	-.08 (.65)	-.05 (.80)	.02 (.95)	-.08 (.70)	-.07 (.70)
	RMSPE: 0.0254													
Austria	.02 (.90)	-.05 (.75)	-.09 (.40)	-.08 (.50)	-.04 (.80)	.02 (.85)	-.03 (.75)	-.03 (.80)	-.07 (.75)	-.10 (.60)	-.08 (.65)	-.10 (.65)	-.10 (.65)	-.10 (.65)
	RMSPE: 0.0582													
Belgium	.03 (.80)	-.03 (.75)	-.07 (.55)	-.10 (.45)	-.12 (.40)	-.07 (.55)	-.16 (.35)	-.21 (.30)	-.22 (.35)	-.23 (.35)	-.19 (.50)	-.11 (.65)	-.16 (.45)	-.22 (.45)
	RMSPE: 0.0965													
Canada	.03 (.75)	.01 (.95)	.03 (.70)	.01 (.95)	-.02 (.85)	.03 (.85)	.01 (.90)	.01 (.90)	.01 (.95)	.01 (1.0)	.00 (.95)	-.01 (1.0)	-.07 (.80)	-.09 (.65)
	RMSPE: 0.0260													
Finland	-.07 (.55)	-.12 (.35)	-.19 (.20)	-.15 (.40)	-.05 (.70)	.08 (.50)	-.00 (1.0)	-.21 (.30)	-.04 (.85)	-.09 (.65)	-.20 (.40)	-.19 (.50)	-.15 (.45)	-.26 (.40)
	RMSPE: 0.0888													
France	.12 (.20)	-.02 (.90)	-.07 (.55)	-.08 (.45)	-.03 (.80)	-.02 (.85)	-.10 (.50)	-.15 (.45)	-.16 (.40)	-.16 (.50)	-.11 (.60)	-.03 (.95)	-.17 (.45)	-.20 (.50)
	RMSPE: 0.1737													
Germany	.05 (.70)	.02 (.95)	.04 (.70)	.01 (1.0)	-.04 (.80)	-.04 (.75)	-.10 (.50)	-.15 (.50)	-.14 (.50)	-.19 (.40)	-.14 (.55)	-.10 (.70)	-.23 (.45)	-.25 (.40)
	RMSPE: 0.1546													
Great Britain	.01 (.90)	-.01 (1.0)	.01 (.95)	-.01 (.95)	-.07 (.60)	-.06 (.70)	-.08 (.55)	-.11 (.60)	-.09 (.60)	-.14 (.50)	-.11 (.65)	-.12 (.65)	-.18 (.45)	-.25 (.40)
	RMSPE: 0.1087													
Italy	.01 (.90)	-.03 (.75)	-.05 (.60)	-.08 (.50)	-.04 (.80)	-.02 (.85)	-.07 (.55)	-.10 (.60)	-.11 (.50)	-.11 (.60)	-.11 (.60)	-.13 (.60)	-.20 (.45)	-.19 (.50)
	RMSPE: 0.0440													
Japan	-.03 (.80)	-.03 (.80)	-.04 (.65)	-.07 (.65)	-.08 (.40)	-.07 (.55)	-.06 (.60)	-.08 (.65)	-.07 (.75)	-.08 (.70)	-.11 (.60)	-.16 (.55)	-.16 (.45)	-.14 (.50)
	RMSPE: 0.0726													
Netherlands	-.04 (.70)	-.10 (.45)	-.11 (.30)	-.13 (.45)	-.08 (.45)	-.07 (.55)	-.10 (.50)	-.17 (.45)	-.18 (.40)	-.15 (.50)	-.10 (.60)	-.05 (.95)	-.08 (.70)	-.12 (.55)
	RMSPE: 0.0538													
Norway	.10 (.25)	.23 (.25)	.14 (.30)	.20 (.25)	.09 (.40)	.24 (.25)	.20 (.25)	.18 (.45)	.25 (.35)	.24 (.35)	.45 (.25)	.49 (.20)	.66 (.15)	.36 (.40)
	RMSPE: 0.0756													
Portugal	.02 (.80)	.04 (.75)	-.08 (.40)	-.13 (.35)	-.07 (.40)	-.18 (.25)	-.18 (.25)	-.20 (.30)	-.33 (.15)	-.32 (.15)	-.35 (.15)	-.34 (.20)	-.44 (.15)	-.44 (.15)
	RMSPE: 0.0516													
Spain	.03 (.80)	.06 (.60)	.04 (.70)	.03 (.85)	.10 (.40)	.10 (.45)	.12 (.35)	.12 (.50)	.09 (.50)	.11 (.60)	.04 (.80)	-.04 (.95)	-.16 (.45)	-.13 (.50)
	RMSPE: 0.1025													
Sweden	.10 (.40)	.03 (.85)	-.02 (.80)	-.07 (.65)	.11 (.40)	.07 (.55)	-.04 (.70)	-.15 (.50)	-.17 (.40)	-.18 (.45)	-.10 (.65)	-.06 (.95)	-.09 (.70)	-.06 (.85)
	RMSPE: 0.2002													

Note: The Table contains treatment effect estimates for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year 1997 (the year of adoption of the KP) as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Table 10: Treatment in 1997: Estimates for Treatment Effects based on US states

Country	Year													
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia	.02 (.60)	-.04 (.30)	-.05 (.30)	-.06 (.35)	.01 (.80)	.01 (.85)	-.02 (.75)	.01 (.75)	.05 (.45)	.05 (.70)	.11 (.30)	.19 (.10)	.12 (.35)	.14 (.30)
	RMSPE: 0.0207													
Austria	-.02 (.60)	-.07 (.15)	-.05 (.35)	.00 (1.0)	.05 (.45)	.12 (.20)	.10 (.30)	.12 (.20)	.13 (.25)	.04 (.85)	.06 (.50)	.07 (.75)	.15 (.30)	.15 (.30)
	RMSPE: 0.0435													
Belgium	.04 (.25)	.01 (.65)	-.02 (.65)	.02 (.80)	-.07 (.30)	-.03 (.70)	-.05 (.35)	-.09 (.35)	-.10 (.30)	-.15 (.25)	-.11 (.30)	-.01 (.90)	-.02 (.95)	-.08 (.50)
	RMSPE: 0.0576													
Canada	.09 (.05)	.07 (.15)	.05 (.35)	.09 (.35)	.06 (.35)	.11 (.20)	.07 (.35)	.08 (.35)	.05 (.50)	.05 (.65)	.06 (.50)	.12 (.35)	.05 (.70)	.02 (.90)
	RMSPE: 0.0172													
Finland	-.05 (.15)	-.07 (.20)	-.13 (.05)	-.04 (.40)	.02 (.60)	.08 (.30)	.05 (.35)	-.15 (.10)	.09 (.30)	.04 (.75)	-.05 (.80)	-.04 (.90)	.12 (.50)	.02 (.90)
	RMSPE: 0.0642													
France	.10 (.05)	-.00 (.90)	-.06 (.30)	.04 (.40)	.00 (.95)	.05 (.45)	.05 (.35)	.06 (.45)	.16 (.25)	.11 (.30)	.17 (.25)	.16 (.30)	.13 (.40)	.12 (.40)
	RMSPE: 0.0792													
Germany	.01 (.65)	-.02 (.50)	-.06 (.25)	.01 (.85)	-.04 (.45)	.03 (.70)	.00 (1.0)	-.00 (.90)	.19 (.20)	.12 (.30)	.23 (.15)	.13 (.50)	.11 (.55)	.16 (.30)
	RMSPE: 0.0473													
Great Britain	-.00 (.90)	-.03 (.45)	-.04 (.40)	.01 (.90)	-.03 (.55)	-.00 (.90)	-.01 (.85)	-.02 (.70)	.07 (.40)	-.00 (1.0)	.04 (.85)	.07 (.80)	.10 (.55)	.02 (.90)
	RMSPE: 0.0329													
Italy	-.03 (.45)	-.06 (.20)	-.01 (.80)	-.01 (.85)	-.00 (.95)	.02 (.75)	.01 (.85)	-.00 (.90)	.02 (.85)	-.03 (.95)	.00 (1.0)	-.05 (.80)	-.04 (.75)	-.03 (.70)
	RMSPE: 0.0241													
Japan	-.05 (.15)	-.07 (.20)	-.08 (.10)	-.09 (.35)	-.06 (.35)	-.02 (.75)	-.06 (.35)	-.09 (.35)	-.09 (.25)	-.11 (.25)	-.09 (.30)	-.14 (.25)	-.12 (.30)	-.09 (.40)
	RMSPE: 0.0254													
Netherlands	-.03 (.40)	-.07 (.20)	-.09 (.10)	-.04 (.40)	-.02 (.60)	-.01 (.85)	-.02 (.65)	-.05 (.50)	-.08 (.35)	-.08 (.40)	-.03 (.90)	.02 (.90)	.06 (.70)	.02 (.90)
	RMSPE: 0.0373													
Norway	.06 (.10)	.19 (.05)	.09 (.10)	.20 (.10)	.05 (.45)	.19 (.05)	.17 (.10)	.18 (.10)	.31 (.05)	.32 (.05)	.55 (.05)	.57 (.05)	.86 (.05)	.55 (.05)
	RMSPE: 0.0535													
Portugal	.05 (.15)	.21 (.05)	.19 (.05)	.20 (.10)	.30 (.05)	.19 (.05)	.22 (.10)	.23 (.10)	.12 (.25)	.11 (.25)	.09 (.30)	.11 (.35)	.03 (.85)	-.03 (.75)
	RMSPE: 0.0723													
Spain	.00 (.95)	.05 (.20)	.10 (.10)	.12 (.25)	.19 (.05)	.19 (.05)	.24 (.05)	.30 (.05)	.29 (.05)	.28 (.05)	.17 (.10)	.07 (.60)	-.02 (.90)	.01 (.90)
	RMSPE: 0.0295													
Sweden	.04 (.20)	.01 (.85)	-.06 (.25)	.01 (.85)	.11 (.20)	.12 (.25)	.09 (.30)	.05 (.50)	.19 (.20)	.12 (.25)	.25 (.10)	.10 (.60)	.28 (.10)	.35 (.05)
	RMSPE: 0.0882													

Note: The Table contains treatment effect estimates for each Annex B country under investigation using US states to construct the synthetic counterfactuals and considering the year 1997 (the year of adoption of the KP) as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

A.3 KP's entering into force as the treatment event

Table 11: Treatment in 2004: Estimates for Treatment Effects based on non-Annex B Countries (left) and US states (right)

Country	non-Annex B countries							US states						
	Year							Year						
	2005	2006	2007	2008	2009	2010	2011	2005	2006	2007	2008	2009	2010	2011
Australia	.03 (.70)	.05 (.65)	.04 (.85)	-.09 (.65)	.18 (.40)	.12 (.55)	.11 (.60)	.03 (.25)	.06 (.15)	.07 (.20)	.12 (.10)	.23 (.05)	.14 (.05)	.13 (.10)
	RMSPE: 0.0330							RMSPE: 0.0222						
Austria	.01 (1.0)	-.02 (.80)	-.07 (.65)	-.10 (.60)	-.13 (.50)	-.09 (.65)	-.10 (.65)	.03 (.25)	.00 (.95)	-.05 (.35)	-.02 (.65)	-.01 (.85)	.03 (.80)	.02 (.85)
	RMSPE: 0.0587							RMSPE: 0.0392						
Belgium	-.04 (.55)	-.08 (.45)	-.09 (.55)	-.11 (.55)	-.06 (.80)	-.01 (.95)	-.15 (.50)	-.04 (.20)	-.00 (.95)	-.06 (.25)	-.02 (.70)	.05 (.55)	.06 (.55)	-.00 (1.0)
	RMSPE: 0.0847							RMSPE: 0.0593						
Canada	.00 (1.0)	-.01 (.90)	-.01 (.90)	-.03 (.85)	-.03 (.95)	-.08 (.70)	-.09 (.85)	.00 (.75)	-.03 (.45)	-.04 (.50)	-.01 (.85)	-.01 (.85)	-.06 (.55)	-.06 (.40)
	RMSPE: 0.0330							RMSPE: 0.0185						
Finland	-.20 (.05)	-.02 (.80)	-.10 (.50)	-.20 (.20)	-.22 (.35)	-.12 (.50)	-.20 (.30)	-.18 (.05)	.06 (.15)	.03 (.65)	-.07 (.20)	-.03 (.85)	.12 (.20)	.03 (.75)
	RMSPE: 0.0726							RMSPE: 0.0541						
France	-.02 (.70)	-.06 (.50)	-.05 (.75)	-.10 (.60)	-.05 (.80)	-.06 (.75)	-.16 (.45)	-.00 (.90)	.06 (.15)	.02 (.80)	.06 (.25)	.08 (.30)	.06 (.70)	.03 (.75)
	RMSPE: 0.0978							RMSPE: 0.0669						
Germany	-.06 (.35)	-.04 (.70)	-.08 (.60)	-.03 (.85)	.01 (1.0)	-.09 (.65)	-.11 (.60)	-.00 (.90)	.19 (.05)	.12 (.05)	.23 (.05)	.13 (.10)	.11 (.20)	.16 (.15)
	RMSPE: 0.1516							RMSPE: 0.0453						
Great Britain	-.02 (.70)	-.04 (.80)	-.03 (.85)	-.07 (.75)	-.10 (.65)	-.08 (.70)	-.20 (.35)	-.02 (.40)	.06 (.15)	-.01 (.95)	.03 (.60)	.10 (.25)	.12 (.20)	.03 (.80)
	RMSPE: 0.0862							RMSPE: 0.0286						
Italy	-.03 (.65)	-.05 (.65)	-.05 (.75)	-.15 (.30)	-.19 (.40)	-.18 (.45)	-.21 (.30)	-.00 (1.0)	-.00 (1.0)	-.04 (.50)	-.03 (.65)	-.04 (.65)	-.07 (.50)	-.07 (.35)
	RMSPE: 0.0424							RMSPE: 0.0183						
Japan	-.04 (.55)	-.04 (.75)	-.03 (.90)	-.06 (.75)	-.10 (.60)	-.04 (.75)	-.04 (.90)	-.02 (.40)	-.03 (.40)	-.04 (.35)	-.04 (.35)	-.09 (.20)	-.06 (.55)	-.04 (.65)
	RMSPE: 0.0410							RMSPE: 0.0165						
Netherlands	-.04 (.55)	-.08 (.45)	-.06 (.70)	-.06 (.75)	-.02 (1.0)	.04 (.75)	-.05 (.90)	-.03 (.25)	-.06 (.15)	-.06 (.25)	.01 (.80)	.07 (.35)	.10 (.20)	.04 (.65)
	RMSPE: 0.0516							RMSPE: 0.0343						
Norway	-.01 (.85)	.04 (.70)	.05 (.75)	.20 (.25)	.19 (.40)	.41 (.25)	.14 (.55)	-.01 (.70)	.04 (.30)	.04 (.35)	.20 (.05)	.29 (.05)	.49 (.05)	.23 (.05)
	RMSPE: 0.0775							RMSPE: 0.0579						
Portugal	-.01 (.95)	-.17 (.10)	-.14 (.40)	-.25 (.05)	-.23 (.35)	-.27 (.30)	-.33 (.25)	-.01 (.75)	.03 (.40)	.05 (.30)	.03 (.65)	.06 (.55)	.01 (.85)	.03 (.75)
	RMSPE: 0.0515							RMSPE: 0.0747						
Spain	.03 (.70)	-.02 (.80)	-.09 (.55)	-.15 (.30)	-.22 (.35)	-.29 (.25)	-.27 (.25)	.04 (.20)	.02 (.70)	.03 (.65)	-.05 (.25)	-.09 (.20)	-.16 (.05)	-.14 (.10)
	RMSPE: 0.0616							RMSPE: 0.0337						
Sweden	-.08 (.30)	-.10 (.35)	-.10 (.55)	-.06 (.75)	-.11 (.60)	.05 (.75)	.01 (1.0)	-.06 (.15)	.01 (.90)	-.05 (.50)	.03 (.65)	-.05 (.65)	.10 (.20)	.15 (.10)
	RMSPE: 0.1333							RMSPE: 0.0750						

Note: The Table contains treatment effect estimates for each Annex B country under investigation using non-Annex B countries and US states to construct the synthetic counterfactuals and considering the year 2004 (when the KP entered into force) as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

A.4 CO₂ intensity as dependent variable

Table 12: Estimates for Treatment Effects (CO₂ per GDP) based on non-Annex B Countries

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.07 (.25)	-.06 (.80)	-.00 (1.0)	.02 (.85)
	RMSPE: 0.0760									
Austria	.01 (.95)	.11 (.30)	.09 (.50)	.14 (.50)	.12 (.60)	.08 (.75)	.09 (.85)	.05 (.85)	.04 (.95)	.05 (.85)
	RMSPE: 0.1019									
Belgium	-.08 (.30)	.01 (.90)	-.02 (.95)	-.06 (.75)	.01 (.95)	-.03 (.90)	.01 (.95)	.07 (.80)	.00 (1.0)	-.02 (.95)
	RMSPE: 0.0926									
Canada		.06 (.45)	.03 (.75)	.03 (.80)	-.01 (1.0)	-.01 (.95)	-.02 (.90)	-.01 (.95)	-.10 (.55)	-.10 (.60)
	RMSPE: 0.0760									
Finland	.08 (.25)	.23 (.15)	.16 (.20)	-.06 (.75)	.17 (.35)	.08 (.75)	-.05 (.95)	-.03 (.90)	.10 (.70)	-.00 (1.0)
	RMSPE: 0.1536									
France	-.01 (.95)	.09 (.50)	.08 (.55)	.07 (.75)	.07 (.80)	.00 (1.0)	.05 (.95)	.15 (.65)	.00 (1.0)	-.00 (1.0)
	RMSPE: 0.1442									
Germany	-.00 (1.0)	.04 (.80)	-.00 (.95)	-.03 (.90)	-.01 (.95)	-.10 (.70)	-.04 (.95)	.17 (.65)	-.13 (.65)	-.09 (.75)
	RMSPE: 0.2135									
Great Britain	-.02 (.80)	.03 (.90)	-.00 (.95)	.00 (1.0)	.04 (.85)	-.00 (1.0)	.07 (.90)	.13 (.70)	-.02 (1.0)	-.04 (.85)
	RMSPE: 0.1951									
Italy	.06 (.30)	.13 (.30)	.15 (.20)	.17 (.35)	.17 (.40)	.11 (.70)	.07 (.90)	.02 (.95)	.02 (1.0)	.02 (.95)
	RMSPE: 0.0568									
Japan	.03 (.70)	.05 (.80)	.05 (.70)	.07 (.75)	.09 (.65)	.08 (.75)	.06 (.95)	.05 (.80)	.03 (.95)	.06 (.80)
	RMSPE: 0.0767									
Netherlands	.04 (.55)	.11 (.35)	.11 (.35)	.01 (.95)	.01 (.95)	-.01 (1.0)	.06 (.90)	.20 (.40)	.08 (.80)	.09 (.75)
	RMSPE: 0.1505									
Norway	-.05 (.45)	.12 (.30)	.05 (.70)	.07 (.75)	.15 (.45)	.16 (.45)	.46 (.15)	.31 (.20)	.49 (.05)	.38 (.20)
	RMSPE: 0.1035									
Portugal	.07 (.30)	-.00 (1.0)	.03 (.80)	.02 (.95)	-.07 (.70)	-.05 (.75)	-.08 (.85)	-.10 (.70)	-.19 (.50)	-.20 (.45)
	RMSPE: 0.0904									
Spain	.03 (.65)	.07 (.60)	.11 (.30)	.13 (.50)	.09 (.65)	.05 (.75)	-.03 (.95)	-.12 (.65)	-.18 (.50)	-.15 (.50)
	RMSPE: 0.0856									
Sweden	.13 (.20)	.12 (.30)	.06 (.70)	.04 (.90)	.01 (.95)	-.08 (.75)	.02 (.95)	.17 (.65)	.02 (.95)	.09 (.70)
	RMSPE: 0.2093									

Note: The Table contains treatment effect estimates for CO₂ emissions per GDP for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

A.5 CO₂ per capita as dependent variable

Table 13: Estimates for Treatment Effects (CO₂ per capita) based on non-Annex B Countries

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.04 (.50)	.11 (.35)	.00 (.95)	-.02 (.80)
	RMSPE: 0.0340									
Austria	.01 (.80)	.08 (.40)	.05 (.80)	.09 (.65)	.06 (.95)	.00 (1.0)	.01 (.95)	-.03 (1.0)	-.01 (.85)	-.03 (.90)
	RMSPE: 0.0570									
Belgium	-.06 (.35)	.00 (1.0)	-.05 (.75)	-.08 (.65)	-.08 (.90)	-.12 (.75)	-.09 (.85)	-.01 (1.0)	-.03 (.80)	-.11 (.60)
	RMSPE: 0.0592									
Canada		.04 (.70)	.02 (.85)	.02 (1.0)	.03 (.90)	.04 (.90)	.08 (.85)	.03 (.95)	-.02 (1.0)	.00 (1.0)
	RMSPE: 0.0483									
Finland	.07 (.30)	.22 (.15)	.16 (.40)	-.05 (.75)	.16 (.40)	.09 (.85)	-.01 (1.0)	-.01 (1.0)	.05 (.65)	-.06 (.75)
	RMSPE: 0.0838									
France	-.01 (.80)	.01 (1.0)	-.00 (1.0)	-.03 (.80)	-.05 (.95)	-.09 (.85)	-.08 (.85)	-.06 (.90)	-.13 (.45)	-.15 (.50)
	RMSPE: 0.0848									
Germany	-.02 (.70)	-.00 (1.0)	-.04 (.85)	-.08 (.65)	-.06 (.95)	-.10 (.80)	-.06 (.90)	-.02 (1.0)	-.11 (.50)	-.10 (.60)
	RMSPE: 0.0930									
Great Britain	-.04 (.60)	-.02 (.85)	-.04 (.85)	-.04 (.75)	-.02 (.95)	-.07 (.90)	-.05 (.95)	-.07 (.85)	-.11 (.50)	-.18 (.45)
	RMSPE: 0.0776									
Italy	-.01 (.80)	.02 (.75)	.01 (1.0)	.02 (.85)	.02 (.95)	-.02 (.90)	-.04 (.95)	-.10 (.65)	-.13 (.45)	-.14 (.55)
	RMSPE: 0.0329									
Japan	-.00 (1.0)	.02 (.80)	.02 (.95)	.02 (.85)	.02 (.95)	.01 (.95)	-.01 (1.0)	-.05 (.90)	-.06 (.65)	-.04 (.80)
	RMSPE: 0.0559									
Netherlands	.02 (.75)	.04 (.65)	.03 (.85)	.00 (1.0)	-.01 (.95)	-.00 (1.0)	.05 (.95)	.11 (.65)	.12 (.50)	.06 (.75)
	RMSPE: 0.0561									
Norway	-.10 (.15)	.02 (.85)	-.01 (1.0)	-.02 (.90)	.03 (.95)	.03 (.90)	.17 (.45)	.13 (.60)	.32 (.25)	.07 (.75)
	RMSPE: 0.0917									
Portugal	.06 (.35)	-.05 (.60)	-.04 (.85)	-.03 (.80)	-.16 (.35)	-.15 (.50)	-.19 (.35)	-.17 (.30)	-.28 (.25)	-.30 (.25)
	RMSPE: 0.0583									
Spain	.04 (.60)	.04 (.70)	.06 (.65)	.08 (.65)	.03 (.95)	.02 (.90)	-.06 (.90)	-.14 (.50)	-.24 (.25)	-.23 (.40)
	RMSPE: 0.0532									
Sweden	.12 (.15)	.09 (.40)	.05 (.75)	-.02 (.90)	-.04 (.95)	-.09 (.85)	-.03 (.95)	-.04 (1.0)	.04 (.75)	.07 (.75)
	RMSPE: 0.0909									

Note: The Table contains treatment effect estimates for CO₂ emissions per capita for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Table 14: Estimates for Treatment Effects (CO₂ per capita) based on US States

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.07 (.05)	.17 (.05)	.07 (.20)	.09 (.20)
	RMSPE: 0.0220									
Austria	.06 (.15)	.14 (.05)	.13 (.05)	.17 (.05)	.14 (.10)	.06 (.40)	.07 (.45)	.04 (.75)	.10 (.40)	.11 (.35)
	RMSPE: 0.0440									
Belgium	-.06 (.10)	.01 (.90)	-.07 (.10)	-.08 (.15)	-.15 (.05)	-.18 (.05)	-.13 (.25)	-.05 (.65)	-.06 (.50)	-.14 (.25)
	RMSPE: 0.0460									
Canada		.04 (.10)	-.01 (.65)	-.01 (.55)	.00 (.90)	-.01 (.90)	.01 (.90)	.04 (.70)	.00 (1.0)	-.00 (.90)
	RMSPE: 0.0207									
Finland	.07 (.10)	.15 (.05)	.10 (.10)	-.09 (.15)	.17 (.05)	.13 (.15)	.05 (.50)	.03 (.85)	.23 (.05)	.11 (.35)
	RMSPE: 0.0688									
France	.01 (.90)	.02 (.80)	-.02 (.60)	-.00 (1.0)	-.01 (.95)	-.03 (.50)	.01 (.90)	.09 (.50)	.08 (.50)	.02 (.80)
	RMSPE: 0.0745									
Germany	-.01 (.85)	.01 (.85)	-.01 (.75)	-.03 (.55)	-.00 (1.0)	-.05 (.40)	.01 (.80)	.10 (.45)	.15 (.25)	.13 (.35)
	RMSPE: 0.0516									
Great Britain	-.01 (.85)	.00 (.95)	-.00 (.95)	-.02 (.85)	.02 (.60)	-.03 (.50)	.01 (.95)	.02 (.85)	.06 (.60)	-.02 (.85)
	RMSPE: 0.0264									
Italy	.04 (.30)	.06 (.15)	.05 (.25)	.05 (.30)	.03 (.55)	.01 (.85)	.01 (.90)	-.04 (.75)	-.06 (.50)	-.05 (.60)
	RMSPE: 0.0212									
Japan	.02 (.50)	.05 (.35)	.04 (.30)	.03 (.55)	-.00 (1.0)	.01 (.90)	-.02 (.70)	-.06 (.60)	-.05 (.60)	-.03 (.65)
	RMSPE: 0.0250									
Netherlands	.02 (.50)	.06 (.20)	.03 (.35)	.02 (.70)	-.04 (.45)	-.02 (.60)	.06 (.50)	.10 (.30)	.16 (.15)	.09 (.40)
	RMSPE: 0.0354									
Norway	-.04 (.30)	.03 (.40)	.00 (.85)	.01 (.95)	.08 (.25)	.08 (.35)	.25 (.05)	.25 (.05)	.46 (.05)	.23 (.15)
	RMSPE: 0.0575									
Portugal	.13 (.05)	.02 (.80)	.06 (.20)	.08 (.15)	-.03 (.55)	-.03 (.40)	-.05 (.50)	-.01 (.90)	-.12 (.25)	-.14 (.25)
	RMSPE: 0.1361									
Spain	.11 (.05)	.11 (.05)	.13 (.05)	.19 (.05)	.13 (.10)	.10 (.20)	.02 (.60)	-.06 (.65)	-.17 (.05)	-.12 (.25)
	RMSPE: 0.0400									
Sweden	.15 (.05)	.10 (.05)	.07 (.10)	.02 (.85)	-.00 (1.0)	-.05 (.40)	.01 (.80)	.05 (.65)	.26 (.05)	.24 (.15)
	RMSPE: 0.0835									

Note: The Table contains treatment effect estimates for CO₂ emissions per capita for each Annex B country under investigation using US states to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

A.6 Accounting for other GHGs

Table 15: Estimates for Treatment Effects based on non-Annex B Countries for all GHGs

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							-.02 (.65)	-.03 (.80)	-.07 (.50)	.07 (.50)
	RMSPE: 0.0371									
Austria	.02 (.90)	.11 (.35)	.06 (.80)	.08 (.50)	.03 (.85)	-.01 (.95)	.00 (1.0)	-.09 (.70)	-.06 (.70)	-.10 (.55)
	RMSPE: 0.0768									
Belgium	-.05 (.50)	-.00 (1.0)	-.07 (.70)	-.08 (.50)	-.12 (.50)	-.15 (.50)	-.13 (.45)	-.18 (.40)	-.14 (.40)	-.14 (.45)
	RMSPE: 0.0319									
Canada		.04 (.45)	.01 (.95)	.03 (.85)	.01 (.80)	.04 (.80)	.04 (.65)	.02 (.95)	-.02 (.75)	-.01 (.95)
	RMSPE: 0.0186									
Finland	.03 (.70)	.12 (.30)	.02 (.90)	-.13 (.45)	-.03 (.85)	-.05 (.70)	-.16 (.45)	-.18 (.40)	-.08 (.50)	-.19 (.35)
	RMSPE: 0.0636									
France	-.02 (.95)	.00 (1.0)	-.04 (.90)	-.02 (.80)	-.06 (.70)	-.08 (.70)	-.09 (.70)	-.13 (.50)	-.14 (.40)	-.21 (.35)
	RMSPE: 0.0943									
Germany	-.03 (.70)	-.03 (.90)	-.08 (.65)	-.10 (.50)	-.08 (.60)	-.14 (.60)	-.11 (.65)	-.15 (.45)	-.15 (.35)	-.19 (.35)
	RMSPE: 0.1604									
Great Britain	-.03 (.65)	-.03 (.85)	-.08 (.65)	-.09 (.50)	-.12 (.50)	-.13 (.60)	-.20 (.35)	-.26 (.35)	-.21 (.25)	-.29 (.20)
	RMSPE: 0.1333									
Italy	.02 (.95)	.06 (.65)	.03 (.90)	.04 (.75)	.00 (.95)	-.02 (.85)	-.06 (.80)	-.15 (.45)	-.16 (.30)	-.20 (.35)
	RMSPE: 0.0723									
Japan	.03 (.65)	.03 (.85)	.01 (.95)	.01 (1.0)	-.00 (1.0)	-.01 (.95)	-.06 (.80)	-.07 (.70)	-.08 (.55)	-.05 (.80)
	RMSPE: 0.0645									
Netherlands	-.00 (1.0)	.00 (1.0)	-.03 (.90)	-.05 (.70)	-.08 (.60)	-.07 (.70)	-.07 (.75)	-.08 (.70)	-.05 (.70)	-.11 (.55)
	RMSPE: 0.0685									
Norway	-.01 (.95)	.05 (.65)	.03 (.90)	.01 (1.0)	.03 (.85)	.01 (.90)	.02 (.90)	.04 (.90)	.03 (.95)	-.00 (1.0)
	RMSPE: 0.0814									
Portugal	.05 (.50)	-.01 (.90)	-.03 (.90)	-.00 (1.0)	-.11 (.50)	-.12 (.60)	-.19 (.40)	-.20 (.40)	-.29 (.15)	-.31 (.20)
	RMSPE: 0.0398									
Spain	.04 (.55)	.07 (.55)	.06 (.70)	.10 (.50)	.07 (.70)	.07 (.70)	-.02 (.90)	-.12 (.50)	-.21 (.25)	-.24 (.25)
	RMSPE: 0.0602									
Sweden	.03 (.65)	.02 (.90)	-.02 (.90)	-.08 (.50)	-.12 (.50)	-.17 (.45)	-.20 (.40)	-.20 (.40)	-.14 (.40)	-.20 (.35)
	RMSPE: 0.1361									

Note: The Table contains treatment effect estimates for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Table 16: Estimates for Treatment Effects based on US States

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.00 (1.0)	.08 (.15)	.01 (.85)	.23 (.05)
	RMSPE: 0.0220									
Austria	.02 (.55)	.06 (.15)	.05 (.15)	.05 (.25)	.07 (.35)	-.00 (.95)	.06 (.50)	.01 (.95)	.07 (.55)	.08 (.50)
	RMSPE: 0.0427									
Belgium	-.06 (.05)	-.04 (.40)	-.09 (.10)	-.13 (.10)	-.14 (.05)	-.20 (.05)	-.10 (.20)	-.13 (.10)	-.09 (.50)	-.07 (.55)
	RMSPE: 0.0537									
Canada		.02 (.50)	-.02 (.50)	-.02 (.50)	-.02 (.70)	.02 (.80)	.07 (.30)	.06 (.45)	.03 (.70)	.08 (.50)
	RMSPE: 0.0092									
Finland	.02 (.45)	.11 (.10)	.06 (.15)	-.11 (.10)	.09 (.25)	.02 (.75)	-.04 (.65)	-.02 (.85)	.08 (.55)	.02 (.85)
	RMSPE: 0.0419									
France	-.04 (.20)	-.05 (.20)	-.06 (.15)	-.08 (.25)	-.05 (.40)	-.10 (.25)	-.03 (.70)	-.02 (.85)	-.02 (.85)	-.06 (.65)
	RMSPE: 0.0308									
Germany	-.01 (.85)	-.04 (.40)	-.04 (.25)	-.10 (.20)	-.03 (.65)	-.11 (.20)	-.06 (.55)	.13 (.30)	.17 (.20)	.07 (.65)
	RMSPE: 0.0430									
Great Britain	-.03 (.45)	-.04 (.40)	-.03 (.45)	-.07 (.25)	-.00 (1.0)	-.07 (.35)	-.06 (.55)	.05 (.55)	.08 (.55)	-.02 (.85)
	RMSPE: 0.0393									
Italy	.04 (.35)	.04 (.40)	.02 (.55)	.01 (.95)	.02 (.70)	-.04 (.65)	-.04 (.65)	-.06 (.50)	-.07 (.55)	-.05 (.65)
	RMSPE: 0.0157									
Japan	.02 (.45)	.02 (.70)	-.00 (1.0)	-.02 (.60)	-.03 (.60)	-.06 (.35)	-.11 (.20)	-.07 (.40)	-.07 (.50)	-.01 (.90)
	RMSPE: 0.0117									
Netherlands	-.01 (.75)	-.01 (.75)	-.01 (.90)	-.06 (.25)	-.04 (.50)	-.08 (.25)	-.04 (.65)	.04 (.60)	.07 (.55)	.01 (.95)
	RMSPE: 0.0189									
Norway	-.00 (.90)	.05 (.20)	.03 (.40)	.01 (.85)	.02 (.65)	-.01 (.90)	.01 (.90)	.03 (.70)	.05 (.55)	.05 (.65)
	RMSPE: 0.0329									
Portugal	.10 (.05)	.02 (.65)	.02 (.55)	.06 (.25)	-.01 (.85)	-.04 (.60)	-.08 (.40)	-.04 (.65)	-.15 (.20)	-.10 (.45)
	RMSPE: 0.0265									
Spain	.10 (.05)	.11 (.10)	.13 (.10)	.17 (.10)	.14 (.05)	.12 (.15)	.07 (.40)	.02 (.85)	-.07 (.55)	-.03 (.85)
	RMSPE: 0.0283									
Sweden	.02 (.55)	.02 (.60)	-.01 (.80)	-.06 (.25)	-.08 (.25)	-.13 (.10)	-.12 (.20)	-.04 (.70)	.00 (1.0)	-.09 (.50)
	RMSPE: 0.0403									

Note: The Table contains treatment effect estimates for each Annex B country under investigation using US states to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

A.7 Accounting for flexibility mechanisms' emission credits

Table 17: Estimates for Treatment Effects based on non-Annex B Countries including emission credits from flexibility mechanisms

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.03 (.70)	.07 (.65)	-.04 (.70)	.01 (.95)
	RMSPE: 0.0371									
Austria	.00 (1.0)	.07 (.60)	.02 (.95)	.03 (.85)	-.00 (1.0)	-.06 (.90)	.04 (.85)	-.36 (.25)	-.16 (.50)	-.23 (.55)
	RMSPE: 0.0659									
Belgium	-.07 (.35)	.01 (.95)	-.06 (.80)	-.11 (.60)	-.12 (.60)	-.17 (.60)	-.02 (.95)	-.02 (.95)	-.11 (.65)	-.14 (.75)
	RMSPE: 0.0813									
Canada		.05 (.45)	.01 (1.0)	.02 (.95)	.00 (.95)	.00 (1.0)	.01 (1.0)	.01 (.90)	-.09 (.75)	-.10 (.70)
	RMSPE: 0.0472									
Finland	.06 (.45)	.20 (.20)	.14 (.55)	-.11 (.60)	.08 (.90)	.01 (.95)	-.06 (.85)	-.02 (.95)	-.07 (.75)	-.17 (.60)
	RMSPE: 0.0921									
France	-.02 (.95)	.02 (.95)	-.03 (.95)	-.08 (.70)	-.12 (.60)	-.18 (.55)	-.15 (.65)	-.13 (.85)	-.19 (.40)	-.32 (.35)
	RMSPE: 0.1124									
Germany	-.05 (.60)	-.03 (.75)	-.09 (.70)	-.13 (.45)	-.12 (.60)	-.17 (.60)	-.17 (.55)	-.12 (.90)	-.25 (.35)	-.28 (.50)
	RMSPE: 0.1758									
Great Britain	-.04 (.60)	-.02 (.95)	-.06 (.80)	-.09 (.70)	-.08 (.90)	-.13 (.65)	-.19 (.45)	-.09 (.90)	-.36 (.15)	-.58 (.10)
	RMSPE: 0.0774									
Italy	-.01 (.95)	.02 (.95)	-.00 (1.0)	-.03 (.85)	-.03 (.95)	-.07 (.90)	-.11 (.70)	-.12 (.85)	-.17 (.50)	-.15 (.70)
	RMSPE: 0.0456									
Japan	-.00 (1.0)	.01 (.95)	-.00 (1.0)	-.03 (.85)	-.02 (.95)	-.03 (.90)	-.06 (.85)	-.16 (.75)	-.10 (.65)	-.05 (.85)
	RMSPE: 0.0418									
Netherlands	.03 (.80)	.05 (.70)	.02 (.95)	-.03 (.85)	-.04 (.95)	-.04 (.90)	-.08 (.85)	.07 (.90)	-.06 (.90)	-.25 (.55)
	RMSPE: 0.0579									
Norway	-.10 (.20)	.02 (.95)	-.00 (1.0)	-.02 (.90)	.04 (.95)	.04 (.90)	.20 (.40)	.14 (.75)	.06 (.85)	-.07 (.85)
	RMSPE: 0.0849									
Portugal	.04 (.60)	-.08 (.35)	-.09 (.70)	-.08 (.70)	-.21 (.25)	-.20 (.50)	-.20 (.35)	-.27 (.30)	-.29 (.20)	-.32 (.20)
	RMSPE: 0.0510									
Spain	.05 (.50)	.06 (.60)	.08 (.70)	.07 (.75)	.01 (.95)	.01 (1.0)	-.08 (.80)	-.14 (.75)	-.23 (.35)	-.34 (.20)
	RMSPE: 0.0524									
Sweden	.11 (.15)	.06 (.60)	.03 (.95)	-.04 (.85)	-.06 (.90)	-.11 (.75)	-.06 (.85)	-.01 (.95)	.01 (.95)	.04 (.95)
	RMSPE: 0.2339									

Note: The Table contains treatment effect estimates for each Annex B country under investigation using non-Annex B countries to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Table 18: Estimates for Treatment Effects based on US States

Country	Year									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Australia							.07 (.05)	.17 (.05)	.07 (.20)	.15 (.10)
	RMSPE: 0.0220									
Austria	.04 (.30)	.12 (.05)	.11 (.10)	.14 (.05)	.12 (.10)	.03 (.75)	.11 (.15)	-.30 (.05)	.02 (.85)	-.07 (.45)
	RMSPE: 0.0427									
Belgium	-.09 (.05)	-.04 (.30)	-.08 (.10)	-.11 (.10)	-.12 (.10)	-.17 (.05)	-.04 (.65)	-.09 (.25)	-.07 (.45)	-.11 (.20)
	RMSPE: 0.0537									
Canada		.04 (.10)	-.01 (.65)	-.01 (.55)	.00 (.90)	-.01 (.90)	.01 (.90)	.04 (.70)	.00 (1.0)	-.00 (.90)
	RMSPE: 0.0207									
Finland	.08 (.10)	.19 (.05)	.17 (.05)	-.06 (.30)	.20 (.05)	.13 (.05)	.05 (.60)	.05 (.60)	.14 (.10)	.02 (.70)
	RMSPE: 0.0670									
France	-.03 (.30)	-.00 (.85)	.00 (1.0)	-.00 (.85)	.06 (.25)	.01 (.85)	.03 (.75)	-.07 (.55)	.11 (.35)	-.08 (.40)
	RMSPE: 0.0724									
Germany	-.05 (.15)	.01 (.85)	-.02 (.45)	-.02 (.55)	.15 (.10)	.08 (.30)	.11 (.30)	.03 (.80)	.01 (.90)	.00 (.95)
	RMSPE: 0.0457									
Great Britain	-.04 (.25)	-.05 (.30)	-.03 (.20)	-.06 (.35)	.00 (1.0)	-.05 (.40)	-.13 (.10)	.03 (.80)	-.17 (.10)	-.47 (.05)
	RMSPE: 0.0298									
Italy	.02 (.70)	.05 (.30)	.02 (.40)	.02 (.55)	.03 (.50)	.00 (.95)	-.03 (.75)	-.04 (.75)	-.03 (.80)	-.02 (.70)
	RMSPE: 0.0238									
Japan	.01 (.95)	.03 (.40)	.01 (.85)	-.01 (.75)	-.03 (.55)	-.05 (.40)	-.07 (.35)	-.17 (.10)	-.12 (.10)	-.08 (.30)
	RMSPE: 0.0194									
Netherlands	.00 (.95)	.01 (.65)	.00 (1.0)	-.03 (.55)	-.07 (.25)	-.06 (.35)	-.10 (.25)	.02 (.85)	-.03 (.65)	-.24 (.05)
	RMSPE: 0.0373									
Norway	-.10 (.05)	.02 (.60)	-.00 (.90)	-.01 (.80)	.04 (.40)	.04 (.70)	.20 (.05)	.21 (.05)	.10 (.30)	-.01 (.70)
	RMSPE: 0.0607									
Portugal	.15 (.05)	-.00 (1.0)	-.06 (.15)	-.07 (.15)	.02 (.85)	.04 (.70)	.07 (.35)	-.01 (.85)	.07 (.50)	.11 (.35)
	RMSPE: 0.0803									
Spain	.07 (.10)	.07 (.15)	.07 (.10)	.11 (.10)	.10 (.10)	.11 (.05)	.00 (1.0)	-.07 (.50)	-.08 (.35)	-.23 (.05)
	RMSPE: 0.0297									
Sweden	.10 (.05)	.10 (.05)	.08 (.10)	.03 (.55)	.16 (.10)	.09 (.25)	.20 (.05)	.15 (.10)	.23 (.05)	.31 (.05)
	RMSPE: 0.0814									

Note: The Table contains treatment effect estimates for each Annex B country under investigation using US states to construct the synthetic counterfactuals and considering the year of ratification as the time of treatment. For each country, we display the yearly treatment effects in percent and the respective probabilities of finding such an effect in parenthesis. Finally, we report the (pre-treatment) RMSPE for each country.

Analyzing the Effectiveness of International Environmental Policies: The Case of the Kyoto Protocol

Online Appendix

This Online Appendix contains the graphs of actual and counterfactual emissions paths and the corresponding placebo studies for all 15 countries under investigation for the two main specifications and all robustness checks discussed in the paper.

Ratification of the KP as the treatment event (Main specification)

Figure 1: Actual and synthetic CO₂ emissions for non-Annex B countries in donor pool

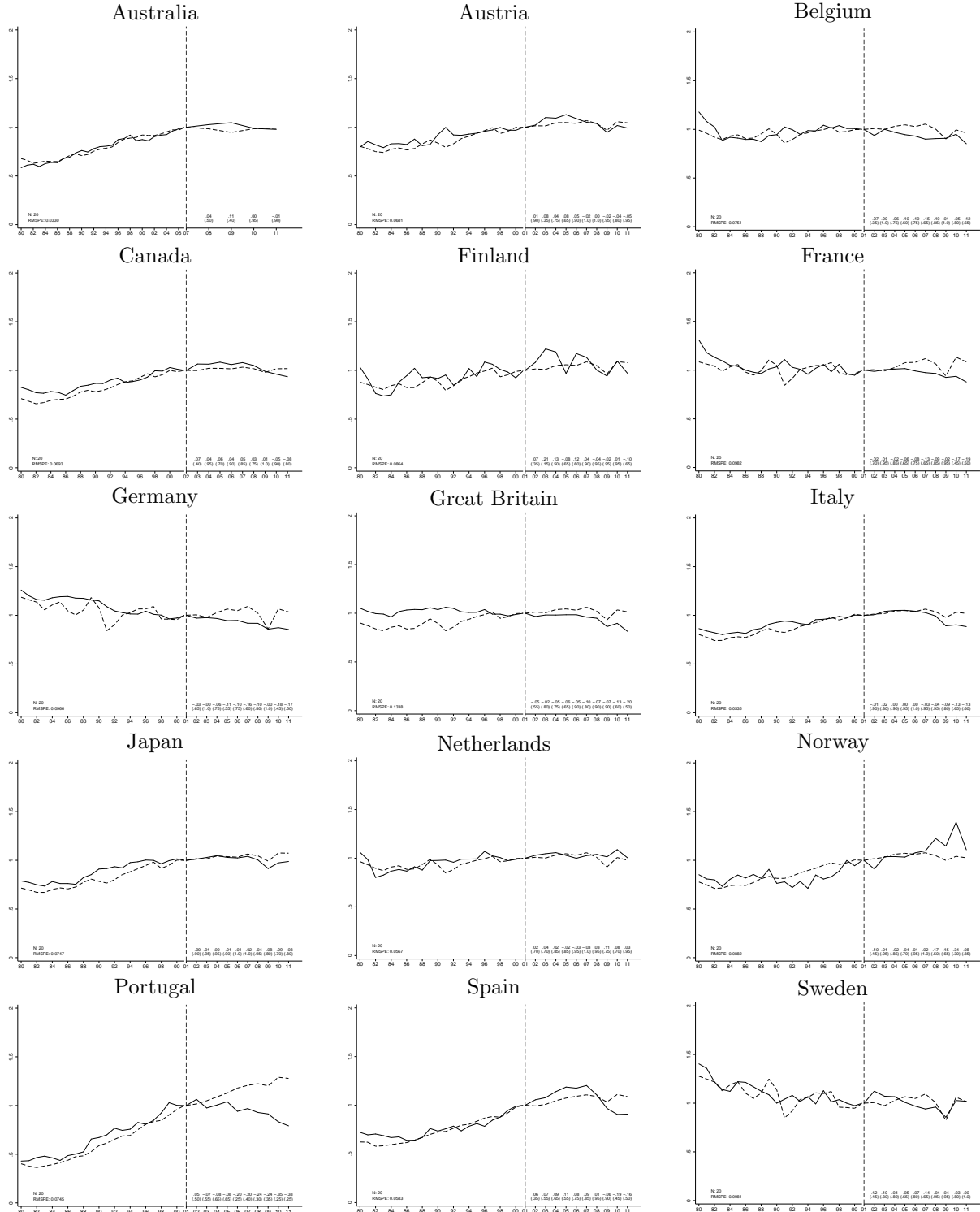


Figure 2: Placebo studies for non-Annex B countries in donor pool

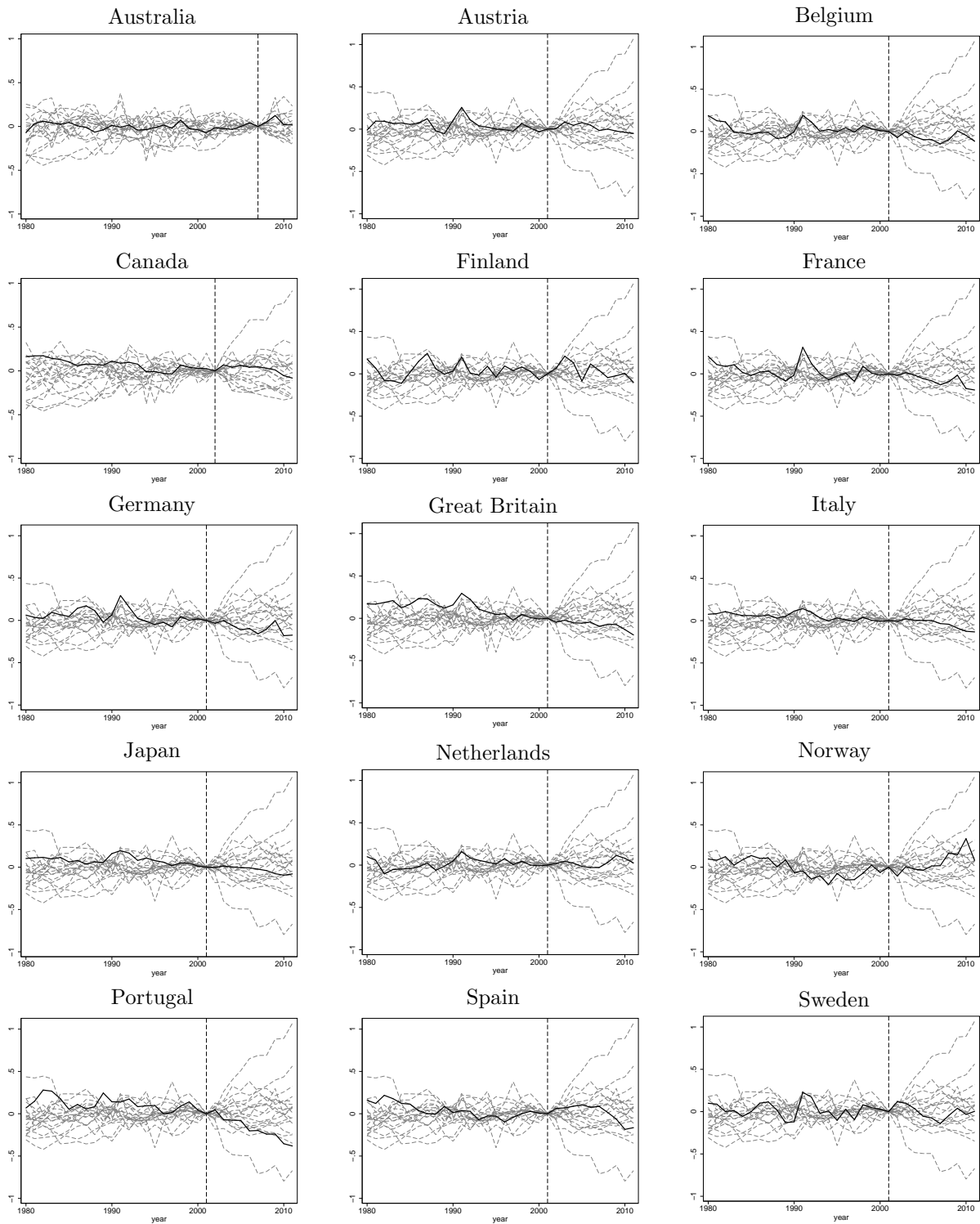


Figure 3: Actual and synthetic CO₂ emissions for US states in donor pool

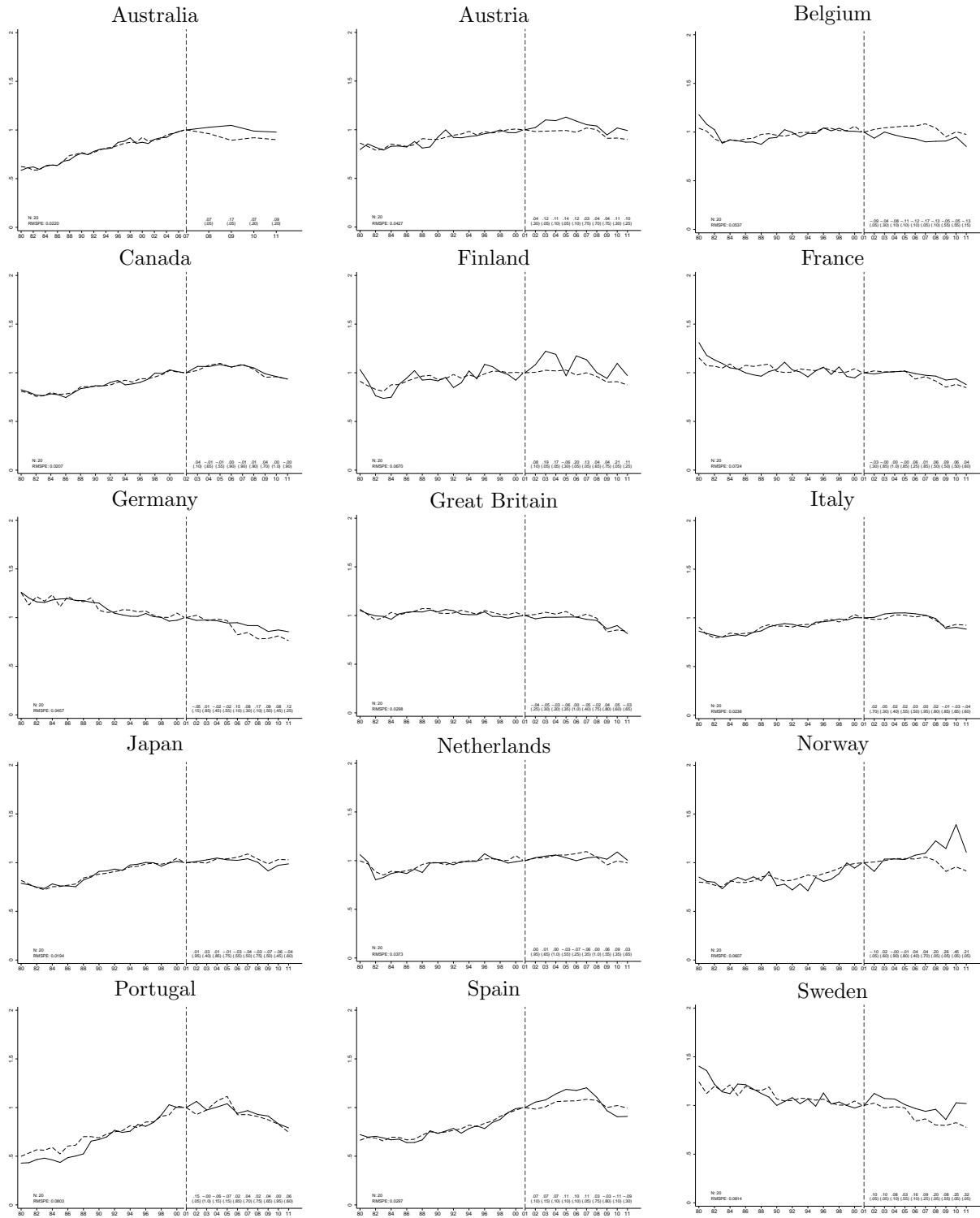
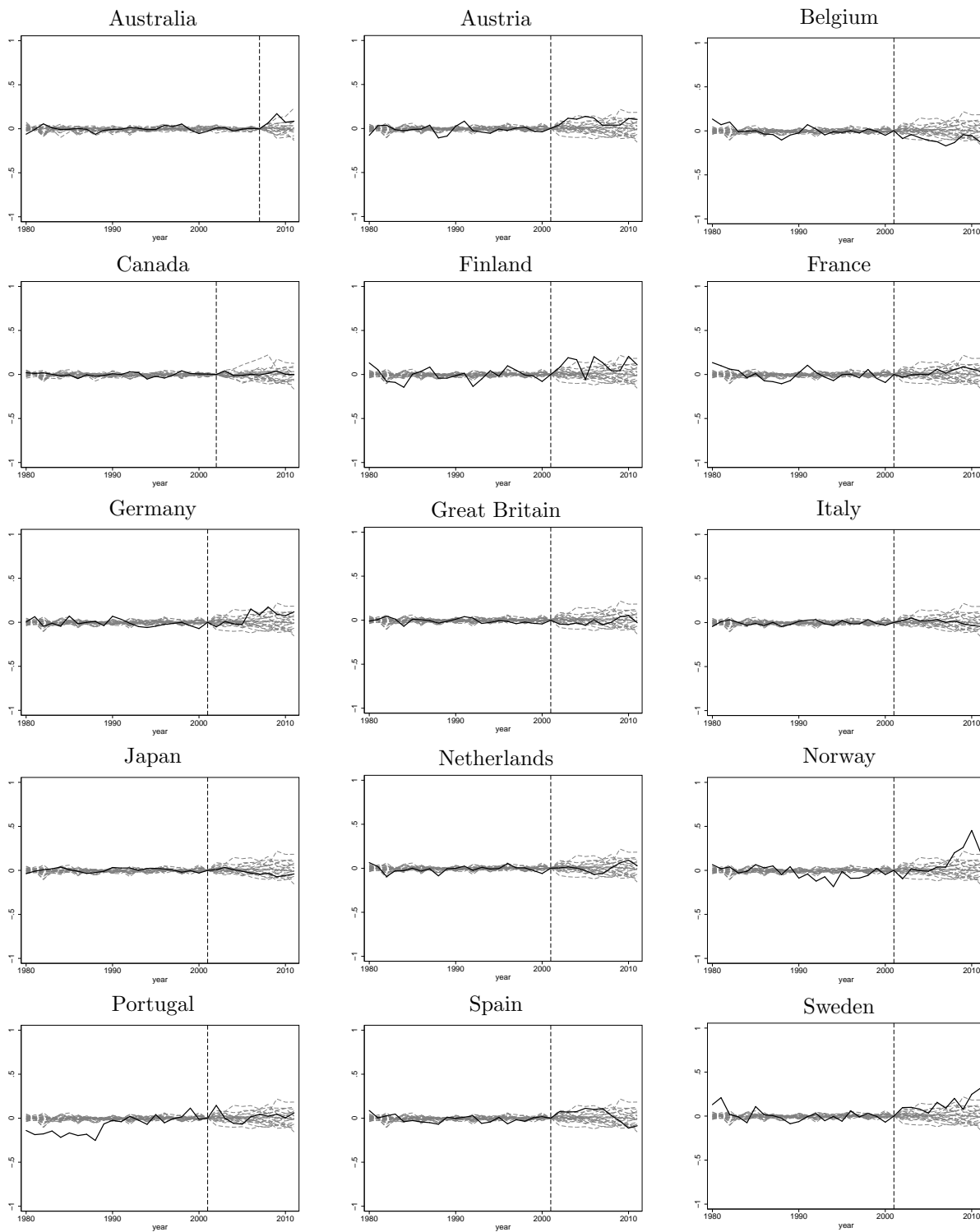


Figure 4: Placebo studies for US states in donor pool



Adoption of the KP as the treatment event

Figure 5: Actual and synthetic CO₂ emissions for non-Annex B countries in donor pool

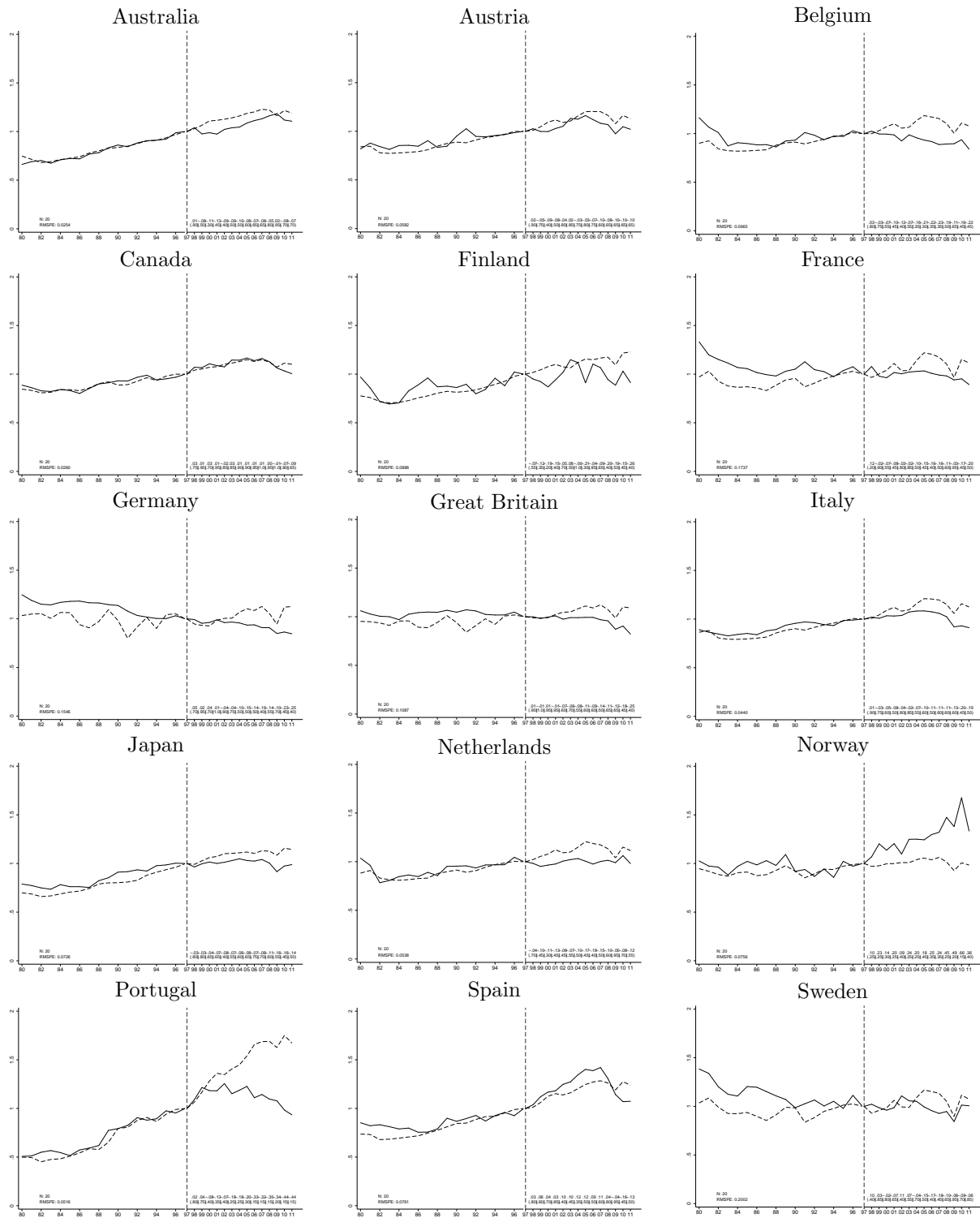


Figure 6: Placebo studies for non-Annex B countries in donor pool

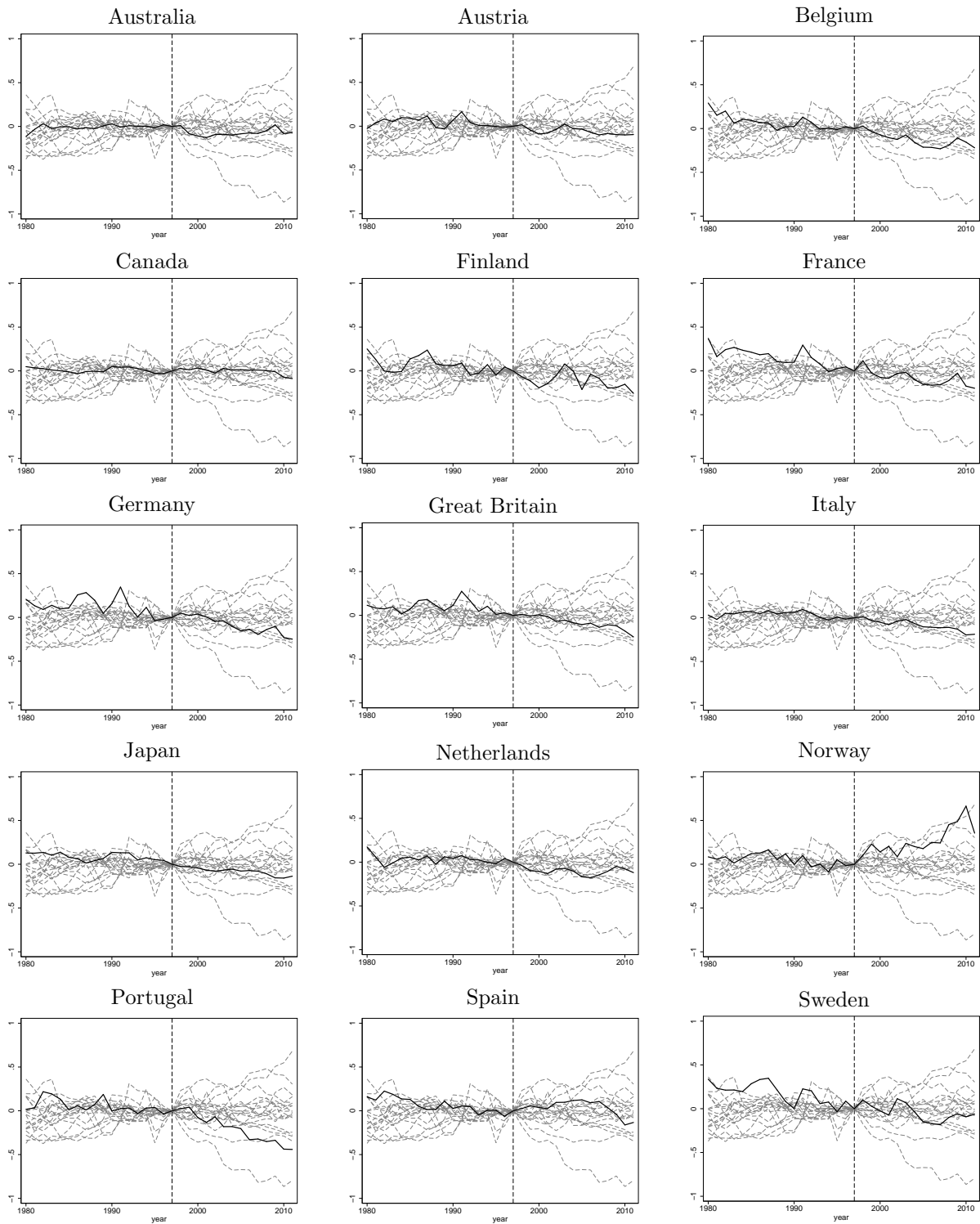


Figure 7: Actual and synthetic CO₂ emissions for US states in donor pool

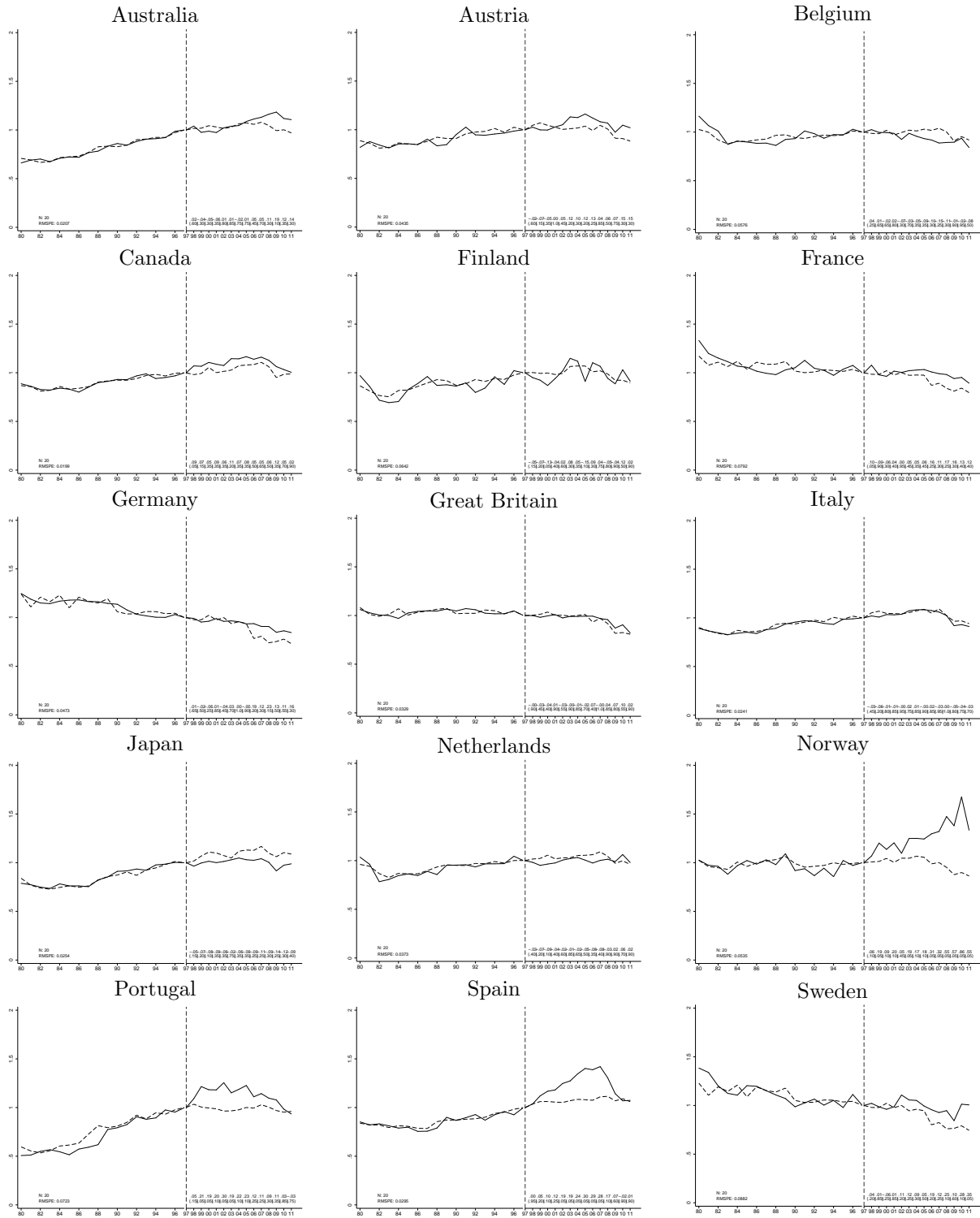
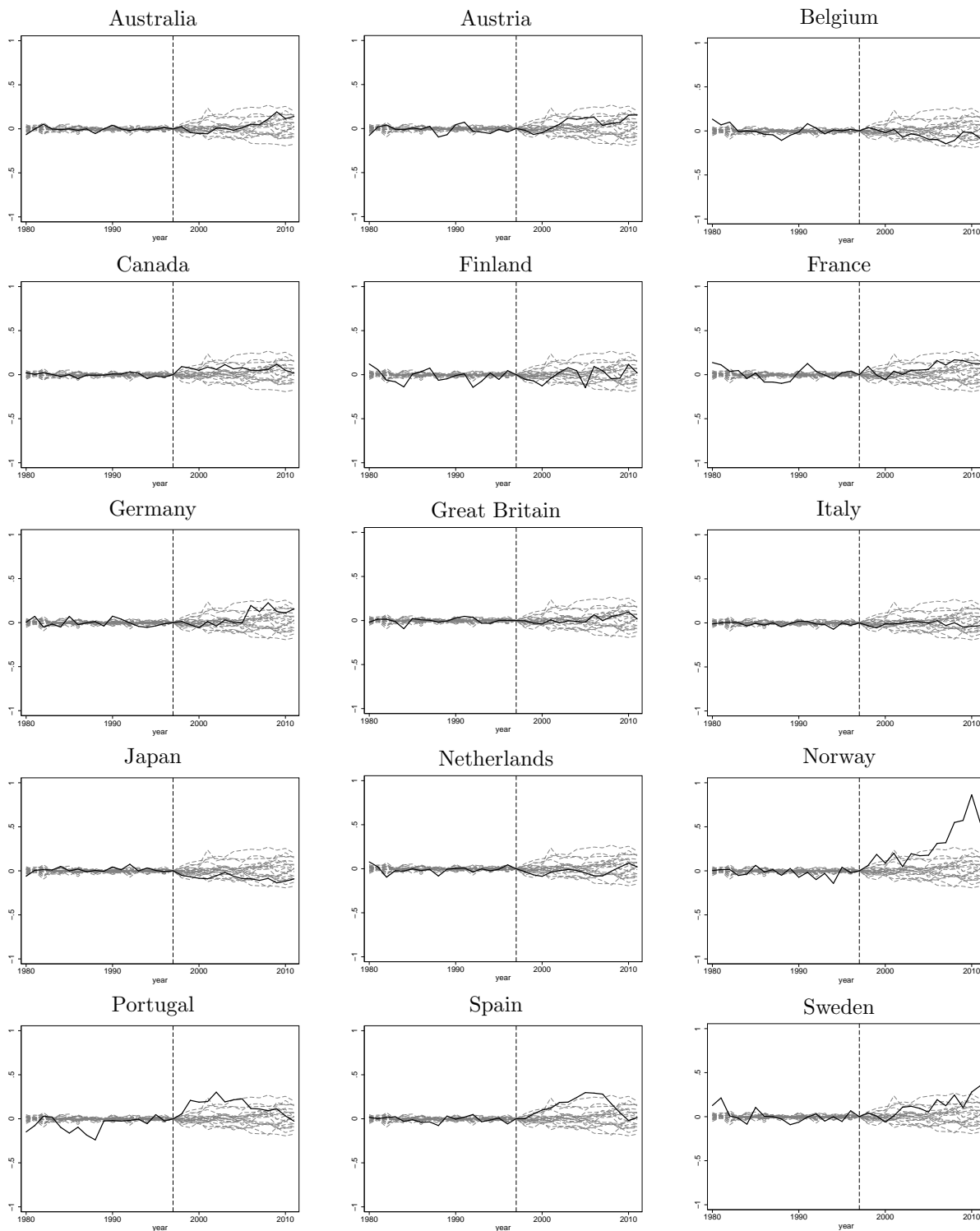


Figure 8: Placebo studies for US states in donor pool



KP's entering into force as the treatment event

Figure 9: Actual and synthetic CO₂ emissions for non-Annex B countries in donor pool

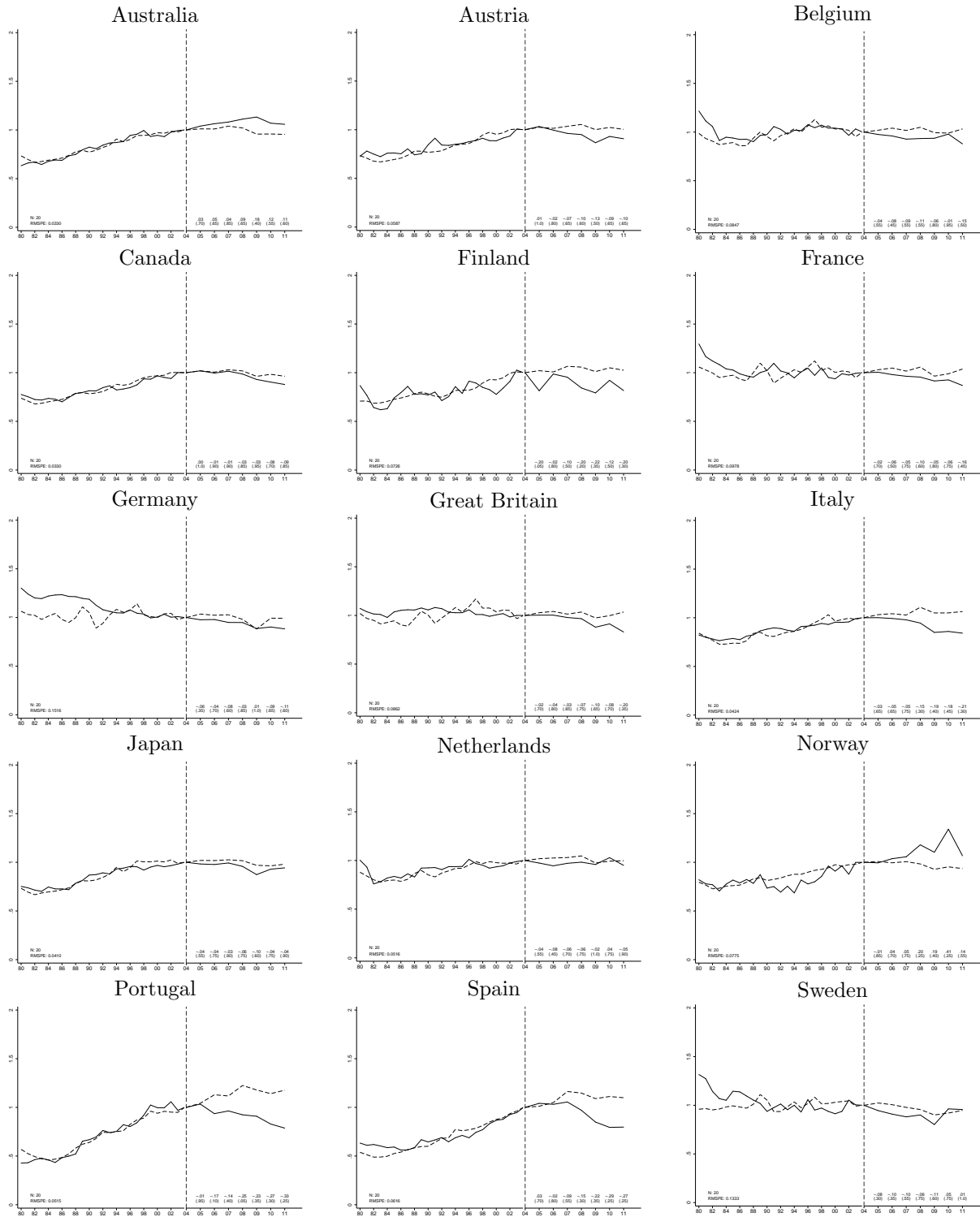


Figure 10: Placebo studies for non-Annex B countries in donor pool

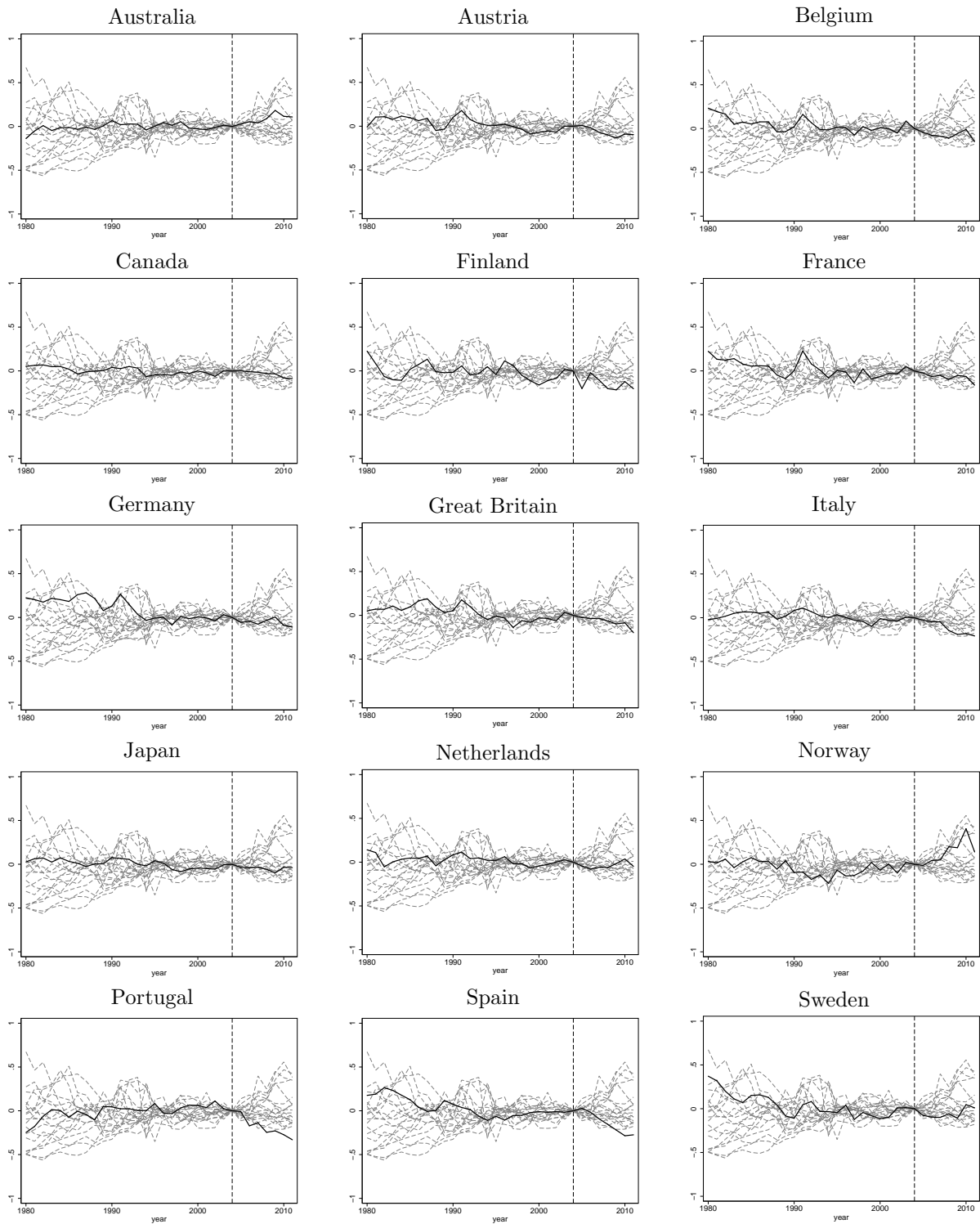


Figure 11: Actual and synthetic CO₂ emissions for US states in donor pool

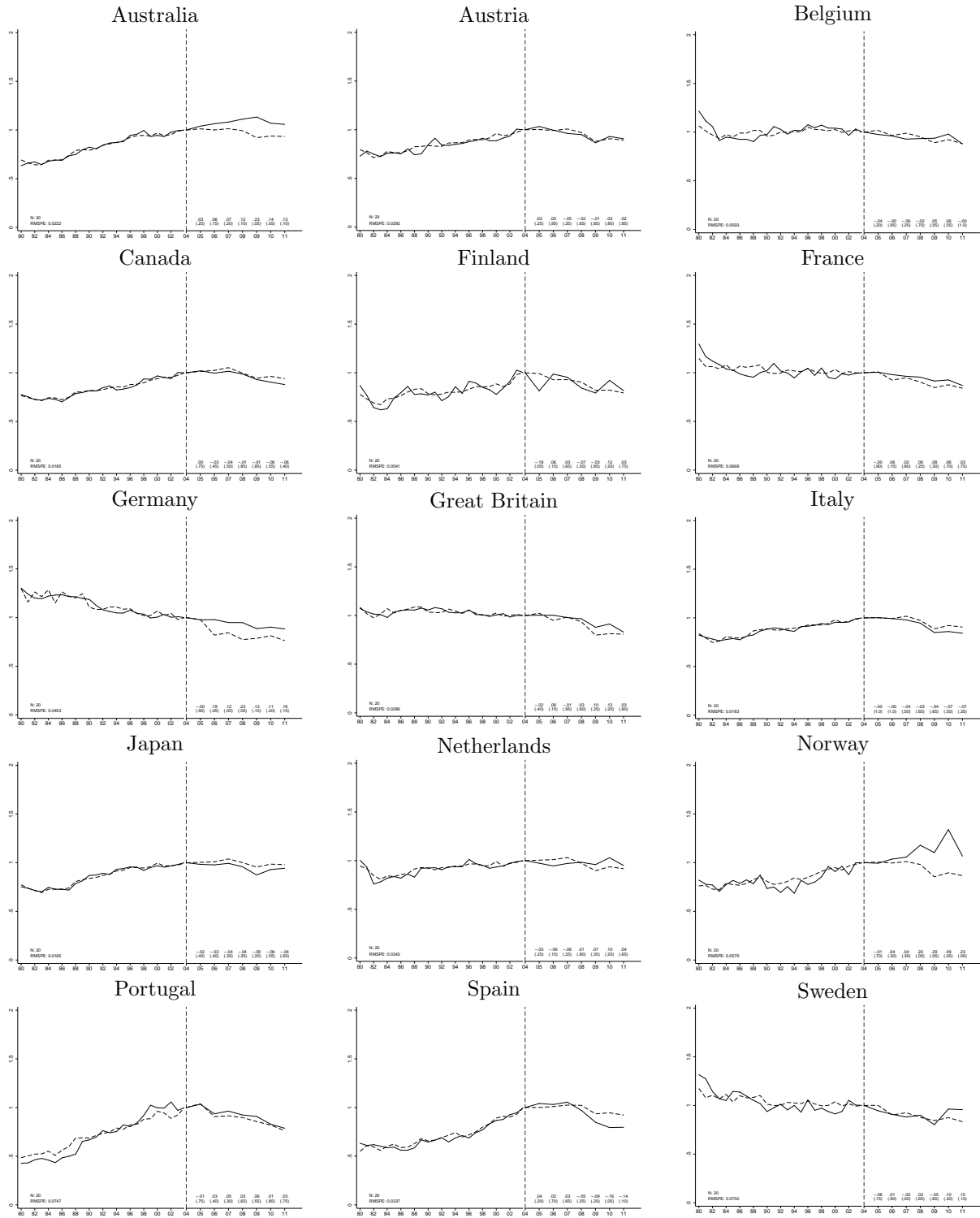
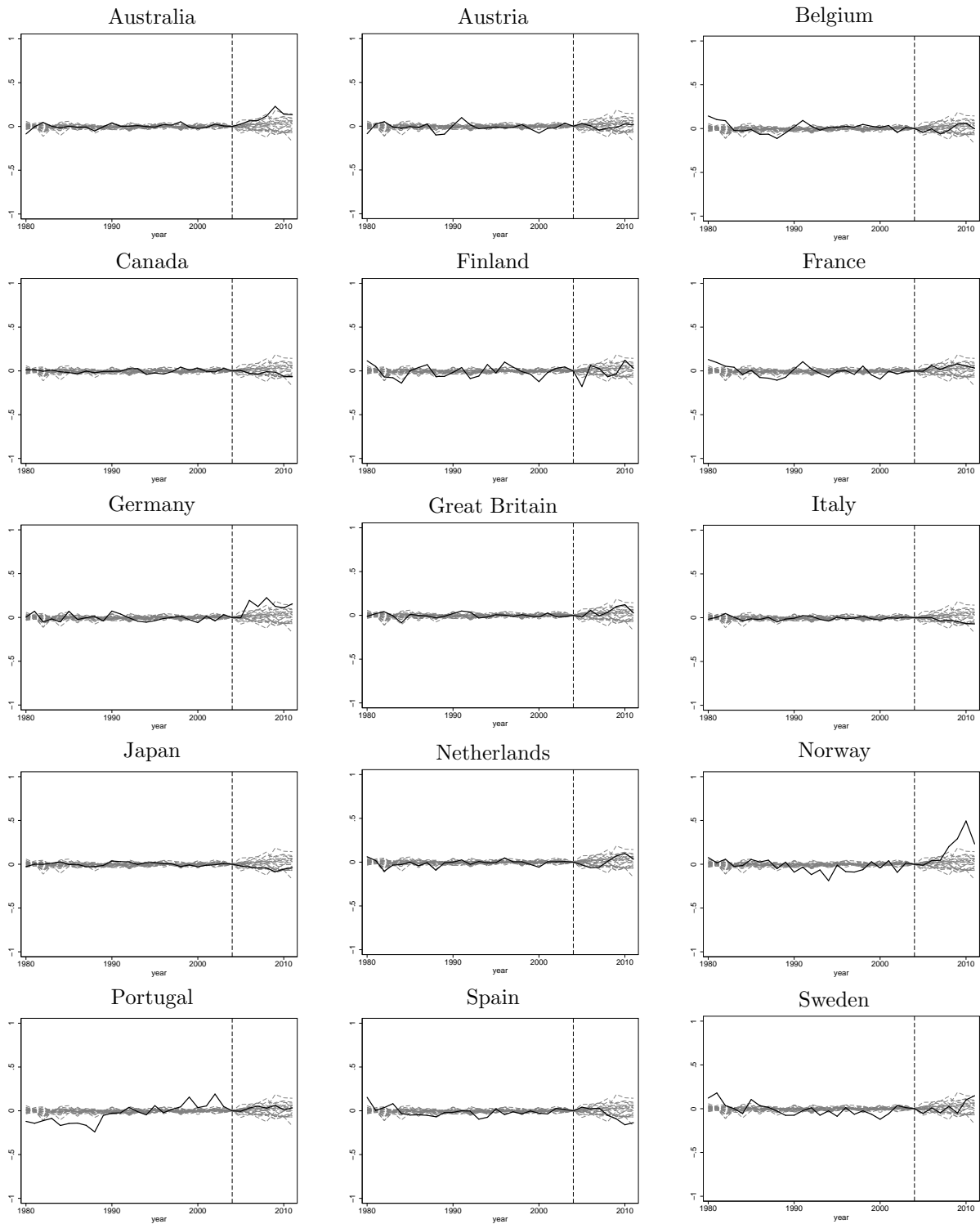


Figure 12: Placebo studies for US states in donor pool



CO₂ intensity as dependent variable

Figure 13: Actual and synthetic CO₂ emissions per GDP for non-Annex B countries

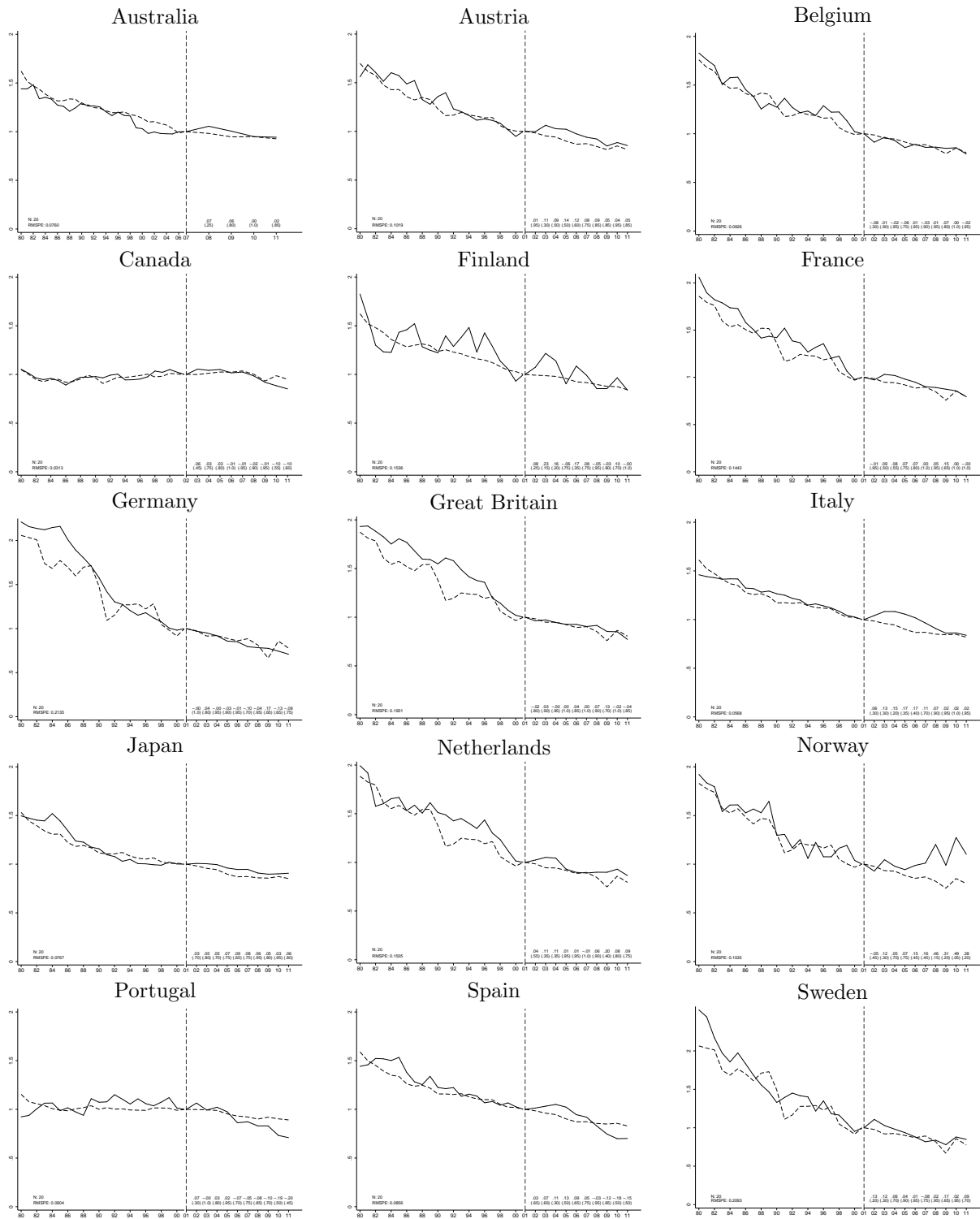
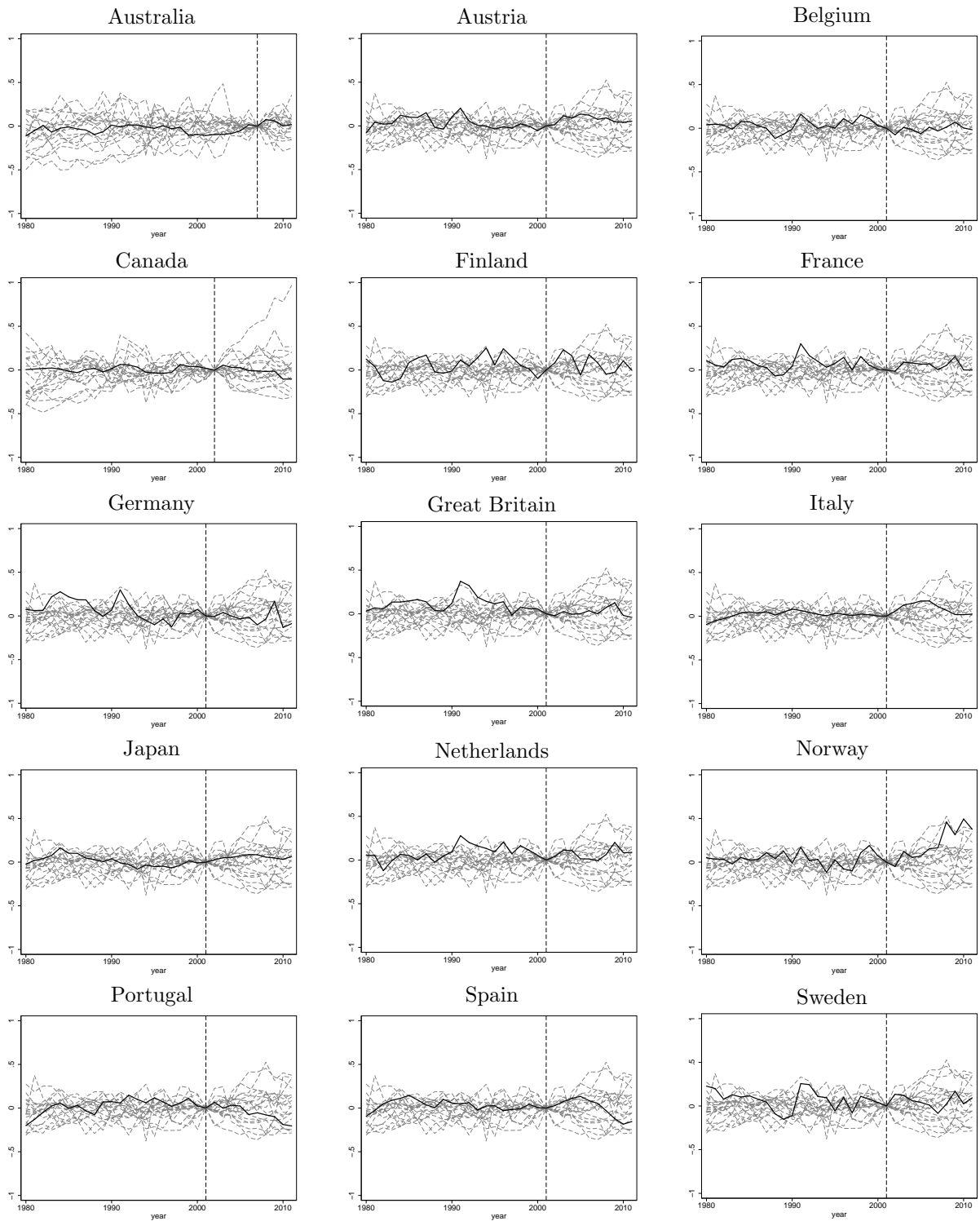


Figure 14: Placebo studies for non-Annex B countries in donor pool



CO₂ per capita as dependent variable

Figure 15: Actual and synthetic CO₂ emissions p.c. for non-Annex B countries

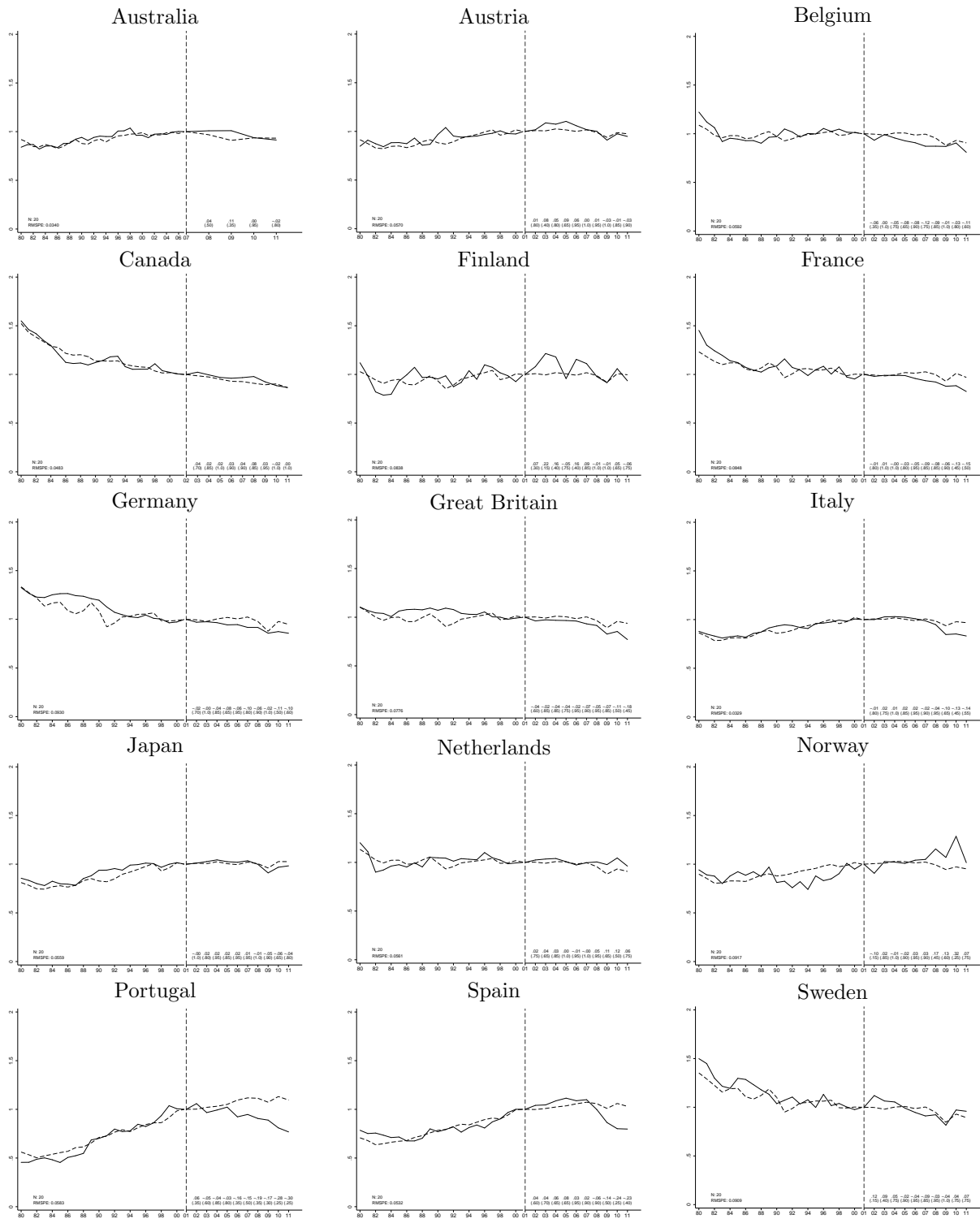


Figure 16: Placebo studies for non-Annex B countries in donor pool

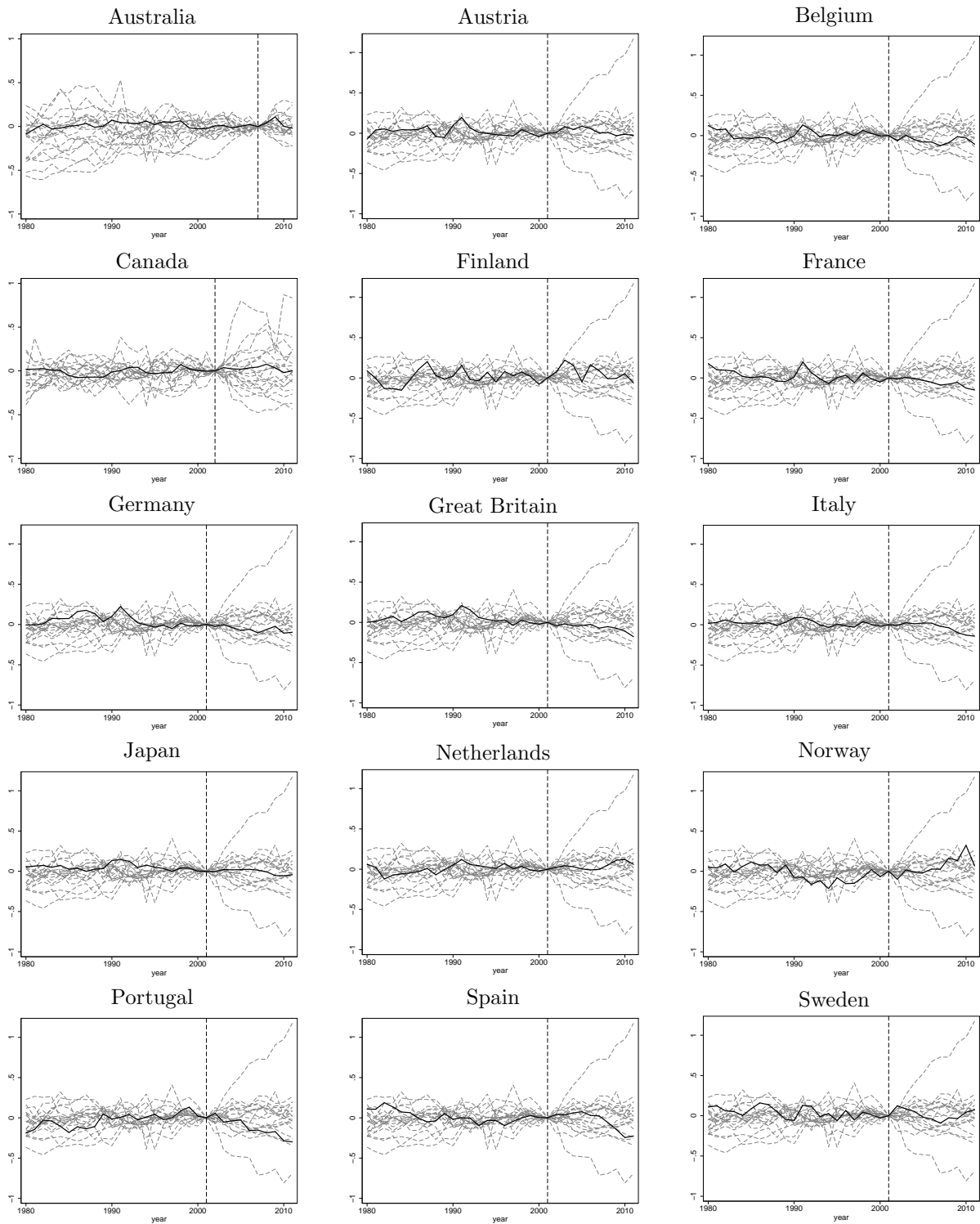


Figure 17: Actual and synthetic CO₂ emissions p.c. for US states

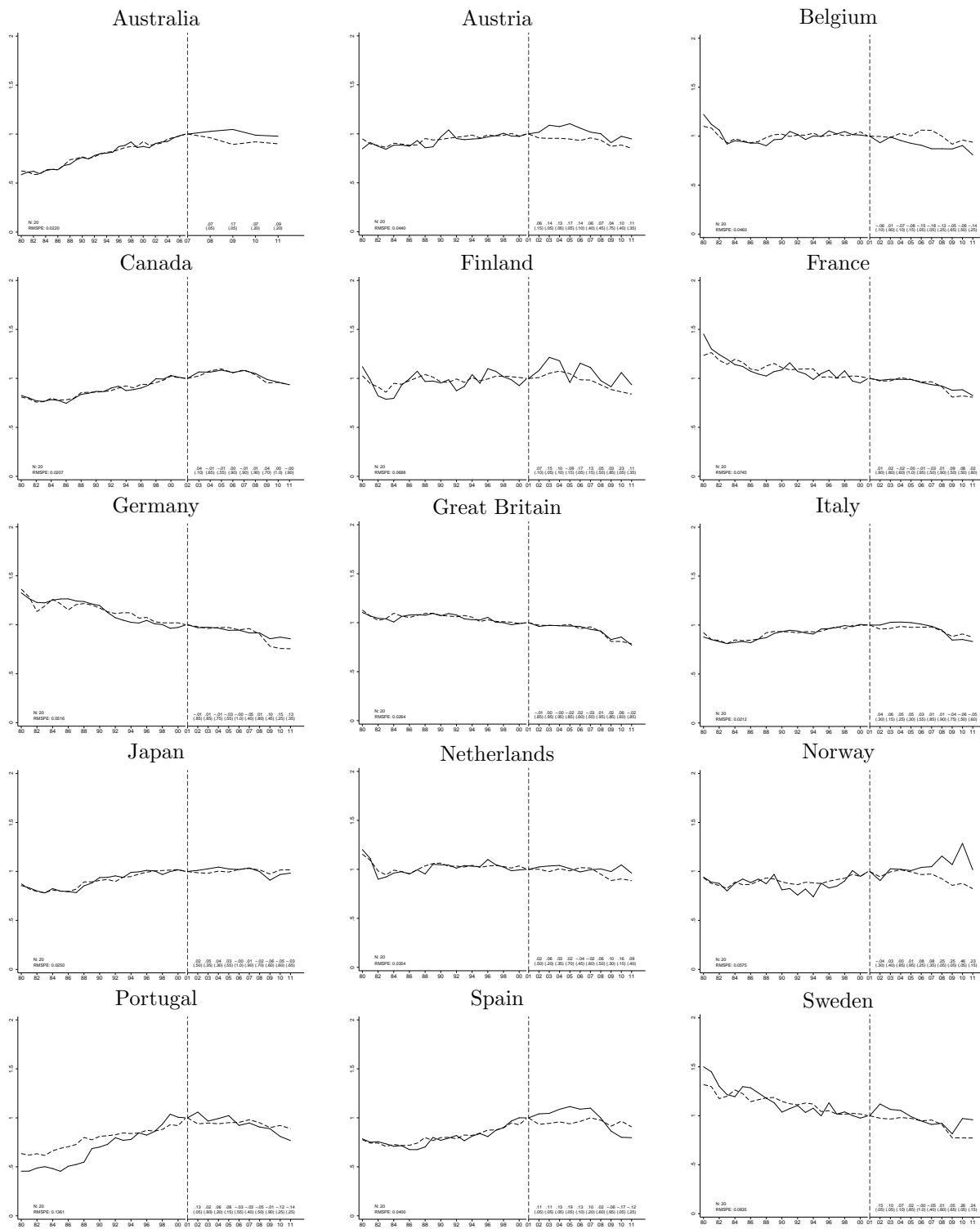
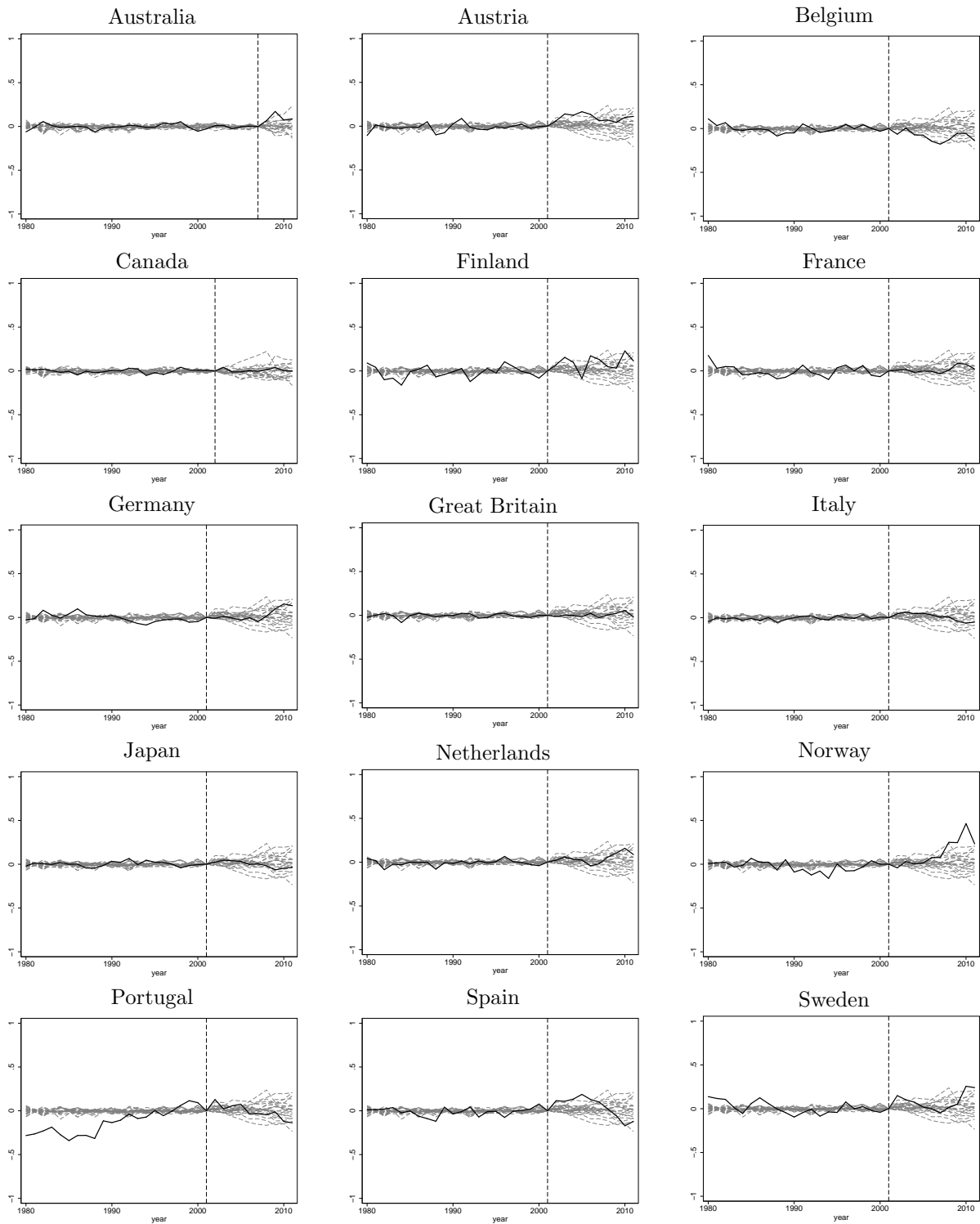


Figure 18: Placebo studies for US states in donor pool



GHGs as dependent variable

Figure 19: Actual and synthetic GHG emissions for non-Annex B countries in donor pool

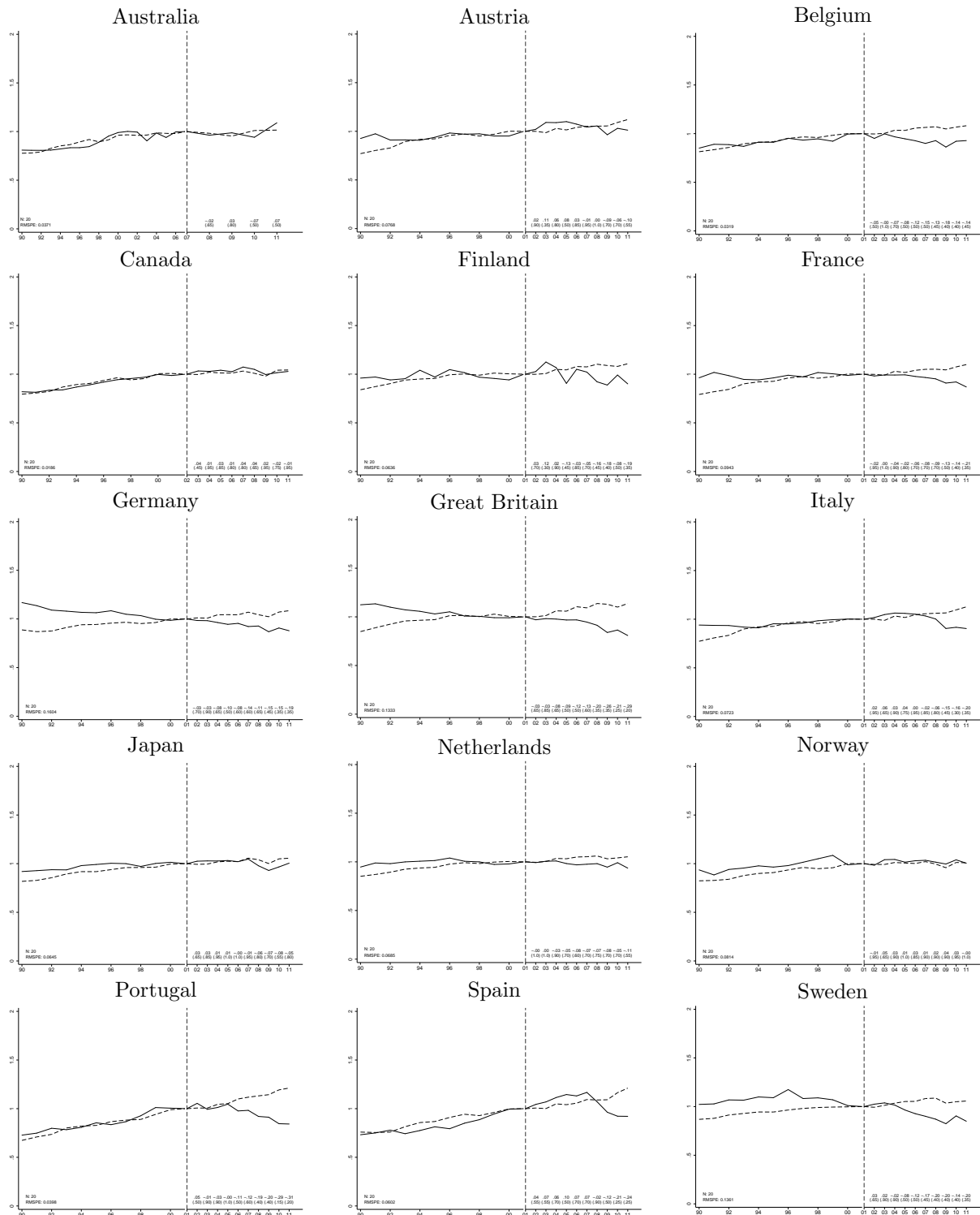


Figure 20: Placebo studies for non-Annex B countries in donor pool

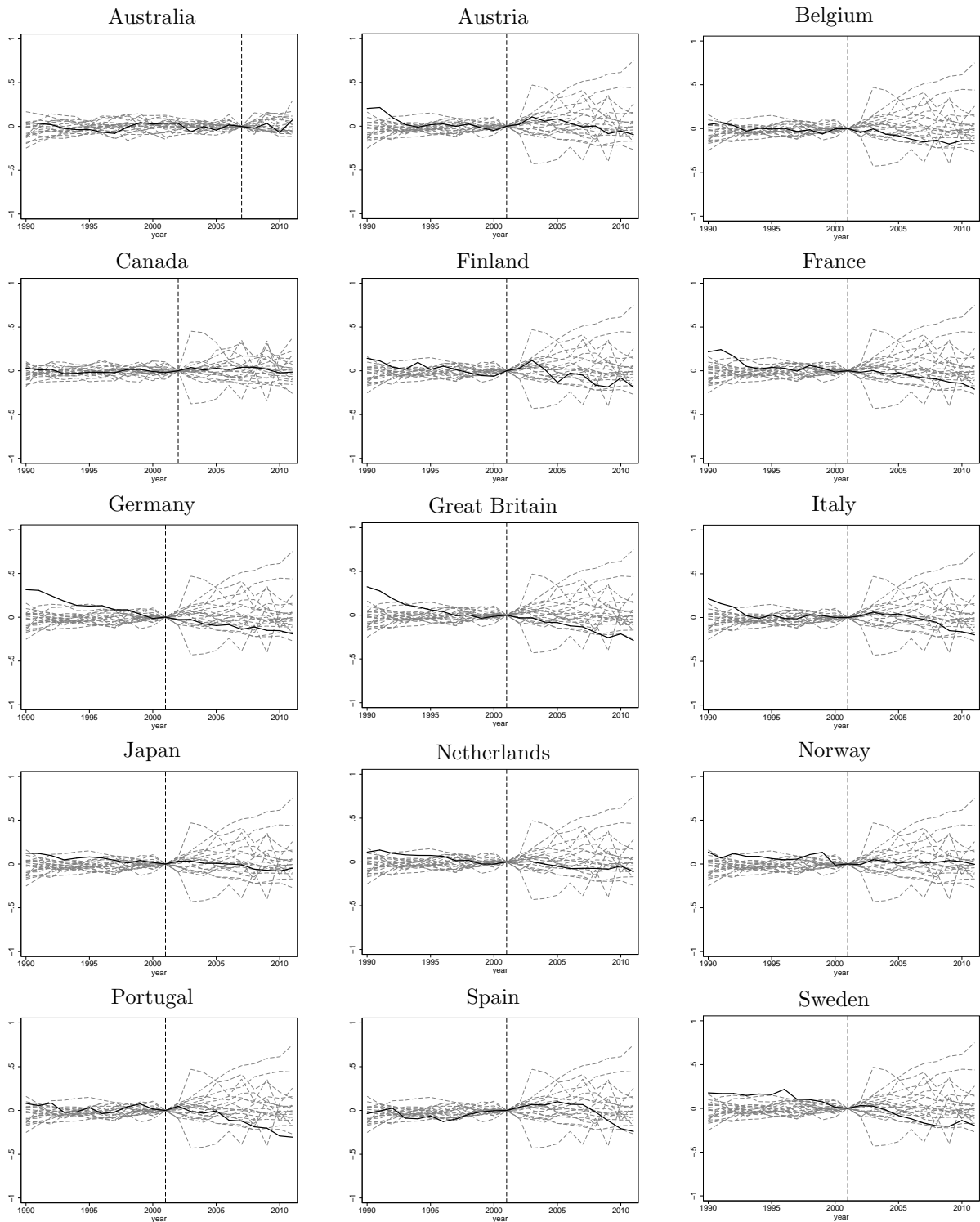


Figure 21: Actual and synthetic GHG emissions for US states in donor pool

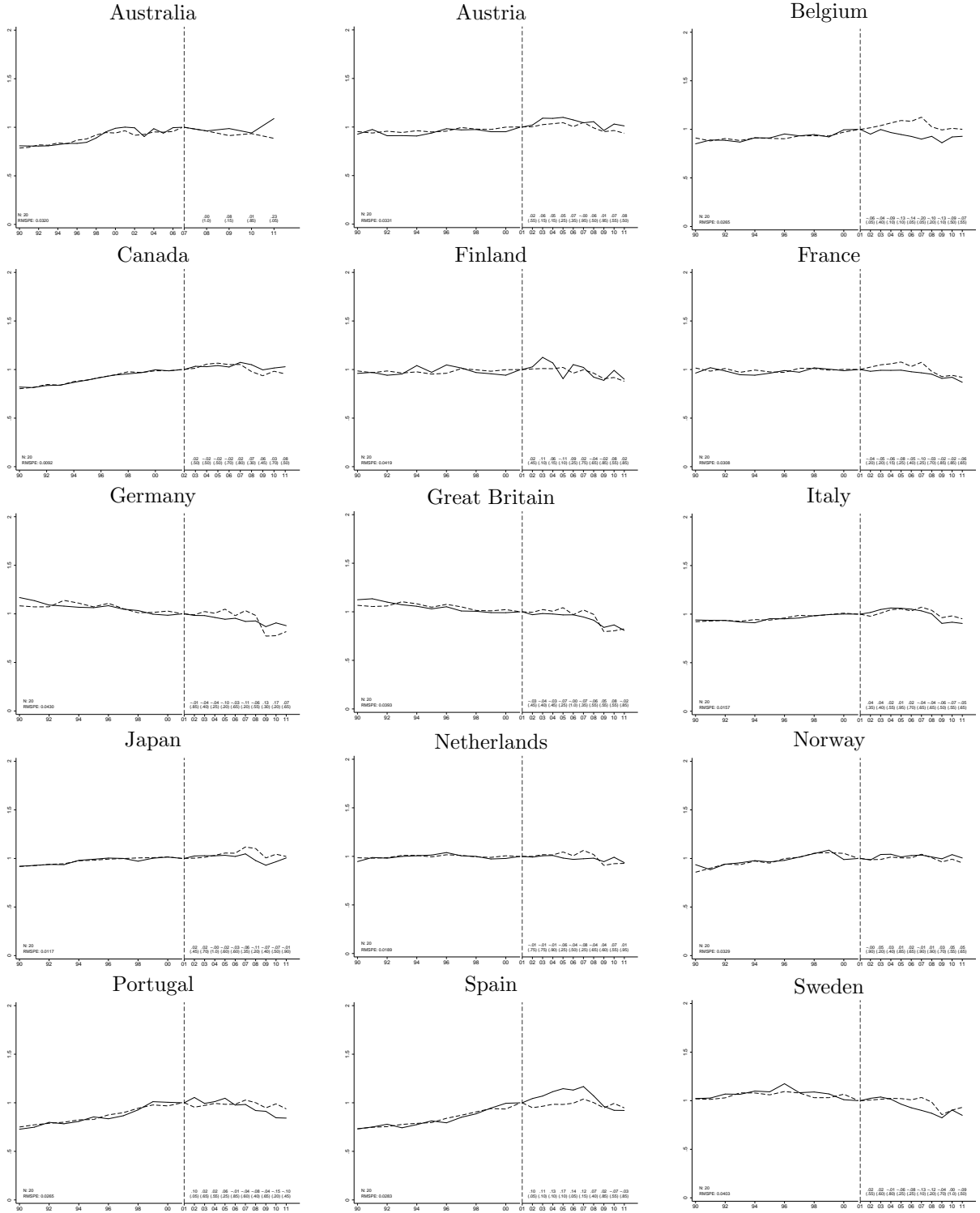
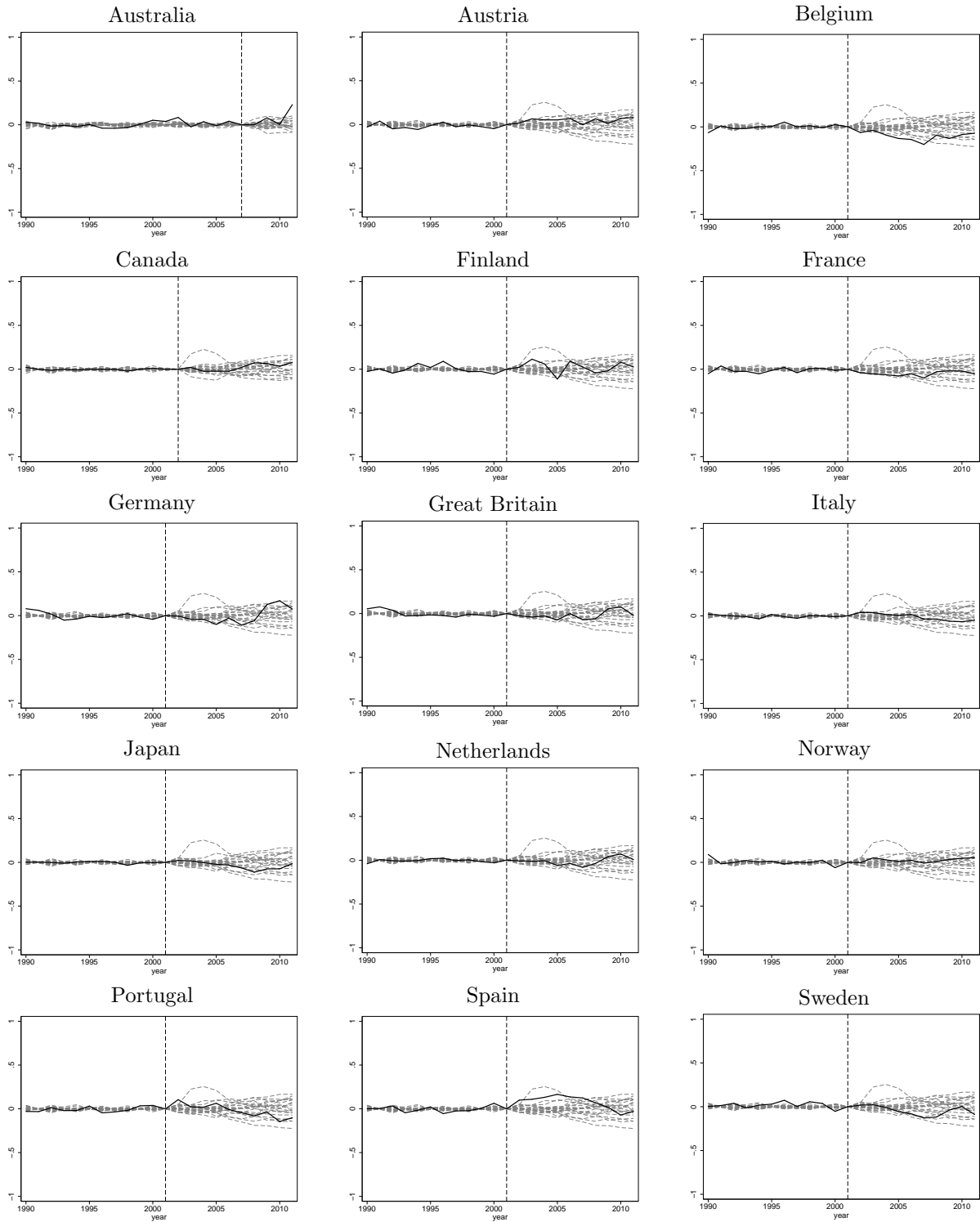


Figure 22: Placebo studies for US states in donor pool



Accounting for flexibility mechanisms' emission credits

Figure 23: Actual and synthetic CO₂ emissions for non-Annex B countries in donor pool

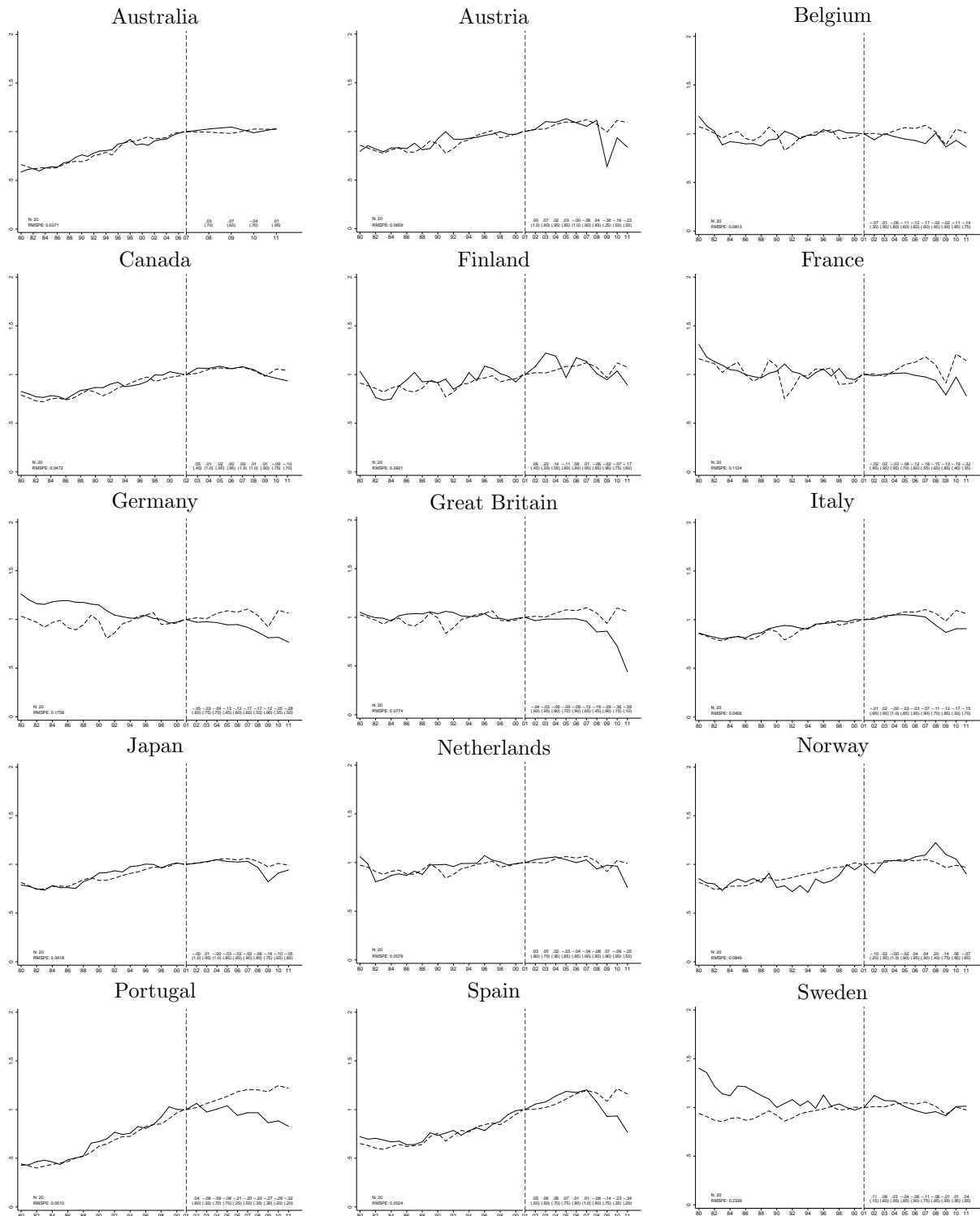


Figure 24: Placebo studies for non-Annex B countries in donor pool

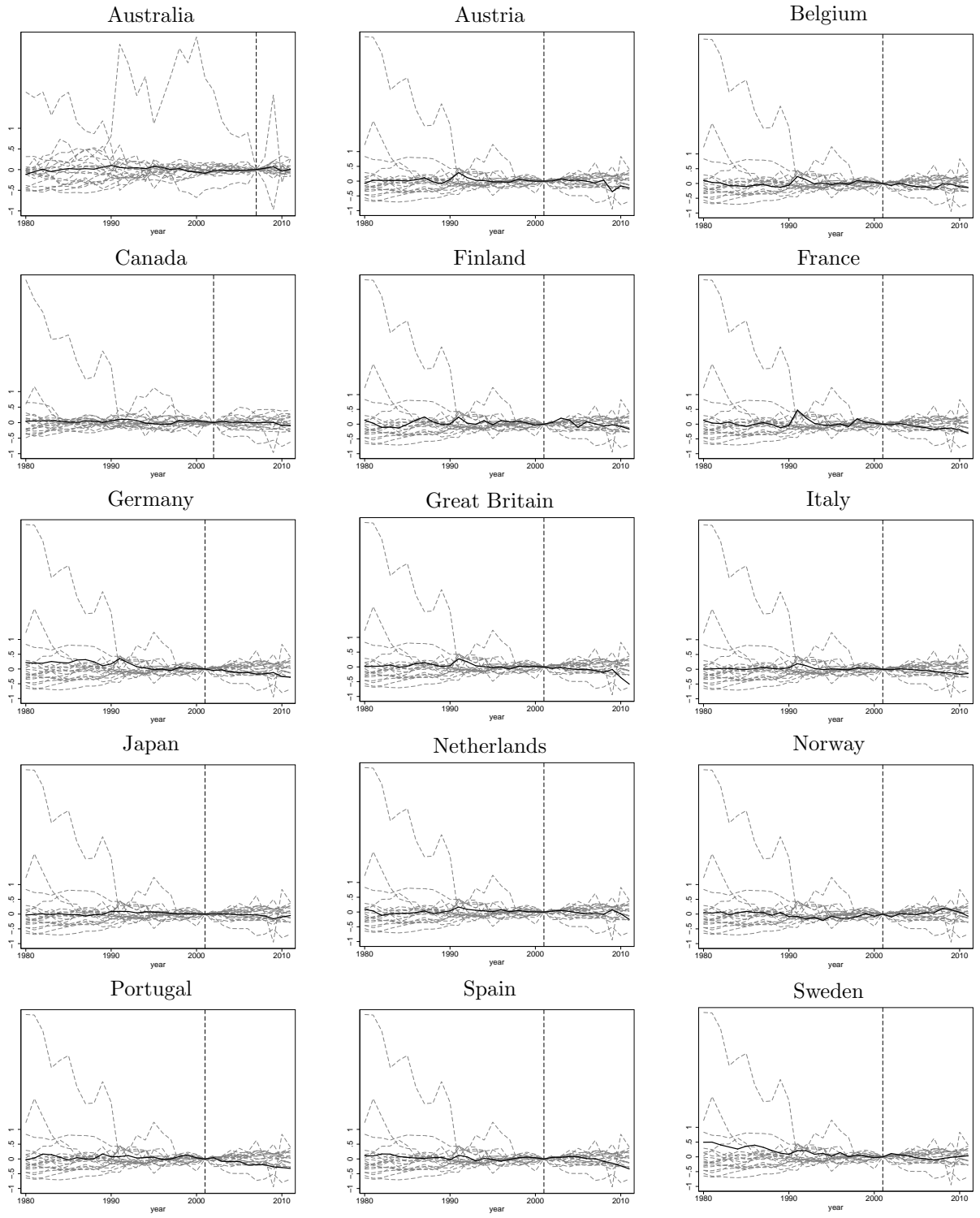


Figure 25: Actual and synthetic CO₂ emissions for US states in donor pool

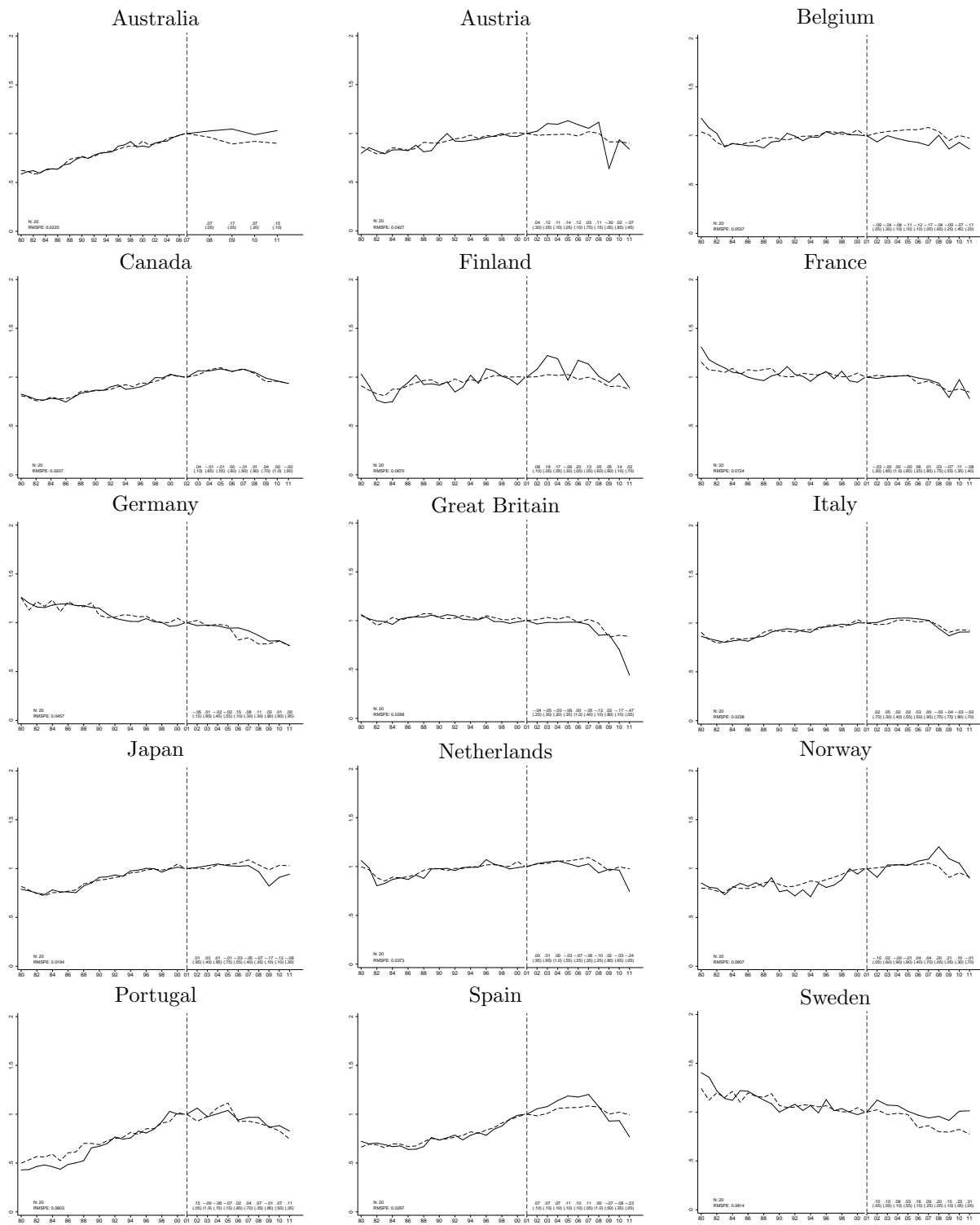


Figure 26: Placebo studies for US states in donor pool

