Quasi-Experimental Evidence on Carbon Pricing

Kasper Vrolijk and Misato Sato

A growing literature suggests that carbon emissions are most efficiently reduced by carbon pricing. The evidence base on the effectiveness of market-based mechanisms, however, faces three key limitations: studies often (a) predict, rather than evaluate effects, (b) show large difference in findings, and (c) cannot always infer causal relations. Quasi-experimental studies can address these challenges by using variation in policies over time, space, or entities. This paper systematically reviews this new literature, outlines the benefits and caveats of quasi-experimental methodologies, and verifies the reliability and value of quasi-experimental estimates. The overall evidence base documents a causal effect between carbon pricing and emission reductions, with ambiguous effects on economic outcomes, and there are important gaps and inconsistencies. This review underscores that estimates should be interpreted with care because of: (a) inappropriate choice of method, (b) incorrect implementation of empirical analysis (e.g., violate identifying assumptions), and (c) data limitations. More cross-learning across studies and use of novel empirical strategies is needed to improve the empirical evidence base going forward.

JEL Codes: H23, Q58, C21
Keywords: proposition: carbon pricing, empirical evidence, policy impact, quasi-experimental designs.

Introduction

A growing economic literature suggests that the reduction of greenhouse emissions, and carbon dioxide emissions in particular, is most efficiently achieved through market-based mechanisms, either by putting a tax on emissions or introducing an emissions trading scheme. Given that climate change is increasingly accelerating and materializing in large-scale environmental and economic damage, it is important to understand the effectiveness of such marked-based policies. Moreover, because

The World Bank Research Observer

https://doi.org/10.1093/wbro/lkad001 0:1–36

carbon taxation and carbon trade mechanisms likely create substantial shocks throughout the economy, influencing economic behavior and outcomes, a comprehensive evidence base on the economic effects of such mechanisms is key to improving the effectiveness of policies and relevant institutions. Particularly for climate change mitigation in least developed and emerging countries, robust evidence is essential to minimize negative economic effects.

Despite the urgency to advance climate action, there are three main challenges that limit our understanding of the effects of carbon taxation and cap-and-trade mechanisms (henceforth carbon pricing) on emissions and economic outcomes. First, a large part of the literature uses *ex-ante* approaches to predict, rather than evaluate *ex-post*, the effect of carbon pricing policies. This is inevitable given that carbon pricing is relatively new. The critique against such integrated assessment models (IAM) is well established (Pindyck 2013, 2017). Second, the quality of the available evidence is at times questionable due to limitations in data and empirical methods that are often descriptive in nature rather than causal. Third, there is seemingly limited agreement among the findings of the empirical studies that evaluate carbon pricing impacts, depending on the empirical design, policy setting, and framing of carbon pricing (see Fried 2018; Andersson 2019).

Tackling this evidence gap, a wave of new empirical research explores alternative approaches in search of more robust results. This research utilizes innovative quasi-experimental designs to examine the economic and environmental effects of carbon pricing. By exploiting variation in policy over time, space, or entities, these studies seek to estimate whether a causal relation exists between carbon pricing, emissions, and various economic outcomes, including firm output, employment, and innovation.[1] This research has not only provided a range of new insights on the effectiveness and consequences of carbon pricing policies, but also insights beyond the field of environmental economics, uncovering economic mechanisms within public economics and providing a deeper understanding of useful empirical methods for policy evaluation.[2]

Against this background, the paper has two goals. First, it performs a systematic review of the empirical studies on carbon pricing that use quasi-experimental designs to obtain results, including synthetic control, regression discontinuity, difference-in-differences, and instrumental variable estimation.[3] Quasi-experimental methods have flourished across multiple domains in economics because they address important limitations of a standard cross-sectional regression.[4] Second, the paper examines the benefits and drawbacks of using quasi-experimental methodologies, exploring potential (methodological) caveats in each of these studies and how they may bias results. While quasi-experimental designs have proliferated, they come with strict identifying assumptions that need to be met to claim causal effects.[5] The main aim is to offer an assessment of how reliable and therefore how informative the quasi-experimental estimates are for understanding carbon pricing impacts. To our knowledge, this is the first paper to systematically review the empirical carbon pricing literature from a methodological perspective.[6]

Methodologically, we document several threats to internal and external validity of estimates in the quasi-experimental carbon pricing literature, including: (a) inappropriate choice of method, (b) incorrect implementation of empirical analysis (e.g., violating the main identifying assumptions), and (c) data limitations. Quasi-experimental designs rely on several important assumptions such as the common trends assumption; i.e., in absence of treatment, average outcomes would have followed common trends for the treatment and control groups, and a multitude of other challenges. For example, coinciding (policy) shocks can prevent researchers from obtaining true estimates of carbon pricing effects. A major challenge is availability of quality, disaggregated data on emissions and economic variables. Often, there is a trade-off between the level of coverage (e.g., on firms, sectors) and the level of detail (e.g., stratified sample). Disaggregated microdata may be hard to access due to confidentiality issues, whereas aggregated data can lead to bias and measurement error. Often, emissions data is missing or in poor quality for periods before a policy intervention, or for subjects outside of a policy intervention, making it difficult to have a credible counterfactual. As a result, the effects of carbon pricing on household consumption and behavior have been understudied, while more attention has been paid to large emitters in energy and industrial sectors.

One main conclusion is that providing truly credible estimates of carbon pricing effects is a non-trivial task that only a few studies have achieved. This means that while carbon pricing is seen as most efficient in reducing carbon emissions, a robust causal evidence base remains scant. Those studies that we deem credible provide empirical and causative evidence that carbon taxation and cap-and-trade reduce emissions, a finding consistent across cross- and within-country analyses.[7] There is less agreement on whether and to what extent emission reductions lead to negative economic outcomes, such as firm employment, revenue, or innovation, although if any such effects are found they tend to be small.

The paper is structured as follows. The next section describes the systematic review procedure. The following section then evaluates the empirical studies that employ quasi-experimental research designs to estimate the effects of carbon pricing on emissions and economic outcomes. The paper then brings together main findings and provides concluding remarks, offering ways in which methodological caveats and literature gaps can be addressed. The Appendices provide an overview of the review sample (Appendix A1), the characteristics, methods, and data used in studies reviewed in this paper (Appendix A2), and the treatments, findings, and methodological caveats of the main studies discussed in this review (Appendix A3).

Systematic Review

To ensure a comprehensive literature review, studies were collected using the search, screening, and data extraction procedure by Grubb et al. (2021). Our scope covered the two main forms of carbon pricing: Carbon taxation and cap-and-trade policies. We searched for studies that ask how carbon pricing influences emissions and economic outcomes of firms and households. To this end, our search strategy was to review (un)published, quantitative, academic, English-language (working) papers using Google Scholar. It excluded unpublished theses. We applied all combinations of keywords between types of carbon pricing ('carbon tax', 'cap-and-trade', 'carbon price') and quasi-experimental methods ('difference-in-difference', 'instrumental variable', 'regression discontinuity').[8] We omitted studies on supplementary policies to carbon pricing, e.g., border carbon adjustments (e.g., Böhringer et al. 2022). Further, the review excluded studies that examine changes in energy prices as a proxy for carbon taxes (e.g., Marin and Vona 2021), studies on broader environmental taxes, and studies that examine carbon pricing design (e.g., the effect of permit allocation on emissions, Fowlie and Perloff (2013); Rafaty, Dolphin, and Pretis (2020)). To check for omissions, references in retrieved studies were used to obtain additional studies. During this stage, it became clear that some papers evaluate policies other than carbon pricing, but offer relevant insights on carbon pricing and were therefore included, e.g., Flues and Lutz (2015) study electricity taxation, Fowlie, Holland, and Mansur (2012) examine NO_x emissions. The pool of studies obtained using this method was screened against the above criteria, first by title, then abstract, and finally the full text. Studies that did not match these criteria were dropped from the sample (Appendix A1 lists the number of studies retrieved and included). The final pool includes 47 studies published between 2012–2022; 79 percent are in the OECD context and 11 percent are non-OECD (only China), 66 percent of which are journal publications and the remainder working papers. During the data extraction phase, relevant information from each study was collected and used to populate the table in Appendix A2.

New Empirics of Carbon Pricing

This section reviews the recent empirical evidence on the causal effects of carbon pricing on emissions and economic outcomes. We discuss each quasi-experimental design separately and review its aim, ideas, assumptions, and key features that determine the validity of results. Within the discussion we use examples of studies that epitomize best-practices and those that typify major methodological issues. At the end of the section we summarize the evidence on carbon pricing from studies that are deemed to provide reliable results. Treatments, findings, and methodological caveats of each study are summarized in Appendix A3.

Synthetic Control

The aim of a synthetic control design is to assess effects of an intervention at the aggregate level. One novel study on carbon taxation by Andersson (2019) studies the effect of carbon taxation and a value-added tax on transport fuels in Sweden. Instead of using an arbitrarily selected control country, an algorithm is applied to find a "synthetic" control group from a weighted combination of control units that best approximates values of predictors of pre-treatment outcomes.

In Andersson's (2019) case, the donor pool is drawn from 14 other OECD countries that did not implement carbon taxes. In contrast to a normal difference-in-difference (DID), the data-driven approach to control-group selection allows effects of unobserved confounders to vary over time. That is, it relaxes the assumption of "common trends" where outcomes should be similar in the treatment and control unit prior to a policy change. With that it addresses an important assumption in quasi-experimental designs that is often hard to enforce in DID designs; that treated and control units have on average similar characteristics and that units only differ in treatment assignment (or the unconfoundedness or selection-on-observables assumption). With a large donor pool including only one neighboring country, it is unlikely that the stable-unit treatment value assumption (SUTVA) is violated due to spillover effects in Andersson (2019). Leroutier's (2022) study of the effect of carbon taxation on emissions in the United Kingdom power sector also uses a synthetic control design but explicitly accounts for spillovers by assessing electricity imports and the "waterbed effect", i.e., the policy would reduce demand for permits in the EU Emissions Trading Scheme (ETS) (both which bias her estimates downwards).

A key feature that determines if estimates are valid in a synthetic design is the construction of the control group. Andersson (2019) does this reasonably well; he excludes those countries that were affected by a similar event or are in characteristics different from the treated unit Sweden. In contrast, Runst and Thonipara (2019), who study the effect of carbon taxation on emissions in the residential sector in Sweden, do not properly specify predictors used to create the synthetic control group, nor the inclusion criteria for control countries. aus dem Moore, Brehm, and Gruhl (2022) include Austria and Luxembourg in the control group, although these countries are characterized by "fuel tourism" and may therefore differ considerably from Sweden.

The validity of synthetic control estimates also lies in the robustness checks that are implemented on the synthetic control pool. The results by Andersson (2019) are robust to a range of checks that are mandatory in the synthetic control approach; shift treatment year, assigning treatment to other countries, omit countries from the control group, or lag outcome variables. However, Andersson (2019) does not check the validity of control group predictors. The latter is important because the choice of predictors (to establish the control weights) is a contested issue in the synthetic approach (Ferman, Pinto, and Possebom 2020). In particular the choice of how to include the lagged outcome variable (i.e., emissions) is important because while relevant to include, it may render the other (economic) predictors irrelevant and bias results (Kaul et al. 2022). In Andersson (2019), different combinations of lagged CO_2 emissions are included. Leroutier (2022) only opts for one combination, but aus dem Moore et al. (2022) perform a range of checks and in the final analysis restrict to average pre-treatment outcome lags (the proposed solution by Kaul et al. 2022).

Regression Discontinuity

Regression discontinuity (RD) designs estimate a local average treatment effect and are useful in settings where a particular threshold determines whether units are treated or untreated.[9] Flues and Lutz (2015) use a sharp discontinuity in electricity taxation on German manufacturing firms (i.e., above a certain level of energy consumption firms paid reduced marginal taxes) to test the effect on firm sales, exports, value added, investment, and employment.[10] The intuition of the approach is to compare units close to the eligibility threshold to units that are close to the threshold but ineligible, assuming they are similar. With that it addresses the unconfoundedness or selection-on-observables assumption. The strength of this approach is that it requires minimal assumptions, but it estimates a local effect around the threshold only, hence there may be limited external validity.

One important assumption in RD is that the intervention (or treatment level) was similar to all units. To this end, Flues and Lutz (2015) show that there were no exemptions to particular firms or sectors prior to the introduction of the tax (in contrast, many sectors in other settings see some exemption). They also show that firms could not self-select into treatment given that firms could not precisely manipulate electricity usage due to production complexity and exogenous factors. What remains is whether the treatment spilled over to other firms (the other part of SUTVA). Flues and Lutz (2015) do this indirectly, but to a limited extent (the DID study by Fowlie et al. (2012), discussed below, offers suggestions on how to evaluate spillover effects in detail).

In an RD approach a key feature is also to confirm the parallel trends (pre-trends) assumption, which Flues and Lutz (2015) credibly do by showing that outcome variables developed continuously with the treatment variable in proximity of the threshold. Finally, in an RD design it is relevant to check the robustness of the RD parameters. Martin, Muûls, and Wagner (2013), who examine the effect of the ETS on firm innovation in six EU economies, review the results against different bandwidths and add quadratic and interaction terms to study potential non-linear functional forms, which, when included, they find reduces their estimates. However, Martin et al. (2013) do not check if all other variables that determine outcomes are continuous around the threshold (pre-trend assumption). Most studies using RD do not show the representative of the units around the cut-off and therefore the external validity of results.[11]

A variation to the common RD approach is the RD "in time" design, which captures changes in policy over time, and which is useful when cross-sectional variation in policy implementation (which is a requirement for DID) is absent.[12] In the RD in-time (RDiT) approach, the threshold includes a discontinuous policy adjustment (or "jump") and units are compared before and after each jump. As in the RD, cofounders are also assumed to change smoothly across the threshold (which is the date of policy change) and that characteristics of treated and untreated units are identical. Gugler, Haxhimusa, and Liebensteiner (2020) examine the effect of sharp adjustments in carbon tax on emissions in the United Kingdom power sector. In their setting, the RDiT is useful to absorb any time-varying factors, such as the EU's ETS that operated alongside, which are likely continuous across the event. While useful when time variance is available, the RDiT design comes with a few potential pitfalls (Hausman and Rapson 2018). Because of the absence of cross-sectional variation, it is necessary to expand the sample around the threshold, which increases the precision of estimates but also raises bias due to potentially unobserved shocks. To this end, Gugler et al. (2020) select only one year, because shorter windows are not able to control for seasonal variation (Chen and Whalley 2012). Also, the time-series nature means that one needs to account for time-varying effects and autoregressive processes, which Gugler et al. (2020) credibly do by using local polynomials in time and other control variables (e.g., day-of-the-week effects). Finally, in the RDiT design it is impossible to directly test for sorting or selection issues (which is part of the SUTVA assumption), but Gugler et al. (2020) indirectly show this is not the case, documenting that events affected *all* British thermal power plants and anticipatory effects were unlikely because electricity cannot be stored at large scale (i.e., to produce before and sell after the event).[13]

Difference in Difference

The objective of the difference-in-difference (DID) design is to obtain average treatment effects of a policy intervention. The intuition is to compare treated and untreated units that are identical in characteristics. Martin, De Preux, and Wagner (2014) examine the effect of a carbon tax on electricity consumption and economic outcomes of firms in the manufacturing sector in the United Kingdom. Given the ease of implementation, DID designs have been abundant, but a drawback has been a difficulty to select valid counterfactuals and account for potentially confounding factors. Martin et al. (2014) address these concerns well and confirm that the treated and untreated units are comparable; they assess the common-trends assumption visually and statistically by including, amongst other things, a time-invariant eligibility dummy in the regression (as also done in Jaraite-Kažukauske and Di Maria 2016; Liu, Ma, and Xie 2020; Pretis 2022).

Less convincing is the DID design applied by Lin and Li (2011). They test the effect of carbon taxation in Finland, Denmark, Sweden, and the Netherlands on total carbon emissions at the country level, by comparing against a range of control countries which did not implement a carbon tax as a control group (e.g., Austria, Luxembourg). Clearly the common trends assumption is implausible here, and the paper fails to show a parallel evolution before the carbon tax implementation. Similarly, in a study of the effect of regional emissions trading schemes on emissions in Japan, Sadayuki and Arimura (2021) do not adequately test pre-carbon tax trends in outcomes as they have insufficient data, such that the robustness of results is reduced.[14] In Lin and Li (2011), there is also likely to be variation in treatment intensity across countries, for example due to varying exemption rules across countries to certain sectors or firms.[15] The coarse analysis at country-level aggregates all treated units, violating the identifying assumption that all units receive the same treatment. Furthermore, carbon price levels varied across countries, which implies heterogeneity in treatment levels, thus violating the SUTVA.

In addition to main DID assumptions, in studies using within-country variation, it is important to study self-selection into treatment. Martin et al. (2014) argue that this is likely in their setting; in the United Kingdom the carbon tax was part of a "Climate Change Levy" (CCL) that also included voluntary energy reduction agreements in return for reduced tax rates. Therefore, Martin et al. (2014) combine a DID with an instrumental variable (IV) approach. They suggest as an instrument a dummy on whether plants were eligible for tax exemption, which is a predictor of a firm being subject to carbon tax because participation in the voluntary agreement is linked to the tax regime (under which CCL plants saw highly discounted tax liability). Martin et al. (2014) also credibly argue that the instrument is independent, given that the reduced tax rates only depended on the eligibility for exemption.

In most DID and many other quasi-experimental designs, treatment spillovers can affect estimates, but spillovers often cannot be directly measured. If unaccounted for they violate the first component of the SUTVA; that there is no inference between units. Leroutier (2022) offers some useful suggestions for how to verify spillovers, as discussed, by assessing electricity imports and the "waterbed effect." Fowlie et al. (2012), who investigate the effect of an emissions trading scheme on NO_x emissions in California, is another notable study using DID (together with a matching approach, which is discussed in detail below) by comparing industrial facilities that were regulated under the RECLAIM emissions trading scheme to those that were exempt. The paper shows that the unconfoundedness assumption is satisfied. It also shows that treatment spillover is unlikely because dropping the nearby facilities from the control group does not significantly change the results. These approaches could be used in other studies to limit bias from spillovers.[16] Data is a key limitation in testing for spillovers. Wagner et al. (2014), using French plant-level information, are able to test if the impact of the ETS was different between *firms* that consisted of ETS and non-ETS plants (i.e., they had opportunity to shift emissions to non-ETS plants) and firms that had only ETS plants.[17]

Many of the above violations in the DID design are particularly apparent in a range of studies on China that aim to reduce the evidence gap on lower-income countries. Results from these results are unlikely to be robust given a range of issues (of which some are specific to the context). First, studies frequently do not describe why specific control variables are included. Second, in the case of China, firms could have influenced treatment status, because the policy was publicly announced, and therefore results may be upwards biased (if not controlled for).[18] Third, the quota allocation, price, and coverage were different across treated provinces (Liu et al. 2020), which means treatments were heterogeneous and the SUTVA assumption may be violated. Fourth, various other policies occurred alongside the ETS, including the 2013 Air Pollution Prevention Action Plan and 2015 Energy Use Transaction Pilot (Tang et al. 2021). If these initiatives were effective, estimates may capture those effects and results are then upwards biased. Finally, cross-regional spillovers are plausible in the China setting, but frequently not accounted for. Studies in which these concerns are prevalent include B. Zhou et al. (2019) and Hu et al. (2020). As described above, some of these concerns can be addressed with better data. Liu et al. (2020) analyze the effect of the emissions trading scheme on energy demand across *cities* in treated and untreated provinces in China, using a host of validity checks on assumptions and spillovers proposed by Pretis (2022) and Fowlie et al. (2012).[19] An alternative solution to address these features in the China setting is to include sector and "announcement" dummies (to capture announcement instead of start year), and deploy a difference-in-difference-in-differences (DDD) approach (Cui, Zhang, and Zheng 2018).[20]

DID with Matching

One way to address the issue of control group validity in DID design is to introduce a matching procedure. However, the limitation of the approach is that a researcher cannot balance unobserved confounders and a lack of common support means you cannot identify an average treatment effect of the treated (ATT), therefore leading to low external validity. The intuition is to compare only treatment with control units that are the same or similar in characteristics. Petrick and Wagner (2014), who review the effect of the EU ETS on emissions and economic performance in the German manufacturing sector, perform the mandatory checks on unconfoundedness and pre-trends, and in addition introduce a propensity score matching (PSM) approach. The main idea is to use firm characteristics to generate scores to firms on the propensity to participate in the ETS and match firms that have similar scores. An important feature (as in the synthetic control and RD design) is to check the robustness of this procedure (e.g., check the balancing of observed characteristics). Matching and results from Petrick and Wagner (2014) are credible given that they extensively evaluate PSM across different base years, stratifications, and matching algorithms. Fowlie et al. (2012) and Muûls et al. (2022) are also notable studies in this regard and review well the robustness of the matching approach (e.g., inclusion and exclusion of covariates). Results by Calel (2020), who examines the effect of the EU ETS in the United Kingdom, and Calel and Dechezleprêtre (2016), who study the effects of the ETS on low-carbon innovation among firms in the EU, both provide largely credible estimates given that they thoroughly perform checks on identifying assumptions, deploy the necessary robustness checks on matching procedures, and use several (novel) strategies to test whether estimates can be extrapolated to the population of firms. Finally, an important issue of the matching approach is external validity. Dechezleprêtre, Nachtigall, and Venmans (2018), who explore the effect of the EU ETS on emissions and economic outcomes among firms in France, the Netherlands, Norway, and the United Kingdom, use a matching procedure, which increases the precision of their causal inferences, but also reduce the sample size (in their case from 8,200 to 1,787 firms), potentially reducing external validity of their estimates.

Instrumental Variables

The aim of the instrumental variable approach is to study the effect of an intervention at local level. It is useful when in a cross-sectional regression the carbon pricing variable is potentially correlated with the error term and thus creates endogeneity. For example, it may be that unobserved variables, e.g., institutional quality, affect both the independent and outcome variable. Rivers and Schaufele (2015) estimate the effect of a carbon tax on petrol demand in British Colombia using personal and corporate income tax revenues as an instrument for carbon taxation. The key assumption of the IV approach is that the instrument is relevant (i.e., highly correlated to the treatment variable) and independent (i.e., affects outcomes only through the treatment variable). Rivers and Schaufele (2015) show that their instrument correlates strongly with the carbon taxation variable, but whether it is exogenous is questionable given that income tax revenues may indirectly be a determinant of petrol demand through equilibrium effects. Sen and Vollebergh (2018), who study the effect of energy taxes on emissions from energy consumption, show that the exogeneity of taxes to demand is not always valid and is only applicable when evaluating short-run demand effects. They propose an improved instrument (which is the tax rates of *neighboring* countries), and show it is relevant (high correlation to energy taxes) and independent (taxes are unlikely related to energy demand of neighboring states through another channel than correlation between tax levels of the neighboring country). The general difficulty with the IV approach is that the relevancy assumption is testable, but the independence assumption is not and can only be inferred. A key feature of the IV approach for researchers is to argue sufficiently the underlying assumptions in each instrument. For example, Antweiler and Gulati (2016), who study effects of taxation on gasoline consumption, borrow an instrument for gasoline price (crude oil prices) from Li, Linn, and Muehlegger (2014), but do not verify sufficiently the assumptions that make crude oil prices an independent instrument.[21]

Results Summary

This section describes the causative evidence on carbon pricing for those studies deemed to offer valid results on the basis of the above discussion. Evidence on carbon taxation and cap-and-trade are discussed separately.

Carbon Taxation

The reliable quasi-experimental evidence suggests that carbon taxation reduces carbon emissions without affecting economic outcomes among firms (with lacking data on the non-industrial sector and households). Andersson (2019) finds that carbon taxation in Sweden reduced carbon emissions in the transport sector by 6.3 percent on average per year between 1990–2005 and by 10.9 percent when combined with a value-added tax. This is significant, because in Sweden the transport sector is responsible for the highest share of carbon emissions. The large magnitude may be explained by the fact that in contrast to carbon taxes in other sectors, no exemptions were applied. These results are convincing, given that the DID and synthetic control are well-implemented, although estimates may potentially change when different predictors, including lagged outcome variables, are applied (and thus different weights in the synthetic control group), or spillovers are accounted for. These results are robust to the main DID checks, although the synthetic donor pool includes incorrect units (e.g., Luxembourg), but this may not affect estimates significantly.[22] Pretis (2022), who studies carbon taxation in British Colombia, finds no effect of carbon taxation on aggregate emissions, but a reduction in transport emissions of 5 percent, which is comparable to estimates from Andersson (2019). Pretis' (2022) estimates are largely credible, although the study cannot account for potential spillovers from treated to untreated units.

In terms of economic outcomes, Flues and Lutz (2015) find that electricity taxation in Germany had no negative effects on firm turnover, exports, value added, investment, or employment. Their results are robust to most RD assumptions, although a more thorough check on spillovers between firms and on the unconfoundedness assumption may change estimates. In the United Kingdom, Martin et al. (2014) find that carbon taxation dropped energy intensity by 18.1 percent and electricity use by 22.6 percent among manufacturing firms, and that it did not negatively affect plant employment, revenue, productivity, or exit.[23] Their results are convincing given that all mandatory statistics tests and robustness checks for DID are performed.

Gugler et al. (2020) examine effects of carbon taxation on emissions in the United Kingdom power sector and find that the reduction in emissions three years after the intervention (which was around 26 percent) could for 60 percent be explained by changes in the carbon tax. They also find that the introduction of the tax led firms to substituted coal for (lower-carbon) gas. These results are valid given that the authors credibly address most assumptions and caveats of the RDiT approach. In the same sector and setting, Leroutier (2022) finds a 21–26 percent emissions reduction per year and finds similar abatement effects to Gugler et al. (2020). Her estimates are largely credible, although may be influenced somehow if more stringent robustness checks are performed on the predictors in the synthetic control method.

On the effect of carbon taxation on household demand, Rivers and Schaufele (2015) find that a US$ five-cent increase in the carbon tax generated a 12.5 percent reduction in petrol demand (a similar increase in the price of petrol generated a reduction of 1.8 percent). Their estimates likely show short-run effects because their instrument may not be entirely independent. In the context of the EU, Sen and Vollebergh (2018) find that a EUR 1 rise in energy taxes (in t/CO_2) reduced CO_2 emissions from energy consumption by 0.73 percent in the long-run (an average elasticity of 0.3). In Canada, Antweiler and Gulati (2016) find that a 1 percent rise in carbon taxes reduced gasoline demand by 1.3 percent and raised the purchase of fuel-efficient vehicles.[24] Estimates are largely credible, although underlying identifying assumptions of the instrument are not confirmed, which if not met may change results.

Cap-and-Trade

The available and credible quasi-experimental evidence on cap-and-trade suggests that it reduces emissions but has ambiguous results on firm outcomes, but generally effects have been moderated due to historically low carbon prices and generous free allocation especially to heavy industry. Wagner et al. (2014) find that in France plants reduced emissions by 15 percent and employment by 7 percent after the introduction of the EU ETS (although with effects significant only in Phase II). These results in France are credible because all identifying assumptions of the DID are verified. Similar robust DID evidence is offered by Petrick and Wagner (2014), who find that in Germany the EU ETS contributed to a emissions reduction of 20 percent (although reductions were only significant in Phase II of the ETS), and did not reduce employment, turnover, or exports. Dechezleprêtre et al. (2018) study the EU ETS in France, the Netherlands, Norway, and the United Kingdom, and provide credible estimates that the policy reduced emissions by respectively 6 and 15 percent during Phase I and II, without reducing firm profit or employment. Firm revenue and assets *increased* in the range of 7–18 percent and 6–10 percent. They also find that reductions in emissions were largest for firms that had highest emissions.

On low-carbon investment, Calel (2020) finds that the EU ETS increased patenting and R&D spending among treated firms in the United Kingdom, although it did not reduce carbon intensity of output (i.e., firms pursued innovation instead of technology adoption). Using a much larger sample on 18 EU countries, Calel and Dechezleprêtre (2016) show that the ETS raised low-carbon innovation among treated firms by 10 percent, while not crowding out innovation in other technologies. Both studies offer largely credible estimates given they thoroughly perform checks on identifying assumptions and show estimates to have external validity to the entire population of firms. There is comparable evidence on China, for which Cui et al. (2018) document that the ETS increased innovation in low-carbon technologies, and that effects were larger in areas and sectors with higher carbon prices. The study's results are credible given rigorous assumption and robustness checks. Liu et al. (2020) find that in China the ETS incentivized firms to substitute away from carbon-intensive inputs and that the policy positively affected innovation activities, with results being robust to standard DID assumptions and potential between-city spillovers. In what is so far the most comprehensive study, Muûls et al. (2022) study the ETS in France and find it to have reduced emissions by 8–12 percent, with no effect on firm performance, and no leakage through international markets, but increased low-carbon technologies investments, which is in line with most evidence above. Finally, Fowlie et al. (2012) show for the United States that emissions at RECLAIM facilities in California fell on average 20 percent in comparison to control facilities in California where command-and-control policies were applied. They do not find these effects to vary systematically across neighborhood demographic characteristics, which mitigates potential effects from differences in political and economic characteristics between the treatment and control group. Estimates are valid, in particular because of the study's credible matching procedure and robustness checks on treatment spillovers.

Concluding Remarks

This paper systematically evaluated the small but growing quasi-experimental evidence on carbon pricing, exploring potential (methodological) caveats and how they may bias results. Empirically, estimates that we deem to be credible show a causal effect between carbon pricing and reductions in carbon emissions, with ambiguous effects on economic outcomes. On carbon taxation, studies find emissions reductions in the range of 5–6 percent in the transport sector (in Sweden and Canada), 16–26 percent emissions reductions in the power sector (in the United Kingdom), and 18–22 percent reductions in energy intensity and electricity consumption (in France). Carbon taxation does not seem to influence firm outcomes, but low-carbon investment increased in the transport sector (in Sweden), although evidence is limited. Studies suggest that carbon taxation reduces gasoline demand, although evidence is thin and estimates range between 1.3–12.3 percent. In terms of cap-and-trade, the EU trading scheme (which has received most scrutiny), is shown to cut emissions in the range of 6–20 percent (mainly in Phase II), and had largely no effects on firm outcomes, with the exception of *increased* firm revenue and assets by 6–10 percent (in France, the Netherlands, Norway, the United Kingdom) and *reductions* in employment by 7 percent (in France). An ETS boosts low-carbon investment, according to evidence from both the EU and China, by around 10 percent. In general, there is a lack of studies and a limited geographical and sector scope in the literature.

Methodologically, the review shows that in the quasi-experimental carbon pricing literature, estimates should be considered with care because of three threats to internal and external validity: (a) inappropriate choice of method, (b) incorrect implementation of analysis (e.g., violating main identifying assumptions), and (c) data limitations. Counterfactual selection, pre-trends analysis, treatment spillovers, and a description of parallel policies and the general limitations of each quasi-experimental designs are some of the recurring areas in which studies fail to provide sufficient checks and documentation. A major challenge is the availability of quality, disaggregated data on emissions and economic variables. This leads carbon pricing effects on household consumption and behavior to have been understudied while more attention has been paid to large emitters in energy and industrial sectors. Fortunately, there is increased access to microdata across various geographies and there are several studies (which we highlight in this paper) that offer best practices in terms of selecting appropriate study settings and methods, and the procedures necessary to infer causal effects. More cross-learning across studies will improve the empirical evidence base going forward.

In addition, novel empirical strategies enable the obtaining of causal estimates on carbon pricing in settings where variation or an appropriate natural experiment is lacking. First, *prospective* policies may be a promising identification strategy. If policies are perceived as sufficiently large and lasting by agents, using changes in behavior from prospective policies as a proxy for potential behavior might provide useful insights on carbon pricing. This is particularly relevant for examining carbon pricing effects in contexts in which such policies have not yet been frequently implemented, e.g., lower-income countries. Second, the regression discontinuity "in time" approach enables the gathering of evidence when cross-sectional variation is lacking or observations are far away from the threshold. The approach is also helpful when other quasi-experimental designs can be used but the researcher wants to test for heterogeneous treatment effects and the validity of the control group, although there are some limitations that should be considered when adopting a RDiT design.

The emphasis of this paper was on reviewing quasi-experimental evidence on carbon pricing and the extent to which methodological caveats bias estimates. By doing so, it did not evaluate the potential equilibrium effects of carbon pricing, nor the extent to which models and reduced-form evidence deviate in its predictions. Also, the paper looked at the effects of *de facto* implementation of carbon pricing, yet price levels and permit allocation seem to matter for policy effectiveness (Fowlie and Perloff 2013; Rafaty et al. 2020). Further, there is value in evaluating whether and how estimates are affected when adjacent policies, such as border carbon adjustments, are enforced alongside carbon pricing policies. These are meaningful avenues for future research.

Notes

German Institute of Development and Sustainability (IDOS), Bonn, Germany; Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

Suggestions from the editor (Peter Lanjouw) and two anonymous referees substantially improved the paper. We are grateful for excellent research assistance from David Stoffel, Christoph Oberthür, and Carla Wolf. Kasper Vrolijk gratefully acknowledges financial support by the German Ministry for Economic Cooperation and Development. Misato Sato gratefully acknowledges support from the Grantham Research Institute on Climate Change and the Environment, at the London School of Economics, and the Economic and Social Research Council grants Centre for Climate Change Economics and Policy (CCCEP) (ref. ES/R009708/1) and PRINZ (ES/W010356/1).

1. Around 30 percent of published environmental economics papers in prominent economics journals use quasi-experimental designs (Deschenes and Meng 2018).

2. For a detailed review of the general quasi-experimental literature, or natural experiments or "program evaluation" literature, see Angrist and Krueger (1999) and Imbens and Wooldridge (2009). For a review of quasi-experimental designs in environmental economics, see Deschenes and Meng (2018) and Greenstone and Gayer (2009). For a comparative review of quasi-experimental design and structural modelling and other randomized and non-randomized approaches, see Rosenzweig and Wolpin (2000), Timmins and Schlenker (2009), Angrist and Pischke (2010), Heckman (2010), and Nevo and Whinston (2010). For a review of specific quasi-experimental designs and their methodological options and exclusion restrictions, see Abadie (2005) on difference-in-differences, Angrist, Imbens, and Rubin (1996) on instrumental variables, Lee and Lemieux (2010) on regression discontinuity, and Abadie (2021) on synthetic control design.

3. Synthetic control design is applied in settings where there is an absence of a credible control group. It uses a weighted group of units with similar characteristics that did not receive treatment to create a counterfactual. Regression discontinuity can be applied in settings where some units receive treatment when the value of an observed variable is above a certain threshold, while units with values below the threshold do not receive treatment. Difference-in-differences estimations are applied to contexts where some units (e.g., firms) experience a change in treatment status over time, while other similar units do not. Instrumental variables are useful in settings where the treatment variable is correlated with the error term.

4. The main difficulty with the cross-sectional approach is that while reverse causality can be tested, the omitted variable bias is untestable. Experimental and quasi-experimental research designs, which provide randomization in treatment, can address this difficulty, because when treatment is randomly assigned, treatment and control groups should be statistically similar across all dimensions, except exposure to the treatment. This means that, unlike the cross-section estimation, it is not necessary to specify and control for all confounding variables.

5. The first main assumption is that there is no interference between units i (for example, when treatment spills over to other units) and that there is no hidden variation in treatment across units i (for

instance, some units receive different levels of treatment than others). This is the stable-unit treatment value assumption (SUTVA) (Rubin 1980). The second assumption is that treated and control units have on average similar characteristics and that units only differ in terms of treatment assignment, i.e., treatment assignment is independent of potential outcomes conditional on the set of covariates X. This is the unconfoundedness or selection-on-observables assumption (Rubin 1990). Additional to these two common assumptions, each design has different assumptions based on their purpose.

6. For a general review of the carbon pricing literature, see Metcalf (2019) and Arlinghaus (2015). For an review of the empirical literature on environmental regulation and competitiveness, see Dechezleprêtre and Sato (2017).

7. Given the focus on quasi-experimental designs, estimates discussed in this paper do not account for and thus may reflect equilibrium responses (such as changes in firm behavior and investment) that follow from carbon pricing policies and that may indirectly alter emissions and economic outcomes.

8. Search term truncations were used to adapt for flexible word permutations (e.g., "pric").

9. In comparison to other designs, the RD approach has mild identifying assumptions and causal inferences are potentially more credible (Lee and Lemieux 2010).

10. The electricity tax represents a carbon tax in that it priced carbon content in line with international standards at EUR 44.4 per tCO_2 and raised electricity prices substantially by 27 percent and 15 percent in 2002 and 2005 respectively.

11. The study by Ivanov, Kruttli, and Watugala (2021), who study the effect of a prospective cap-and-trade policy on access to credit for GHG emitting firms in the United States, shows the potential magnitude of this issue. In their discussion of results, they show large standard deviations in outcome variables, suggesting that estimates might differ across the firm population (and therefore results have limited external validity). For example, loan maturity is on average 35 months at the threshold, but 14 months and 56 months at the 10^{th} and 90^{th} percentile.

12. For a review of "regression discontinuity in time", and the differences in approach and assumptions from traditional regression discontinuity design, see Hausman and Rapson (2018).

13. An alternative approach to check for anticipatory effects is to omit observations prior to the event, which does not affect the results of Gugler et al. (2020).

14. Olale, Yiridoe, Ochuodho, and Lantz (2019), who investigate the effect of carbon taxation on farm income and production costs in British Colombia, also do not provide evidence that the untreated units are valid counterfactuals for treated units, and assume common trends but do not statistically confirm them. Löfgren, Wråke, Hagberg, and Roth (2014), who study the effect of the EU ETS on investments in clean technologies in Sweden, similarly lack data on the announcement period, and can thus not examine pre-trends and check for announcement effects. Jaraite-Kažukauske and Di Maria (2016), who examine the effect of the EU ETS on emissions and economic outcomes in Lithuania, have access to data only for Phase I, while most other studies find significant effects on the ETS only in Phase II, which may explain their insignificant results.

15. Martinsson, Sajtos, Strömberg, and Thomann (2020) use sector-specific exemptions as a source of variation in their study on Sweden, but such variation may not be exogenous (e.g., lobbying of government by firms).

16. Petrick and Wagner (2014) deploy most relevant DID checks, but only deploy one spillover strategy by Fowlie et al. (2012). Liu et al. (2020), who analyzes the effect of the emissions trading scheme on energy demand across cities in treated and untreated provinces in China, uses all three strategies by Fowlie et al. (2012). Results from Klemetsen, Rosendahl, and Jakobsen (2020), who examine the impact of the ETS on environmental and economic performance of manufacturing firms in Norway, are questionable because they run no robustness checks regarding spillover effects (nor the parallel trends assumption, or the assumption that treatment variables are independent of the unobserved plant-specific fixed effects).

17. H. S. Chan, Li, and Zhang (2013), who study the effect of the ETS on firm costs, employment and revenues across European countries using a DID approach, cannot test for spillovers because the study uses *firm* rather than *plant* data and thus cannot identify whether ETS plants reallocated emissions

to non-ETS plants within the same firm. Generally, spillover effects seem to be significant. Sadayuki and Arimura (2021) show that regional carbon trading schemes in Japan reduced emissions at treated plants as well as non-treated plants of the same firm. Bartram, Hou, and Kim (2022) find that the California cap-and-trade program led financially-constrained firms to relocate emissions and output from California to other (non-regulated) states.

18. Results from Ott and Weber (2022), who study the effect of a carbon tax on household expenditure in Switzerland, are likewise questionable because the tax was publicly announced, meaning that households could have selected into the control group.

19. Liu et al. (2020) also provide a best practice study on China's emissions trading pilots and perform most robustness checks from Fowlie et al. (2012).

20. Several studies, including Tang et al. (2021) and D. Zhou et al. (2020), use a DDD but do not include the sector and announcement dummies, thus not improving on the regular DID approach.

21. They are that (a) the demand shocks in the panel are not correlated with demand or supply shocks in the baseline year of the panel, and (b) province-level demand shocks are uncorrelated with crude oil prices (Li et al. 2014, p. 20).

22. Andersson (2019) shows that excluding such countries does not change estimates in his study.

23. Note that because the control group consists of plants that joined the voluntary reduction agreements, the estimates capture the effect of carbon taxation *relative to* emissions targets.

24. Li et al. (2014), on which Antweiler and Gulati (2016) build their instrument, find for the United States that a 5-cent increase in the gasoline tax reduced gasoline consumption by 0.86 percent.

References

Abadie, A. 2005. "Semiparametric Difference-in-Differences Estimators." *Review of Economic Studies*, 72 (1): 1–19.

Abadie, A. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, 59 (2): 391–425.

Andersson, J. J. 2019. "Carbon Taxes and CO 2 Emissions: Sweden as a Case Study." *American Economic Journal: Economic Policy*, 11 (4): 1–30.

Angrist, J. D., G. W. Imbens, and D. B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91 (434): 444–55.

Angrist, J. D., and A. B Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*, (Vol. 3), edited by Orley C. Ashenfelter and David Card, 1277–366. Elsevier.

Angrist, J. D., and J.-S Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic perspectives*, 24 (2): 3–30.

Antweiler, W., and S Gulati. 2016. "Frugal Cars or Frugal Drivers? How Carbon and Fuel Taxes Influence the Choice and Use of Cars." SSRN Working Paper No. 2778868.

Arlinghaus, J. 2015. "Impacts of Carbon Prices on Indicators of Competitiveness: A Review of Empirical Findings." OECD Environment Working Papers, No. 87, OECD Publishing, Paris, https://doi.org/10.1787/5js37p21grzq-en.

Brehm, J., N. aus dem Moore, and H Gruhl. 2022. "Driving Innovation?–Carbon Tax Effects in the Swedish Transport Sector." Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg.

Bartram, S. M., K. Hou, and S Kim. 2022. "Real Effects of Climate Policy: Financial Constraints and Spillovers." *Journal of Financial Economics*, 143 (2): 668–96.

Böhringer, C., C. Fischer, K. E. Rosendahl, and T. F. Rutherford. 2022. "Potential Impacts and Challenges of Border Carbon Adjustments." *Nature Climate Change*, 12 (1): 22–29.

Calel, R. 2020. “Adopt or Innovate: Understanding Technological Responses to Cap-and-trade.” *American Economic Journal: Economic Policy*, 12 (3): 170–201.

Calel, R., and A. Dechezleprêtre. 2016. “Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market.” *Review of Economics and Statistics*, 98 (1): 173–91.

Carattini, S., and S Sen. 2019. *Carbon taxes and stranded assets: Evidence from Washington state*. SSRN Working Paper No. 3434841.

Chan, H. S., S. Li, and F. Zhang. 2013. “Firm Competitiveness and the European Union Emissions Trading Scheme.” *Energy Policy*, 63: 1056–64.

Chan, N. W., and J. W. Morrow. 2019. “Unintended Consequences of Cap-and-trade? Evidence from the Regional Greenhouse Gas Initiative.” *Energy Economics*, 80: 411–22.

Chen, Y., and A. Whalley. 2012. “Green Infrastructure: The Effects of Urban Rail Transit on Air Quality.” *American Economic Journal: Economic Policy*, 4 (1): 58–97.

Cui, J., J. Zhang, and Y. Zheng. 2018. “*Carbon pricing induces innovation: Evidence from China's regional carbon market pilots.*” Paper presented at the AEA Papers and Proceedings.

Davis, L. W., and L. Kilian. 2011. “Estimating the Effect of a Gasoline Tax on Carbon Emissions.” *Journal of Applied Econometrics*, 26 (7): 1187–214.

Dechezleprêtre, A., D. Nachtigall, and F. Venmans. 2018. “The Joint Impact of the European Union Emissions Trading System on Carbon Emissions and Economic Performance.” OECD Economics Department Working Papers, No. 1515, OECD Publishing, Paris, https://doi.org/10.1787/4819b016-en.

Dechezleprêtre, A., and M. Sato. 2017. “The Impacts of Environmental Regulations on Competitiveness.” *Review of Environmental Economics and Policy*, 11 (2): 183–206.

Deschenes, O., and K. C. Meng. 2018. “Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges.” In *Handbook of Environmental Economics* (Vol. 4), edited by Partha Dasgupta, Subhrendu K. Pattanayak and V. Kerry Smith, 285–332. Elsevier.

Fell, H., and P. Maniloff. 2018. “Leakage in Regional Environmental Policy: The Case of the Regional Greenhouse Gas Initiative.” *Journal of Environmental Economics and Management*, 87: 1–23.

Ferman, B., C. Pinto, and V. Possebom. 2020. “Cherry Picking with Synthetic Controls.” *Journal of Policy Analysis and Management*, 39 (2): 510–32.

Flues, F., and B. J. Lutz. 2015. “The Effect of Electricity Taxation on the German Manufacturing Sector: A Regression Discontinuity Approach.” *ZEW-Centre for European Economic Research Discussion Paper* (15-013).

Fowlie, M., S. P. Holland, and E. T. Mansur. 2012. “What Do Emissions Markets Deliver and to Whom? Evidence from Southern California's NOx Trading Program.” *American Economic Review*, 102 (2): 965–93.

Fowlie, M., and J. M. Perloff. 2013. “Distributing Pollution Rights in Cap-and-trade Programs: Are Outcomes Independent of Allocation? *Review of Economics and Statistics*, 95 (5): 1640–52.

Fried, S. 2018. “Climate Policy and Innovation: A Quantitative Macroeconomic Analysis.” *American Economic Journal: Macroeconomics*, 10 (1): 90–118.

Germeshausen, R. 2020. “The European Union Emissions Trading Scheme and Fuel Efficiency of Fossil Fuel Power Plants in Germany.” *Journal of the Association of Environmental and Resource Economists*, 7 (4): 751–77.

Greenstone, M., and T. Gayer. 2009. “Quasi-experimental and Experimental Approaches to Environmental Economics.” *Journal of Environmental Economics and Management*, 57 (1): 21–44.

Grubb, M., P. Drummond, A. Poncia, W. McDowall, D. Popp, S. Samadi, C. Penasco, K. T. Gillingham, S. Smulders, M. Glachant, G. Hassall, E. Mizuno, E. S. Rubin, A. Dechezleprêtre, and G. Pavan. 2021. “Induced Innovation in Energy Technologies and Systems: A Review of Evidence and Potential Implications for CO2 Mitigation.” *Environmental Research Letters*, 16 (4): 043007.

Gugler, K. P., A. Haxhimusa, and M. Liebensteiner. 2020. “Carbon Pricing and Emissions: Causal Effects of Britain’s Carbon Tax.” Available at SSRN: https://ssrn.com/abstract=4116240 or http://dx.doi.org/10.2139/ssrn.4116240.

Hausman, C., and D. S. Rapson. 2018. “Regression Discontinuity in Time: Considerations for Empirical Applications.” *Annual Review of Resource Economics*, 10: 533–52.

Heckman, J. J. 2010. “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy.” *Journal of Economic Literature*, 48 (2): 356–98.

Hintermann, B., and M. Žarković. 2021. “A Carbon Horse Race: Abatement Subsidies vs. Permit Trading in Switzerland.” *Climate Policy*, 21 (3): 290–306.

Hu, Y., S. Ren, Y. Wang, and X. Chen. 2020. “Can Carbon Emission Trading Scheme Achieve Energy Conservation and Emission Reduction? Evidence from the Industrial Sector in China.” *Energy Economics*, 85: 104590.

Imbens, G. W., and J. M. Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature*, 47 (1): 5–86.

Ivanov, I., M. S. Kruttli, and S. W. Watugala. 2021. *Banking on Carbon: Corporate Lending and Cap-and-Trade Policy*. SSRN Working Paper No. 3650447.

Jaraite-Kažukauske, J., and C. Di Maria. 2016. “Did the EU ETS Make a Difference? An Empirical Assessment Using Lithuanian Firm-level Data.” *Energy Journal*, 37 (1): 1–24.

Kaul, A., S. Klößner, G. Pfeifer, and M. Schieler. 2022. “Standard Synthetic Control Methods: The Case of Using all Preintervention Outcomes Together with Covariates.” *Journal of Business & Economic Statistics*, 40 (3): 1362–76.

Klemetsen, M., K. E. Rosendahl, and A. L. Jakobsen. 2020. “The Impacts of the EU ETS on Norwegian Plants’ Environmental and Economic Performance.” *Climate Change Economics*, 11 (1): 2050006.

Lee, D., and T. Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature*, 48 (2): 281–355.

Leroutier, M. 2022. “Carbon Pricing and Power Sector Decarbonization: Evidence from the UK.” *Journal of Environmental Economics and Management*, 111: 102580.

Li, S., J. Linn, and E. Muehlegger. 2014. “Gasoline Taxes and Consumer Behavior.” *American Economic Journal: Economic Policy*, 6 (4): 302–42.

Lin, B., and X. Li. 2011. “The Effect of Carbon Tax on per Capita CO2 Emissions.” *Energy Policy*, 39 (9): 5137–46.

Liu, C., C. Ma, and R. Xie. 2020. “Structural, Innovation and Efficiency Effects of Environmental Regulation: Evidence from China’s Carbon Emissions Trading Pilot.” *Environmental and Resource Economics* 1–28.

Löfgren, Å., M. Wråke, T. Hagberg, and S. Roth. 2014. “Why the EU ETS Needs Reforming: An Empirical Analysis of the Impact on Company Investments.” *Climate Policy*, 14 (5): 537–58.

Löschel, A., B. J. Lutz, and S Managi. 2016. “The Impacts of the EU ETS on Efficiency–An Empirical Analyses for German Manufacturing Firms.” *ZEW-Centre for European Economic Research Discussion Paper* (16-089).

Lutz, B. J. 2016. “Emissions Trading and Productivity: Firm-level Evidence from German Manufacturing.” *ZEW-Centre for European Economic Research Discussion Paper* (16-067).

Marin, G., and F. Vona. 2021. “The Impact of Energy Prices on Socioeconomic and Environmental Performance: Evidence from French Manufacturing Establishments, 1997–2015.” *European Economic Review*, 135: 103739.

Martin, R., L. B. De Preux, and U. J. Wagner. 2014. “The Impact of a Carbon Tax on Manufacturing: Evidence from Microdata.” *Journal of Public Economics*, 117: 1–14.

Martin, R., M. Muûls, and U. Wagner. 2013. *Carbon markets, carbon prices and innovation: Evidence from interviews with managers*. Paper presented at the Annual Meetings of the American Economic Association, San Diego.

Martinsson, G., L. Sajtos, P. Strömberg, and C. Thomann. 2020. "Carbon Pricing and Firm-Level CO2 Abatement: Evidence from a Quarter of a Century-Long Panel." *Manuscript, Swedish House of Finance, Stockholm*.

Meng, K. C. 2017. "Using a Free Permit Rule To Forecast the Marginal Abatement Cost of Proposed Climate Policy." *American Economic Review*, 107 (3): 748–84.

Metcalf, G. E. 2019. "On the Economics of a Carbon Tax for the United States." *Brookings Papers on Economic Activity*, 2019 (1): 405–84.

Muûls, M., R. Martin, J. Colmer, and U. Wagner. 2022. "Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence from the European Union Emissions Trading Scheme." CEPR Discussion Paper No. DP16982.

Nevo, A., and M. D. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24 (2): 69–82.

Olale, E., E. K. Yiridoe, T. O. Ochuodho, and V. Lantz. 2019. "The Effect of Carbon Tax on Farm Income: Evidence from a Canadian Province." *Environmental and Resource Economics*, 74 (2): 605–23.

Ott, L., and S. Weber. 2022. "How Effective Is Carbon Taxation on Residential Heating Demand? A Household-level Analysis." *Energy Policy*, 160: 112698.

Peng, H., S. Qi, and J. Cui. 2021. "The Environmental and Economic Effects of the Carbon Emissions Trading Scheme in China: The Role of Alternative Allowance Allocation." *Sustainable Production and Consumption*, 28: 105–15.

Petrick, S., and U. J. Wagner. 2014. "The Impact of Carbon Trading on Industry: Evidence from German Manufacturing Firms." *Available at SSRN 2389800*.

Pindyck, R. S. 2013. "Climate Change Policy: What Do the Models Tell Us? *Journal of Economic Literature*, 51 (3): 860–72.

Pindyck, R. S. 2017. "The Use and Misuse of Models for Climate Policy." *Review of Environmental Economics and Policy*, 11 (1): 100–14.

Pretis, F. 2022. "Does a Carbon Tax Reduce CO2 Emissions? Evidence from British Columbia." *Environmental and Resource Economics*, 83 (1): 115–44.

Rafaty, R., G. Dolphin, and F. Pretis. 2020. "Carbon Pricing and the Elasticity of CO2 Emissions." Cambridge Working Papers in Economics 20116.

Rivers, N., and B. Schaufele. 2015. "Salience of Carbon Taxes in the Gasoline Market." *Journal of Environmental Economics and Management*, 74: 23–36.

Rosenzweig, M. R., and K. I. Wolpin. 2000. "Natural" Natural Experiments" in Economics." *Journal of Economic Literature*, 38 (4): 827–74.

Rubin, D. B. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment." *Journal of the American Statistical Association*, 75 (371): 591–3.

Rubin, D. B. 1990. "Formal Models of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference*, 25 (3): 279–92.

Runst, P., and A. Thonipara. 2019. "Why the Scope of the Carbon Tax Matters–Evidence from the Swedish Residential Sector (September 11, 2019)." USAEE Working Paper No. 19-416. Available at SSRN: https://ssrn.com/abstract=3452019 or http://dx.doi.org/10.2139/ssrn.3452019.

Sadayuki, T., and T. H. Arimura. 2021. "Do Regional Emission Trading Schemes Lead to Carbon Leakage within Firms? Evidence from Japan." *Energy Economics*, 104: 105664.

Sen, S., and H. Vollebergh. 2018. "The Effectiveness of Taxing the Carbon Content of Energy Consumption." *Journal of Environmental Economics and Management*, 92: 74–99.

Tang, K., Y. Zhou, X. Liang, and D. Zhou. 2021. “The Effectiveness and Heterogeneity of Carbon Emissions Trading Scheme in China.” *Environmental Science and Pollution Research*, 28 (14): 17306–18.

Timmins, C., and W. Schlenker. 2009. “Reduced-form versus Structural Modeling in Environmental and Resource Economics.” *Annual Review of Resource Economics*, 1 (1): 351–80.

Wagner, U. J., M. Muûls, R. Martin, and J. Colmer. 2014. *The causal effects of the European Union Emissions Trading Scheme: Evidence from French manufacturing plants. Paper presented at the Fifth World Congress of Environmental and Resources Economists*, Istanbul, Turkey.

Yamazaki, A. 2017. “Jobs and Climate Policy: Evidence from British Columbia’s Revenue-neutral Carbon Tax.” *Journal of Environmental Economics and Management*, 83: 197–216.

Yamazaki, A. 2019. *Who Bears More Burdens of Carbon Taxes? Heterogeneous Employment Effects within Manufacturing Plants*. https://akioyamazaki.weebly.com/uploads/5/6/9/8/56981769/carbon_tax_pl-npl_v3.pdf.

Yan, J. 2021. “The Impact of Climate Policy on Fossil Fuel Consumption: Evidence from the Regional Greenhouse Gas Initiative (RGGI).” *Energy Economics*, 100: 105333.

Zhou, B., C. Zhang, H. Song, and Q. Wang. 2019. “How Does Emission Trading Reduce China’s Carbon Intensity? An Exploration Using a Decomposition and Difference-in-differences Approach.” *Science of the Total Environment*, 676: 514–23.

Zhou, D., X. Liang, Y. Zhou, and K. Tang. 2020. “Does Emission Trading Boost Carbon Productivity? Evidence from China’s Pilot Emission Trading Scheme.” *International Journal of Environmental Research and Public Health*, 17 (15): 5522.