Overlapping Climate Policies*

Grischa Perino[1] Robert A. Ritz[2] Arthur A. van Benthem[3,4]

October 2023

Abstract

Countries around the world are enacting climate policies such as coal phase-outs, aviation taxes, and renewable energy support. These policies often overlap with a wider multi-jurisdictional carbon-pricing system like the EU's Emissions Trading System. We develop a general framework to study how effectively such "overlapping climate policies" can help combat climate change—depending on their design, location and timing. We find that some policies are truly complementary while others backfire by raising aggregate emissions. At a conceptual level, our model encompasses the market design of most carbon-pricing systems used in practice and a wide range of popular unilateral climate policies.

Keywords: overlapping policy, internal carbon leakage, waterbed effect, cap-and-trade, carbon pricing, carbon markets, hybrid regulation

JEL codes: H23 (externalities), Q54 (climate)

*We thank Jan Abrell, Severin Borenstein, Dallas Burtraw, Jim Bushnell, Meredith Fowlie, Don Fullerton, Reyer Gerlagh, Andrew Hinchberger, Marten Ovaere, Sebastian Rausch, Mar Reguant, Knut Einar Rosendahl, Herman Vollebergh, Maximilian Willner and Yuting Yang for helpful comments, and Heski Bar-Isaac (Editor) and three referees for their review and constructive feedback. We also thank seminar and conference participants at the Cambridge EPRG 2019 Spring Seminar, EAERE 2019 Annual Meeting, AERE 2021 Summer Conference, Mannheim 2021 Conference on Energy and the Environment, NBER 2021 Summer Institute, Toulouse School of Economics, the SWEEEP seminar at ZEW Mannheim, Florence School of Regulation at the European University Institute, and the University of Pennsylvania. We are grateful to Andrew Hinchberger for excellent research assistance. Perino's research was funded by the DFG (German Research Foundation) under Germany's Excellence Strategy, cluster EXC 2037 "Climate, Climatic Change, and Society" (project 390683824) and ARIADNE (BMBF project 03SFK5S0). Van Benthem thanks the National Science Foundation (award SES1530494), the Kleinman Center for Energy Policy at the University of Pennsylvania, Penn Global, the Analytics at Wharton Data Science and Business Analytics Fund, and the Wharton Global Initiatives Research Program for support.

[1]Department of Socioeconomics and Center for Earth System Research and Sustainability (CEN), University of Hamburg, Germany. Email: Grischa.Perino@uni-hamburg.de. [2]Energy Policy Research Group and Judge Business School, University of Cambridge, United Kingdom. Email: rar36@cam.ac.uk. [3]The Wharton School, University of Pennsylvania, 327 Vance Hall, 3733 Spruce Street, Philadelphia, PA 19104, United States. Phone: +1 (215) 898-3013. Fax: +1 (215) 898-7635. Email: arthurv@wharton.upenn.edu. [4]National Bureau of Economic Research.

1 Introduction

Under increasing pressure to combat climate change, jurisdictions around the world are enacting a plethora of policies to support decarbonisation. A notable example is the UK's Carbon Price Support which added a unilateral carbon fee of £18/tCO_2 to the allowance price already faced by its power generators under the European Union's Emissions Trading System (EU ETS) (House of Commons Library, 2018), and has been hailed as "perhaps the clearest example in the world of a carbon tax leading to a significant cut in emissions" (New York Times, 2019). Other examples include Spain's large-scale renewables support, extra carbon levies on air travel in several European countries, and Germany's phase-out of coal-fired power generation by 2038. These policies share a common feature: they are enacted by an individual jurisdiction for an individual sector—like electricity or aviation—that is also covered by the EU ETS, a multi-jurisdictional cap-and-trade system that spans 30 countries and covers power generation, industrials and domestic aviation.

In this paper, we refer to these as "overlapping climate policies" and ask a simple question: what is the climate benefit of a policy that overlaps an existing carbon-pricing system? As it is a global public good, any mitigation of climate change will be driven solely by the policy's impact on aggregate emissions.

To illustrate the different effects at work, suppose that an overlapping policy succeeds in reducing the jurisdiction's own domestic demand for emissions allowances in an ETS. How this translates, in equilibrium, into a reduction in aggregate emissions depends on the responses the policy induces in the product market and in the carbon market.

The first effect, in the product market, is what we term "internal carbon leakage". Consider an overlapping policy on, say, power generation that, holding fixed the system-wide carbon price, reduces *domestic* allowance demand by 1 ton of CO_2. If this emissions reduction is exactly offset by additional electricity imports that increase emissions by 1 ton of CO2 in *other* jurisdictions within the ETS, then the policy has no climate benefit: its rate of internal carbon leakage is 100% and it does not alter the aggregate demand for allowances across the ETS as a whole.

The second effect, in the carbon market, is the "waterbed effect" (Fankhauser et al., 2010; Böhringer, 2014). Suppose instead that the policy reduces ETS-wide allowance demand—as its internal carbon leakage is less than 100%. For a textbook cap-and-trade system with a fixed aggregate emissions cap (Montgomery, 1972; Baumol and Oates, 1988), this reduction is exactly offset by increased demand for emissions as the system's carbon price adjusts downwards. That is, the waterbed effect in the carbon market is 100% and no overlapping policy has a climate benefit—regardless of its internal carbon leakage.[1] By contrast, a carbon tax does not cap emissions and has no waterbed effect.

[1]Within a single-sector ETS, this equilibrium adjustment necessarily happens within the same sector;

Real-world carbon markets are more complicated than these textbook cases: they now often involve hybrid designs that combine elements of both price and quantity regulation. North American carbon markets—like the Regional Greenhouse Gas Initiative (RGGI) for power generation across eleven states in the northeastern US—use price floors and ceilings to contain the variability of the allowance price. The 2018 EU ETS reform enabled its "Market Stability Reserve" to cancel allowances under certain market circumstances, partly based on the idea that "the reserve will also enhance synergy with other climate and energy policies"(European Parliament and Council, 2015). As the emissions cap is no longer fixed, the waterbed effect is "punctured" (Perino, 2018) so overlapping policies in Europe *can* now reduce aggregate emissions.

This paper contributes to the literature in three ways. Our first contribution is to develop a general conceptual framework to study the climate-effectiveness of overlapping policies depending on their design, location and timing. Our framework nests a wide range of carbon-market designs and popular unilateral climate policies, and delineates internal carbon leakage and the waterbed effect. This approach is sufficiently general to encompass earlier literature that has examined instances of internal leakage or the waterbed effect for a specific overlapping policy in a particular carbon market.

Our second contribution is a novel result on the extent of the waterbed effect that unifies different hybrid carbon-market designs, and connects it to basic economic principles on pass-through from the classic literature on tax incidence (Jenkin, 1872; Weyl and Fabinger, 2013). Our model spans flexibility mechanisms based on allowance prices (including price ceilings and floors) (Roberts and Spence, 1976; Pizer, 2002; Newell et al., 2005; Borenstein et al., 2019; Karp and Traeger, 2021; Burtraw et al., 2022) and those, as in the EU ETS, that are based on intertemporal banking of allowances (Perino, 2018; Gerlagh et al., 2021). We show that, for price-based schemes, the waterbed effect lies between zero and 100%—and is independent of the overlapping policy—while, for quantity-based schemes, it is potentially very sensitive to the policy's timing and spans a wider range of values. We derive a unifying result that shows how a (dynamic) quantity-based flexibility mechanism can be represented by a (static) supply function.

Our third contribution is a new set of results on internal carbon leakage. We label as "supply-side" overlapping policies those that reduce the supply of dirty products, e.g., by unilaterally raising the carbon price for emissions-intensive production or directly limiting production at targeted firms as in a coal phase-out.[2] We show that supply-side overlapping

---

for example, if allowance demand from power generation falls for a given carbon price, it rises by the same amount again at equilibrium with an endogenous carbon price due to the fixed emissions cap. Within a multi-sector ETS, like the EU ETS, the induced increase in allowance demand may span several sectors.

[2]Our use of the term "supply-side" policy differs from the literature on the market for fossil resources (Sinn, 2008; Harstad, 2012)—our reference point is the market for goods produced by a polluting industry.

policies have positive internal carbon leakage—sometimes in excess of 100%—as they raise emissions demand in other jurisdictions that seek to "fill the gap" due to lower domestic production and the increased product price in the targeted sector.[3]

We label policies that reduce the (residual) demand for emissions-intensive production, e.g., by promoting renewables or energy efficiency, as "demand-side" policies. These policies have negative internal carbon leakage: the induced dip in domestic demand for emissions-intensive production reduces the product price and thereby reduces emissions from both domestic firms and firms in other jurisdictions. This result is the flip side of the widely-discussed "merit-order effect" through which renewable electricity generation with near-zero marginal cost reduces the wholesale electricity price (Sensfuß et al., 2008; Borenstein, 2012; Acemoglu et al., 2017; Antweiler and Muesgens, 2021).[4]

A caveat is that we do not endogenise a policy's fiscal implications—how costs are recovered for demand-side policies and how revenues are redistributed to consumers for supply-side policies—and how that might affect their leakage properties; our model assumes that these operate as lump-sum transfers or via general taxation.[5] The predicted signs of internal carbon leakage from our theory are consistent with empirical findings for supply- and demand-side policies (Vollebergh, 2018; Abrell et al., 2019a; Gerarden et al., 2020) that also abstract from fiscal implications.

Section 2 begins with a model-independent conceptual framework that provides a mapping—–in terms of internal carbon leakage and the waterbed effect—–from the domestic emissions cut an overlapping policy achieves to its equilibrium impact on aggregate emissions. Section 3 presents a theory of internal carbon leakage in the product market that yields simple leakage formulae for supply- and demand-side overlapping policies.[6]

[3]Our focus in this paper differs from "external" carbon leakage to jurisdictions *outside* a carbon-pricing system. Prior literature has examined the global impacts of unilateral policy in industrial sectors such as cement and steel where the scope of the product market is wider than that of the carbon price (Martin et al., 2014; Aldy and Pizer, 2015; Fowlie et al., 2016; Fowlie and Reguant, 2018). We here explore leakage among jurisdictions *inside* the system because (i) it is less well-understood in the literature, in part because it did not matter in systems with an 100% waterbed effect like the pre-2018 EU ETS; and (ii) it has received much less policy attention, despite likely being more important than its external cousin for sectors such as airlines and electricity. Internal carbon leakage as a result of overlapping policies has also been studied outside of the context of a carbon-pricing system; see, e.g., Goulder and Stavins (2011) and Goulder et al. (2012) on interactions between federal and state-level policies in the United States.

[4]Our finding of negative (internal) leakage of demand-side overlapping policies is distinct from negative (external) leakage via input-market effects in the general-equilibrium model of Baylis et al. (2013).

[5]See Section 3.4 for further discussion of this point. For example, if the investment cost of renewables support is instead funded via a levy on the retail electricity prices, then its overall price impact becomes ambiguous (e.g., Jarke and Perino, 2017). Our analysis, in effect, assumes that the direct effect—here, the merit-order effect—outweighs the fiscal effect. Our analysis also abstracts from the effects of carbon regulation on other—potentially unregulated—pollutants (Novan, 2017).

[6]While our main interest in this paper is "intra-industry" leakage that occurs in the sector targeted by an overlapping policy, our approach also allows for the presence of "inter-industry" leakage to other sectors that are part of a multi-sector ETS. For example, the reduction in the electricity price due to renewables support may also induce a change in the emissions of a sector, such as aluminium or steel, that

Section 4 presents a theory of the waterbed effect in the carbon market.

Section 5 illustrates the empirical usefulness of the framework with examples from Europe as well as North American carbon-pricing systems such as RGGI, the California-Québec carbon market, and Canada's federal minimum carbon price. We find a wide range of climate benefits—see Figure 1 for a visual summary. Section 6 concludes and suggests future research avenues.

In sum, while we do not attempt to quantify cost-effectiveness or broader welfare impacts, our results show how a policy's ability to combat climate change varies enormously depending on its design, location and timing—and the carbon market it overlaps with.

2 Conceptual framework

We begin by setting out a (static) conceptual framework that encompasses a wide range of carbon-market designs and delineates internal carbon leakage and the waterbed effect. We consider a multi-jurisdiction carbon-pricing system that covers multiple sectors, like the EU ETS. A single-sector system like RGGI is nested as a special case of our framework. Denote the system-wide carbon price by τ.

In general, an "overlapping climate policy" is any unilateral policy that targets a subset of jurisdictions (or subset of sectors) of a wider carbon-pricing system (ETS). Our main interest is in policies enacted by an individual jurisdiction for an individual sector, as when jurisdiction i introduces a coal phase-out policy in the electricity sector—independently of any other ETS jurisdictions, denoted by j.

We assume that, holding fixed the carbon price τ, the overlapping policy is successful at reducing i's *domestic* demand for emissions in the targeted sector, $\Delta e_i < 0$. We let ΔE denote the change in aggregate emissions demand across all jurisdictions and all sectors.

Our main question is, what is the overlapping policy's impact on *equilibrium* aggregate emissions ΔE^* with an equilibrium τ? This is the critical issue for climate change.

Our framework answers this question using three concepts. First, the rate of "intra-industry" internal carbon leakage in the product market captures emissions displacement due to i's policy within its targeted sector for a given system-wide carbon price:

$$L_i \equiv -\Delta e_j / \Delta e_i \text{ (fixed } \tau),$$

where Δe_j is the change in the emissions demand in the targeted sector of other jurisdictions j.[7] For example, this reflects how a coal phase-out by i may induce higher emissions

uses electricity as an input and is also part of the ETS. We obtain conditions under which intra-industry leakage is the dominant effect that determines the sign of internal carbon leakage.

[7] Notice that this is akin to the standard definition of external carbon leakage (e.g., IPCC, 2007) that

from power generation in j due to its impact on electricity prices—even for a fixed τ.

Second, for a multi-sector ETS, an additional form of internal carbon leakage may arise: emissions displacement to non-targeted sectors for which the targeted sector supplies an input. We define the rate of "inter-industry" internal leakage associated with i's policy, again for a fixed system-wide carbon price, as:

$$\ell_i \equiv -\Delta\widehat{E}/\Delta e_i \text{ (fixed } \tau\text{)},$$

where $\Delta\widehat{E}$ is the induced change in the emissions demand of non-targeted sectors. For example, this reflects how an industrial sector such as steel or aluminium uses electricity as a factor of production and may adjust its use both of electricity and of other inputs in response to a change in the price of electricity due to an overlapping policy.

In sum, the (net) system-wide change in emissions demand $\Delta E = \Delta e_i + \Delta e_j + \Delta\widehat{E}$ in a multi-sector ETS can therefore also be written as $\Delta E = [(1 - L_i) - \ell_i]\Delta e_i$ (fixed τ).

Third, the waterbed effect in the carbon market captures the emissions impacts arising from induced changes to the equilibrium system-wide carbon price τ:

$$W \equiv 1 - \Delta E^*/\Delta E \text{ (endogenous } \tau\text{)}.$$

This translates the system-wide change in emissions demand ΔE due to i's policy into an equilibrium change in emissions ΔE^* (Eichner and Pethig, 2019; Osorio et al., 2020; Böhringer and Rosendahl, 2022).[8] For example, the waterbed effect captures how a coal phase-out induces a reduction in the system's carbon price which then induces further emissions changes in the electricity sector—and in any other sectors covered by the ETS. Cap-and-trade with a fixed emissions cap has $W = 1$ (as $\Delta E^* \equiv 0$) while a carbon tax has $W = 0$ (so $\Delta E^* = \Delta E$); we will show how real-world carbon markets like the EU ETS or RGGI typically feature punctured waterbeds $W \in (0, 1)$.

We can now state the central equation of our conceptual framework. To distinguish clearly between the two cases, we use $\kappa \in \{0, 1\}$ as a binary indicator for whether the ETS is multi-sector ($\kappa = 1$) or single-sector ($\kappa = 0$) without inter-industry internal leakage.

Lemma 1 *The equilibrium change in aggregate emissions due to an overlapping policy satisfies:*

$$\Delta E^* = \Big[[1 - \widetilde{L}_i(\kappa)](1 - W)\Big]\, \Delta e_i \textit{ (endogenous } \tau\textit{)}, \tag{1}$$

relates to shifting of emissions to jurisdictions *outside* the system.

[8] Like other literature, we address the impact of an overlapping policy on aggregate emissions *given* the rules of the ETS. We acknowledge that jurisdictions make periodic adjustments to emissions caps and flexibility mechanisms based on past market outcomes that are—among other things—affected by the accumulation of overlapping policies. Such rule changes can affect the magnitude of the waterbed effect (Perino, 2018). Incorporating them would require a model of the political process driving ETS design.

where $\widetilde{L}_i(\kappa) \equiv [L_i + \kappa\ell_i]$ *is the aggregate rate of internal carbon leakage that combines intra-industry and inter-industry leakages.*

Lemma 1 incorporates the equilibrium carbon price path via the waterbed effect. It shows how internal carbon leakage and the waterbed effect together drive the sign and magnitude of the overlapping policy's equilibrium impact on aggregate emissions. Letting $R_i \equiv [1 - \widetilde{L}_i][1 - W]$, policies for which leakage and waterbed effects are such that $R_i \geq 1$ are complementary (or super-additive) policies while those with $R_i < 1$ are substitutes (or sub-additive). If $R_i < 0$, substitutability is so strong that aggregate emissions rise ($\Delta E^* > 0$) even though local emissions fall ($\Delta e_i < 0$).[9]

A key advantage of the decomposition in Lemma 1 is that it enables our analysis to proceed sequentially. First, we derive intra-industry leakage for a single-sector ETS (L_i) and inter-industry leakage for a multi-sector ETS (ℓ_i) for different overlapping policies. This yields the aggregate rate of internal carbon leakage $\widetilde{L}_i$: this maps Δe_i to ΔE. Second, we derive the extent of the waterbed effect W under different carbon-market designs: this maps ΔE to ΔE^*. In Appendix A.1, we develop a multi-period generalisation of this conceptual framework and show how the basic structure of Lemma 1 is preserved.

While other emissions decompositions are possible, we propose Lemma 1 as the simplest and analytically clearest framework. Other work estimates a "meta" version of internal leakage $L_i^{\mathrm{M}} = 1 - R_i$ that obscures two distinct phenomena—internal carbon leakage and the waterbed effect—that should be clearly delineated, not least because of their different policy implications. For example, if $L_i^{\mathrm{M}} = 1$ because $W = 1$ then no overlapping policy can combat climate change—thus helping explain the EU ETS's 2018 MSR reform; by contrast, if $L_i^{\mathrm{M}} = 1$ because $\widetilde{L}_i = 1$ then this as such reveals the limitations only of a single policy. In Appendix A.2, we explain how different leakage concepts relate to ours, and why our sequential approach is cleanest.

3 Internal carbon leakage

We next present a theory of internal carbon leakage, with a view to obtaining intuitive formulae for the equilibrium rate of intra-industry carbon leakage L_i in the sector targeted by an overlapping policy. We then discuss the robustness of the results to different modelling assumptions in a single-sector ETS ($\kappa = 0$) and additional results on inter-industry leakage ℓ_i within a multi-sector ETS ($\kappa = 1$).

[9] We do not attempt to explain the overlapping policy's size—that is, its impact on i's domestic emissions demand, $\Delta e_i < 0$—rather we are interested in understanding the equilibrium aggregate impact ΔE^* of a given policy-driven local impact Δe_i.

We denote the overlapping policy implemented by i as λ_i. In what follows, this policy will reduce i's domestic emissions demand, $de_i/d\lambda_i < 0$ in the targeted sector, but may also alter j's emissions. In the main text, we focus on a marginal policy change, for which intra-industry leakage $L_i = (-de_j/d\lambda_i)/(de_i/d\lambda_i)$. As per our conceptual framework, the system-wide carbon price τ is held fixed in this leakage analysis.

3.1 Model setup

We begin with a very simple model to make our main points about intra-industry carbon leakage. A representative firm in each jurisdiction k produces output x_k $(k = i, j)$. Firm k's emissions are $e_k = \theta_k x_k - a_k$ where a_k is abatement and its baseline emissions without any abatement, $e_k|_{a_k=0} = \theta_k x_k$, have an emissions intensity θ_k.

In general, firm k's cost function $G_k(x_k, a_k)$ depends on its output and abatement. For expositional convenience, we focus in the main text on the separable case, $G_k(x_k, a_k) = C_k(x_k) + \phi_k(a_k)$.[10] For a well-behaved solution, we assume $C_k(0) = C_k'(0) = 0$, $C_k'(x_k) > 0$ for $x_k > 0$, and $C_k''(x_k) > 0$ as well as $\phi_k'(a_k) > 0$ for $a_k > 0$ and $\phi_k''(a_k) > 0$.

The firms face an inverse demand function $p(X)$ for their product, where $X \equiv x_i + x_j$ is total output and $\varepsilon^D \equiv -p(\cdot)/Xp'(\cdot) > 0$ is the price elasticity of demand. We interpret $p(X)$ to represent consumers in jurisdiction i, served partly by domestic production and partly by imports from j. Internal carbon leakage then captures the extent to which i's consumers, due to an overlapping policy, are increasingly served by j's production. This form of leakage receives perhaps the most attention in the policy debate.[11]

To obtain a first set of benchmark results, we assume perfect competition in the product market. Firm k faces a carbon price $\tau_k(\tau)$ on each unit of emissions, which depends on the system-wide carbon price τ (held fixed, as per Lemma 1) and may also depend on the particular kind of overlapping policy (as detailed below).

To maximise profits, firm k solves $\max_{x_k,a_k} \Pi_k = px_k - G_k(x_k, a_k) - \tau_k(\theta_k x_k - a_k)$, where its emissions satisfy $e_k = \theta_k x_k - a_k$. The first-order conditions write as:

$$p = C_k'(x_k) + \tau_k\theta_k \equiv \widehat{C}_k'(x_k) \tag{2}$$

$$\tau_k = \phi_k'(a_k), \tag{3}$$

so the product price is equal to k's marginal cost of output plus per-unit carbon costs based

[10] A separable cost function can be interpreted as an end-of-pipe technology which cleans up production *ex post*. Examples include carbon capture and storage and the purchase of carbon offsets.

[11] The demand function $p(X)$ could also be interpreted as aggregate demand across consumers in i and j. Our interpretation is closer to policy concerns around carbon leakage and existing empirical work. Appendix B.3 shows that our key insights are robust to a multi-market formulation of internal leakage where each jurisdiction has its own demand function.

on its baseline emissions intensity, and the carbon price equals the marginal abatement cost.[12] Define the cost elasticity $\eta_k^S \equiv x_k \widehat{C}_k''(x_k)/\widehat{C}_k'(x_k) > 0$, noting that $\widehat{C}_k''(x_k) \equiv C_k''(x_k)$. By (2), k's supply curve is upward-sloping $x_k'(p) = 1/C_k''(x_k) > 0$, so $\varepsilon_k^S \equiv px_k'(p)/x_k(p) > 0$ is k's price elasticity of supply and, at the optimum, $\eta_k^S = 1/\varepsilon_k^S$. By (3), abatement rises with the domestic carbon price, $da_k/d\tau_k = 1/\phi_k''(\cdot) > 0$ but, due to cost separability, here does not affect the product-market outcome.

In our extensions, summarized in Section 3.4, we relax the assumptions of cost separability, a single product market, and a marginal policy change.

3.2 Supply-side overlapping climate policies

We begin with two "supply-side" policies that reduce the supply of dirty products, e.g. by unilaterally raising the carbon price for emissions-intensive production or directly limiting production of targeted firms as in a coal phase-out.

Our first overlapping policy λ_i imposes an additional carbon price only in jurisdiction i. Formally, i's firm now faces a carbon price $\tau_i = \tau_i(\tau, \lambda_i)$, where $\frac{d}{d\lambda_i}\tau_i(\tau, \lambda_i) > 0$. A leading example is a unilateral carbon price floor that "tops up" the system-wide carbon price, $\tau_i = \tau + \lambda_i$, like Great Britain's Carbon Price Support for power generation that ran alongside the EU ETS (and continues in the UK ETS). Firm j is subject only to the system-wide carbon price, $\tau_j = \tau$.

This policy leads to an asymmetric cost shock, inducing i's firm to cut output and emissions, $dx_i/d\lambda_i < 0$ and $de_i/d\lambda_i < 0$, but raising the "competitiveness" of its rival in j. Since τ_j remains unchanged, j's abatement decision also stays unchanged so $de_j/d\lambda_i = \theta_j(dx_j/d\lambda_i)$, and any change in its emissions is driven solely by output. Hence the policy's rate of internal leakage will be signed by j's output response.

Our second policy has jurisdiction i institute a unilateral reduction in carbon-intensive production. A topical example is the phase-out of coal-fired power generation, which a number of European countries have individually committed to—alongside these plants being covered by the EU ETS. Formally, we suppose that i's policy λ_i directly imposes a (marginal) reduction in i's output, $dx_i/d\lambda_i < 0$. In contrast to the previous policy, the carbon price faced by i's firm remains unchanged, so $\tau_k = \tau$ for $k = i, j$, and so i's abatement decision here also is unchanged.

In the benchmark case without abatement, internal carbon leakage satisfies $L_i = (\theta_j/\theta_i)(-dx_j/d\lambda_i)/(dx_i/d\lambda_i)$, where the first term is jurisdictions' "relative dirtiness" and the second term is output leakage $L_i^{\mathrm{O}} \equiv (-dx_j/dx_i)$. With abatement, using $e_k = \theta_k x_k -$

[12] To guarantee an interior solution for outputs, assume $p(0) > \max_k\{C_k'(0) + \tau_k\theta_k\}$. The analysis can easily accommodate a fixed cost of abatement, $\phi_k(0) > 0$. If abatement occurs despite the fixed cost, then the following results apply directly. If this fixed cost makes abatement unprofitable, then $a_k = 0$—a case which is nested in our results where $\phi_k''(\cdot) \to 0$).

a_k, we have $L_i = (\theta_j/\theta_i)[(dx_j/d\lambda_i)/(-dx_i/d\lambda_i)+(da_i/d\lambda_i)/\theta_i)]$. Defining k's market share $\sigma_k \equiv x_k/X \in (0,1)$, we obtain:

Proposition 1 *A supply-side overlapping policy by jurisdiction i increases the product price $dp/d\lambda_i > 0$ and has intra-industry internal carbon leakage to jurisdiction j in the targeted sector of:*

$$L_i = \frac{\theta_j}{\theta_i}\left[\frac{\sigma_j}{(\sigma_j + \varepsilon^D/\varepsilon_j^S)}\right]\frac{1}{(1+\gamma\Omega_i)} > 0,$$

where $\gamma \in \{0,1\}$ equals zero (one) for a unilateral reduction in carbon-intensive production (unilateral carbon price), and $\Omega_i \equiv \frac{C_i''}{\theta_i^2\phi_i''}\left(1+\frac{(1-\sigma_j)\varepsilon_i^S/\varepsilon_j^S}{(\sigma_j+\varepsilon^D/\varepsilon_j^S)}\right) \geq 0$ is an abatement effect.

Proposition 1 provides a simple formula to quantify internal carbon leakage at the sectoral level (see Appendix B.2 for proof). For both kinds of supply-side policies, carbon leakage is always positive as the underlying output leakage is positive: i's firm loses market share to j's either because it incurs an asymmetric cost shock or has its production directly reduced. The flip side is that i's supply-side overlapping policies lead to an increase in the product price—as the rate of output leakage is always less than 100% in equilibrium. However, carbon leakage can exceed 100% if j's firm is sufficiently dirtier.

To understand the result, consider the unilateral cut in carbon-intensive production ($\gamma = 0$). The comparative statics are intuitive: output leakage $L_i^{\mathrm{O}} = \sigma_j/(\sigma_j + \varepsilon^D/\varepsilon_j^S)$ is more pronounced where: (i) j's market share is larger (higher σ_j), (ii) demand is relatively inelastic (lower ε^D), and (iii) j's firm is more supply-responsive, e.g., because of significant spare capacity (higher ε_j^S). In short, j's firm more aggressively "fills the gap" in market supply due to the policy when it is larger and more responsive. Output leakage then maps into carbon leakage by way of the relative emissions intensity θ_j/θ_i.

For a unilateral carbon price ($\gamma = 1$), internal carbon leakage is mitigated by abatement, as represented by $\Omega_i \geq 0$, which breaks the direct link between output and emissions: for a given output contraction by i—and resulting competitive gain by j—domestic emissions fall by more. If the marginal abatement cost curve is almost flat, with $\phi_i''(\cdot) \to 0$, then it easy to abate more and carbon leakage tends to zero, $L_i \to 0$ as $\Omega_i \to \infty$. By contrast, if abatement is infeasible, as $\phi_i''(\cdot) \to \infty$, representing a Leontief technology for which emissions are proportional to output, then $\Omega_i \to 0$ and so carbon leakage becomes identical to that under a unilateral production cut ($\gamma = 0$).

Note also that the formula for L_i does not depend on the precise functional form of i's policy $\tau_i = \tau_i(\tau, \lambda_i)$; at the margin, this matters for the *absolute* output and emissions impacts but not for the *relative* effects—which is what our leakage rate captures.

From a policy perspective, Proposition 1 formalises the rationale for a regional coalition within the EU introducing a carbon price floor for electricity generation (Newbery et al.,

2019) or a coordinated phase-out of coal-fired generation. Coordinated policies combine greater market share and reduced supply-responsiveness of those not part of the coalition than single-country action and thereby contain internal leakage.

To illustrate, suppose that the demand elasticity $\varepsilon^D = .5$ and that j has market share $\sigma_j = 20\%$ with supply-responsiveness $\eta_j^S = .2 \Leftrightarrow \varepsilon_j^S = 5$. With identical emissions intensities $\theta_i = \theta_j$ and no abatement ($\phi_i''(\cdot) \to \infty$ or $\gamma = 0$), $L_i = 67\%$ is driven by output leakage. If instead j's technology is less responsive $\eta_j^S = 1 \Leftrightarrow \varepsilon_j^S = 1$ or demand is more elastic $\varepsilon^D = 2.5$, then leakage falls to $L_i = 28\%$. Conversely, if instead j's firm is twice as dirty then leakage doubles to $L_i = 133\%$. If i has significant abatement opportunity faced with a unilateral carbon price, as implied by a cost function with $\frac{C_i''}{\theta_i^2 \phi_i''} = \frac{1}{3}$,[13] and also letting $\varepsilon_i^S = \varepsilon_j^S = 5$ and $\frac{\theta_j}{\theta_i} = 2$, this yields $L_i = 60\%$, illustrating how abatement can bring forth an aggregate emissions cut.

3.3 Demand-side overlapping climate policies

We now turn to three "demand-side" policies that reduce the residual demand for emissions-intensive production: promoting zero-carbon renewables, an energy-efficiency program, and a carbon-consumption tax. We model an overlapping policy λ_i by jurisdiction i and write $p(X; \lambda_i)$ where $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i) < 0$ so the overlapping policy reduces demand for both i and j's firms. Both firms continue to face the common carbon price τ.

The policies fit into this setup as follows. First, for the renewables program, we write demand as $p(X; \lambda_i) = p(X + \lambda_i)$ where λ_i is the volume of zero-carbon electricity supported by the policy. Second, for the energy-efficiency program, we write direct demand as $D(p; \lambda_i) = (1 - \lambda_i) D(p)$ so it reduces demand by a fraction $\lambda_i < 1$ (for a given p) and hence $p(X; \lambda_i) = D^{-1}(X/(1 - \lambda_i))$. Third, for the carbon-consumption tax, we write $p(X; \lambda_i) = [p(X) - \lambda_i \theta_i]$ where the tax λ_i is levied on consumption according to i's baseline emissions intensity θ_i. In all three cases, $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i) < 0$ at an interior equilibrium.

Proposition 2 *A demand-side overlapping policy by jurisdiction i—(i) a renewables support program that brings in additional zero-carbon production, or (ii) an energy-efficiency program that reduces demand for carbon-intensive production, or (iii) a carbon-consumption tax—decreases the product price $dp/d\lambda_i < 0$ and has intra-industry internal carbon leakage to jurisdiction j in the targeted sector of:*

$$L_i = -\frac{\theta_j}{\theta_i} \frac{\sigma_j}{(1 - \sigma_j)} \frac{\varepsilon_j^S}{\varepsilon_i^S} < 0.$$

[13] For a quadratic cost function, C_i'' and ϕ_i'' are constant so $\frac{C_i''}{\theta_i^2 \phi_i''}$ depends only on exogenous parameters.

Demand-side overlapping policies have negative internal carbon leakage: j's firm is now directly affected by the policy and responds by also cutting output and emissions (see Appendix B.2 for proof). As imported emissions decrease, the product price is reduced and the aggregate emissions reduction is more pronounced than the local reduction.

For renewables support programs, the result is consistent with the well-known "merit-order effect" through which renewable electricity generation with near-zero marginal cost reduces the wholesale power price (Sensfuß et al., 2008; Borenstein, 2012; Cludius et al., 2014; Acemoglu et al., 2017; Antweiler and Muesgens, 2021). In the present analysis, the flip side of the merit-order effect is that electricity imports to i also decline—and so internal carbon leakage is negative.

Akin to Proposition 1, internal carbon leakage is more strongly negative where j's firm is dirtier, more supply-responsive and has greater market share. In addition, it is more pronounced if i's own supply-responsiveness is weaker; then i's output contraction is smaller relative to j's. As the system-wide carbon price τ remains fixed (as per Lemma 1), unilateral action here brings no extra abatement ($da_k/d\lambda_i = 0$ for $k = i, j$).

Proposition 2's internal leakage rate does not depend on any demand characteristics, including the precise form of $p(X; \lambda_i)$ and the demand elasticity ε^D. To first order, for a marginal policy, the reduction in i's production—and hence also of i's emissions—is proportional to $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i)$. This is also true, to first order, for the changes in j's production and emissions. So the relative magnitude of emissions changes, as captured by the leakage rate, does not depend on $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i)$—and so all three demand-side overlapping policies have identical leakage properties.

To illustrate, again using $\sigma_j = 20\%$, $\theta_i = \theta_j$, and $\varepsilon_i^S = \varepsilon_j^S$ yields internal carbon leakage of $L_i = -25\%$. If, instead, j's firms are twice as dirty or twice as supply-responsive as i's, leakage doubles in absolute terms to $L_i = -50\%$. With both $\theta_j/\theta_i = 2$ and $\varepsilon_j^S/\varepsilon_i^S = 2$, internal leakage becomes $L_i = -100\%$, and so the aggregate reduction in emissions demand is now twice the size of the local reduction.

3.4 Discussion of robustness and extensions

Single-sector ETS For a single-sector ETS ($\kappa = 0$), like RGGI, we can already draw a number of conclusions as the rate of aggregate internal carbon leakage $\widetilde{L}_i(0) = L_i$ is then driven solely by intra-industry leakage. One immediate conclusion, combining Lemma 1 and Proposition 1, is that supply-side policies are necessarily sub-additive with $R_i < 1$ for *any* waterbed effect $W \in [0, 1]$. By contrast, by the same logic and Proposition 2, demand-side policies may be genuinely complementary, with $R_i > 1$, as long as the waterbed effect is not too pronounced.

These insights on intra-industry leakage from the simplified model are robust in a number of directions. First, in Appendix B.1, we solve the model with general cost functions $G_k(x_k, a_k)$, and show that—while non-separability between production costs and abatement costs creates additional effects—the key insights from the separable case carry over (Propositions 1A and 2A). Second, in Appendix B.3, we show that our conclusions extend to a model of multi-market internal carbon leakage in the targeted sector in which both firms now serve both markets i and j—so there is an additional channel of leakage in that i's exports to j may also be affected by the overlapping policy (Propositions 1M and 2M). Third, in Appendix B.4, we confirm that our conclusions also apply for larger, non-marginal changes in i's policy.

Multi-sector ETS For a multi-sector ETS ($\kappa = 1$), aggregate internal carbon leakage is $\widetilde{L}_i(1) = [L_i + \ell_i]$, by Lemma 1, where $\ell_i \equiv -\Delta\widehat{E}/\Delta e_i$ is "inter-industry" internal carbon leakage. In Appendix C, we present an extension that captures how *non-targeted* ETS sectors adjust their emissions because they purchase an input from the targeted sector; for example, industrial sectors such as aluminium or steel in the EU ETS use electricity which is subject to an overlapping policy such as a coal phase-out.

We derive simple formulae to characterize inter-industry internal leakage (Propositions 5 and 6). One important question is when aggregate internal leakage follows intra-industry leakage in that $\text{sign}\{\widetilde{L}_i(1)\} = \text{sign}\{L_i\}$. A grossly sufficient condition for robustness in this sense is that inter-industry ℓ_i is positive (negative) for a supply-side (demand-side) overlapping policy. Appendix C shows that this property holds in a range of cases for production technologies such as Cobb-Douglas; one example is under constant returns to scale with price-inelastic demand in the non-targeted sector—which is a common empirical finding for emissions-intensive industrial sectors such as cement and steel (e.g., Mathiesen and Maestad, 2004; Szabo et al., 2006). We also obtain conditions for intra-industry leakage to dominate when it has the opposite sign to inter-industry leakage.

Fiscal implications A caveat is that our modelling of internal carbon leakage does not endogenise an overlapping policy's fiscal implications. For example, if the investment cost for a renewables support program is recouped as a lump-sum charge to consumers (or via general taxation), then our result of negative internal carbon leakage from Proposition 2 applies. If it is instead funded via a levy on the retail electricity price, then its overall price impact—and hence the sign of internal leakage—becomes ambiguous (e.g., Jarke and Perino, 2017) if the retail price rises even though the wholesale price declines. Our analysis, in effect, assumes that the merit-order effect outweighs any such fiscal effect.

Our Proposition 2 is consistent with empirical evidence, summarized further in Section 5, on the cross-border impacts of demand-side policies. Also abstracting from fiscal implications, Abrell et al. (2019a) document negative leakage for policy support for wind and solar power in Germany and Spain. However, when controlling for levy-based funding of the renewable-energy subsidies, they find that retail electricity prices still decline for wind power but now *increase* for (relatively more costly) solar. This suggests that demand-side overlapping policies could, in some cases, have significantly higher leakage rates when accounting for how abatement costs are paid for.[14] Subsequently, in July 2022, Germany switched to funding its renewables support out of general taxation so our Proposition 2 again applies directly.

Conversely, we do not capture how the fiscal revenue raised by a unilateral carbon levy—one of our supply-side overlapping policies—is used. If it is returned to consumers as a lump-sum payment, then our result from Proposition 1 holds. If it is instead used to *subsidize* consumer electricity prices then this introduces an additional effect that may push the leakage rate downwards—but is beyond the scope of our model. Again, the results from our theory are consistent with prior empirical work which finds positive—and often high—internal leakage rates from supply-side overlapping policies (see Section 5). As neither our theory nor prior empirical work consider their fiscal implications, it is implicit that these run via general taxation—as in Proposition 1.

In sum, the details on how abatement costs are recovered for demand-side policies, and how revenues are redistributed for supply-side policies, have the potential to affect the leakage ordering between these two types of policies. Choices about cost recovery and revenue redistribution are therefore also an integral part of climate-policy design.

4 The waterbed effect

We now turn to the other building block of our conceptual framework of Section 2: the waterbed effect $W = 1 - \Delta E^*/\Delta E$, for which the carbon price τ is now endogenous. Again we focus on a "marginal" policy change so $W = 1 - (dE^*/d\lambda_i)/(dE/d\lambda_i)$.

We consider a stylised model of an allowance market, and denote the *inverse* aggregate demand function for allowances $\rho(E, \lambda_i)$, where E is aggregate emissions and $\partial\rho/\partial E < 0$. The impact of an overlapping policy on allowance demand $\partial\rho/\partial\lambda_i$ is negative (positive) if its aggregate internal carbon leakage $\widetilde{L}_i$ is below (above) 100%.

4.1 Flexibility mechanisms based on allowances prices

Most real-world carbon markets feature flexibility mechanisms based on allowance prices, which can be represented by an aggregate allowance supply $s(\tau)$ with $\partial s/\partial\tau \geq 0$. An

[14] Abrell et al. (2019a) assume a uniform fee for all consumers. In practice, however, German industrials with a high electricity demand paid a much lower rate than retail consumers (specifically, only 20% of the consumer rate or at most 0.5% of gross value added).

ETS with a fixed cap or any vertical section of an allowance supply curve are represented by $\partial s/\partial\tau = 0$. A carbon tax or a horizontal section of an allowance supply curve or price corridor as in RGGI (Burtraw et al., 2022), the UK ETS and the German cap-and-trade system for fossil fuels outside the EU ETS are represented by $s(\tau)$ being perfectly price elastic at a particular $\bar{\tau}$. We also capture allowance supply curves with $\partial s/\partial\tau > 0$, e.g., those that trace a social marginal damage curve of emissions (Roberts and Spence, 1976).

The equilibrium conditions of this carbon-market design jointly determine E and τ:

$$\rho(E, \lambda_i) - \tau = 0 \tag{4}$$

$$E - s(\tau) = 0, \tag{5}$$

where the former balances the marginal costs of abatement with the carbon price while the latter is market clearing for the allowance market. These conditions show how the overlapping policy changes the equilibrium carbon price (see Appendix D.1):

$$\frac{\partial\tau}{\partial\lambda_i} = \frac{\frac{dE}{d\lambda_i}}{\left(\frac{\partial s}{\partial\tau} - \frac{\partial E}{\partial\tau}\right)} \Longrightarrow \text{sign}\left(\frac{\partial\tau}{\partial\lambda_i}\right) = \text{sign}\left(\frac{dE}{d\lambda_i}\right) = \text{sign}\left(\frac{\partial\rho}{\partial\lambda_i}\right) = \text{sign}\left(\widetilde{L}_i - 1\right). \tag{6}$$

where $\frac{\partial E}{\partial\tau} < 0$ is the slope of the aggregate allowance demand curve and the overlapping policy reduces allowance demand ($\frac{dE}{d\lambda_i} \leq 0$) precisely when $\frac{\partial\rho}{\partial\lambda_i} \leq 0$. So any policy with aggregate internal leakage $\widetilde{L}_i$ below 100% also reduces the system-wide carbon price.

Using (6), we find that adjustments in total equilibrium emissions E^* are "spatially blind", i.e., independent of how the overlapping policy is spread over space:

$$\frac{dE^*}{d\lambda_i} = \frac{\partial s}{\partial\tau}\frac{\partial\tau}{\partial\lambda_i} = \frac{dE}{d\lambda_i}\frac{\frac{\partial s}{\partial\tau}}{\left(\frac{\partial s}{\partial\tau} - \frac{\partial E}{\partial\tau}\right)} = \frac{dE}{d\lambda_i}\frac{\omega^S}{(\omega^S - \omega^D)}, \tag{7}$$

where $\omega^D \equiv \frac{\partial E}{\partial\tau}\frac{\tau}{E} < 0, \omega^S \equiv \frac{\partial s}{\partial\tau}\frac{\tau}{s} \geq 0$ are the elasticities of allowance demand and supply.

Proposition 3 *The waterbed effect for an overlapping policy under a price-based flexibility mechanism in the carbon market is:*

$$W = 1 - \frac{\omega^S}{(\omega^S - \omega^D)} = 1 - m \in [0, 1], \tag{8}$$

which is independent of the specifics of the overlapping policy and where $m = \omega^S/(\omega^S - \omega^D) \in [0, 1]$ *is the rate of cost pass-through in the carbon market.*

Proposition 3 shows that the waterbed effect depends only on elasticities of allowance demand and supply—and is independent of the type of overlapping policy, notably the sign and extent of its internal leakage $\widetilde{L}_i$. Special cases include a carbon tax that leaves the quantity of emissions fully flexible ($\partial s/\partial\tau, \omega^S \rightarrow \infty$, so $W = 0$), thus negating any waterbed effect, and a textbook cap-and-trade system with a fixed emissions cap ($\partial s/\partial\tau = \omega^S = 0$, so $W = 1$). So the extent of the waterbed effect does not depend on how a particular overlapping policy affects the system's carbon price as per (6).

We thus uncover a natural connection between the waterbed effect and classic principles of tax incidence under perfect competition (Jenkin, 1872; Weyl and Fabinger, 2013). In that literature, pass-through determines the incidence of an excise tax, driven by elasticities of demand and supply. In the same way, Proposition 3 shows how a carbon-market design with higher pass-through m has less of a waterbed effect. Intuitively, a more tax-like system corresponds to a flatter supply (i.e., marginal-cost) curve—which exhibits stronger pass-through, and thus a lower W as quantity can adjust more flexibly.

For marginal policies that induce relatively small shifts in allowance demand, Proposition 3 applies also to step-wise allowance supply functions featured in the California-Québec system, RGGI, the German ETS, and the UK ETS. If the initial equilibrium is in a vertical (horizontal) section of the supply curve, the waterbed effect is 100% (zero).

The expected waterbed effect of marginal changes is in the intermediate range if, at the time of passing legislation for an overlapping policy, future market outcomes are still uncertain (Borenstein et al., 2019). Ex ante, if the probability that the equilibrium is in any of the horizontal sections of the allowance supply curve is π, then $W = 1 - \pi$. Ex post, the waterbed effect is either zero or 100%.

For larger interventions—where allowance demand moves across one or several kinks in the step-wise supply schedule—none of the extreme cases appropriately capture the impact on supply. The average waterbed effect of a large-scale policy can be computed by integrating over the marginal effects. In Appendix D.4, we extend Proposition 3 more formally to non-marginal policies.[15]

A dynamic price-based flexibility mechanism Because most cap-and-trade systems allow for banking of allowances across several periods, we show next that the waterbed effect for a dynamic price-based flexibility mechanism is also "temporally blind". This means that, given that an overlapping policy is fully anticipated, its impact on cumulative emissions is independent of how the cumulative shift in allowance demand is distributed over the periods linked by banking.

Consider a two-period model where, given the interest rate r, intertemporal arbitrage implies $\tau_2 = (1 + r)\tau_1$. (For the remainder of this section, indices represent time periods

[15]For non-marginal changes, the size of the demand shift (ΔE) matters as it determines the extent of movement of, and along, the allowance demand curve. So now the waterbed W is affected by the size of the discrete change in emissions induced in jurisdiction i and also by internal leakage $\widetilde{L}_i$ (as $\Delta E = (1 - \widetilde{L}_i(\kappa))\Delta e_i$). Nevertheless the analytical separation of $\widetilde{L}_i$ and W still makes sense: even for non-marginal policies, W only depends on the net shift of the demand curve—not on how it comes about.

$t = 1, 2.$) *Cumulative* allowance supply $s(\tau_1) \equiv s_1(\tau_1) + s_2((1+r)\tau_1)$, where $\frac{\partial s_t}{\partial \tau_1} \geq 0$ $(t = 1, 2)$. The three equilibrium conditions of this dynamic carbon market are:

$$\begin{aligned} \rho_1(E_1, \lambda_i) - \tau_1 &= 0 && (9) \\ \rho_2(E_2, \lambda_i) - (1+r)\tau_1 &= 0 && (10) \\ E_1 + E_2 - s(\tau_1) &= 0, && (11) \end{aligned}$$

where period-specific emissions E_1 and E_2 and the carbon price τ_1 are endogenous.

Proposition 3 extends to this dynamic setting, simply by reinterpreting ω^S and ω^D as long-run elasticities that capture the cumulative effect over both periods (see Appendix D.1) and defining the corresponding period-specific supply and demand elasticities $\omega_t^S \equiv (\partial s_t / \partial \tau_1)(\tau_1 / s_t) \geq 0$ and $\omega_t^D \equiv (\partial E_t / \partial \tau_1)(\tau_1 / E_t) < 0$.

Lemma 2 *The waterbed effect for an anticipated policy overlapping a two-period cap-and-trade system with a price-based flexibility mechanism is given by:*

$$W = 1 - \frac{\omega_1^S E_1 + \omega_2^S E_2}{\omega_1^S s_1 + \omega_2^S s_2 - \left(\omega_1^D E_1 + \omega_2^D E_2\right)} = 1 - \frac{\omega^S}{\left(\omega^S - \omega^D\right)} = 1 - m \in [0, 1]. \tag{12}$$

Given that intertemporal arbitrage is efficient and overlapping policies are anticipated, the carbon price in period 1 becomes a sufficient statistic for the "state of the market". As a result, the timing of an overlapping policy and its price-induced adjustment in allowance supply do not additionally matter. So, for any given $dE^*/d\lambda_i$, all $dE_1^*/d\lambda_i$, $dE_2^*/d\lambda_i$ and all $s_1(\tau_1)$, $s_2(\tau_2)$ that yield the same $s(\tau_1)$ are equivalent; that is, the allowance market is also temporally blind with respect to both anticipated changes in demand and supply.

Representing multi-period cap-and-trade systems and overlapping policies that differ in how shifts in allowance demand are distributed over time in a simple static setting is straightforward and—with respect to the waterbed effect—without loss of generality.

4.2 Flexibility mechanisms based on allowance banking

With the 2018 EU ETS reform, namely the introduction of cancellations within the Market Stability Reserve, a new form of flexibility mechanism gained prominence. Here we present a stylised two-period version of such a mechanism, and show how its economics contrasts markedly with a price-based flexibility mechanism.

The flexibility mechanism adjusts a *cumulative* cap $s(b) = s_1 + s_2(b)$ based on the number of allowances banked for future use in earlier periods, where $b = s_1 - E_1$ is banking at the end of period 1 and $\partial s_2 / \partial b \in [-1, 0]$.[16] A plain cap-and-trade system is

[16]Restricting $\partial s_2 / \partial b \geq -1$ captures the entire range of values relevant for both the EU ETS and

again nested as $\partial s_2/\partial b = 0$. Assuming that any constraints on banking and borrowing do not bind, intertemporal arbitrage again implies $\tau_2 = (1+r)\tau_1$. Any overlapping policy is announced at the beginning of period 1 and perfectly anticipated by market participants.

Analogous to the dynamic price-based flexibility mechanism, we have:

$$\rho_1(E_1, \lambda_i) - \tau_1 = 0 \tag{13}$$

$$\rho_2(E_2, \lambda_i) - (1+r)\tau_1 = 0 \tag{14}$$

$$E_1 + E_2 - s_1 - s_2(s_1 - E_1) = 0. \tag{15}$$

These equilibrium conditions yield the response of short-run equilibrium emissions to the overall change in allowance demand (see Appendix D.2):

$$\frac{\partial E_1^*}{\partial \lambda_i} = \frac{\frac{dE}{d\lambda_i}}{\left(1+\frac{\partial s_2}{\partial b}\xi\right)} (\beta - \xi) \Longrightarrow \text{sign}\left(\frac{\frac{\partial E_1^*}{\partial \lambda_i}}{\frac{dE}{d\lambda_i}}\right) = \text{sign}\,(\beta - \xi)\,. \tag{16}$$

Given the impact of the overlapping policy on total allowance demand ($dE/d\lambda_i$), the direction of the policy's impact on equilibrium emissions in period 1 ($dE_1^*/d\lambda_i$) depends on two effects. First, $\xi \equiv \frac{\partial E_1}{\partial \tau_1}/\frac{\partial E}{\partial \tau_1}$ measures the share of the price-responsiveness of long-run allowance demand that is due to the price-responsiveness of allowance demand in period 1. As a higher carbon price induces lower emissions in both periods, we always have $\xi \in (0,1)$ and so $(1+\frac{\partial s_2}{\partial b}\xi) > 0$. Second, $\beta \equiv \frac{dE_1}{d\lambda_i}/\frac{dE}{d\lambda_i}$ is the fraction of the overlapping policy's impact on total allowance demand that occurs in period 1. This is a measure of the timing of the overlapping policy; those with higher β are more "front-loaded".

Shifting the allowance demand curve to the left in period 1 ($dE_1/d\lambda_i < 0$), ceteris paribus, reduces first-period equilibrium emissions. The price drop triggered by the decrease in overall scarcity induces a movement along the demand curve and, ceteris paribus, increases first-period equilibrium emissions. The direct demand-shifting effect β (Perino, 2018) and the indirect price-mediated effect ξ (Rosendahl, 2019) are hence antagonistic.

Policy timing is therefore critical under a banking-based flexibility mechanism. First-period equilibrium emissions E_1^* decrease if a policy is sufficiently front-loaded, with $\beta \geq \xi$, in terms of its impact on allowance demand—but they increase for back-loaded policy.

This dependence on timing directly carries over to the change in total equilibrium emissions E^*. This occurs via adjustment of the cumulative cap, where the banking of

RGGI. RGGI used to have a banking-based design element: the number of banked allowances at the end of 2011, 2013 and 2020 were deducted from the baseline cap of future years (Regional Greenhouse Gas Initiative, 2017) implying $\partial s_2/\partial b = -1$. The same holds after the most recent EU ETS reform in April 2023 for immediate changes in allowance demand (Perino et al., 2022). For future shifts in allowance demand, in expected terms $\partial s_2/\partial b \in (-1, 0)$ holds. Appendix D.3 extends the results to cases where $\partial s_2/\partial b < -1$.

allowances $b = s_1 - E_1^*$ mirrors the change in first-period equilibrium emissions as the first-period cap s_1 is fixed:

$$\frac{dE^*}{d\lambda_i} = \frac{ds_2}{d\lambda_i} = \frac{\partial s_2}{\partial b}\frac{\partial b}{\partial E_1^*}\frac{\partial E_1^*}{\partial \lambda_i} = -\frac{\partial s_2}{\partial b}\frac{\partial E_1^*}{\partial \lambda_i} \Longrightarrow \text{sign}\left(\frac{dE^*}{d\lambda_i}\right) = \text{sign}\left(\frac{\partial E_1^*}{\partial \lambda_i}\right). \quad (17)$$

Policies that mainly reduce allowance demand early on (i.e., $\beta \geq \xi$) reduce the cumulative cap as firms respond to the shift in the first-period allowance demand curve by emitting less and banking more. The increase in the bank induces additional reductions in allowance supply via the flexibility mechanism.

By contrast, policies that reduce allowance demand in the distant future tend to increase the cumulative cap. As firms anticipate the drop in demand, they have less incentive to bank allowances and therefore emit more in the first period. The reduction in the size of the bank results in a smaller reduction in the cap (relative to the reference point without the anticipated demand reduction induced by the overlapping policy).[17]

Proposition 4 *The waterbed effect for an anticipated overlapping policy under a flexibility mechanism based on allowance banking is:*

$$W = \frac{1 + \frac{\partial s_2}{\partial b}\beta}{1 + \frac{\partial s_2}{\partial b}\xi} \quad (18)$$

so an overlapping policy that (i) is front-loaded, i.e., effective only in period 1 (with $\beta = 1$), has a punctured waterbed $W \in [0, 1]$; (ii) is sufficiently back-loaded (with $\beta \leq \xi$) has $W \geq 1$; (iii) reduces allowance demand in period 1 but increases it sufficiently strongly in period 2 according to $\beta \geq (-\frac{\partial s_2}{\partial b})^{-1} \geq 1$ has $W \leq 0$.

(See Appendix D.2 for a proof.) Cases (i) and (ii) highlight the direct and the price-mediated indirect effect on cumulative emissions. In case (i), for policies affecting aggregate demand early on, the price effect is of second order so the direct effect dominates and cumulative emissions decrease.[18]

In case (ii), however, policies affecting aggregate demand only in the far future such as an anticipated coal phase-out have no immediate emissions-demand impact—and so the price-driven effect dominates. Here, anticipation of a future reduction in relative scarcity leads to a lower carbon price in both periods. First-period emissions increase which, in combination with the fixed cap in period 1, induces a drop in the bank. This "Rosendahl effect" (Rosendahl, 2019; Bruninx and Ovaere, 2022; Gerlagh et al., 2021) in turn increases the cumulative cap and thus raises aggregate emissions.

[17] While the EU ETS's Market Stability Reserve technically can only cancel allowances, canceling fewer than without the impact of an overlapping policy reflects a cap increase relative to the reference point.

[18] With a fixed emissions cap, $\frac{\partial s_2}{\partial b} = 0$, we obtain $W = 1$ irrespective of the policy's timing.

An example of case (iii) is an amendment of a previously enacted coal phase-out plan that shuts down old inefficient plants earlier but grants new, highly-efficient plants a longer grace period, leading to a negative waterbed effect. This implies that the equilibrium reduction in emissions is larger than the net shift in long-run allowance demand induced by the amendment. This is most apparent if the net reduction in demand is small compared to the shift in demand from period 1 to 2. The latter here induces additional banking and hence additional cancellations that exceed the net change in allowance demand.

4.3 Representing banking-based flexibility mechanisms as implicit supply functions

Proposition 4 revealed that banking-based flexibility mechanisms can induce surprising responses to demand shocks: changes in equilibrium emissions that are larger than the initial demand shock or even point in the opposite direction.

Here we provide intuition by representing the response of a banking-based flexibility mechanism as an "equilibrium expansion path" that mimics a case-specific, implicit allowance supply curve ($s(\tau_1|\beta)$) for overlapping policies of different stringency ($dE/d\lambda_i$) but identical timing of impacts (β). This allows to use standard comparative statics to identify the response to a shift in the allowance demand curve induced by an overlapping climate policy (Gerlagh et al., 2021).

The representation as an implicit allowance supply function also reveals that—despite substantial differences between price- and quantity-based flexibility mechanisms—there is an economically intuitive link between Propositions 3 and 4 (see Appendix D.2 for a proof and a graphical illustration):

Corollary 1 *Propositions 3 and 4 are equivalent when considering the equilibrium expansion path as an implicit allowance supply function that is specific to the overlapping policy under consideration:*

$$\left.\frac{ds}{d\tau_1}(\beta)\right|_{equilibrium} = \frac{\frac{ds_2}{d\lambda_i}}{\frac{\partial\tau_1}{\partial\lambda_i}} = \frac{\partial E}{\partial\tau_1}\left(\frac{\beta-\xi}{\beta+\left(\frac{\partial s_2}{\partial b}\right)^{-1}}\right). \tag{19}$$

and defining $\omega^S = \frac{ds}{d\tau_1}(\beta)\frac{\tau_1}{s(\tau_1|\beta)}$ as the elasticity of the implicit allowance supply curve to be substituted into (8) in Proposition 3.

In contrast to the allowance supply function of a price-based flexibility mechanism, the equilibrium expansion path (19) is specific to the timing of the overlapping policy under consideration (β). The slope of the implicit allowance supply function is proportional to the slope of the long-run allowance demand function ($\frac{\partial E}{\partial\tau_1} < 0$) when holding the relation between the slope of first-period and long-run demand (ξ) fixed. It increases in the responsiveness of the flexibility mechanism ($|\partial s_2/\partial b|$) and decreases in the share of the price-responsiveness of long-run allowance demand originating form period 1 (ξ).

The key point is that implicit supply curves are *downward-sloping*, $\omega^S < 0$, whenever the policy's waterbed effect is above 100% or negative, i.e., if $\beta \notin [\xi, (-\frac{\partial s_2}{\partial b})^{-1}]$, providing an intuition for the surprising responses identified by Proposition 4.[19]

The connection between the waterbed effect under price- and banking-based flexibility mechanisms identified by Corollary 1 is novel. It shows how a banking-based flexibility mechanism can be represented by a (case-specific) allowance supply function; by contrast, previous academic (Abrell et al., 2019b) and regulatory (European Commission, 2021, p. 6) contributions had to resort to ad-hoc assumptions about its shape, which misses its dependence on the timing of the demand shock.

Corollary 1 can also be applied, as a unifying result, to shifts in allowance demand due to technological change (Bruninx et al., 2020), pandemics (Bruninx and Ovaere, 2022; Gerlagh et al., 2020), stimulus packages (Bruninx and Ovaere, 2022), and business cycles (Kollenberg and Taschini, 2016).

5 Empirical illustrations

5.1 Moving from theory to empirics

We now illustrate how real-world policies that overlap with carbon-pricing systems fit into our conceptual framework from Section 2. Our main outcome of interest is a policy's "effective emissions reduction" rate $R_i \equiv [1 - \widetilde{L}_i(\kappa)][1 - W]$ which we compute using a combination of sources. The objective here is to leverage our theory and prior literature to obtain "ballpark" estimates of the likely climate-effectiveness of overlapping policies in Europe and North America; we do not attempt any original empirical work.

In line with prior literature, our empirical approach restricts attention to intra-industry internal carbon leakage in the policy's targeted sector. For single-sector cap-and-trade ($\kappa = 0$) like RGGI, this assumption is always met; for a multi-sector ETS ($\kappa = 1$) like the EU ETS, it is implicit in the literature that we build on for our illustrations (e.g., Klobasa and Sensfuss, 2016; Frontier Economics, 2018; Abrell et al., 2019a; Schnaars, 2022), where effectively $\ell_i \approx 0$ is assumed so that aggregate leakage $\widetilde{L}_i(1) \approx L_i$. We are not aware of any prior empirical work that estimates both intra- and inter-industry internal leakages.

The empirical literature also relies on different concepts of internal carbon leakage.

[19] Karp and Traeger (2021) derive conditions for socially-optimal allowance supply curves to be downward-sloping; their conditions differ substantially from those established above. In our model, the downward-sloping implicit supply curve does not induce multiplicity of equilibria given that $\partial s_2/\partial b \geq -1$.

Some papers hold fixed the system-wide carbon price—and hence correspond directly to our L_i from Lemma 1. Other papers estimate "total" internal leakage L_i^{T}, which does not control for the carbon price; yet others only estimate "meta" leakage $L_i^{\mathrm{M}} = 1 - R_i$ (see Appendix A.2). We indicate below which leakage concept is estimated in each study.[20]

Figure 1 is a visual preview of this section (details are below). It plots the contour lines of R in (L, W)-space along with our policy illustrations. This is a novel way to graphically summarise the climate-effectiveness of a rich array of overlapping policies. Policies in the green regions are highly effective, with those in the bottom-left being complementary ($R > 1$). Policies in the light orange regions have limited effect, while those in the dark orange regions backfire by increasing aggregate emissions ($R < 0$). (For ease of exposition, we usually drop the overlapping policy's i subscript and, as neither prior empirical work nor our theory provide time-granular estimates of internal leakage, we drop any t subscript unless they matter for the waterbed effect.)

5.2 Overlapping climate policies and the EU ETS

Waterbed effect

The waterbed effect in the EU ETS is driven by its Market Stability Reserve (MSR). While Section 4.2 analysed a simplified two-period version of this flexibility mechanism, Appendix E.1 presents the full details of its multi-period operation. In brief, if the MSR's allowance bank exceeds 833 million at the end of a given year, then the number of EU ETS allowances auctioned in the subsequent year is reduced by a (time-varying) percentage of the size of the bank, and surplus allowances are placed in the MSR. Conversely, when the bank drops below 833 million at time $t_{B=833}$, the MSR stops taking in allowances.

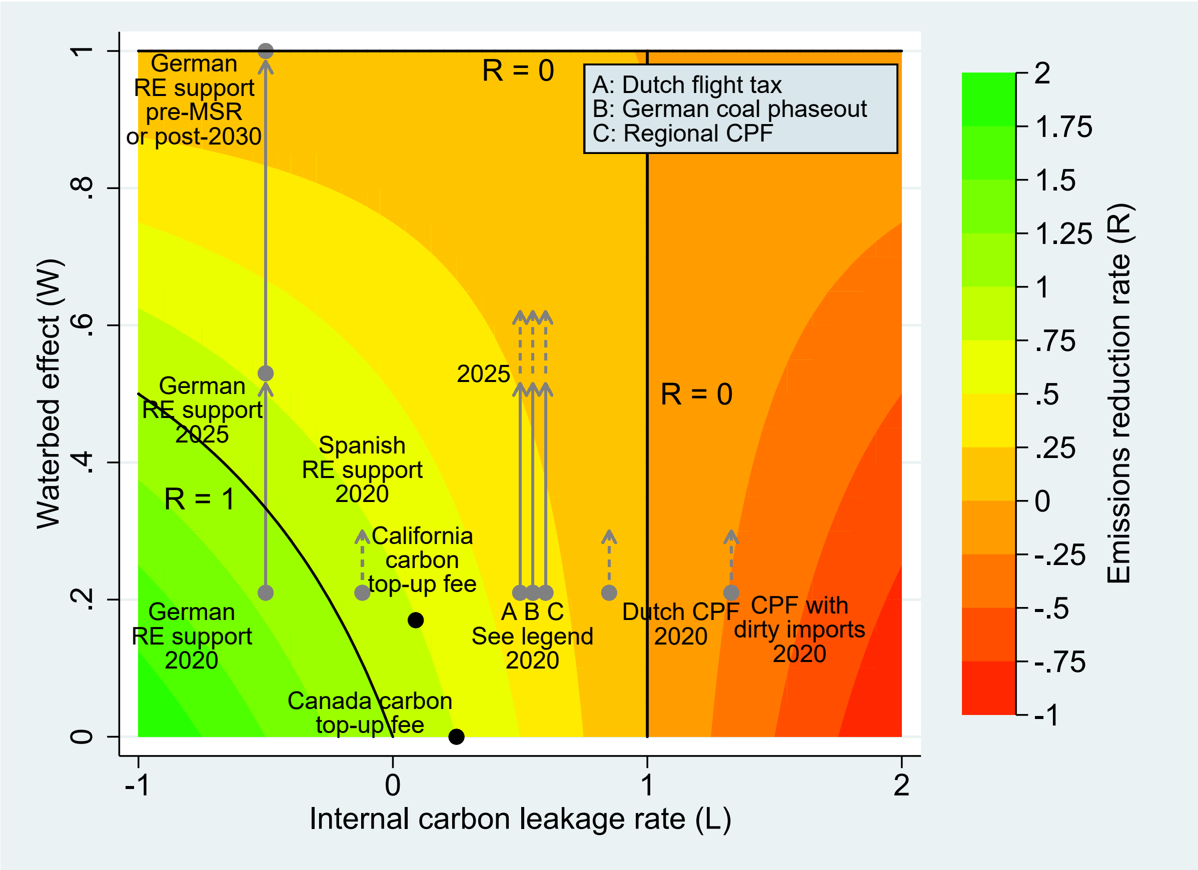
The waterbed effect is punctured because allowances held by the MSR in excess of an upper bound are permanently cancelled. In particular, these rules imply a *time-varying instantaneous waterbed effect* $\hat{W}_t$—the eventual impact of a marginal change in the allowance bank in year t on aggregate EU ETS emissions. Given that the year in which an overlapping policy is announced is held constant, $\hat{W}_t$ depends both on the year t in which allowance demand shifts and the year at which the MSR stops cancelling allowances $t_{B=833}$ as it determines the number of times the annual marginal intake rate of the MSR

[20] For a multi-sector ETS, we expect that the former two concepts give similar results, $L_i^{\mathrm{T}} \approx L_i$, as long as the policy is small relative to the ETS, and rely on this in our calibrations (unless stated otherwise). To see this, note from Appendix A.2 that total internal carbon leakage (endogenous τ) can be written as:

$$L_i^{\mathrm{T}} \equiv \frac{\frac{de_j}{d\lambda_i} + \frac{de_j}{d\tau}\frac{d\tau}{d\lambda_i}}{-\left(\frac{de_i}{d\lambda_i} + \frac{de_i}{d\tau}\frac{d\tau}{d\lambda_i}\right)} = \left[1 + \frac{\frac{de_i}{d\tau}}{-\left(\frac{de_i}{d\lambda_i} + \frac{de_i}{d\tau}\frac{d\tau}{d\lambda_i}\right)}\frac{d\tau}{d\lambda_i}\right] L_i + \frac{\frac{de_j}{d\tau}}{-\left(\frac{de_i}{d\lambda_i} + \frac{de_i}{d\tau}\frac{d\tau}{d\lambda_i}\right)}\frac{d\tau}{d\lambda_i},$$

where typically $de_k/d\tau < 0$ $(k = i, j)$. For cap-and-trade systems, $d\tau/d\lambda_i \neq 0$, so there may be a wedge between L_i^{T} and L_i but if i's policy is marginal, we expect $d\tau/d\lambda_i \simeq 0$ so that $L_i^{\mathrm{T}} \approx L_i$.

Figure 1: Overlapping policies facing internal carbon leakage and a waterbed effect



Notes: Figure shows the contour plot of the effective emissions reduction rate $R_{it} = (1 - L_{it})(1 - W)$ of various policies discussed in this section. Solid black lines indicate the contour lines where $R_{it} = 0$ (when $L = 1$ or $W = 1$) and $R_{it} = 1$ (bottom left). For EU ETS policies, we plot the instantaneous waterbed effect $\hat{W}_t$ for a fixed carbon-price path conditional on the year of the demand reduction. Dashed grey arrows indicate that, in the EU ETS, a policy's $\hat{R}_{it}$ moves towards zero as t approaches $t_{B=833}$ and $\hat{W}_t \to 1$. We assume $t_{B=833} = 2030$. Solid grey arrows show specific shifts in time for the German renewable energy support systems and for a proposed regional carbon price floor. German RE support: L = -0.50; W = 0.21 (2020), 0.53 (2025), 1 (2030) (Abrell et al., 2019a; Klobasa and Sensfuss, 2016). Spanish RE support: L = -0.12; W = 0.21 (2020) (Abrell et al., 2019a). CA top-up fee: L = 0.09; W = 0.17 (Caron et al., 2015; Borenstein et al., 2017). Canada top-up fee: L = 0.25; W = 0 (source: authors' assumptions). Dutch flight tax: L = 0.50; W = 0.21 (2020), 0.53 (2025) (Gordijn and Kolkman, 2011). German coal phase-out: L = 0.55; W = 0.21 (2020), 0.53 (2025) (Pahle et al., 2019). Dutch carbon price floor (CPF): L = 0.85; W = 0.21 (2020) (Frontier Economics, 2018). CPF with dirty imports: L = 1.33; W = 0.21 (2020) (source: authors' assumptions).

applies (for details see Appendix E.1). After time $t_{B=833}$, we are back to a fixed cap, i.e. $\hat{W}_t = 1$. Therefore, the effective emissions reduction rate for an overlapping policy itself changes over time, $\hat{R}_{it} = (1 - L_i)(1 - \hat{W}_t)$. The change in $\hat{W}_t$ is illustrated by the grey arrows in Figure 1. We use $t_{B=833} = 2030$ as a lower mid-range value,[21] and contrast policies acting in years $t = 2020, 2025, 2030$. In Figure 1, as time moves on, $\hat{W}_t$ increases from 0.21 (2020) to 0.53 (2025) to 1 (post-2030) and all European policies move north, as indicated by the dotted grey arrows.[22,23]

Supply-side overlapping policies

Electricity

The UK's Carbon Price Support for power generators from 2013 to 2020 ran alongside the EU ETS. Similarly, the Dutch government approved a national carbon price floor (CPF) for the electricity sector which increases from EUR 14.90/tCO_2 in 2022 to EUR 31.90/tCO_2 in 2030—as of mid-2023, however, the policy is not binding as the EU ETS carbon price exceeds the Dutch CPF.[24]

Proposition 1 shows that such supply-side policies, if binding, suffer from intra-EU leakage as domestic electricity generation gets replaced with imports; we expect high leakage for small countries (high σ_j) that are strongly interconnected to neighbours with flexible yet dirty supply (high $\varepsilon_j^S, \theta_j/\theta_i$).

Consistent with this theoretical prediction, quantitative estimates for the Dutch CPF find $L \simeq 0.85$ (Frontier Economics, 2018).[25] Such CPFs in small interconnected countries

[21] This date is subject to substantial uncertainty, with estimates ranging from 2022 (Perino and Willner, 2017) to the second half of the 2030s (Quemin and Trotignon, 2021), and $t_{B=833} = 2030$ as a mid-range value (Vollebergh, 2018).

[22] The specific values for $\hat{W}_t$ can be calculated using Equation (A.21) in Appendix E.1, evaluated at $t_{B=833} = 2030$ and $t = 2020, 2025, 2030$.

[23] The figure abstracts from the Rosendahl effect—the anticipated future reduction in relative scarcity of allowances—and other effects driven by adjustments in the carbon-price path. In our sensitivity analysis in Appendix E.2, we also consider $t_{B=833} = 2048$, as estimated in Gerlagh et al. (2021), and show that $\hat{W}_t$ increases for policies that are announced before they take effect, especially for the years close to $t_{B=833}$, and that, as shown in Proposition 4, $\hat{W}_t$ can increase above 100% for such anticipated policies.

[24] Source: `https://www.eerstekamer.nl/wetsvoorstel/35216_wet_minimum_co2_prijs`. The Netherlands has also adopted a CPF for its industrial sectors, starting at EUR 30.48 in 2021 and increasing to EUR 128.71 by 2030. See `https://www.emissieautoriteit.nl/onderwerpen/tarieven-co2-heffing`.

[25] Table 1 in Frontier Economics (2018) estimates that the Dutch price floor will reduce domestic emissions by 26 million tCO_2 in 2030, but the net EU-wide emissions reduction is only 4 million tCO_2, implying $L = 0.85$. These results are from a power dispatch model and correspond to our definition of L; the EU ETS allowance market is not modelled so the carbon price is implicitly held fixed. Vollebergh (2018) estimates L to be 85% for the Dutch price floor and 61% for a larger regional CPF including the Benelux, France and Germany, but this analysis uses the WorldScan CGE model that includes allowance supply, banking, and the Market Stability Reserve with the carbon price responding endogenously to the CPF—in other words, it measures $L^M = 1 - R$. There is no decomposition into L and W.

are unlikely to reduce EU-wide emissions by much, with $\hat{R}_{2020} = 0.12$ ($\hat{W}_{2020} = 0.21, L = 0.85$) even under the punctured waterbed (see Figure 1).[26] As more countries join the CPF, $\hat{R}_{2020}$ rises to 0.31 ($\hat{W}_{2020} = 0.21, L = 0.61$). Furthermore, the solid grey arrow shows that the regional CPF's $\hat{R}$ decreases to 0.18 by 2025 when $\hat{W}_{2025} = 0.53$, so early action is preferable in this sense.

Cost-raising policies can backfire if imports are substantially dirtier than domestic production (see Proposition 1). We plot a hypothetical "CPF with dirty imports" (assuming $\theta_j/\theta_i = 2$, $\varepsilon_j^S = 5 \Leftrightarrow \eta_j^S = 0.2$, $\sigma_j = 0.2$, $\varepsilon^D = 0.5$ and no abatement $\phi_i''(\cdot) \to \infty$) for which $L = 1.33$ such that EU-wide emissions *increase*, $R < 0$. Since this policy lies to the right of the $R = 0$ contour line, the negative effect gets *weaker* over time as the waterbed effect gets stronger.

The Powering Past Coal Alliance groups national and sub-national governments, including twelve EU countries, committed to phasing out coal. Examples include the British and Dutch policies to close their remaining coal-fired power plants by 2025 and 2030, respectively. Germany has also passed regulation to phase out coal by 2038.[27] This would lead to reduced demand for allowances both before and after this date, relative to the counterfactual. The policy has been estimated to have an internal carbon leakage rate of 55% in 2020 (Pahle et al., 2019), so $\hat{R}_{2020} = 0.36$ ($\hat{W}_{2020} = 0.21, L = 0.55$) and decreasing to zero by 2030. The leakage estimate $L = 0.55$ corresponds to L^T as the allowance price is allowed to adjust in Pahle et al. (2019); we assume it is a good approximation of our definition of L holding carbon prices fixed.

Post-2030, $\hat{W}_t = 1$, so all overlapping policies within the EU ETS end up at $R = 0$. We also note that the above policies in the power sector likely have negligible external carbon leakage to regions outside the EU ETS, justifying our focus on internal leakage.

Aviation

Several European countries, such as Austria, Germany, Norway and Sweden, have aviation taxes; others, such as Denmark, Ireland and the Netherlands, abolished them after initial implementation. Such policies are prone to leakage: when the Netherlands adopted an aviation tax in July 2008 at a rate of EUR 11.25 for short-haul flights and EUR 45 for long-haul flights, about 50% of the decline in passengers at Dutch airports was offset by increased passenger volumes at nearby airports in Belgium and Germany (Gordijn and Kolkman, 2011).[28] This intra-EU leakage rate of 50%—which we interpret "in spirit" as

[26] We expect internal carbon leakage to have been lower for Great Britain's carbon fee under the EU ETS as import supply is more inelastic due to interconnection constraints to continental Europe.

[27] Sources: `https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2020/20200703-final-decision-to-launch-the-coal-phase-out.html` (press release), `http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl120s1818.pdf` (coal phase-out law), `https://www.wired.co.uk/article/germany-coal-electricity-spremberg` (press coverage).

[28] Gordijn and Kolkman (2011) estimate that the tax accounted for nearly two million fewer passengers

holding carbon prices fixed—is in line with L in Proposition 1. As a result, the Dutch government abolished the tax in July 2009—but then reintroduced a modest ticket tax of EUR 7 on all flights starting in 2021 (Forbes, 2020a). Assuming the same internal leakage rate as in 2008-9, we estimate $\hat{R}_{2020} = 0.40$ ($\hat{W}_{2020} = 0.21, L = 0.50$).

There is broader evidence that aviation taxes are most likely in countries where leakage is mitigated—e.g., in high-population countries such as France, Germany, Italy and the UK (low σ_j) as well as countries such as Norway and Sweden whose population is far away from low-tax airports abroad (high ε_j^S) (PricewaterhouseCoopers, 2017).

Demand-side overlapping policies

Under the EU's 2009 Renewables Directive, each member state developed a national action plan aimed at increasing the share of renewables in its energy mix. Germany and Spain have adopted some of the world's most ambitious incentives for wind and solar energy, which include feed-in tariffs and market premium programs.

Abrell et al. (2019a) estimate negative carbon (and output) leakage (holding carbon prices fixed by controlling for EUA prices in their regressions) for renewables support in Germany and Spain as additional zero-carbon energy depresses wholesale electricity prices—via the merit-order effect—and offsets imported gas- and coal-fired electricity in Germany ($L = -0.50$) and Spain ($L = -0.12$). Their leakage estimates correspond to our L in Lemma 1.[29] Similarly, a German government report finds $L = -0.54$ (Klobasa and Sensfuss, 2016), also holding carbon prices fixed.[30] Figure 1 shows that, at least in the year 2020, the German renewable support scheme reduces system-wide emissions by *more* than the domestic emissions reduction in Germany ($\hat{W}_{2020} = 0.21, L = -0.50, \hat{R}_{2020} = 1.19$)—and is truly a complementary policy in this sense. As time passes, W increases and eventually the puncture is sealed, reducing R to zero from 2030 onwards.

Proposition 2 shows equivalence between renewables support and other demand-side policies such as energy-efficiency programs and a carbon-consumption tax. Therefore, we

from Amsterdam's Schiphol Airport during the period over which the tax was in effect, while an extra one million Dutch passengers flew from foreign airports. There is no mention of carbon prices.

[29]In their Table 3, Abrell et al. (2019a) report d(import quantity)/d(policy) and d(domestic quantity)/d(policy), from which we calculate output leakage as -78%, -77%, -7% and -21% for German wind, German solar, Spanish wind and Spanish solar, respectively. Similarly, we compute carbon leakage from their Table 5: -49%, -50%, -6% and -19%, respectively. Averaged over wind and solar, we use $L = -0.50$ for Germany and $L = -0.12$ for Spain in Figure 1. Schnaars (2022) provides another negative carbon leakage rate estimate (controlling for EUA prices, hence corresponding directly to our L) of -51% for renewable energy incentives in Germany, further bolstering the case for negative leakage. The differences between output and emissions leakage in Germany and Spain suggest that the marginal unit of output reduction in Germany is approximately 50% more carbon intensive than the marginal reduction for its trading partners; for Spain the emissions intensities of these marginal units are about equal. Abrell et al. (2019a) show that the German power mix is indeed dirtier than Spain's.

[30]They estimate leakage while holding the carbon price fixed between the baseline and the renewable support scenarios, hence the leakage estimates correspond to our L.

expect negative internal leakage also for these policies but are not aware of any empirical estimates, so do not include them in Figure 1.

5.3 Overlapping climate policies in North America

California-Québec carbon market

California and Québec have a joint carbon market with a price-based flexibility mechanism, namely an auction price floor ($22.21 in 2023)[31] and a price ceiling ($81.50 in 2023) (Politico, 2018). Before the hard price cap is reached, two soft price caps create horizontal segments in the allowance supply function: up to some limit, allowances will be offered at $51.92 and at $66.71 before the market could reach the hard price cap. Borenstein et al. (2017) estimate that, by 2030, the probability that the equilibrium will occur on any of the horizontal sections of the allowance supply curve equals $\pi = 0.83$—therefore, by Proposition 3, the expected waterbed effect $W = 1 - \pi = 0.17$ is punctured.[32]

The California-Québec carbon market is known to cause *external* leakage to other states that are interconnected in the electricity market (Fowlie, 2009; Caron et al., 2015). We now consider a counterfactual Western Climate Initiative (WCI) in which states surrounding California join the carbon market.[33] If California then imposed a unilateral carbon top-up fee, this would lead to "intra-WCI" carbon leakage to neighbouring states. Thus *external* leakage under the current system gets transformed into *internal* leakage under a counterfactual WCI, allowing us to rely on existing estimates from the literature. Fowlie (2009) finds that a carbon price in California that exempts out-of-state producers achieves only 25-35% of the total emissions reductions achieved under complete regulation (Arizona, Nevada, New Mexico, Oregon, Utah and Washington) so that $L = 0.65\text{-}0.75$. Caron et al. (2015) provide a relevant leakage estimate of $L = 0.09$ (holding carbon prices fixed[34]) for California's cap-and-trade program assuming that—as the current market rules specify—there is a border-tax adjustment and "resource shuffling" is banned.[35] Fig-

[31]The auction price floor was binding in various auctions in the year 2016. In addition, in many other quarterly auctions, the markets cleared only slightly above the price ceiling. See `https://ww3.arb.ca.gov/cc/capandtrade/capandtrade.htm` for details.

[32]Borenstein et al. (2017)'s calculation is based on values of the price floor, steps, and cap that differ somewhat from the eventually-implemented level, but we expect this to have a minor impact on their estimate of π.

[33]The WCI (`http://www.wci-inc.org/`) started in 2007 as an initiative by the governors of Arizona, California, New Mexico, Oregon and Washington with a goal to develop a regional multi-sector cap-and-trade market. Most states left during the economic downturn in the early 2010s but the idea of regional carbon trading has resurfaced in discussions among states.

[34]They estimate leakage from California to out-of-state producers that face no carbon price. So implicitly the carbon price is held fixed, hence the leakage estimates correspond to our L.

[35]Resource shuffling is defined as "any plan, scheme, or artifice to receive credit based on emissions reductions that have not occurred, involving the delivery of electricity to the California grid" (Caron et al., 2015). For example, out-of-state generators could reconfigure transmission so that low-carbon

ure 1 plots the hypothetical California carbon top-up fee using $L = 0.09$, as this estimate corresponds most closely to California's current market rules. Given these values, the overlapping policy would be reasonably climate effective: for every ton of carbon saved in California, system-wide emissions decrease by $R = 0.76$ tons ($W = 0.17, L = 0.09$). (These values are illustrative only given the profound changes in the Western power grid since Caron et al. (2015) was written.)

Regional Greenhouse Gas Initiative

RGGI caps CO_2 emissions from electricity in eleven Northeastern states. It has a flexibility mechanism based on allowances prices, with a 'hard' price floor and a 'soft' price cap that offers up to 10 million allowances at a fixed price ($13 in 2021; increasing at 7% per year). Once these allowances are exhausted then prices would continue to rise. From 2021, in addition to the price floor, the program features an Emissions Containment Reserve that removes up to 10% of the annual allowance budget from circulation if the price falls below $6, increasing by 7% thereafter. The price floor ($2.38 in 2021) was binding during 2010-2012;[36] the states decided to retire unsold allowances. The soft price cap was triggered in 2014 and 2015. Effectively, this produces an upward-sloping step-function allowance supply function which fits our analysis of Section 4.1. Non-marginal interventions—for which allowance demand moves across one or several steps in the supply schedule—have an expected waterbed effect between zero and 100%.

Several RGGI states have floated the idea of unilateral policies. Most notably, New York has proposed an additional carbon fee equal to the difference between the social cost of carbon and the RGGI allowance price (Forbes, 2020b). Shawhan et al. (2019) model the power market and RGGI allowance market, and estimate $L^M = 1 - R$—the combined effect of internal leakage and RGGI's waterbed effect—to other RGGI states that results from New York's policy at $R = 0.42$.[37] We do not plot New York's carbon fee in Figure 1 as Shawhan et al. (2019) do not decompose R into L and W and we are not aware of a direct estimate of L or an estimate of the fraction of the time that the system is expected to trade at the price floor or ceiling, so W is also missing. As RGGI is a single-sector ETS, $L^M = L^T = 1 - R$ (see Appendix A.2) so $R = 0.42$ would place the policy in the

electricity is diverted to California and high-carbon electricity is sold to other states.

[36] See `https://fas.org/sgp/crs/misc/R41836.pdf`

[37] New York's carbon-pricing policy differs somewhat from our theory. First, a border tax applies to imported electricity from other RGGI states. Second, there is scope for nontrivial *external* leakage to non-RGGI states. Shawhan et al. (2019) estimate this external carbon leakage to be substantially negative—an increase in renewable power in New York reduces dirty imports from non-RGGI to RGGI states. This underscores that external and internal leakage are distinct phenomena that can even have different signs. Fell and Maniloff (2018) find positive *external* leakage of 51% from the introduction of RGGI as a whole. As this is a very different policy than New York's proposed carbon price we have no reason to expect that external leakage rates would be similar.

light-orange region of Figure 1, so "medium climate-effective."

Canada's national minimum carbon tax

Canada adopted a national minimum carbon tax of \$20 per ton starting in 2019, increasing to \$50 by 2022. Some provinces, such as Alberta and British Columbia, already had in place carbon taxes with a price above the national minimum level. By Proposition 3, such unilateral carbon taxes face no waterbed effect but, by Proposition 1, they may suffer from internal leakage to other provinces. Though we are not aware of direct leakage estimates, Murray and Rivers (2015) and Yamazaki (2017) find that British Columbia's carbon tax has had negligible or modest effects on the aggregate economy, suggesting leakage is modest, and so Figure 1 plots this policy assuming $L = 0.25$ and $W = 0$, leaving a higher carbon tax in British Columbia reasonably climate-effective ($R = 0.75$).

6 Conclusion

This paper has presented a new modelling framework to understand overlapping climate policies within a wider carbon-pricing system. Design matters in that different popular policies have very different properties in terms of their internal carbon leakage. Space matters as leakages can differ substantially across industries and jurisdictions. Time matters as it can affect the magnitude of the waterbed effect. The issues we have highlighted extend beyond policy-making in Europe and North America and are critical for the design of new climate policies like China's national emissions trading system.[38]

We hope that our analysis will be useful to policymakers. It yields simple formulae for internal leakage and the waterbed effect that lend themselves to "back-of-the-envelope" calculations that can be extremely valuable in a real-time policy context. Demand-side overlapping policies have negative internal carbon leakage and with a hybrid carbon-market design can indeed be truly complementary in that they induce further emissions reductions across the system—and our analysis suggests that some renewables support in the EU ETS has met these conditions. Supply-side overlapping policies, like the UK's Carbon Price Support, can be very successful at reducing domestic emissions (Abrell

[38] Our framework has deeper connections to the literature on fixed stocks of fossil resources. Their eventual exhaustion (Sinn, 2008; Eichner and Pethig, 2011; Van der Ploeg, 2016) impedes the climate effectiveness of policies that reduce fossil demand—and corresponds to a 100% waterbed effect in cap-and-trade. There are close parallels with the design of allowance supply functions in carbon markets in our framework. For example, Harstad (2012) shows how a coalition, by buying but then not exploiting specific non-coalition fossil resources, can create a vertical section in the aggregate resource supply function—which then makes fully effective its domestic resource-conservation policy. This becomes equivalent to individual countries inside a multi-jurisdiction carbon market with a fixed cap (like the pre-2018 EU ETS) pursuing overlapping policies that involve cancelling allowances.

et al., 2022), but are never complementary—aggregate emissions fall by less than local carbon emissions, and those with high internal carbon leakage may backfire.

Potential caveats to our results include the fiscal implications of overlapping policies. These may drive a wedge between retail and wholesale product prices and thereby affect the economics of internal carbon leakage, notably for renewables support programs that are funded via retail charges. Another caveat is "external" carbon leakage to jurisdictions (or sectors) outside the carbon-pricing system; this is often negligible for overlapping policies on electricity and aviation but in some cases both internal and external leakages may play an important role. While a full policy analysis will always have to be case-by-case, we believe that the insights from our conceptual framework will be widely applicable.

Future research could extend our work in a number of directions. Theory work could examine how market power and product differentiation affect the extent of internal carbon leakage.[39] Policy work should pursue a welfare analysis that incorporates abatement cost-effectiveness and distributional impacts; trade-offs may arise: while our results show that supply-side overlapping policies are (much) more prone to internal leakage than demand-side policies, they may have lower abatement costs (Gugler et al., 2021).[40] Empirical work on a multi-sector ETS should estimate both intra-industry and inter-industry internal carbon leakage.

References

Abrell, Jan, Mirjam Kosch, and Sebastian Rausch, "Carbon Abatement with Renewables: Evaluating Wind and Solar Subsidies in Germany and Spain," *Journal of Public Economics*, 2019, *169*, 172–202.

_ , _ , and _ , "How effective is carbon pricing? A machine learning approach to policy evaluation," *Journal of Environmental Economics and Management*, 2022, *112*, 102589.

_ , Sebastian Rausch, and Hidemichi Yonezawa, "Higher Price, Lower Costs? Minimum Prices in the EU Emissions Trading Scheme," *Scandinavian Journal of Economics*, 2019, *121* (2), 446–481.

Acemoglu, Daron, Ali Kakhbod, and Asuman Ozdaglar, "Competition in electricity markets with renewable energy sources," *The Energy Journal*, 2017, *38*, 137–155.

Aldy, Joseph E. and William A. Pizer, "The Competitiveness Impacts of Climate Change Mitigation Policies," *Journal of the Association of Environmental and Resource Economists*, 2015, *2* (4), 565–595.

[39] We expect that as domestic and foreign production become strongly differentiated, output leakage—and hence also internal (intra-industry) carbon leakage—would tend to zero (but also not change sign).

[40] In special cases, such as an aggregate welfare function with a constant social marginal damage of emissions σ, the environmental analysis can be separated analytically from the wider welfare analysis. Our framework then quantifies the (monetised) environmental benefit of i's policy as $-\sigma[1-\widetilde{L}_i(\kappa)](1-W)\Delta e_i$.

Antweiler, Werner and Felix Muesgens, "On the Long-Term Merit Order Effect of Renewable Energies," *Energy Economics*, 2021, *99*, 105275.

Baumol, William and Wallace Oates, *The Theory of Environmental Policy*, Cambridge University Press, 1988. 2nd edition.

Baylis, Kathy, Don Fullerton, and Daniel H. Karney, "Leakage, Welfare and Cost-Effectiveness of Carbon Policy," *American Economic Review: Papers & Proceedings*, 2013, *103* (3), 332–337.

Böhringer, Christoph, "Two Decades of European Climate Policy: A Critical Appraisal," *Review of Environmental Economics and Policy*, 2014, *8* (1), 1–17.

_ and Knut Einar Rosendahl, "Europe Beyond Coal – An Economic and Climate Impact Assessment," *Journal of Environmental Economics and Management*, 2022, *113*, 102658.

Borenstein, Severin, "The private and public economics of renewable electricity generation," *Journal of Economic Perspectives*, 2012, *26*, 67–92.

_ , James Bushnell, and Frank A. Wolak, "California's Cap-and-Trade Market Through 2030: A Preliminary Supply/Demand Analysis," *Energy Institute at Haas Working Paper 281*, 2017.

_ , _ , _ , and Matthew Zaragoza-Watkins, "Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design," *American Economic Review*, 2019, *109* (11), 3953–3977.

Borghesi, Simone, Michael Pahle, Grischa Perino, Simon Quemin, and Maximilian Willner, "The Market Stability Reserve in the EU Emissions Trading System: A Critical Review," *Annual Review of Resource Economics*, 2023, *15.*

Bruninx, Kenneth and Marten Ovaere, "COVID-19, Green Deal & the Recovery Plan Permanently Change Emissions and Prices in EU ETS Phase IV," *Nature Communications*, 2022, *13*, 1–10.

_ , _ , and Erik Delarue, "The Long-Term Impact of the Market Stability Reserve on the EU Emission Trading System," *Energy Economics*, 2020, *89*, 104746.

Burtraw, Dallas, Charles Holt, Karen Palmer, and William Shobe, "Price-responsive allowance supply in emissions markets," *Journal of the Association of Environmental and Resource Economists*, 2022, *9* (5), 851–884.

Caron, Justin, Sebastian Rausch, and Niven Winchester, "Leakage from Sub-National Climate Policy: The Case of California's Cap-and-Trade Program," *The Energy Journal*, 2015, *36* (2), 167–190.

Cludius, Johanna, Hauke Hermann, Felix Chr. Matthes, and Verena Graichen, "The Merit Order Effect of Wind and Photovoltaic Electricity Generation in Germany 2008–2016: Estimation and Distributional Implications," *Energy Economics*, 2014, *44*, 302–313.

Eichner, Thomas and Rüdiger Pethig, "Carbon Leakage, the Green Paradox, and Perfect Future Markets," *International Economic Review*, 2011, *52* (3), 767–805.

_ **and** _ , "EU-type Carbon Regulation and the Waterbed Effect of Green Energy Promotion," *Energy Economics*, 2019, *80*, 656–679.

European Commission, "Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL Amending Directive 2003/87/EC Establishing a System for Greenhouse Gas Emission Allowance Trading within the Union, Decision (EU) 2015/1814 Concerning the Establishment and Operation of a Market Stability Reserve for the Union Greenhouse Gas Emission Trading Scheme and Regulation (EU) 2015/757, Annex 7: Legal Review of the Market Stability Reserve," Technical Report COM(2021) 551 final, European Commission 2021.

_ , "Publication of the Total Number of Allowances in Circulation in 2021 for the Purposes of the Market Stability Reserve under the EU Emissions Trading System Established by Directive 2003/87/EC and of the Number of Unallocated Allowances During the Period 2013-2020," *Official Journal of the European Union*, 2022, (2022/C 195/02)).

European Parliament and Council, "DECISION (EU) 2015/1814 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 October 2015 Concerning the Establishment and Operation of a Market Stability Reserve for the Union Greenhouse Gas Emission Trading Scheme and Amending Directive 2003/87/EC," *Official Journal of the European Union*, 2015, *9.10.2015*, L264/1 – L264/5.

_ , "Directive (EU) 2018/410 of the European Parliament and of the Council of 14 March 2018 Amending Directive 2003/87/EC to Enhance Cost-Effective Emission Reductions and Low-Carbon Investments, and Decision (EU) 2015/1814," *Official Journal of the European Union*, 2018, *19.3.2018*, L76/4 – L76/27.

Fankhauser, Samuel, Cameron Hepburn, and Jisung Park, "Combining Multiple Climate Policy Instruments: How Not to Do It," *Climate Change Economics*, 2010, *1* (3), 209–225.

Fell, Harrison and Peter Maniloff, "Leakage in Regional Environmental Policy: The Case of the Regional Greenhouse Gas Initiative," *Journal of Environmental Economics and Management*, 2018, *87*, 1–23.

Forbes, "The Netherlands Keeps Going Upstream When It Comes to Aviation Tax," April 18th 2020. Available at `https://www.forbes.com/sites/emanuelabarbiroglio/2020/04/18/the-netherlands-keeps-going-upstream-when-it-comes-to-aviation-tax/#552d54136e86`.

_ , "New York Power Grid Proposes Adding Carbon Costs to Market Prices for Electricity," February 20th 2020. Available at `https://www.forbes.com/sites/peterdetwiler/2020/02/20/in-a-path-breaking-approach-new-yorks-grid-operator-proposes-inclusion-of-carbon-costs-in-market-prices/`.

Fowlie, Meredith L., "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage," *American Economic Journal: Economic Policy*, 2009, *1* (2), 72–112.

_ **and Mar Reguant**, "Challenges in the Measurement of Leakage Risk," *American Economic Review: Papers & Proceedings*, 2018, *108*, 124–129.

_ , _ , **and Stephen P. Ryan**, "Market-Based Emissions Regulation and Industry Dynamics," *Journal of Political Economy*, 2016, *124* (1), 249–302.

Frontier Economics, "Research on the Effects of a Carbon Price Floor," 2018. Available at https://www.frontier-economics.com/media/2240/reasearch-effects-carbon-price-floor.pdf.

Gerarden, Todd D., W. Spencer Reeder, and James H. Stock, "Federal Coal Program Reform, the Clean Power Plan, and the Interaction of Upstream and Downstream Climate Policies," *American Economic Journal: Economic Policy*, 2020, *12* (1), 167–199.

Gerlagh, Reyer, Roweno J.R.K. Heijmans, and Knut Einar Rosendahl, "COVID-19 Tests the Market Stability Reserve," *Environmental and Resource Economics*, 2020, *76* (4), 855–865.

_ **, Roweno JRK Heijmans, and Knut Einar Rosendahl**, "An Endogenous Emissions Cap Produces a Green Paradox," *Economic Policy*, 2021, *36*, 485–522.

Gordijn, Hugo and Joost Kolkman, "Effects of the Air Passenger Tax: Behavioral Responses of Passengers, Airlines and Airports," 2011. KiM Netherlands Institute for Transport Policy Analysis. Available at http://publicaties.minienm.nl/documenten/effects-of-the-air-passenger-tax-behavourial-responses-of-passen.

Goulder, Lawrence H. and Robert N. Stavins, "Challenges from State-Federal Interactions in US Climate Change Policy," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 253–257.

_ **, Mark R. Jacobsen, and Arthur A. van Benthem**, "Unintended Consequences from Nested State and Federal Regulations: The Case of the Pavley Greenhouse-Gas-per-Mile Limits," *Journal of Environmental Economics and Management*, 2012, *63* (2), 187–207.

Gugler, Klaus, Adhurim Haxhimusa, and Mario Liebensteiner, "Effectiveness of climate policies: Carbon pricing vs. subsidizing renewables," *Journal of Environmental Economics and Management*, 2021, *106*, 102405.

Harstad, Bard, "Buy Coal! A Case for Supply-Side Environmental Policy," *Journal of Political Economy*, 2012, *120* (1), 77–115.

House of Commons Library, "Carbon Price Floor (CPF) and the price support mechanism," 2018. Briefing Paper Number 05927, 8 January 2018. Available at https://commonslibrary.parliament.uk/research-briefings/sn05927.

IPCC, "Fourth Assessment Report: Climate Change 2007—Working Group III: Mitigation of Climate Change," 2007. Chapter 11.7: International Spillover Effects, Intergovernmental Panel on Climate Change: Geneva, Switzerland.

Jarke, Johannes and Grischa Perino, "Do Renewable Energy Policies Reduce Carbon Emissions? On Caps and Inter-Industry Leakage," *Journal of Environmental Economics and Management*, 2017, *84*, 102–124.

Jenkin, Fleeming, "3. On the Principles which Regulate the Incidence of Taxes," *Proceedings of the Royal Society of Edinburgh*, 1872, *7*, 618–631.

Karp, Larry and Christian Traeger, "Smart Cap," *CEPR Discussion Paper No. DP15941*, 2021.

Klobasa, Marian and Frank Sensfuss, "CO2-Minderung im Stromsektor durch den Einsatz Erneuerbarer Energien in den Jahren 2012 und 2013," 2016. Umweltbundesamt. Available at `https://www.umweltbundesamt.de/publikationen/co2-minderung-im-stromsektor-durch-den-einsatz`.

Kollenberg, Sascha and Luca Taschini, "Emissions Trading Systems with Cap Adjustments," *Journal of Environmental Economics and Management*, 2016, *80*, 20–36.

Martin, Ralf, Mirabelle Muûls, Laure B. de Preux, and Ulrich J. Wagner, "Industry Compensation under Relocation Risk: A Firm-Level Analysis of the EU Emissions Trading Scheme," *American Economic Review*, 2014, *104* (8), 2482–2508.

Mathiesen, Lars and Ottar Maestad, "Climate Policy and the Steel Industry: Achieving Global Emission Reductions by an Incomplete Climate Agreement," *The Energy Journal*, 2004, *25* (4), 91–114.

Montgomery, W. David, "Markets in Licences and Effluent Control Programs," *Journal of Economic Theory*, 1972, *5* (1), 395–418.

Murray, Brian and Nicholas Rivers, "British Columbia's Revenue-Neutral Carbon Tax: A Review of the Latest "Grand Experiment" in Environmental Policy," *Energy Policy*, 2015, *86*, 674–683.

New York Times, "These Countries Have Prices on Carbon. Are They Working?," April 2nd 2019. Available at `https://www.nytimes.com/interactive/2019/04/02/climate/pricing-carbon-emissions.html`.

Newbery, David M., David M. Reiner, and Robert A. Ritz, "The Political Economy of a Carbon Price Floor for Power Generation," *The Energy Journal*, 2019, *40* (1).

Newell, Richard, William Pizer, and Jiangfeng Zhang, "Managing Permit Markets to Stabilize Prices," *Environmental and Resource Economics*, 2005, *31* (2), 133–157.

Novan, Kevin, "Overlapping Environmental Policies and the Impact on Pollution," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (S1), S1–S288.

Osorio, Sebastian, Robert C. Pietzcker, Michael Pahle, and Ottmar Edenhofer, "How to Deal with the Risks of Phasing out Coal in Germany," *Energy Economics*, 2020, *87*, 104730.

Pahle, Michael, Ottmar Edenhofer, Robert Pietzcker, Oliver Tietjen, Sebastian Osorio, and Christian Flachsland, "Die Unterschätzten Risiken des Kohleausstiegs," *Energiewirtschaftliche Tagesfragen*, 2019, *69* (6), 1–4.

Perino, Grischa, "New EU ETS Phase 4 Rules Temporarily Puncture Waterbed," *Nature Climate Change*, 2018, *8* (4), 262–264.

_ , "Reply: EU ETS and the Waterbed Effect," *Nature Climate Change*, 2019, *9* (10), 736.

_ **and Maximilian Willner**, "EU-ETS Phase IV: Allowance Prices, Design Choices and the Market Stability Reserve," *Climate Policy*, 2017, *17* (7), 936–946.

_ **,** _ **, Simon Quemin, and Michael Pahle**, "Policy Brief - The EU ETS Market Stability Reserve: Does It Stabilize or Destabilize the Market?," *Review of Environmental Economics and Policy*, 2022, *16* (2), 338–345.

Pizer, William A., "Combining Price and Quantity Controls to Mitigate Global Climate Change," *Journal of Public Economics*, 2002, *85* (3), 409–434.

Politico, "California Pro Preview: Carbon Prices," November 16th 2018. Available at `https://www.politico.com/states/california/newsletters/politico-california-pro-preview/2018/11/16/carbon-prices-136174`.

PricewaterhouseCoopers, "The Economic Impact of Air Taxes in Europe: Germany," 2017. Available at `https://www.bdl.aero/wp-content/uploads/2018/10/The-economic-impact-of-air-taxes-in-Europe-Germany-004.pdf`.

Quemin, Simon and Raphaël Trotignon, "Emissions Trading with Rolling Horizons," *Journal of Economic Dynamics and Control*, 2021, *125*, 104099.

Regional Greenhouse Gas Initiative, "2017 Model Rule—Part XX CO2 Budget Trading Program," Technical Report, Regional Greenhouse Gas Initiative 2017.

Roberts, Marc J. and Michael Spence, "Effluent Charges and Licenses under Uncertainty," *Journal of Public Economics*, 1976, *5* (3-4), 193–208.

Rosendahl, Knut Einar, "EU ETS and the Waterbed Effect," *Nature Climate Change*, 2019, *9* (10), 734–735.

Schnaars, Philip, "The Real Substitution Effect of Renewable Electricity: An Empirical Analysis for Germany," *The Electricity Journal*, 2022, *35* (1), 107074.

Sensfuß, Frank, Mario Ragwitz, and Massimo Genoese, "The Merit-Order Effect: A Detailed Analysis of the Price Effect of Renewable Electricity Generation on Spot Market Prices in Germany," *Energy Policy*, 2008, *36* (8), 3086–3094.

Shawhan, Daniel, Paul Picciano, and Karen Palmer, "Benefits and Costs of Power Plant Carbon Emissions Pricing in New York," 2019. Resources for the Future. Available at https://www.rff.org/publications/reports/benefits-and-costs-of-the-new-york-independent-system-operators-carbon-pricing-initiative/.

Sinn, Hans-Werner, "Public Policies Against Global Warming: A Supply Side Approach," *International Tax and Public Finance*, 2008, *15* (4), 360–394.

Szabo, Laszlo, Ignacio Hidalgo, Juan Carlos Ciscar, and Antonio Soria, "CO2 Emission Trading Within the European Union and Annex B Countries: The Cement Industry Case," *Energy Policy*, 2006, *34* (1), 72–87.

Van der Ploeg, Frederick, "Second-Best Carbon Taxation in the Global Economy: the Green Paradox and Carbon Leakage Revisited," *Journal of Environmental Economics and Management*, 2016, *78*, 85–105.

Vollebergh, Herman, "National Measures Complementary to EU ETS," 2018. PBL Netherlands Environmental Assessment Agency. Available at https://www.eprg.group.cam.ac.uk/wp-content/uploads/2018/12/H.-Vollebergh_final.pdf.

Weyl, E. Glen and Michal Fabinger, "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 2013, *121* (3), 528–583.

Yamazaki, Akio, "Jobs and Climate Policy: Evidence from British Columbia's Revenue-Neutral Carbon Tax," *Journal of Environmental Economics and Management*, 2017, *83*, 197–216.

Appendix A: Extensions of the conceptual framework

A.1. Multi-period generalisation

We here extend our conceptual framework from Section 2 to $T \geq 2$ periods—and hence also formally to a multi-period waterbed effect. Denote the system's carbon-price path by $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_T)$, and suppose that i's overlapping policy reduces its emissions demand in each period by $\Delta e_{i1}, \Delta e_{i2}, ..., \Delta e_{iT} < 0$ (fixed $\boldsymbol{\tau}$) and by $\Delta e_i = \sum_{t=1}^{T} \Delta e_{it} < 0$ over time. Define $\Gamma_{it} \equiv \Delta e_{it} / \Delta e_i \in (0, 1)$ as the fraction of the change in i's cumulative emissions demand that occurs in period t (where $\sum_{t=1}^{T} \Gamma_{it} \equiv 1$).

Generalizing Section 2, the key metric for climate change is the equilibrium impact on cumulative aggregate emissions $\Delta E^* = \sum_{t=1}^{T} \Delta E_t^*$ (endogenous $\boldsymbol{\tau}$). Intra-industry internal carbon leakage in period t is $L_{it} \equiv -\Delta e_{jt} / \Delta e_{it}$ (fixed $\boldsymbol{\tau}$) while per-period inter-industry leakage is $\ell_{it} \equiv -\Delta \widehat{E}_t \Big/ \Delta e_{it}$ (fixed $\boldsymbol{\tau}$). Define $L_i \equiv \sum_{t=1}^{T} \Gamma_{it} L_{it}$ and $\ell_i \equiv \sum_{t=1}^{T} \Gamma_{it} \ell_{it}$ as the weighted-average intra- and inter-industry leakage rates across the T periods. The multi-period waterbed effect $W = 1 - \Delta E^* / \Delta E$ is as before.

This yields a clean generalization of Lemma 1: at equilibrium, the change in cumulative emissions due to i's policy satisfies $\Delta E^* = [1 - \widetilde{L}_i(\kappa)][1 - W]\Delta de_i$, where $\widetilde{L}_i(\kappa) \equiv [L_i + \kappa \ell_i]$ is (multi-period) aggregate internal carbon leakage. To see why, note that $\Delta E^* = [1 - W]\Delta E$ (endogenous $\boldsymbol{\tau}$) by construction while the cumulative net change in emissions demand due to i's policy satisfies $\Delta E = \sum_{t=1}^{T} [1 - L_{it}]\Delta e_{it} + \kappa \sum_{t=1}^{T} \Delta \widehat{E}_t$ (fixed $\boldsymbol{\tau}$) where $\kappa = 1$ ($\kappa = 0$) for a multi-sector (single-sector) ETS. Using the definition of Γ_{it}, this rewrites as $\Delta E = \left[\sum_{t=1}^{T} [1 - L_{it}]\Gamma_{it} + \kappa \sum_{t=1}^{T} \ell_{it} \Gamma_{it}\right] \Delta e_i$ and the result follows from the definitions of the different leakage rates (and recalling that $\sum_{t=1}^{T} \Gamma_{it} \equiv 1$).

A.2. Alternative concepts of internal carbon leakage

This appendix explains how different concepts of internal carbon leakage relate to our conceptual framework of Section 2, and why our sequential approach that delineates the waterbed effect is the simplest and analytically clearest framework.

Two other definitions of internal leakage have been used in prior work. First, some empirical papers estimate the "meta" version of internal leakage $L_i^{\mathrm{M}} = 1 - R_i$ that bundles our $\widetilde{L}_i$ with the waterbed effect (e.g., Vollebergh, 2018). Second, others define a "total" intra-industry form that incorporates the emissions response to any induced change to the system-wide carbon price, $L_i^{\mathrm{T}} \equiv -\Delta e_j / \Delta e_i$ (endogenous τ), but ignores inter-industry internal leakage (e.g., Shawhan et al., 2019).[41]

[41] Section 5 in the main text discusses different strands of empirical work in detail.

ETS sectors	**Fixed carbon price**	**Equilibrium carbon price**
ETS sector targeted by overlapping policy	A	C
Other non-targeted ETS sectors	B	D

Table A.1: Decomposition of equilibrium emissions change due to an overlapping policy

Table A.1 clarifies the rationale for our approach by decomposing the impact on equilibrium aggregate emissions of i's policy into four parts. First, A captures our notion of intra-industry internal carbon leakage in the sector targeted by the overlapping policy, for a fixed carbon price: $A = [1 - L_i]\Delta e_i$. Second, B captures inter-industry internal leakage to non-targeted ETS sectors, for a fixed carbon price: $B = \kappa \ell_i \Delta e_i$. Finally, $C + D$ reflects the waterbed effect with an endogenous carbon price: $C + D = -W[1 - L_i - \kappa \ell_i]\Delta e_i$. Taken together, this decomposition corresponds to Lemma 1 in that $A + B + C + D = [1 - \widetilde{L}_i][1 - W]\Delta e_i$, where $\widetilde{L}_i \equiv L_i + \kappa \ell_i$ is aggregate internal carbon leakage.

Now consider alternative definitions of internal carbon leakage with a multi-sector ETS ($\kappa = 1$), like the EU ETS. The concept of "total" internal leakage L_i^{T} captures only the impacts $A + C$ that occur in the targeted sector of the overlapping policy itself. An equivalent result to Lemma 1 based on L_i^{T} is necessarily more complex and less intuitive: it would require (i) re-including the waterbed effect W and making adjustments for the "missing D" as well as (ii) incorporating any inter-industry leakage as per B. As a result, it would also no longer be possible to proceed sequentially—from internal carbon leakage to the waterbed effect—as in our analysis.

The concept of "meta" internal carbon leakage $L_i^{\mathrm{M}} = 1 - R_i$, where $R_i = [1 - L_i][1 - W]$, by contrast, has the immediate drawback that it obscures the diverging policy implications that stem from internal carbon leakage compared with those from the waterbed effect. To illustrate, suppose that the analysis shows that an overlapping policy has $L_i^{\mathrm{M}} = 1$, and therefore induces no change in equilibrium aggregate emissions. If this stems from $\widetilde{L}_i = 1$, then this as such reveals only the limitations of this particular overlapping policy. However, if instead $W = 1$ then the carbon-market design does not allow *any* overlapping policy to affect aggregate emissions. This distinction is central to our conceptual framework—but is missed by instead using only L_i^{M}.

For a single-sector ETS ($\kappa = 0$), like power generation in RGGI, there is a direct correspondence between our framework and L_i^{T}—precisely because then $B \equiv 0$ as well as $D \equiv 0$. In this special case, $A + C = [1 - L_i^{\mathrm{T}}]\Delta e_i = [1 - L_i][1 - W]\Delta e_i$ so now also $L_i^{\mathrm{T}} = L_i^{\mathrm{M}} = 1 - R_i$. That is, an analysis based on L_i^{T} is then just as correct as ours—but is still subject to the critique that it obscures the different policy implications, e.g., of whether equilibrium emissions do not fall because $W = 1$ or because $L_i = 1$.

Appendix B: Proofs for intra-industry internal leakage

First, we derive two generalised results, Propositions 1A–2A, on intra-industry internal carbon leakage using a general non-separable cost function. Second, we obtain Propositions 1–2 from the main text as corollaries and discuss how the key insights from the separable case are robust. Third, we extend the baseline model to multi-market settings. Fourth, we discuss robustness for overlapping policies that are not marginal.

B.1. General results with non-separable cost functions

Firm k's emissions are $e_k = \theta_k x_k - a_k$ where a_k is abatement. Its general cost function is $G_k(x_k, a_k)$, with standing assumptions $G_k^x, G_k^a > 0$ and $G_k^{xx}, G_k^{aa} > 0$ so $G_k^{aa} \to \infty$ means that additional abatement is infeasible. The stability condition is $G_k^{xx} G_k^{aa} - G_k^{xa} G_k^{ax} > 0$.[42]

To maximise profits, firm k ($k = i, j$) solves $\max_{x_k, a_k} \Pi_k = p x_k - G_k(x_k, a_k) - \tau_k(\theta_k x_k - a_k)$. The two first-order conditions are:

$$p = G_k^x + \tau_k \theta_k \text{ and } \tau_k = G_k^a. \tag{A.1}$$

Let $M_k(x_k; a_k) \equiv [G_k^x + \theta_k G_k^a]$ be k's optimal marginal cost of output, given its optimal choice of abatement with $\tau_k = G_k^a$. We assume $M_k^a(x_k; a_k) \equiv [G_k^{xa} + \theta_k G_k^{aa}] > 0$, or equivalently that:

$$\delta_k \equiv \left(1 + \frac{G_k^{ax}}{\theta_k G_k^{aa}}\right) > 0.$$

This condition is trivially met for a separable cost function ($G_k^{xa} = 0$) and, more generally, is satisfied if $G_k^{xa} \geq 0$ or $G_k^{xa} < 0$ but not too negative. Intuitively, it limits the degree of cost complementarity between output and abatement so there is "no free lunch."

It will also be useful to define an index of non-separability of k's cost function:

$$\psi_k \equiv \frac{G_k^{xa}}{G_k^{xx}} \frac{G_k^{ax}}{G_k^{aa}} \in [0, 1).$$

The separable case ($G_k^{xa} = G_k^{ax} = 0$) is nested where $\psi_k = 0$ while $\psi_k < 1$ follows by stability. A key metric to characterise output responses in the general model will be:

$$\mu_k \equiv \frac{-p'}{[-p' + G_k^{xx}(1 - \psi_k)]} \in (0, 1)$$

where $\mu_k < 1$ is satisfied because of stability of equilibrium, $\psi_k < 1$. Armed with these preliminaries, we now derive generalisations of the results from the main text.

[42] The model from the main text with a separable cost function is nested where $G_k^{xa} = G_k^{ax} = 0$.

Supply-side overlapping policies

Proposition 1A. *With general cost functions, a supply-side overlapping policy by jurisdiction i increases the product price $dp/d\lambda_i > 0$ and has intra-industry internal carbon leakage to jurisdiction j in the targeted sector of:*

$$L_i = \frac{\theta_j}{\theta_i}\mu_j\frac{\delta_j}{\delta_i}\frac{1}{[1+\gamma\Omega_i^G]} > 0,$$

where the rate of output leakage is $L_i^O = \mu_j \in (0,1)$, $\gamma = 0$ for a unilateral reduction in carbon-intensive production, $\gamma = 1$ for a unilateral carbon price, and $\Omega_i^G \equiv \frac{G_i^{aa}}{M_i^a}\frac{G_i^{xx}}{M_i^a}\left[(1-\psi_i)+\mu_j(1-\psi_j)\frac{G_j^{xx}}{G_i^{xx}}\right] \geq 0$ is an abatement effect.

Proof of Proposition 1A. We begin with i's unilateral carbon price ($\gamma = 1$) for which $\tau_i = \tau_i(\tau, \lambda_i)$, and then obtain the unilateral reduction in carbon-intensive production ($\gamma = 0$) as a special case. Differentiating i's two first-order conditions from A.1 yields:

$$p'(X)\left(\frac{dx_i}{d\lambda_i}+\frac{dx_j}{d\lambda_i}\right) - G_i^{xx}\frac{dx_i}{d\lambda_i} - G_i^{xa}\frac{da_i}{d\lambda_i} - \theta_i\frac{d\tau_i}{d\lambda_i} = 0$$

$$\frac{d\tau_i}{d\lambda_i} - G_i^{ax}\frac{dx_i}{d\lambda_i} - G_i^{aa}\frac{da_i}{d\lambda_i} = 0 \Longrightarrow \frac{da_i}{d\lambda_i} = \frac{1}{G_i^{aa}}\left[\frac{d\tau_i}{d\lambda_i} - G_i^{ax}\frac{dx_i}{d\lambda_i}\right]. \tag{A.2}$$

As j's carbon price $\tau_j = \tau$ is fixed, differentiating j's first-order conditions from A.1 yields:

$$p'(X)\left(\frac{dx_i}{d\lambda_i}+\frac{dx_j}{d\lambda_i}\right) - G_j^{xx}\frac{dx_j}{d\lambda_i} - G_j^{xa}\frac{da_j}{d\lambda_i} = 0$$

$$- G_j^{ax}\frac{dx_j}{d\lambda_i} - G_j^{aa}\frac{da_j}{d\lambda_i} = 0 \Longrightarrow \frac{da_j}{d\lambda_i} = -\frac{G_j^{ax}}{G_j^{aa}}\frac{dx_j}{d\lambda_i}. \tag{A.3}$$

We derive the result in two steps. First, combining the two previous expressions for j from A.3 shows that the firms' output changes are related according to:

$$p'\frac{dx_i}{d\lambda_i} = \left[-p' + G_j^{xx}(1-\psi_j)\right]\frac{dx_j}{d\lambda_i}.$$

The same approach for i yields:

$$p'\frac{dx_j}{d\lambda_i} = \theta_i\delta_i\frac{d\tau_i}{d\lambda_i} + \left[-p' + G_i^{xx}(1-\psi_i)\right]\frac{dx_i}{d\lambda_i}.$$

using the definitions of ψ_k and δ_k. Writing this two-equation system in more compact form using the definition of μ_k gives:

$$-\mu_j\frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i} \text{ and } -\mu_i\frac{dx_j}{d\lambda_i} = \mu_i\frac{\theta_i\delta_i}{(-p')}\frac{d\tau_i}{d\lambda_i} + \frac{dx_i}{d\lambda_i}.$$

Hence solving for the equilibrium output responses yields:

$$\frac{dx_i}{d\lambda_i} = -\left[\frac{\mu_i}{(1-\mu_i\mu_j)}\frac{\theta_i\delta_i}{(-p')}\right]\frac{d\tau_i}{d\lambda_i} < 0 \text{ and } \frac{dx_j}{d\lambda_i} = \left[\frac{\mu_i\mu_j}{(1-\mu_i\mu_j)}\frac{\theta_i\delta_i}{(-p')}\right]\frac{d\tau_i}{d\lambda_i} > 0. \qquad \text{(A.4)}$$

Therefore the rate of internal output leakage is $L_i^{\mathrm{O}} \equiv \frac{dx_j/d\lambda_i}{-dx_i/d\lambda_i} = \mu_j \in (0,1)$, which is always positive but less than 100% by stability. The change in the equilibrium price satisfies $\frac{dp}{d\lambda_i} = -[-p'(X)]\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right)$ which, using A.4, yields:

$$\frac{dp}{d\lambda_i} = -[-p'(X)]\left(-\frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\frac{\theta_i\delta_i}{[-p'(X)]}\frac{d\tau_i}{d\lambda_i}\right) = \frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\theta_i\delta_i\frac{d\tau_i}{d\lambda_i} > 0. \qquad \text{(A.5)}$$

Second, recall that emissions changes and output changes are related according to $\frac{de_k}{d\lambda_i} = \theta_k\frac{dx_k}{d\lambda_i} - \frac{da_k}{d\lambda_i}$. Using j's equilibrium output response from A.4 and its abatement response from A.3 we obtain:

$$\frac{de_j}{d\lambda_i} = \theta_j\delta_j\frac{dx_j}{d\lambda_i} = \theta_i\theta_j\frac{\mu_i\mu_j}{(1-\mu_i\mu_j)}\frac{\delta_i\delta_j}{(-p')}\frac{d\tau_i}{d\lambda_i} > 0.$$

We similarly obtain for i:

$$\frac{de_i}{d\lambda_i} = \theta_i\delta_i\frac{dx_i}{d\lambda_i} - \frac{1}{G_i^{aa}}\frac{d\tau_i}{d\lambda_i} = -\theta_i^2\left[\frac{\mu_i}{(1-\mu_i\mu_j)}\frac{\delta_i^2}{(-p')} + \frac{1}{\theta_i^2 G_i^{aa}}\right]\frac{d\tau_i}{d\lambda_i} < 0.$$

Therefore the rate of internal carbon leakage due to the unilateral carbon price satisfies:

$$L_i \equiv \frac{de_j/d\lambda_i}{-de_i/d\lambda_i} = \frac{\theta_j}{\theta_i}\mu_j\frac{\frac{\mu_i}{(1-\mu_i\mu_j)}\frac{\delta_i\delta_j}{(-p')}}{\left[\frac{\mu_i}{(1-\mu_i\mu_j)}\frac{\delta_i^2}{(-p')} + \frac{1}{\theta_i^2 G_i^{aa}}\right]} = \frac{\theta_j}{\theta_i}\mu_j\frac{\delta_j}{\delta_i}\frac{1}{\left[1 + \frac{(-p')}{\delta_i^2\theta_i^2 G_i^{aa}}\left[\frac{1}{\mu_i} - \mu_j\right]\right]}.$$

Now rewrite the last term recalling the definition $\mu_k \equiv (-p')/\left[-p' + G_k^{xx}(1-\psi_k)\right]$:

$$\begin{aligned}
\left[(-p')\left[\frac{1}{\mu_i} - \mu_j\right]\right. &= \left[\frac{G_i^{xx}(1-\psi_i)}{-p'} + \frac{G_j^{xx}(1-\psi_j)}{\left[-p' + G_j^{xx}(1-\psi_j)\right]}\right](-p') \\
&= G_i^{xx}\left[(1-\psi_i) + \mu_j(1-\psi_j)\frac{G_j^{xx}}{G_i^{xx}}\right]
\end{aligned}$$

Also recalling that $M_i^a(x_k; a_k) \equiv [G_i^{xa} + \theta_i G_i^{aa}] > 0$, we have:

$$\frac{G_i^{xx}}{\delta_i^2\theta_i^2 G_i^{aa}} = \frac{1}{\left(\theta_i + \frac{G_i^{ax}}{G_i^{aa}}\right)^2\frac{G_i^{aa}}{G_i^{xx}}} = \frac{G_i^{aa}G_i^{xx}}{\left(G_i^{ax} + \theta_i G_i^{aa}\right)^2} = \frac{G_i^{aa}}{M_i^a}\frac{G_i^{xx}}{M_i^a}.$$

Using these terms in the expression for L_i yields the result for the unilateral carbon prices.

Now consider the unilateral reduction in carbon-intensive production ($\gamma = 0$), for which $dx_i/d\lambda_i < 0$ while $\tau_i = \tau_j = \tau$. Now i's output change is determined by policy directly rather than induced in equilibrium by a unilateral carbon price but the other choices—abatement by i and output and abatement by j—remain optimal.

Hence differentiating i's remaining first-order condition for abatement from A.1 yields:

$$-G_i^{ax}\frac{dx_i}{d\lambda_i} - G_i^{aa}\frac{da_i}{d\lambda_i} = 0 \Longrightarrow \frac{da_i}{d\lambda_i} = -\frac{G_i^{ax}}{G_i^{aa}}\frac{dx_i}{d\lambda_i}.$$

Differentiating j's two first-order conditions, also from A.1, yields:

$$p'(X)\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - G_j^{xx}\frac{dx_j}{d\lambda_i} - G_j^{xa}\frac{da_j}{d\lambda_i} = 0$$

$$-G_j^{ax}\frac{dx_j}{d\lambda_i} - G_j^{aa}\frac{da_j}{d\lambda_i} = 0 \Longrightarrow \frac{da_j}{d\lambda_i} = -\frac{G_j^{ax}}{G_j^{aa}}\frac{dx_j}{d\lambda_i}.$$

Writing these conditions in more compact form, using the definitions of ψ_j and μ_j, gives:

$$-\mu_j\frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i} > 0 \Longrightarrow L_i^{\mathrm{O}} \equiv \frac{dx_j/d\lambda_i}{-dx_i/d\lambda_i} = \mu_j \in (0,1),$$

which is exactly as for the unilateral carbon price.

Emissions changes and output changes are again related according to $\frac{de_k}{d\lambda_i} = \theta_k\frac{dx_k}{d\lambda_i} - \frac{da_k}{d\lambda_i}$. Given the policy's $dx_i/d\lambda_i < 0$, using the above results on firms' equilibrium output and abatement responses and the definition of δ_k, we obtain:

$$\frac{de_i}{d\lambda_i} = \theta_i\delta_i\frac{dx_i}{d\lambda_i} < 0 \text{ and } \frac{de_j}{d\lambda_i} = -\theta_j\delta_j\mu_j\frac{dx_i}{d\lambda_i} > 0$$

So the equilibrium rate of internal carbon leakage is as claimed:

$$L_i = \frac{\theta_j}{\theta_i}\mu_j\frac{\delta_j}{\delta_i} > 0.$$

Demand-side overlapping policies

Proposition 2A. *With general cost functions, a demand-side overlapping policy by jurisdiction i—(i) a renewables support program that brings in additional zero-carbon production, or (ii) an energy-efficiency program that reduces demand for carbon-intensive production, or (iii) a carbon-consumption tax—decreases the product price $dp/d\lambda_i < 0$ and has intra-industry internal carbon leakage to jurisdiction j in the targeted sector of:*

$$L_i = -\frac{\theta_j}{\theta_i}\left[\frac{\mu_j/(1-\mu_j)}{\mu_i/(1-\mu_i)}\right]\frac{\delta_j}{\delta_i} < 0.$$

Proof of Proposition 2A. As explained in the main text, all three demand-side overlapping policies are modeled via their impact on the demand curve, with $\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) < 0$. The carbon price remains unchanged, $\tau_i = \tau_j = \tau$. Thus differentiating i's first-order conditions from A.1 yields its equilibrium output and abatement responses:

$$\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) + p'(X;\lambda_i)\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - G_i^{xx}\frac{dx_i}{d\lambda_i} - G_i^{xa}\frac{da_i}{d\lambda_i} = 0$$

$$-G_i^{ax}\frac{dx_i}{d\lambda_i} - G_i^{aa}\frac{da_i}{d\lambda_i} = 0 \Longrightarrow \frac{da_i}{d\lambda_i} = -\frac{G_i^{ax}}{G_i^{aa}}\frac{dx_i}{d\lambda_i}.$$

Differentiating j's first-order conditions from A.1 yields symmetrically:

$$\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) + p'(X;\lambda_i)\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - G_j^{xx}\frac{dx_j}{d\lambda_i} - G_j^{xa}\frac{da_j}{d\lambda_i} = 0$$

$$-G_j^{ax}\frac{dx_j}{d\lambda_i} - G_j^{aa}\frac{da_j}{d\lambda_i} = 0 \Longrightarrow \frac{da_j}{d\lambda_i} = -\frac{G_j^{ax}}{G_j^{aa}}\frac{dx_j}{d\lambda_i}.$$

We again proceed in two main steps. First, combining these two expressions for j's equilibrium output and abatement responses and using the definition of ψ_j shows that firms' output changes are related according to:

$$\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) + p'\frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i}\left[-p' + G_j^{xx}(1-\psi_j)\right].$$

The same approach for i yields:

$$\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) + p'\frac{dx_j}{d\lambda_i} = \frac{dx_i}{d\lambda_i}\left[-p' + G_i^{xx}(1-\psi_i)\right].$$

Writing this two-equation system using the definition of μ_k gives:

$$\frac{dx_i}{d\lambda_i} = -\mu_i\left[\frac{dx_j}{d\lambda_i} - \frac{1}{(-p')}\frac{\partial}{\partial \lambda_i} p(X;\lambda_i)\right] \text{ and } \frac{dx_j}{d\lambda_i} = -\mu_j\left[\frac{dx_i}{d\lambda_i} - \frac{1}{(-p')}\frac{\partial}{\partial \lambda_i} p(X;\lambda_i)\right].$$

Solving for equilibrium output responses yields:

$$\frac{dx_i}{d\lambda_i} = \frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\frac{1}{(-p')}\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) < 0 \text{ and } \frac{dx_j}{d\lambda_i} = \frac{\mu_j(1-\mu_i)}{(1-\mu_i\mu_j)}\frac{1}{(-p')}\frac{\partial}{\partial \lambda_i} p(X;\lambda_i) < 0.$$

So the rate of internal output leakage is:

$$L_i^{\mathrm{O}} \equiv \frac{dx_j/d\lambda_i}{-dx_i/d\lambda_i} = -\frac{\mu_j(1-\mu_i)}{\mu_i(1-\mu_j)} < 0$$

which is always negative. The change in the equilibrium price satisfies $\frac{dp}{d\lambda_i} = \frac{\partial p}{\partial \lambda_i} - [-p'(X)]\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right)$ which, using the equilibrium output responses, yields:

$$\frac{dp}{d\lambda_i} = \frac{\partial p}{\partial \lambda_i} - \frac{[\mu_i(1-\mu_j) + \mu_j(1-\mu_i)]}{(1-\mu_i\mu_j)}\frac{\partial p}{\partial \lambda_i} = \frac{1}{(1-\mu_i\mu_j)}\frac{\partial p}{\partial \lambda_i} < 0. \tag{A.6}$$

Second, emissions changes and output changes are here related according to $\frac{de_k}{d\lambda_i} = \theta_k \frac{dx_k}{d\lambda_i} - \frac{da_k}{d\lambda_i}$. Using j's equilibrium output and abatement responses, and the definition of δ_k, we obtain:

$$\frac{de_j}{d\lambda_i} = \left(\theta_j + \frac{G_j^{ax}}{G_j^{aa}}\right)\frac{dx_j}{d\lambda_i} = \theta_j\delta_j \frac{\mu_j(1-\mu_i)}{(1-\mu_i\mu_j)}\frac{1}{(-p')}\frac{\partial}{\partial \lambda_i}p(X;\lambda_i) < 0.$$

We similarly obtain for i:

$$\frac{de_i}{d\lambda_i} = \left(\theta_i + \frac{G_i^{xa}}{G_i^{aa}}\right)\frac{dx_i}{d\lambda_i} = \theta_i\delta_i \frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\frac{1}{(-p')}\frac{\partial}{\partial \lambda_i}p(X;\lambda_i) < 0. \tag{A.7}$$

Therefore the equilibrium rate of internal carbon leakage is as claimed:

$$L_i = -\frac{\theta_j}{\theta_i}\left[\frac{\mu_j/(1-\mu_j)}{\mu_i/(1-\mu_i)}\right]\frac{\delta_j}{\delta_i}.$$

B.2. Robustness of results with separable cost functions

We now derive Propositions 1–2 for separable cost functions as direct corollaries of Propositions 1A–2A from Appendix B.1 and discuss how the key insights from the simplified model are robust.

The separable cost function $G_k(x_k, a_k) \equiv [C_k(x_k) + \phi_k(a_k)]$ is nested within the general model where $G_k^{xa} = G_k^{ax} = 0$. The general model then simplifies with $\delta_k = 1$, $\psi_k = 0$ as well as $\mu_k = (-p')/(-p' + C_k'') \in (0,1)$ for $k = i, j$.

We begin by recording two preliminary results. First, using the demand elasticity $\varepsilon^D \equiv -p(\cdot)/Xp'(\cdot) > 0$ and k's elasticity of total marginal cost $\eta_k^S \equiv x_k\widehat{C}_k''(x_k)/\widehat{C}_k'(x_k) > 0$, where $\widehat{C}_k'(x_k) \equiv [C_k'(x_k) + \tau_k\theta_k] = p(X)$ and $\widehat{C}_k''(x_k) \equiv C_k''(x_k)$, we can rewrite this cost term as follows:

$$C_k''(x_k) = \frac{x_k C_k''(x_k)}{\widehat{C}_k'(x_k)}\frac{\widehat{C}_k'(x_k)}{x_k} = \frac{x_k \widehat{C}_k''(x_k)}{\widehat{C}_k'(x_k)}\frac{\widehat{C}_k'(x_k)}{x_k} = \eta_k^S \frac{p(X)}{X}\frac{1}{\sigma_k} = \frac{p(X)}{X}\frac{1}{\sigma_k\varepsilon_k^S}$$

where the last expression uses the definition $\sigma_k \equiv x_k/X \in (0,1)$ and $\eta_k^S = 1/\varepsilon_k^S$ (see

Section 3.1). Second, using the same ingredients, we also obtain that:

$$\mu_k \equiv \frac{-p'}{(-p' + C''_k)} = \frac{\sigma_k}{(\sigma_k + \varepsilon^D/\varepsilon^S_k)} > 0,$$

which will again be the key driver of firms' equilibrium output responses.

Supply-side overlapping policies

Proposition 1. *A supply-side overlapping policy by jurisdiction i increases the product price $dp/d\lambda_i > 0$ and has intra-industry internal carbon leakage to jurisdiction j of:*

$$L_i = \frac{\theta_j}{\theta_i}\left[\frac{\sigma_j}{(\sigma_j + \varepsilon^D/\varepsilon^S_j)}\right]\frac{1}{(1+\gamma\Omega_i)} > 0,$$

where $\gamma = 0$ for a unilateral reduction in carbon-intensive production, $\gamma = 1$ for a unilateral carbon price, and $\Omega_i \equiv \frac{C''_i}{\theta_i^2\phi''_i}\left(1 + \frac{(1-\sigma_j)\varepsilon^S_i/\varepsilon^S_j}{(\sigma_j+\varepsilon^D/\varepsilon^S_j)}\right) \geq 0$ is an abatement effect.

Proof of Proposition 1. For the unilateral carbon price ($\gamma = 1$), the leakage formula from Proposition 1A simplifies to:

$$L_i = \frac{\theta_j}{\theta_i}\mu_j \frac{1}{\left[1 + \frac{G^{aa}_i}{M^a_i}\frac{G^{xx}_i}{M^a_i}\left[1 + \mu_j\frac{G^{xx}_j}{G^{xx}_i}\right]\right]} = \frac{\theta_j}{\theta_i}\mu_j \frac{1}{\left[1 + \frac{C''_i}{\theta_i^2\phi''_i}\left[1 + \mu_j\frac{C''_j}{C''_i}\right]\right]}.$$

Using the two preliminary results, including the property $C''_j/C''_i = \sigma_i\varepsilon^S_i/\sigma_j\varepsilon^S_j$, yields the result. For the unilateral reduction in carbon-intensive production ($\gamma = 0$), Proposition 1A simplifies directly to $L_i = \frac{\theta_j}{\theta_i}\mu_j$, thus establishing the result.

Compared with the general result from Proposition 1A, an obvious difference is the absence of the term δ_j/δ_i, where $\delta_k = (1 + G^{ax}_k/\theta_k G^{aa}_k) > 0$ captures the extent of non-separability in k's cost function. There are two immediate observations. First, all else equal, the two results will be similar—even identical—if non-separability plays out similarly for both firms, with $\delta_i \simeq \delta_j \neq 1$. Second, there is no obvious bias: the simplified result is an overestimate of internal leakage if $\delta_j < \delta_i$ and an underestimate otherwise.

To understand the economics, observe that, if $\delta_k < 1 \Leftrightarrow G^{xa}_k < 0$ $(k = i, j)$, j tends to abate more for a given output increase—which pushes downwards the internal leakage of i's policy. By the same token, however, i's output reduction then undermines its own abatement incentive—which pushes internal leakage upwards. The net effect is therefore ambiguous. The reverse logic applies where $\delta_k > 1 \Leftrightarrow G^{xa}_k > 0$.

A second difference between the two results arises via the rate of output leakage. In particular, recall that $L^{\mathrm{O}}_i = \mu_j \equiv (-p')/\left[-p' + G^{xx}_j(1-\psi_j)\right] \in (0,1)$ in the general case. Hence, from the same starting point, output leakage is more pronounced in the general case ($\psi_j > 0$) than in the separable case ($\psi_j = 0$). Intuitively, if $G_j^{xa} < 0$, then abatement raises the marginal return to output, and vice versa, so, all else equal, j's output increase is more pronounced. The same logic applies in reverse for $G_j^{xa} > 0$: abatement makes output less attractive, and vice versa. Hence, across both cases, non-separability raises j's marginal return to output—so L_i^{O} is greater for $G_j^{xa} \neq 0$ than for $G_j^{xa} = 0$.

The relative emissions intensity θ_j/θ_i plays exactly the same role in both results, and internal carbon leakage exceeds 100% if it is sufficiently pronounced. Finally, the abatement effect also plays a similar role in the general (Ω_i^G) and separable (Ω_i) models for the unilateral carbon price ($\gamma = 1$)—but is irrelevant for the production cut ($\gamma = 0$).

In sum, while the precise numbers may differ, the main insights from the case with separable cost functions hold more generally—most notably that internal leakage from supply-side policies is always positive.

Demand-side overlapping policies

Proposition 2. *A demand-side overlapping policy by jurisdiction i—(i) a renewables support program that brings in additional zero-carbon production, or (ii) an energy-efficiency program that reduces demand for carbon-intensive production, or (iii) a carbon consumption tax—decreases the product price $dp/d\lambda_i < 0$ and has intra-industry internal carbon leakage to jurisdiction j of:*

$$L_i = -\frac{\theta_j}{\theta_i}\frac{\sigma_j}{(1-\sigma_j)}\frac{\varepsilon_j^S}{\varepsilon_i^S} < 0.$$

Proof of Proposition 2. The expression for internal carbon leakage from Proposition 2A simplifies as:

$$L_i = -\frac{\theta_j}{\theta_i}\left[\frac{\mu_j/(1-\mu_j)}{\mu_i/(1-\mu_i)}\right]\frac{\delta_j}{\delta_i} = -\frac{\theta_j}{\theta_i}\frac{C_i''}{C_j''}.$$

Using the relationship $C_k''(x_k) = \frac{p(X)}{X}\frac{1}{\sigma_k \varepsilon_k^S}$ yields the result as claimed.

Comparing this with the general result from Proposition 2A, similar effects are at work as for supply-side policies. A difference is that demand-side policies do not lead to a carbon price-induced abatement effect, neither in the separable nor in the general case.

First, exactly as for supply-side policies, the term δ_j/δ_i is absent in the separable case. However, by the same arguments as before, this effect (i) becomes negligible if non-separability plays out similarly for both firms, with $\delta_i \simeq \delta_j \neq 1$ and (ii) does not lead to any clear-cut bias in the result on internal leakage for the separable case.

Second, for demand-side policies, by contrast, the impact of separability on output leakage is now ambiguous as firms in both jurisdictions experience a direct change on their marginal return to output. In particular, note that $L_i^{\mathrm{O}} = -[\mu_j/(1-\mu_j)]/[\mu_i/(1-\mu_i)] = -G_i^{xx}(1-\psi_i)/G_j^{xx}(1-\psi_j)$ in the general case. This makes clear that, very similar to the previous point, this non-separability additional effect from the general case may be negligible and does not lead to any clear-cut bias in Proposition 2.

Third, the relative emissions intensity θ_j/θ_i again plays an identical role in both results.

In sum, the main insights from the separable case again hold more generally—most notably that internal leakage from demand-side policies is always negative.

B.3. Robustness of results with multi-market internal leakage

The main text considers a model with a single demand curve, interpreted to represent consumers in jurisdiction i. We here show that the main insights extend to multi-market settings with demand from both consumers in i and j.

Firm i sells x_i in its home market and exports y_i to j's market while firm j sells y_j in its home market and exports x_j to i's market. Demand in market i is $p_i(X)$ while demand in market j is $p_j(Y)$ where $X \equiv x_i + x_j$ and $Y \equiv y_i + y_j$. Firm k produces emissions $e_k = \theta_k Q_k$ where $Q_k \equiv x_k + y_k$ $(k = i, j)$ is its total sales across both markets. For simplicity, assume that firms' emissions intensities are fixed (no abatement). An overlapping policy by i can now induce changes across both markets.

The rate of internal carbon leakage now writes as $L_i \equiv -\frac{de_j}{de_i} = -\frac{\theta_i}{\theta_j}\frac{dQ_j}{dQ_i}$, and suppose for now that i indeed cuts back in both markets: $dx_i, dy_i < 0$. It is easy to check that internal leakage can then be re-expressed as a weighted average:

$$L_i = \frac{dx_i}{dQ_i} L_{ii} + \frac{dy_i}{dQ_i} L_{ij} \tag{A.8}$$

where $L_{ii} \equiv \frac{\theta_i}{\theta_j}\left(-\frac{dx_j}{dx_i}\right)$ is the "single-market" leakage rate arising in i's domestic market and $L_{ij} \equiv \frac{\theta_i}{\theta_j}\left(-\frac{dy_j}{dy_i}\right)$ is "cross-market" leakage arising from i's exports to j's home market.

Consider the benchmark case in which the multi-market cost functions $C_k^M(x_k, y_k)$ are separable with $C_i^M(x_i, y_i) = C_i(x_i) + H_i(y_i)$ while $C_j^M(x_j, y_j) = C_j(x_j) + H_j(y_j)$ for firm j. Firm i now has a first-order condition for each market:

$$\frac{\partial \Pi_i}{\partial x_i} = 0 = p_i - C_i'(x_i) - \tau_i\theta_i \text{ and } \frac{\partial \Pi_i}{\partial y_i} = 0 = p_j - H_i'(y_i) - \tau_i\theta_i \tag{A.9}$$

so that the product price, net of non-carbon costs, is equalized across the two markets, with $p_i - C_i'(x_i) = p_j - H_i'(y_i) = \tau_i\theta_i$. Similarly, firm j's first-order conditions are:

$$\frac{\partial \Pi_j}{\partial x_j} = 0 = p_i - C_j'(x_j) - \tau_j\theta_j \text{ and } \frac{\partial \Pi_j}{\partial y_j} = 0 = p_j - H_j'(y_j) - \tau_j\theta_j. \tag{A.10}$$

Market-level and firm-level responses to a policy λ_i relate according to $\frac{dX}{d\lambda_i}+\frac{dY}{d\lambda_i}=\frac{dQ_i}{d\lambda_i}+\frac{dQ_j}{d\lambda_i}$

Generalising the single-market case, define $\mu_{ii}\equiv\frac{-p_i'(X)}{\left[-p_i'(X)+C_i''(x_i)\right]}\in(0,1)$ and $\mu_{ij}\equiv\frac{-p_j'(Y)}{\left[-p_j'(Y)+H_i''(y_i)\right]}\in(0,1)$ as well as $\mu_{ji}\equiv\frac{-p_i'(X)}{\left[-p_i'(X)+C_j''(x_j)\right]}$, $\mu_{jj}\equiv\frac{-p_j'(Y)}{\left[-p_j'(Y)+H_j''(y_j)\right]}\in(0,1)$.

Supply-side overlapping policies

As in the main text, we here consider two overlapping policies: (i) a unilateral carbon price that raises i's carbon price $\tau_i=\tau_i(\tau,\lambda_i)$ according to $\frac{d\tau_i}{d\lambda_i}>0$, and (ii) a unilateral policy that requires a cut $\frac{dQ_i}{d\lambda_i}<0$ in i's overall production (e.g., a coal phase-out).

Proposition 1M. *A supply-side overlapping policy by jurisdiction i has intra-industry internal carbon leakage to jurisdiction j of:*

$$L_i=\frac{\theta_j}{\theta_i}\left[\frac{dx_i}{dQ_i}\mu_{ji}+\frac{dy_i}{dQ_i}\mu_{jj}\right]>0$$

where $\frac{dx_i}{dQ_i},\frac{dy_i}{dQ_i}\in(0,1)$ and $\mu_{ji},\mu_{jj}\in(0,1)$.

Proposition 1M generalizes our results to a multi-market setting: internal carbon leakage is always positive. The economics of output leakage for i's export market is akin to before: faced with a higher carbon price, its exports to market j become less competitive and firm j (partially) "fills the gap" by increasing its own domestic sales. If the two markets are identical in terms of costs and demands, then $L_{ii}=L_{ij}$ and our single-market result remains exact as then also $L_i=L_{ii}$.

Going beyond this simplified model, abatement by i induced by a unilateral carbon price will mitigate—but also not turn negative—internal leakage while a unilateral reduction of fossil production yields no abatement for the same reasons as in the main text. Non-separability of the multi-market cost functions $C_k^M(x_k,y_k)$—such as $C_k^M(x_k+y_k)$—would introduce additional effects as the single-market leakage rates then become interdependent. However, the overarching insight that supply-side overlapping policies have positive multi-market internal carbon leakage applies far beyond this simplified model.

Proof of Proposition 1M. For the unilateral carbon price, differentiating firm i's two first-order conditions from A.9 gives:

$$0=p_i'(X)\left(\frac{dx_i}{d\lambda_i}+\frac{dx_j}{d\lambda_i}\right)-C_i''(x_i)\frac{dx_i}{d\lambda_i}-\theta_i\frac{d\tau_i}{d\lambda_i}\text{ and }0=p_j'(Y)\left(\frac{dy_j}{d\lambda_i}+\frac{dy_i}{d\lambda_i}\right)-H_i''(y_i)\frac{dy_i}{d\lambda_i}-\theta_i\frac{d\tau_i}{d\lambda_i}$$

while for firm j using A.10:

$$0=p_j'(Y)\left(\frac{dy_j}{d\lambda_i}+\frac{dy_i}{d\lambda_i}\right)-H_j''(y_j)\frac{dy_j}{d\lambda_i}\text{ and }0=p_i'(X)\left(\frac{dx_i}{d\lambda_i}+\frac{dx_j}{d\lambda_i}\right)-C_j''(x_j)\frac{dx_j}{d\lambda_i}$$

Recall that the last condition governing j's exports to i is the main driver of leakage in the single-market model. Rearranging this condition gives:

$$\frac{dx_j}{d\lambda_i} = -\mu_{ji}\frac{dx_i}{d\lambda_i}$$

which mirrors the result from the main text. Similarly, rearranging i's output response for market i gives:

$$\frac{dx_i}{d\lambda_i} = -\mu_{ii}\frac{dx_j}{d\lambda_i} - \frac{\theta_i}{[[-p_i'(X)] + C_i''(x_i)]}\frac{d\tau_i}{d\lambda_i}$$

Now solving these last two equations simultaneously yields:

$$\frac{dx_j}{d\lambda_i} = \frac{\mu_{ii}\mu_{ji}}{(1-\mu_{ii}\mu_{ji})}\frac{\theta_i}{[-p_i'(X)]}\frac{d\tau_i}{d\lambda_i} > 0$$

and so the single-market leakage rate of i's policy via its home market i is:

$$L_{ii} = \frac{\theta_j}{\theta_i}\left(-\frac{dx_j}{dx_i}\right) = \frac{\theta_j}{\theta_i}\mu_{ji} > 0.$$

Given separability, the same arguments show that the single-market leakage rate of i's policy via its export market j is:

$$L_{ij} = \frac{\theta_j}{\theta_i}\left(-\frac{dy_j}{dy_i}\right) = \frac{\theta_j}{\theta_i}\mu_{jj} > 0,$$

and the result is immediate in conjunction with A.8.

For a unilateral reduction in fossil production, $\frac{dQ_i}{d\lambda_i} < 0$, using the same arguments as for a unilateral carbon price, output leakages across both markets satisfy:

$$\frac{dx_j}{d\lambda_i} = -\mu_{ji}\frac{dx_i}{d\lambda_i} \text{ and } \frac{dy_j}{d\lambda_i} = -\mu_{jj}\frac{dy_i}{d\lambda_i}$$

so that the single-market internal leakage rates are again $L_{ii} = \frac{\theta_j}{\theta_i}\mu_{ji} > 0$ and $L_{ij} = \frac{\theta_j}{\theta_i}\mu_{jj} > 0$. Moreover, i's profit-maximizing strategy translates $\frac{dQ_i}{d\lambda_i} < 0$ into cutbacks in both its home-destined and export-destined production, $\frac{dx_i}{d\lambda_i}, \frac{dy_i}{d\lambda_i} < 0$ with $\frac{dx_i}{d\lambda_i} + \frac{dy_i}{d\lambda_i} = \frac{dQ_i}{d\lambda_i} < 0$, where it equalizes the marginal profit from each market:

$$p_i'(X)\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - C_i''(x_i)\frac{dx_i}{d\lambda_i} = p_j'(Y)\left(\frac{dy_j}{d\lambda_i} + \frac{dy_i}{d\lambda_i}\right) - H_i''(y_i)\frac{dy_i}{d\lambda_i}$$

which, using j's output leakages and $\frac{dx_i}{d\lambda_i} + \frac{dy_i}{d\lambda_i} = \frac{dQ_i}{d\lambda_i}$, rearranges as:

$$[p_i'(X)(1-\mu_{ji}) - C_i''(x_i)]\frac{dx_i}{d\lambda_i} = \left[p_j'(Y)(1-\mu_{jj}) - H_i''(y_i)\right]\left(\frac{dQ_i}{d\lambda_i} - \frac{dx_i}{d\lambda_i}\right)$$

and so i's output responses satisfy:

$$\frac{\frac{dx_i}{d\lambda_i}}{\frac{dQ_i}{d\lambda_i}} = \frac{\left[(1-\mu_{jj}) + \frac{H_i''(y_i)}{\left[-p_j'(Y)\right]}\right]}{\frac{\left[-p_i'(X)\right]}{\left[-p_j'(Y)\right]}\left[(1-\mu_{ji}) + \frac{C_i''(x_i)}{\left[-p_i'(X)\right]}\right] + \left[(1-\mu_{jj}) + \frac{H_i''(y_i)}{\left[-p_j'(Y)\right]}\right]}$$

which, using the definitions of $\mu_{ii}, \mu_{ij}, \mu_{ji}, \mu_{jj}$, also writes as:

$$\frac{\frac{dx_i}{d\lambda_i}}{\frac{dQ_i}{d\lambda_i}} = \frac{\frac{\mu_{ii}}{(1-\mu_{ii}\mu_{ji})}}{\frac{\mu_{ii}}{(1-\mu_{ii}\mu_{ji})} + \frac{\mu_{ij}}{(1-\mu_{ij}\mu_{jj})}\frac{\left[-p_i'(X)\right]}{\left[-p_j'(Y)\right]}} = \frac{dx_i}{dQ_i} \in (0,1).$$

It is easy to check that these output responses in response to a unilateral reduction in fossil production are identical to those induced by a unilateral carbon price (as they both induce equalized marginal profits across both markets).

Demand-side overlapping policies

As in the main text, we here consider policies that reduce carbon-intensive demand in market i, $p_i(X, \lambda_i)$, with $\frac{\partial p_i}{\partial \lambda_i} < 0$, representing either renewables support, an energy-efficiency program or a carbon consumption tax.

Proposition 2M. *A demand-side overlapping policy by jurisdiction i has intra-industry internal carbon leakage to jurisdiction j of:*

$$L_i = -\frac{\theta_j}{\theta_i}\left[\frac{\mu_{ji}(1-\mu_{ii})}{\mu_{ii}(1-\mu_{ji})}\right] = L_{ii} < 0$$

where $\mu_{ii}, \mu_{ji} \in (0,1)$.

Proposition 2M shows an identical leakage rate to that of the single-market analysis in the main text. The reason is that the overlapping policy here only affects the equilibrium in market i, precisely because it impacts only demand in market i.

Going beyond this simplified model, by the same arguments as in the main text, demand-side overlapping policies do not induce any additional abatement. Moreover, if a renewables support policy in market i also leads to zero-carbon exports to market j then this would displace further fossil production in market j and bring forth an additional channel of negative internal leakage. Like for supply-side policies, non-separability of the multi-market cost functions $C_k^M(x_k, y_k)$ would also introduce additional effects—but our overarching insight that demand-side overlapping policies have negative multi-market internal carbon leakage again applies far beyond this simplified model.

Proof of Proposition 2M. Differentiating firm i's and j's first-order conditions for market i from A.9 and A.10 gives:

$$0 = \frac{\partial p_i}{\partial \lambda_i} + \frac{\partial p_i}{\partial X}\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - C_i''(x_i)\frac{dx_i}{d\lambda_i}$$

$$0 = \frac{\partial p_i}{\partial \lambda_i} + \frac{\partial p_i}{\partial X}\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - C_j''(x_j)\frac{dx_j}{d\lambda_i}$$

These conditions together imply that:

$$C_i''(x_i)\frac{dx_i}{d\lambda_i} = C_j''(x_j)\frac{dx_j}{d\lambda_i}$$

So the single-market rate of internal carbon leakage in market i is:

$$L_{ii} \equiv -\frac{\theta_j}{\theta_i}\frac{dx_j}{dx_i} = -\frac{\theta_j}{\theta_i}\frac{\frac{dx_j}{d\lambda_i}}{\frac{dx_i}{d\lambda_i}} = -\frac{\theta_j}{\theta_i}\frac{C_i''(x_i)}{C_j''(x_j)} = -\frac{\theta_j}{\theta_i}\left[\frac{\mu_{ji}(1-\mu_{ii})}{\mu_{ii}(1-\mu_{ji})}\right] < 0$$

just as in the main text. Since the overlapping policy has no impact on firms' costs or demands for market j, we conclude that $\frac{dy_i}{d\lambda_i} = \frac{dy_j}{d\lambda_i} = 0$ so that multi-market internal carbon leakage is given by:

$$L_i = -\frac{\theta_j}{\theta_i}\frac{dQ_j}{dQ_i} = \frac{\theta_j}{\theta_i}\frac{dx_j}{dx_i} = L_{ii}.$$

B.4. Robustness of results with non-marginal policies

Our results so far have been for a marginal overlapping policy $d\lambda_i$ that shifts emissions by small amounts, de_i and de_j. More generally, policy tightens from an initial level $\underline{\lambda}_i \geq 0$ to a new level $\overline{\lambda}_i$ where $\triangle\lambda_i \equiv (\overline{\lambda}_i - \underline{\lambda}_i)$ is a discrete change. We here make two points on the robustness of our results from the first-order approximation.

The first point is that non-marginal supply-side (demand-side) policies also have positive (negative) internal leakage more generally:

$$L_i \equiv \frac{\triangle e_j}{-\triangle e_i} = \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de_j}{-de_i}\right)\left(\frac{de_i}{d\lambda_i}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de_i}{d\lambda_i}\right)d\lambda_i},$$

showing that it is a weighted average of marginal leakage rates, where $de_i/d\lambda_i < 0$ for all $\lambda_i \in [\underline{\lambda}_i, \underline{\lambda}_i + \triangle\lambda_i]$. Hence $\text{sign}(\frac{\triangle e_j}{-\triangle e_i}) = \text{sign}(\frac{de_j}{-de_i})$, which is unambiguously positive (negative) for supply-side (demand-side) policies, as shown in Propositions 1(A) and 2(A).

The second point is that the marginal approximation implies no obvious bias in the magnitude and, in an important special case, yields an exact result. As we have seen, marginal rates of internal leakage in general depend on first-order derivatives of demand (via the demand elasticity) and second-order derivatives of cost functions (via supply elasticities and abatement opportunities). So the non-marginal leakage rate will be quantitatively similar as long as any second-order demand terms and third-order cost terms are negligible as needed. For supply-side policies, this obtains exactly if the demand curve is linear ($p'(X)$ is constant) and the cost functions are quadratic in output and abatement (in the general case, $G_k^{xx}, G_k^{aa}, G_k^{xa}$ all constant). Then $L_i \equiv \frac{\triangle e_j}{-\triangle e_i} = \frac{de_j}{-de_i}$ since marginal leakage $\frac{de_j}{-de_i}$ is constant over $\lambda_i \in [\underline{\lambda}_i, \underline{\lambda}_i + \triangle\lambda_i]$. By contrast, for demand-side policies, the exact result does not require *any* restrictions on the demand curve. Moreover, the simple marginal formulae contain no obvious bias: they could be an over- or underestimate depending on the precise higher-order properties of demand and cost functions.

Appendix C: Inter-industry internal carbon leakage

This appendix extends the baseline model to capture "inter-industry" internal carbon leakage that occurs as other sectors in a multi-sector ETS ($\kappa = 1$) adjust their emissions because they purchase an input from the targeted sector of an overlapping climate policy.

Similar to our analysis of intra-industry internal leakage, we derive simple formulae in Propositions 5 and 6 below to characterize the rate of inter-industry internal leakage, $\ell_i \equiv -\Delta\widehat{E}/\Delta e_i$ (fixed τ). We obtain sufficient conditions under which the rate of intra-industry internal leakage still gives a directionally correct prediction of the rate of aggregate internal carbon leakage, $\text{sign}\{\widetilde{L}_i(1)\} = \text{sign}\{L_i\}$.

We think of direct demand in the targeted sector as $D(p;\kappa) = \overline{D}(p) + \kappa Y(p)$, where $\overline{D}(p)$ represents final consumer demand and $Y(p)$ represents industrial input demand (say, from aluminium or steel), where we here focus on a multi-sector ETS with $\kappa = 1$. The price elasticity of demand $\varepsilon^D \equiv -\frac{pD'(p)}{D(p)}$ can be written as:

$$\varepsilon^D = \frac{\overline{D}(p)}{D(p)}\left(-\frac{\overline{D}'(p)}{p\overline{D}(p)}\right) + \frac{Y(p)}{D(p)}\left(-\frac{Y'(p)}{pY(p)}\right) = (1-v)\overline{\varepsilon}^D + v\varepsilon^Y \tag{A.11}$$

where $\overline{\varepsilon}^D - \frac{\overline{D}'(p)}{p\overline{D}(p)}$ is the price elasticity of final consumer demand, $\varepsilon^Y \equiv -\frac{Y'(p)}{pY(p)} > 0$ is the price elasticity of input demand from the industrial sector, and $v \equiv \frac{Y(p)}{D(p)} \in [0,1)$ is the share of industrial demand of overall demand in the targeted sector. Both final consumers and industrial demand buy from the targeted sector at a common price p; in our baseline model, this is the product price in the domestic jurisdiction i.

A representative industrial firm produces emissions $\widehat{E}(p) = \widehat{\theta}(p)Z(Y(p))$ where Z is its output and $\widehat{\theta} \equiv \widehat{E}/Z$ is its emissions intensity of output (both $Z(p)$ and $\widehat{\theta}(p)$ may vary with the price p paid to the targeted sector).[43] This representative firm in turn sells its product into a competitive market with an inverse demand curve $g(Z)$, where $g'(Z) < 0$ and the price elasticity of demand is given by $\varepsilon^Z = \frac{g(Z)}{-g'(Z)Z} > 0$.

A key metric will be how the industrial firm's emissions respond to the input price. Defining the input price elasticity of its emissions as $\varepsilon^E \equiv \frac{p\widehat{E}'(p)}{\widehat{E}(p)}$, we obtain:

$$\varepsilon^E = \frac{Z'(Y)Y}{Z}\left(\frac{pY'(p)}{Y}\right) + \frac{p\widehat{\theta}'(p)}{\widehat{\theta}} = -\varphi\varepsilon^Y + \varepsilon^\theta \tag{A.12}$$

where $\varepsilon^\theta \equiv \frac{p\widehat{\theta}'(p)}{\widehat{\theta}}$ is the input price elasticity of its emissions intensity and $\varphi \equiv \frac{Z'(Y)Y}{Z} > 0$ is a measure of its production flexibility in form of the output elasticity of input.

In other words, the input price elasticity of its emissions is driven by two effects. First, a "scale effect" $-\varphi\varepsilon^Y < 0$ that captures how, holding fixed input ratios, the industrial firm will reduce its use of the input factor—and hence also reduce the emissions associated with its own production. Second, a "substitution effect" ε^θ which captures how a higher input price induces substitution to other polluting inputs such as fossil fuels which may increase its emissions. In general, therefore, the sign of ε^E is ambiguous.

To illustrate the balance between these two effects, consider two familiar production functions. First, with a Leontief technology $Z = f(Y, \widehat{E}) = \min\{\alpha Y, \beta\widehat{E}\}$ it produces in fixed proportion to its use of its input factors (including emissions), so $\varphi = 1$ reflecting constant returns to scale, and the emissions intensity $\widehat{\theta}(p) = 1/\beta$ is a constant (with respect to the input price p) so $\varepsilon^\theta = 0$. Hence there is no substitution effect and $\varepsilon^E = -\varepsilon^Y < 0$ is driven solely by the scale effect.

Second, with Cobb-Douglas $Z = f(Y, \widehat{E}) = Y^\alpha\widehat{E}^\beta$, cost-minimization (with input prices p for Y and τ for $\widehat{E}$) yields the two factor demand functions:

$$Y(Z) = \left(\frac{\alpha}{\beta}\frac{\tau}{p}\right)^{\frac{\beta}{\alpha+\beta}} Z^{\frac{1}{\alpha+\beta}} \text{ and } \widehat{E}(Z) = \left(\frac{\beta}{\alpha}\frac{p}{\tau}\right)^{\frac{\alpha}{\alpha+\beta}} Z^{\frac{1}{\alpha+\beta}},$$

and so the representative firm's optimized cost function is:

$$J(Z) = pY(Z) + \tau\widehat{E}(Z) = \left[\left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}}\right] p^{\frac{\alpha}{\alpha+\beta}}\tau^{\frac{\beta}{\alpha+\beta}}Z^{\frac{1}{\alpha+\beta}}.$$

[43] For simplicity, we here model the non-targeted industrial sector as an ETS-wide aggregate, that is, we do not distinguish between jurisdictions i and j.

Hence $\varphi = \alpha + \beta$ (so $\varphi = 1$ under constant returns to scale) while the emissions intensity of output $\widehat{\theta}(p) = \left(\frac{\beta}{\alpha}\frac{p}{\tau}\right)^{\frac{\alpha}{\alpha+\beta}} Z^{\frac{1-\alpha-\beta}{\alpha+\beta}}$ now does depend on the input price p (of Y)—and on the system-wide carbon price τ (of $\widehat{E}$).

The calculation of $\varepsilon^{\theta} \equiv \frac{p\widehat{\theta}'(p)}{\widehat{\theta}(p)}$ then goes as follows:

$$
\begin{aligned}
\varepsilon^{\theta} &= \frac{p}{\widehat{\theta}}\left(\frac{\beta}{\alpha}\frac{1}{\tau}\right)^{\frac{\alpha}{\alpha+\beta}} \frac{d}{dp}\left[p^{\frac{\alpha}{\alpha+\beta}} Z(p)^{\frac{1-\alpha-\beta}{\alpha+\beta}}\right] \\
&= \frac{p}{\widehat{\theta}}\left(\frac{\beta}{\alpha}\frac{1}{\tau}\right)^{\frac{\alpha}{\alpha+\beta}} \left[\left(\frac{\alpha}{\alpha+\beta}\right) p^{-\frac{\beta}{\alpha+\beta}} Z(p)^{\frac{1-\alpha-\beta}{\alpha+\beta}} + p^{\frac{\alpha}{\alpha+\beta}}\left(\frac{1-\alpha-\beta}{\alpha+\beta}\right) Z(p)^{\frac{1-\alpha-\beta}{\alpha+\beta}-1}\frac{dZ}{dp}\right] \\
&= \frac{p}{\widehat{\theta}}\left[\frac{\widehat{\theta}}{p}\left(\frac{\alpha}{\alpha+\beta}\right) + \left(\frac{1-\alpha-\beta}{\alpha+\beta}\right)\frac{\widehat{\theta}}{Z}\frac{dZ}{dp}\right] \\
&= \left(\frac{\alpha}{\alpha+\beta}\right) - (1-\alpha-\beta)\,\varepsilon^{Y},
\end{aligned}
$$

where the third line uses the expression for $\widehat{\theta}$ and the last line uses $\frac{dZ(Y(p))}{dp} = \frac{dZ}{dY}\frac{dY}{dp} \Rightarrow \frac{p}{Z}\frac{dZ}{dp} = \left(\frac{dZ}{dY}\frac{Y}{Z}\right)\left(\frac{p}{Y}\frac{dY}{dp}\right) = -\varphi\varepsilon^{Y} = -(\alpha+\beta)\varepsilon^{Y}$. The relationship $\varepsilon^{E} = -\varphi\varepsilon^{Y} + \varepsilon^{\theta}$ then yields $\varepsilon^{E} = \left(\frac{\alpha}{\alpha+\beta}\right) - \varepsilon^{Y}$. So here the substitution effect dominates with $\varepsilon^{E} \geq 0$ whenever the price elasticity of input demand is sufficiently small, $\varepsilon^{Y} \leq \left(\frac{\alpha}{\alpha+\beta}\right)$.

The result can also be written in terms of the price elasticity of demand, $\varepsilon^{Z} > 0$. In a competitive market equilibrium in the non-targeted sector, price equals marginal cost, that is, $g = J'(Z)$, which implies that also

$$
\frac{dg}{g} = \left[\frac{p\frac{\partial}{\partial p}\left(J'(Z)\right)}{J'(Z)}\right]\frac{dp}{p} = \left(\frac{\alpha+\beta}{\alpha}\right)\frac{dp}{p}.
$$

where the second equality uses the previous expression for the cost function $J(Z)$. The price elasticity of demand can also be written as:

$$
\varepsilon^{Z} \equiv -\frac{g}{dg}\frac{dZ}{Z} = \frac{g}{dg}\left(\frac{dZ}{dY}\frac{Y}{Z}\right)\left(-\frac{dY}{dp}\frac{p}{Y}\right)\frac{dp}{p} = \left(\frac{\alpha+\beta}{\alpha}\right)(\alpha+\beta)\varepsilon^{Y}
$$

showing that $\text{sign}(\varepsilon^{Z}) = \text{sign}(\varepsilon^{Y})$—which reflects the principle that the demand for an input is more elastic when the demand for the final product is more elastic. Hence we also have:

$$
\varepsilon^{E} = \left(\frac{\alpha}{\alpha+\beta}\right)\left[1 - \frac{\varepsilon^{Z}}{(\alpha+\beta)}\right]
$$

where $\varepsilon^{E} \geq 0$ if and only if $\varepsilon^{Z} \leq (\alpha+\beta) = \varphi$. With constant returns to scale ($\varphi = 1$), the condition boils down to the final demand being price-inelastic, $\varepsilon^{Z} \leq 1$.

Intuitively, the scale effect disappears as $\varepsilon^{Y} \to 0, \varepsilon^{Z} \to 0$ as the emissions impact is then driven only by the substitution effect. For example, if a supply-side overlapping policy increases the electricity price, then the industrial firm's input cost rises but if final demand for steel is insensitive to the steel price, with $\varepsilon^Z \to 0$, then its sales remain (approximately) constant so the scale effect is zero but the firm will still substitute from electricity (Y) to fossil fuels ($\widehat{E}$) so its direct emissions go up and so $\varepsilon^\theta > 0, \varepsilon^E > 0$.

Supply-side overlapping policies

Proposition 5. *For a supply-side overlapping policy, inter-industry internal carbon leakage from the targeted sector to the non-targeted sector equals:*

$$\ell_i = \frac{\widehat{\theta}}{\theta_i}\left[\frac{\varepsilon^D/\varepsilon_j^S}{(\sigma_j + \varepsilon^D/\varepsilon_j^S)}\right]\frac{\varepsilon^E}{\varepsilon^D}\frac{Z}{X},$$

so $\ell_i \geq 0$ where the input price elasticity of emissions in the non-targeted sector $\varepsilon^E \geq 0$.

Proposition 5 is the analogue to Proposition 1 for intra-industry leakage, and depends on similar quantities. The emissions intensity of the non-targeted sector relative to i's emissions intensity of the targeted sector $\widehat{\theta}/\theta_i$ acts a scaling factor. The demand and supply elasticities ($\varepsilon^D, \varepsilon_j^S$) and j's market share (σ_j) in the targeted sector drive the increase in the product price due to the supply-side policy (Proposition 1). The ratio Z/X is a measure of relative market sizes: it sets the output of the non-targeted sector Z against the output of the targeted sector X; inter-industry leakage is more pronounced in magnitude if the non-targeted sector is relatively large.

A critical role is played by the input price elasticity of emissions in the non-targeted sector, $\varepsilon^E = (-\varphi\varepsilon^Y + \varepsilon^\theta)$ in that it *signs* inter-industry leakage to the non-targeted sector. By Proposition 1, the supply-side policy raises the product price in the targeted sector. If $\varepsilon^E \geq 0$ then this higher input price leads to an increase in the emissions in the non-targeted sector—and so also $\ell_i \geq 0$. This reflects the dominance of the substitution effect: the higher input price induces substitution to other polluting inputs, such as fossil fuels, which here leads to an increase in emissions because $\varepsilon^\theta > 0$ turns positive the sign of ε^E. As seen above, this is the case with a Cobb-Douglas technology and sufficiently low price elasticity of demand as $\varepsilon^Z \leq \varphi$. Conversely, if $\varepsilon^E \leq 0$ then $\ell_i \leq 0$ and this logic is reversed as the substitution effect is then dominated by the scale effect. This is the case, for instance, with a Leontief technology ($\varphi = 1$ and $\varepsilon^\theta = 0$) for which the substitution effect is zero and so $\varepsilon^E = -\varepsilon^Y < 0$.

By our conceptual framework, aggregate internal carbon leakage satisfies $\widetilde{L}_i(\kappa)\Big|_{\kappa=1} = L_i + \ell_i$. By Proposition 1, intra-industry leakage for a supply-side overlapping policy is positive, $L_i > 0$. It follows that a grossly sufficient condition for aggregate internal leakage to follow intra-industry leakage, sign$\{\widetilde{L}_i(1)\}$ = sign$\{L_i\}$, is that inter-industry leakage is also positive, $\ell_i \geq 0$. By Proposition 5, this holds if and only if $\varepsilon^E \geq 0$ which, in turn, is met for example by Cobb-Douglas with $\varepsilon^Z \leq \varphi$. In other cases, there is a tension between the two leakages so the robustness of our initial result depends on intra-industry leakage quantitatively dominating the calculation. By Proposition 5, this holds, for instance, if the non-targeted sector is already relatively clean ($\widehat{\theta}/\theta_i$ is sufficiently small) or when its size is small relative to the targeted sector (Z/X is sufficiently small).

Proof of Proposition 5. We derive the rate of inter-industry internal carbon leakage, defined as per our conceptual framework, as $\ell_i = \frac{\frac{d\widehat{E}}{d\lambda_i}}{-\frac{de_i}{d\lambda_i}}$ for a supply-side policy. We begin with responses in the targeted sector, and then turn to the non-targeted sector.

First, the change in the equilibrium price in the targeted sector due to i's overlapping policy, with separable cost functions $\psi_k = 0$ and without abatement $\delta_k = 1$ (see Proposition 1), follows from (A.5) as:

$$\frac{dp}{d\lambda_i} = \left[\frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\theta_i\right]\frac{d\tau_i}{d\lambda_i} > 0.$$

Also, from (A.4), the change in equilibrium emissions in the targeted sector then satisfies:

$$\frac{de_i}{d\lambda_i} = \theta_i\frac{dx_i}{d\lambda_i} = -\left[\frac{\mu_i}{(1-\mu_i\mu_j)}\frac{\theta_i^2}{[-p'(X)]}\right]\frac{d\tau_i}{d\lambda_i} < 0.$$

Second, emissions in the non-targeted sector therefore respond to the changed input price according to:

$$\frac{d\widehat{E}}{d\lambda_i} = \frac{d\widehat{E}}{dp}\frac{dp}{d\lambda_i} = \varepsilon^E\frac{\widehat{E}(p)}{p}\frac{dp}{d\lambda_i} = \varepsilon^E\frac{\widehat{E}(p)}{p}\left[\frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\right]\theta_i\frac{d\tau_i}{d\lambda_i}$$

which also uses the definition of the input price elasticity of its emissions $\varepsilon^E \equiv \frac{p\widehat{E}'(p)}{\widehat{E}(p)}$.

Combining the past two expressions, it follows that the rate of inter-industry leakage satisfies:

$$\ell_i = \frac{\frac{d\widehat{E}}{d\lambda_i}}{-\frac{de_i}{d\lambda_i}} = \frac{\widehat{\theta}}{\theta_i}(1-\mu_j)\frac{\varepsilon^E}{\varepsilon^D}\frac{Z}{X}$$

which uses the definition of the price elasticity of demand in the targeted sector $\varepsilon^D = \frac{p}{[-p'(X)]X}$ and that $\widehat{E} = \widehat{\theta}Z$. Recalling that $\mu_j = \sigma_j/(\sigma_j + \varepsilon^D/\varepsilon_j^S)$ from the proof of Proposition 1 now yields the result as claimed.

Demand-side overlapping policies

Proposition 6. *For a demand-side overlapping policy, inter-industry internal carbon leakage from the targeted sector to the non-targeted sector equals:*

$$\ell_i = -\frac{\widehat{\theta}}{\theta_i}\left[\frac{(\sigma_i + \varepsilon^D/\varepsilon_i^S)}{\sigma_i}\frac{(\sigma_j + \varepsilon^D/\varepsilon_j^S)}{\varepsilon^D/\varepsilon_j^S}\right]\frac{\varepsilon^E}{\varepsilon^D}\frac{Z}{X}$$

so $\ell_i \leq 0$ where the input price elasticity of emissions in the non-targeted sector $\varepsilon^E \geq 0$.

Proposition 6's formula for inter-industry leakage is the analogue to Proposition 2 for intra-industry leakage. The relative emissions intensity $\widehat{\theta}/\theta_i$ acts a scaling factor for the relative change $\ell_i \equiv -\Delta\widehat{E}/\Delta e_i$. The demand and supply elasticities ($\varepsilon^D, \varepsilon_k^S$) and market shares (σ_k) in the targeted sector ($k = i, j$) drive the decrease in the product price (Proposition 2). The ratio Z/X is a measure of relative market sizes; inter-industry leakage is more pronounced in magnitude if the non-targeted sector is relatively large.

Once again, the input price elasticity of emissions in the non-targeted sector, $\varepsilon^E = (-\varphi\varepsilon^Y + \varepsilon^\theta)$ signs inter-industry leakage to the non-targeted sector. By Proposition 2, the demand-side policy reduces the product price in the targeted sector. If $\varepsilon^E \geq 0$ then this lower input price leads to lower emissions in the non-targeted sector—and so also $\ell_i \leq 0$. This again reflects the substitution effect dominating the scale effect.

Finally, by Proposition 2, intra-industry leakage for a demand-side overlapping policy $L_i < 0$ so, recalling that $\widetilde{L}_i(\kappa)\Big|_{\kappa=1} = L_i + \ell_i$, a grossly sufficient condition for aggregate internal leakage to follow, with $\text{sign}\{\widetilde{L}_i(1)\} = \text{sign}\{L_i\}$, is again that inter-industry leakage is also negative, $\ell_i \leq 0$ which, by Proposition 6, holds if and only if $\varepsilon^E \geq 0$.

Proof of Proposition 6. We derive the rate of inter-industry internal carbon leakage, defined as per our conceptual framework, as $\ell_i = \frac{\frac{d\widehat{E}}{d\lambda_i}}{-\frac{de_i}{d\lambda_i}}$ for a demand-side policy. We begin with responses in the targeted sector, and then turn to the non-targeted sector.

First, from the proof of Proposition 2 (as per A.6), the change in the equilibrium price in the targeted sector due to i's overlapping policy is given by:

$$\frac{dp}{d\lambda_i} = \frac{1}{(1-\mu_i\mu_j)}\frac{\partial p}{\partial \lambda_i} < 0.$$

Also, from (A.7), the change in equilibrium emissions in the targeted sector then satisfies:

$$\frac{de_i}{d\lambda_i} = \theta_i\frac{dx_i}{d\lambda_i} = \theta_i\left[\frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)}\frac{1}{[-p'(X)]}\right]\frac{\partial p}{\partial \lambda_i} < 0.$$

Second, emissions in the non-targeted sector therefore respond to the changed input price according to:

$$\frac{d\widehat{E}}{d\lambda_i} = \frac{d\widehat{E}}{dp}\frac{dp}{d\lambda_i} = \varepsilon^E \frac{\widehat{E}(p)}{p}\frac{dp}{d\lambda_i} = \varepsilon^E \frac{\widehat{E}(p)}{p}\left[\frac{1}{(1-\mu_i\mu_j)}\right]\frac{d\tau_i}{d\lambda_i}$$

which uses the definition of the input price elasticity of its emissions $\varepsilon^E \equiv \frac{p\widehat{E}'(p)}{\widehat{E}(p)}$.

Combining the past two expressions, it follows that the rate of inter-industry leakage satisfies:

$$\ell_i = \frac{\frac{d\widehat{E}}{d\lambda_i}}{-\frac{de_i}{d\lambda_i}} = -\frac{\widehat{\theta}}{\theta_i}\left[\frac{1}{\mu_i(1-\mu_j)}\right]\frac{\varepsilon^E}{\varepsilon^D}\frac{Z}{X}$$

which again uses $\varepsilon^D = \frac{p}{[-p'(X)]X}$ and $\widehat{E} = \widehat{\theta}Z$. Again recalling that $\mu_k = \sigma_k/(\sigma_k + \varepsilon^D/\varepsilon^S_k)$ yields the result as claimed.

We conclude that our main insights on internal carbon leakage translate to a range of cases in a multi-sector setting. A grossly sufficient condition is $\varepsilon^E \geq 0$, ensures that intra- and inter-industry leakage always have the same sign. More generally, our simple formulae from the main text have no obvious bias.

Appendix D: Proofs for waterbed effects

First, we present proofs for several results for price-based flexibility mechanisms from Section 4, including how the timing of shifts in emissions demand is irrelevant as per Lemma 2. Second, we present proofs for quantity-based flexibility mechanisms, including the link to price-based mechanisms from Corollary 1. Third, we show the robustness of our results to a more responsive banking-based flexibility mechanism. Fourth, we discuss the robustness of our results to overlapping policies that are non-marginal.

D.1. Waterbed effect for price-based flexibility mechanisms

Derivation of Equation (6)

Application of Cramer's rule to the two equilibrium carbon-market conditions $\rho(E, \lambda_i) = \tau$ from (4) and $E = s(\tau)$ from (5) yields:

$$\frac{\partial \tau}{\partial \lambda_i} = \frac{-\frac{\partial \rho}{\partial \lambda_i}}{\left(\frac{\partial \rho}{\partial E}\frac{\partial s}{\partial \tau} - 1\right)} \tag{A.13}$$

while total differentiation of the former condition $\tau = \rho(E, \lambda_i)$ yields

$$\frac{dE}{d\lambda_i} = -\frac{\frac{\partial \rho}{\partial \lambda_i}}{\frac{\partial \rho}{\partial E}} \Longrightarrow \text{sign}\left(\frac{dE}{d\lambda_i}\right) = \text{sign}\left(\frac{\partial \rho}{\partial \lambda_i}\right) = \text{sign}\left(\widetilde{L}_i - 1\right), \tag{A.14}$$

where, from the main text, demand is downward-sloping $\partial\rho/\partial E < 0$ and the overlapping policy affects the inverse allowance demand curve according to $\operatorname{sign}(\partial\rho/\partial\lambda_i) = \operatorname{sign}(\widetilde{L}_i - 1)$.

Extending A.13 with $\left(\frac{\partial\rho}{\partial E}\right)^{-1}$ using (A.14), and substituting the slope of the allowance demand curve $\frac{\partial E}{\partial\tau}$ for the inverse of the slope of the inverse allowance demand curve $(\partial\rho/\partial E)^{-1}$ then yields (6) and the conclusion that $\operatorname{sign}(\partial\tau/\partial\lambda_i) = \operatorname{sign}(\widetilde{L}_i - 1)$.

Independence of timing of allowance demand shifts

Consider the two-period model with a price-based flexibility mechanism with the three equilibrium conditions $\rho_1(E_1, \lambda_i) = \tau_1$ from (9), $\rho_2(E_2, \lambda_i) = (1+r)\tau_1$ from (10), and market clearing in the allowance market $E_1 + E_2 = s(\tau_1)$ from (11).

Application of Cramer's rule to these conditions (9)-(11) yields:

$$\frac{\partial\tau_1}{\partial\lambda_i} = \frac{\frac{\partial\rho_2}{\partial E_2}\frac{\partial\rho_1}{\partial\lambda_i} + \frac{\partial\rho_1}{\partial E_1}\frac{\partial\rho_2}{\partial\lambda_i}}{\frac{\partial\rho_2}{\partial E_2} + (1+r)\frac{\partial\rho_1}{\partial E_1} - \frac{\partial\rho_1}{\partial E_1}\frac{\partial\rho_2}{\partial E_2}\frac{\partial s}{\partial\tau_1}}. \tag{A.15}$$

Total differentiation of $\tau_1 = \rho_1(E_1, \lambda_i)$ and $(1+r)\tau_1 = \rho_2(E_2, \lambda_i)$ (see (9) and (10)) yields

$$\frac{dE_t}{d\lambda_i} = -\frac{\frac{\partial\rho_t}{\partial\lambda_i}}{\frac{\partial\rho_t}{\partial E_t}}. \tag{A.16}$$

for both $t = 1, 2$. Cancelling $-\frac{\partial\rho_1}{\partial E_1}\frac{\partial\rho_2}{\partial E_2}$ from (A.15), using (A.16), and substituting the slope of the allowance demand curve $\frac{\partial E_t}{\partial\tau_t}$ for the inverse of the slope of the inverse allowance demand curve $(\partial\rho_t/\partial E_t)^{-1}$ in the denominator and using $\frac{\partial E_2}{\partial\tau_1} = \frac{1+r}{1+r}\frac{\partial E_2}{\partial\tau_1} = (1+r)\frac{\partial E_2}{\partial\tau_2}$ yields:

$$\frac{\partial\tau_1}{\partial\lambda_i} = \frac{\frac{dE_1}{d\lambda_i} + \frac{dE_2}{d\lambda_i}}{\left(\frac{\partial s_1}{\partial\tau_1} + \frac{\partial s_2}{\partial\tau_1}\right) - \left(\frac{\partial E_1}{\partial\tau_1} + \frac{\partial E_2}{\partial\tau_1}\right)} = \frac{\frac{dE}{d\lambda_i}}{\frac{\partial s}{\partial\tau_1} - \frac{\partial E}{\partial\tau_1}}.$$

where $\partial E/\partial\tau_1 = \frac{\partial E_1}{\partial\tau_1} + \frac{\partial E_2}{\partial\tau_1} < 0$ is the slope of the aggregate allowance demand curve across the two periods. Adjustments in total equilibrium emissions E^* are independent of how a given total shift in demand induced by the overlapping policy is spread over time:

$$\frac{dE^*}{d\lambda_i} = \frac{\partial s}{\partial\tau_1}\frac{\partial\tau_1}{\partial\lambda_i} = \frac{dE}{d\lambda_i}\frac{\frac{\partial s}{\partial\tau_1}}{\frac{\partial s}{\partial\tau_1} - \frac{\partial E}{\partial\tau_1}} = \frac{dE}{d\lambda_i}\frac{\omega^S}{(\omega^S - \omega^D)} \tag{A.17}$$

where $\omega^D < 0$ and $\omega^S \geq 0$ are the long-run elasticities of allowance demand and supply.

Alternatively, one can express (A.17) in terms of per-period elasticities as follows:

$$\frac{dE^*}{d\lambda_i} = \frac{\frac{dE_1}{d\lambda_i} + \frac{dE_2}{d\lambda_i}}{\left(\frac{\partial s_1}{\partial\tau_1} + \frac{\partial s_2}{\partial\tau_1}\right) - \left(\frac{\partial E_1}{\partial\tau_1} + \frac{\partial E_2}{\partial\tau_1}\right)} = \frac{dE}{d\lambda_i}\left(\frac{\omega_1^S s_1 + \omega_2^S s_2}{\omega_1^S s_1 + \omega_2^S s_2 - (\omega_1^D E_1 + \omega_2^D E_2)}\right).$$

D.2. Waterbed effect for quantity-based flexibility mechanisms

Derivation of Equation (16)

Application of Cramer's rule to conditions (13)-(15) yields:

$$\frac{\partial E_1^*}{\partial \lambda_i} = \frac{\frac{\partial \rho_2}{\partial \lambda_i} - (1+r)\frac{\partial \rho_1}{\partial \lambda_i}}{(1+r)\frac{\partial \rho_1}{\partial E_1} - \frac{\partial \rho_2}{\partial E_2}\left(1+\frac{\partial s_2}{\partial b}\right)}.$$

Cancelling $-\frac{\partial \rho_1}{\partial E_1}\frac{\partial \rho_2}{\partial E_2}$, using (A.14), and substituting the slope of the allowance demand curve $\frac{\partial E_t}{\partial \tau_t}$ for the inverse of the slope of the inverse allowance demand curve $(\partial \rho_t/\partial E_t)^{-1}$ yields:

$$\frac{\partial E_1^*}{\partial \lambda_i} = -\frac{\frac{dE_2}{d\lambda_i}\frac{\partial E_1}{\partial \tau_1} - (1+r)\frac{dE_1}{d\lambda_i}\frac{\partial E_2}{\partial \tau_2}}{(1+r)\frac{\partial E_2}{\partial \tau_2} - \frac{\partial E_1}{\partial \tau_1}\left(1+\frac{\partial s_2}{\partial b}\right)}.$$

Using $dE/d\lambda_i = dE_1/d\lambda_i + dE_2/d\lambda_i$, $\frac{\partial E}{\partial \tau_1} = \frac{\partial E_1}{\partial \tau_1} + \frac{\partial E_2}{\partial \tau_1}$ and $\frac{\partial E_2}{\partial \tau_1} = \frac{1+r}{1+r}\frac{\partial E_2}{\partial \tau_1} = (1+r)\frac{\partial E_2}{\partial \tau_2}$ yields

$$\frac{\partial E_1^*}{\partial \lambda_i} = -\frac{\frac{dE}{d\lambda_i}\frac{\partial E}{\partial \tau_1} - \frac{dE_1}{d\lambda_i}\frac{\partial E_2}{\partial \tau_2}}{\frac{\partial E}{\partial \tau_1} + \frac{\partial E_1}{\partial \tau_1}\frac{\partial s_2}{\partial b}}.$$

Factoring out $dE/d\lambda_i$ in the numerator and dividing both numerator and denominator by $\partial E/\partial \tau_1$ and using our definitions $\xi = \frac{\partial E_1}{\partial \tau_1}/\frac{\partial E}{\partial \tau_1}$ and $\beta = \frac{dE_1}{d\lambda_i}/\frac{dE}{d\lambda_i}$ yields (16).

Proof of Proposition 4

The waterbed effect is defined as $W = 1 - (dE^*/d\lambda_i)/(dE/d\lambda_i)$. Using the expressions for $dE/d\lambda_i$ from (16) and for $dE^*/d\lambda_i$ from (17) yields

$$W = 1 - \frac{\frac{\partial s_2}{\partial b}}{1+\frac{\partial s_2}{\partial b}\xi}(\xi - \beta),$$

and some rearranging then yields (18) as claimed.

Proof of Corollary 1

We proceed in three steps. First, we use conditions (13)-(15) and Cramer's rule to compute:

$$\frac{\partial \tau_1}{\partial \lambda_i} = -\frac{\frac{\partial \rho_1}{\partial \lambda_i}\frac{\partial \rho_2}{\partial E_2}\left(1+\frac{\partial s_2}{\partial b}\right) + \frac{\partial \rho_1}{\partial E_1}\frac{\partial \rho_2}{\partial \lambda_i}}{(1+r)\frac{\partial \rho_1}{\partial E_1} - \frac{\partial \rho_2}{\partial E_2}\left(1+\frac{\partial s_2}{\partial b}\right)}.$$

Cancelling $-\frac{\partial \rho_1}{\partial E_1}\frac{\partial \rho_2}{\partial E_2}$, using (A.14), and substituting the slope of the allowance demand curve $\frac{\partial E_t}{\partial \tau_t}$ for the inverse of the slope of the inverse allowance demand curve $(\partial \rho_t/\partial E_t)^{-1}$ yields:

$$\frac{\partial \tau_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{dE_1}{d\lambda_i}\left(1+\frac{\partial s_2}{\partial b}\right)+\frac{dE_2}{d\lambda_i}}{(1+r)\frac{\partial E_2}{\partial \tau_2}-\frac{\partial E_1}{\partial \tau_1}\left(1+\frac{\partial s_2}{\partial b}\right)}.$$

Using $dE/d\lambda_i = dE_1/d\lambda_i + dE_2/d\lambda_i$, $\frac{\partial E}{\partial \tau_1} = \frac{\partial E_1}{\partial \tau_1} + \frac{\partial E_2}{\partial \tau_1}$ and $\frac{\partial E_2}{\partial \tau_1} = \frac{1+r}{1+r}\frac{\partial E_2}{\partial \tau_1} = (1+r)\frac{\partial E_2}{\partial \tau_2}$ yields

$$\frac{\partial \tau_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{dE}{d\lambda_i}+\frac{dE_1}{d\lambda_i}\frac{\partial s_2}{\partial b}}{\frac{\partial E}{\partial \tau_1}+\frac{\partial E_1}{\partial \tau_1}\frac{\partial s_2}{\partial b}}. \tag{A.18}$$

Second, recalling that $\xi = \frac{\partial E_1}{\partial \tau_1}/\frac{\partial E}{\partial \tau_1}$ and $\frac{ds_2}{d\lambda_i} = -\frac{\partial s_2}{\partial b}\frac{\partial E_1^*}{\partial \lambda_i}$ from the main text, we derive the equilibrium expansion path by relating changes in the equilibrium allowance supply to changes in the equilibrium allowance price that are induced by the shift in total allowance demand:

$$\left.\frac{ds}{d\tau_1}\right|_{\text{equilibrium}} \quad = \quad \frac{\frac{ds_2}{d\lambda_i}}{\frac{\partial \tau_1}{\partial \lambda_i}} = \frac{\frac{\partial s_2}{\partial b}}{1+\frac{\partial s_2}{\partial b}\xi}\cdot[\xi-\beta]\cdot(-1)\frac{\frac{\partial E}{\partial \tau_1}\left[1+\frac{\partial s_2}{\partial b}\xi\right]}{\left[1+\frac{\partial s_2}{\partial b}\beta\right]},$$

and some rearranging yields (19) as claimed.

Third, it remains to be shown that Propositions 3 and 4 are, in fact, equivalent. While $W \in [0,1]$ holds for weakly upward-sloping allowance supply curves, this is no longer guaranteed once we substitute in (19)—values below 0 and above 1 are both now possible. We thus now convert (8) from Proposition 3 into (18) from Proposition 4. Substituting $\omega^S = \frac{ds}{d\tau_1}(\beta)\frac{\tau_1}{s(\tau_1|\beta)}$ into (8) and using (19) yields:

$$W \quad = \quad \frac{-\frac{\partial E}{\partial \tau_1}\frac{\tau_1}{E}}{\frac{\partial E}{\partial \tau_1}\frac{\beta-\xi}{\beta+\left(\frac{\partial s_2}{\partial b}\right)^{-1}}\frac{\tau_1}{s(\tau_1|\beta)}-\frac{\partial E}{\partial \tau_1}\frac{\tau_1}{E}}.$$

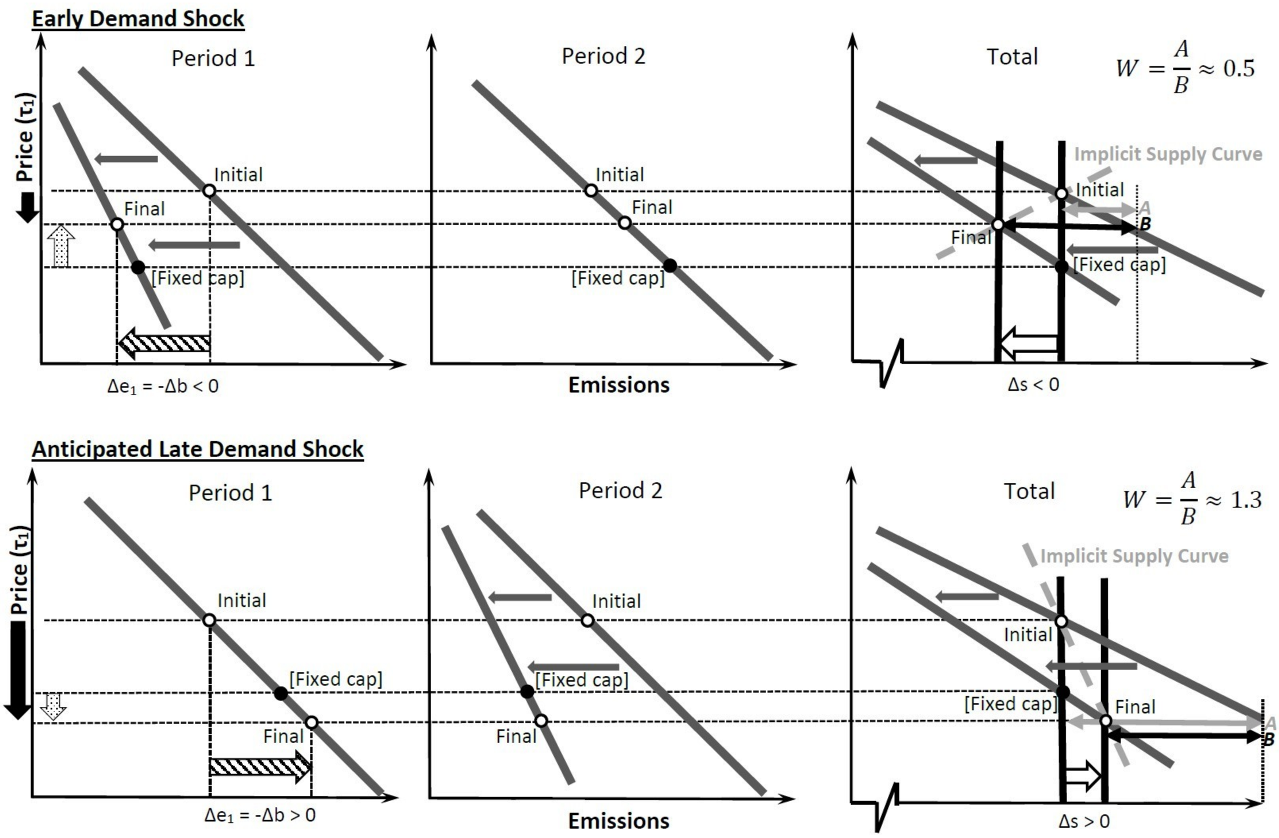
Now, noting that $s(\tau_1|\beta) = E$ in equilibrium, cancelling $-\frac{\partial E}{\partial \tau}\frac{\tau_1}{E}$ from the above equation and expanding it with $1+\frac{\partial s_2}{\partial b}\beta$ gives:

$$W \quad = \quad \frac{1+\frac{\partial s_2}{\partial b}\beta}{\frac{\partial s_2}{\partial b}\cdot[\xi-\beta]+1+\frac{\partial s_2}{\partial b}\beta},$$

and some simplifying then yields (18) as claimed.

Figure A.1 illustrates Proposition 4 and Corollary 1 graphically. The direct impact of the overlapping policy shifts the demand curve for allowances. The indirect effect

Figure A.1: The timing of overlapping policies, waterbed effects and implicit supply curves



Notes: An identical shift in total allowance demand ($\Delta E(\tau_1) = dE/d\lambda_i$, grey single-headed arrows) induced by an overlapping policy occurs either in period 1 (upper panel) or period 2 (lower panel). Direction of change in first-period equilibrium emissions and bank ($\Delta E_1 = -\Delta b$, fasciated arrow, left figures) and hence supply response (Δs, white arrow, right figures) depend on timing of policy. Implicit supply curve (dashed light grey line, right figures) represents response to a continuum of demand shifts of the same timing. Total demand and implicit supply curve jointly determine the waterbed effect W in line with the tax incidence literature (double-headed arrows A and B; representing $\Delta E - \Delta E^*$ and ΔE, respectively). Price response to policy (black single-headed arrow, left figures) is reduced (upper panel) or increased (lower panel) relative to a fixed cap (dotted arrow, left figures).

is represented by movements along a given allowance demand curve and mediated by changes in prices. These two effects jointly drive first-period equilibrium emissions, and their interaction directly determines the direction of the supply adjustment.

D.3. Robustness of results for a more responsive banking-based flexibility mechanism

Here we check the robustness of our findings to a more responsive banking-based flexibility mechanism for which $\frac{\partial s_2}{\partial b} \in (-\frac{1}{\xi}, -1)$, where $\xi \equiv \frac{\partial E_1}{\partial \tau_1} / \frac{\partial E}{\partial \tau_1} \in (0,1)$—in contrast to the mechanism with $\frac{\partial s_2}{\partial b} \in [-1, 0]$ from the main text. In words, the more responsive mechanism has the feature that, for every additional allowance banked, total supply is reduced by *more* than one allowance.[44]

It is easy to verify that the two central formulae (16) and (17) from the main text remain valid. Proposition 4 changes only slightly in that the waterbed of a front-loaded overlapping policy is now strictly smaller than 1 and that $(-\frac{\partial s_2}{\partial b})^{-1}$ is now strictly larger than 1.

Proposition 7. The waterbed effect for an anticipated overlapping policy under a flexibility mechanism based on allowance banking is:

$$W = \frac{1 + \frac{\partial s_2}{\partial b}\beta}{1 + \frac{\partial s_2}{\partial b}\xi} \tag{A.19}$$

so an overlapping policy that (i) is front-loaded, i.e., effective only in period 1 (with $\beta = 1$), has a punctured waterbed $W \in [0, 1)$; (ii) is sufficiently back-loaded (with $\beta \leq \xi$) has $W > 1$; (iii) reduces allowance demand in period 1 but increases it sufficiently strongly in period 2 according to $\beta \geq (-\frac{\partial s_2}{\partial b})^{-1} > 1$ has $W < 0$.

Corollary 1 remains valid—while the higher responsiveness of the flexibility mechanism also has an impact on the slope of the implicit allowance supply curve. Again, the implicit allowance supply curve has a negative slope whenever $W \notin [0, 1]$; the latter is a function of the mechanism's responsiveness and the range of β values for which the slope of supply becomes negative now extends to cases where the waterbed effect lies strictly below 100%.

D.4. Robustness of results with non-marginal policies

We have so far focused on marginal overlapping policies, with $d\lambda_i > 0$, that shift emissions demand at fixed carbon prices by a small amount, dE. This yields a waterbed effect

[44] We do not consider flexibility mechanisms that are even more responsive, i.e., with $\frac{\partial s_2}{\partial b} \leq -\frac{1}{\xi}$, as these would induce infinitely sensitive responses and other highly problematic effects.

$W = 1 - \frac{dE^*}{dE}$ that can be seen as an approximation to a non-marginal rate $W = 1 - \frac{\triangle E^*}{\triangle E}$. More generally, we now consider an overlapping policy that tightens from an initial level $\underline{\lambda}_i \geq 0$ to a new level $\overline{\lambda}_i$ where $\triangle\lambda_i \equiv (\overline{\lambda}_i - \underline{\lambda}_i)$ is a discrete change.

Here we show that Propositions 3 and 4 extend to non-marginal changes in policy. To see why, write the non-marginal waterbed effect as:

$$W = 1 - \frac{\triangle E^*}{\triangle E} = 1 - \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i} + \frac{dE^*}{d\tau}\frac{d\tau}{d\lambda_i}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\right)d\lambda_i} = -\frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE^*}{d\tau}\frac{d\tau}{d\lambda_i}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\right)d\lambda_i}. \tag{A.20}$$

Plugging (6) into (A.20) and using $\omega_t^S = (\partial s_t/\partial\tau_1)(\tau_1/s_t)$ and $\omega_t^D = (\partial E_t/\partial\tau_1)(\tau_1/E_t)$ we get the non-marginal version of (8) in Proposition 3:

$$W = \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\frac{-\omega_t^D}{(\omega_t^S-\omega_t^D)}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\right)d\lambda_i} \in (0,1]$$

Since the slope of the allowances supply curve is assumed to be everywhere weakly positive and that of the cumulative allowance demand curve to be strictly negative, we still have that $W \in (0,1]$ and so Proposition 3 extends to non-marginal changes in overlapping policies. (For non-marginal changes, W depends on $dE/d\lambda_i$ and hence on de_i and $\widetilde{L}_i$ as they determine the size of the change in long-run demand, via $dE/d\lambda_i = [1-\widetilde{L}_i](de_i/d\lambda_i)$.

For quantity-based flexibility mechanisms, we plug (A.18) into (A.20) and simplify:

$$W = \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\frac{1+\frac{\partial s_2}{\partial b}\beta}{1+\frac{\partial s_2}{\partial b}\xi}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{dE}{d\lambda_i}\right)d\lambda_i}$$

If the non-marginal policy change does not affect the demand for allowances in period 2, i.e., if $\beta = 1$ for all $\lambda_i \in [\underline{\lambda}_i, \overline{\lambda}_i]$, then the waterbed effect is weakly smaller than 100% confirming that part (i) of Proposition 4 extends to non-marginal changes in overlapping policies. If the non-marginal policy change does not affect the demand for allowances in period 1, i.e., if $dE_1/d\lambda_i = 0$ for all $\lambda_i \in [\underline{\lambda}_i, \overline{\lambda}_i]$, then the waterbed effect is weakly larger than 100%, i.e. part (ii) of Proposition 4 applies to non-marginal policies, too. By using a continuity argument, part (iii) has to extend to non-marginal changes as well. If $(\partial s_2/\partial b)\beta < 0$ for all $\lambda_i \in [\underline{\lambda}_i, \overline{\lambda}_i]$, then the waterbed effect is negative. Hence, part (iv) also applies to non-marginal policies.

The results from the marginal analysis are identical to the non-marginal analysis if allowance demand and allowance supply adjustments (either in prices or in banks) are all linear; then $\partial E_1/\partial\tau_1$, $\partial E/\partial\tau_1$, $\partial s/\partial\tau_1$ and $\partial s_2/\partial b$ are all constants.

Appendix E: The reformed post-2018 EU ETS

E.1: Multi-period waterbed effect in the post-2018 EU ETS

We here present details on the operation of the post-2018 EU ETS and its Market Stability Reserve (MSR), focusing on subtleties originating from the timing of overlapping policies in a setting with a multi-period waterbed effect.

We show how to compute the sensitivity of the cumulative cap to changes in the size of the bank, $\partial s/\partial b$ (see Section 4.2), when an overlapping policy impacts the bank in more than one period. To this end, we derive the *instantaneous* waterbed effect $\hat{W}_t$ used in Section 5.2 and understand how it varies with policies timing.

The MSR works as follows.[45] If the bank, known as the "total number of allowances in circulation" (TNAC) in the legal language of the EU ETS, exceeds 833 million at the end of a given year (in 2017 or later), then the number of allowances auctioned in the 12 months following September of the next year is reduced by a certain percentage (the "intake rate") of the size of the bank shown in Table A.2. Allowances withheld are placed in the MSR and released in installments of 100 million/year once the bank has dropped below 400 million. We label $t_{B=833}$ the year in which the bank drops below the 833 million threshold and the MSR hence stops taking in allowances.

Year (if bank > 833 million on Dec. 31^{st})	Intake rate (2018 Reform) (%)
2017	16
2018 to 2022	24
2022 to $t_{B=1,096}$	12
$t_{B=1,096}$ to $t_{B=833}$	12

Table A.2: Intake rates for the EU ETS Market Stability Reserve (MSR)

The maximum number of allowances held in the MSR is limited to the number auctioned in the previous year.[46] Allowances in excess of this upper bound are permanently cancelled. At the end of 2021 the MSR contained 2.6 billion allowances (European Commission, 2022). The annually decreasing cancellation threshold in 2023 will be below 1

[45]The MSR rules are in European Parliament and Council (2018) and summarized by Perino (2018). In April 2023 a further reform of the EU ETS—including a revision of the rules of the MSR—was passed (Borghesi et al., 2023). It increases the marginal intake rate to 24% until 2030 and to 100% in any year in which the TNAC is between 833 million and 1,096 million. So the instantaneous waterbed effect drops to zero for all preceding years if this range is ever hit. While our calculations are based on the 2018 EU ETS reform, all these subsequent amendments can be represented within our current framework.

[46]The target share of auctioning in Phase 4 is 57% (European Parliament and Council, 2018) with the remaining allowances being freely allocated. The current reform proposal simplifies the cancellation threshold such that all allowances held in the MSR in excess of 400 million allowances will be cancelled.

billion. Hence, any additional allowance drawn into the MSR, e.g., due to an overlapping policy, is eventually cancelled.

Computing the multi-period waterbed effect (see Appendix A.1) for the EU ETS faces several challenges. First, the MSR's intake rate changes over time (Table A.2). Second, the MSR is active over multiple periods so the cumulative effect of an early shift in allowance demand depends on its impact on the TNAC in all periods up to $t_{B=833}$. Third, $t_{B=833}$ (and also $t_{B=1{,}096}$) is itself determined by market outcomes and hence potentially by the overlapping policy itself. Fourth, the price-mediated "Rosendahl effect" (see Section 4.2) of anticipated future changes in allowance demand depends on the same dynamics. These complexities mean that W can only be estimated by numerical simulation.[47]

We derive an "instantaneous waterbed effect cum Rosendahl effect" $\hat{W}^R(t_a, t, t_{B=833})$ that captures the first two and the last of the four complexities: the MSR's time-varying intake rate and its multi-period nature. It represents the extent of the waterbed effect for a one-off shift in allowance demand in period t that is announced in period $t_a \leq t$ taking into account all future intake rates up to the point when intake by the MSR stops ($t_{B=833} \geq t$) as they jointly determine the share of the demand shift that translates into supply adjustments. The instantaneous waterbed effect $\hat{W}$ from Section 5.2 is a special case of $\hat{W}^R(t_a, t, t_{B=833})$ that drops the fourth complexity again.

As a first step, we derive the sensitivity of the cumulative cap to changes in the bank, $\partial s/\partial b$, as an explicit function of time. An instantaneous change in the number of banked allowances triggers a sequence of transfers to the MSR.[48] This implies that adding one allowance to the bank in year t and with the bank dropping below 833 million allowances in year $t_{B=833}$, the effective sensitivity of the cumulative cap in the EU ETS under the rules of the 2018 reform is given by:

$$\begin{aligned} \frac{\partial}{\partial b} s\left(t, t_{B=833}\right) &= -1 + (1-.16)^{\max[0,\min[2018,t_{B=833}]-\max[2017,t]]} \\ &\quad \times (1-.24)^{\max[0,\min[2023,t_{B=833}]-\max[2018,t]]} \\ &\quad \times (1-.12)^{\max[0,\max[2023,t_{B=833}]-\max[2023,t]]}. \end{aligned} \tag{A.21}$$

This expression follows directly from the numbers in Table A.2 (see also Perino (2018)).[49]

The instantaneous waterbed effect cum Rosendahl effect $\hat{W}^R(t_a, t, t_{B=833})$ results from

[47] See Bruninx and Ovaere (2022); Gerlagh et al. (2021) for simulation results and Rosendahl (2019); Perino (2019) for informal discussions.

[48] A share ν_t of the increase in the bank is transferred in the first year, the remainder $(1-\nu_t)$ adds to the bank in the following year and again induces a transfer at rate ν_{t+1}, i.e., $(1-\nu_t)\nu_{t+1}$, and so on.

[49] The proposed 2023 reform includes a provision that changes the (marginal) intake rate to 100% whenever the TNAC is between 833 million and 1,096 million allowances. If this is the case in at least one year, then the effective sensitivity jumps to $\partial s/\partial b = -1$ in all previous years (Perino et al., 2022). The EU ETS response to overlapping policies would hence mimic that of a carbon tax until the TNAC drops below 833 million when the waterbed effect jumps back to 100%.

plugging the above into (18) taking into account that the sensitivity of the cumulative cap in the numerator is triggered by the demand shift of the overlapping policy directly, i.e. it starts at point t but the response in the denominator is mediated by price changes and starts at t_a. Moreover, we here have that $\beta = \hat{\beta}_t = 1$ while $\hat{\beta}_{-t} = 0$, so that $\hat{\beta}_t$ is essentially a dummy variable that marks the period t for which the instantaneous waterbed effect is computed (by construction, we here look at a demand shift in period t in isolation). Hence:

$$\hat{W}^R\left(t_a, t, t_{B=833}\right) = \frac{1 + \frac{\partial}{\partial b} s\left(t, t_{B=833}\right)}{1 + \frac{\partial}{\partial b} s\left(t_a, t_{B=833}\right) \frac{\frac{\partial E_{t_a}}{\partial \tau_{t_a}}}{\frac{\partial E}{\partial \tau_{t_a}}}}. \tag{A.22}$$

This highlights the triple importance of timing: the year an overlapping policy is announced, t_a, the year it shifts allowance demand, t, and the year the carbon-pricing system stops responding to market outcomes, $t_{B=833}$, jointly determine the size of the instantaneous waterbed effect. (Note that this still ignores the endogeneity of $t_{B=833}$.)

In the main part of the paper, we fix $t_a = 2020$ and $t_{B=833} = 2030$. Moreover, as in Perino (2018), we abstract from the Rosendahl effect, i.e. we ignore the price change induced by the overlapping policy and set the denominator of A.22 equal to 1, and hence:

$$\hat{W}_t \equiv \hat{W}\left(t, 2030\right)|_{\tau \text{ fixed}} = 1 + \frac{\partial}{\partial b} s\left(t, 2030\right). \tag{A.23}$$

The total waterbed effect of an overlapping policy in a multi-period setting announced in period one (as in (18 in the main text) can be decomposed into a linear combination of T instantaneous waterbed effects cum Rosendahl effect, i.e. $W(\beta_1, ..., \beta_T, t_{B=833}) \equiv \sum_{t=1}^{T} \beta_t \hat{W}^R(1, t, t_{B=833})$, where $\beta_t \equiv \frac{dE_t^*/d\lambda_i}{dE^*/d\lambda_i}$ represents the share of the total demand shift due to the overlapping policy that occurs in period t, so that $\sum_t^T \beta_t = 1$.

(While $\hat{\beta}_t$ is a dummy for the period for which the instantaneous waterbed effect is assessed, β_t captures how shifts in total allowance demand are distributed across time.)

E.2: Sensitivity to exogenous changes in $t_{B=833}$ and the Rosendahl effect

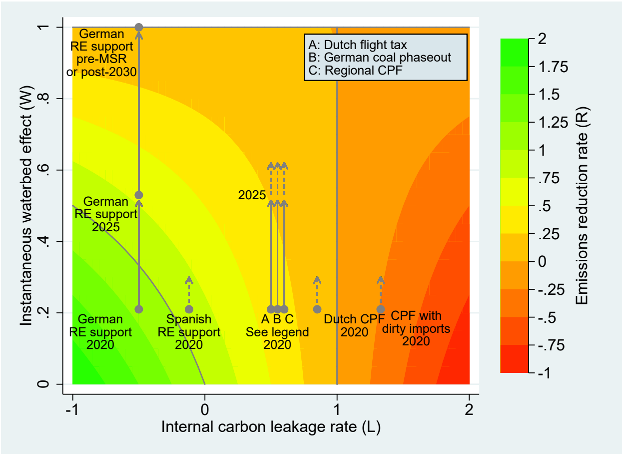
In Section 5.2, we assume the MSR will stop taking in allowances in 2030 ($t_{B=833} = 2030$) (Figure A1, Panel (a)). In Figure A1, Panel (b), we investigate how the effective emissions reduction rate changes when we assume $t_{B=833} = 2048$ (following Gerlagh et al. (2021)). Panel (c) shows the performance of two key policies—renewable energy support and a coal phase-out in Germany—when we consider the instantaneous waterbed effect without holding carbon prices fixed and thus allowing for the Rosendahl effect ($\hat{W}_t^R$, see Equation

(A.22)). We use Gerlagh et al. (2021)'s estimates of the Rosendahl effect but note that estimates in the literature differ and this is a highly active area of research.

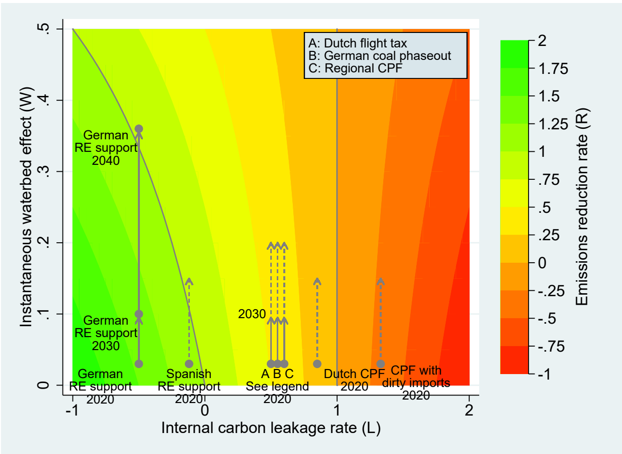
Panel (b) of Figure A1 shows that the instantaneous waterbed effect decreases substantially when $t_{B=833}$ lies further in the future. The waterbed effect can only fall below 100% if the MSR takes in allowances; if allowances still flow into the MSR in the 2030s and 2040s, then $\hat{W}_t < 1$ for many more years over which policies operate. In Panel (a), $\hat{W}_{2030} = 1$; in Panel (b), $\hat{W}_{2030}$ falls by an order of magnitude.

Panel (c) compares $\hat{W}_t$ holding carbon prices fixed (grey arrows and dots) with endogenous allowance prices (black arrows and dots). A black dot should be interpreted as a policy announced in 2020 but expected to reduce the demand for emissions allowances in year $t \geq 2020$. The Rosendahl effect increases $\hat{W}_t^R$ substantially compared to $\hat{W}_t$, especially for years close to $t_{B=833}$. Until the mid-2030s, the waterbed effect is still relatively limited (below 0.5) but in or after the year 2048, the waterbed effect is *larger* than 1. This is consistent with Proposition 4 and highlights the potential unintended consequences of announcing policies that reduce emissions demand far into the future.

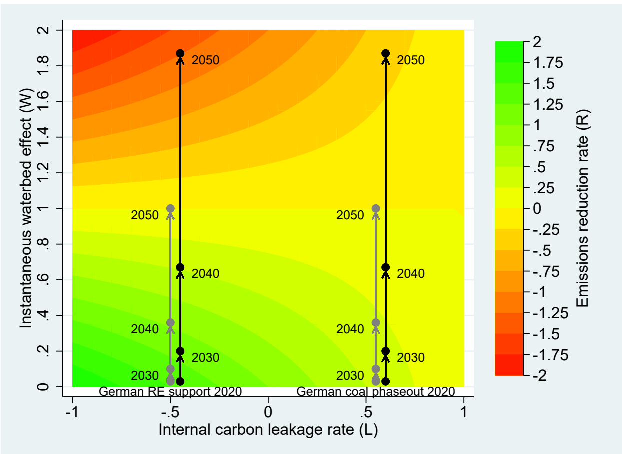
Figure A1: Leakage and waterbed effects in the EU ETS under varying assumptions



(a) $t_{B=833} = 2030$; no Rosendahl effect



(b) $t_{B=833} = 2048$; no Rosendahl effect



(c) $t_{B=833} = 2048$; with Rosendahl effect

Notes: Panel (a) presents Figure 1 excluding policies outside the EU ETS. Panel (b) plots the same policies assuming $t_{B=833} = 2048$ instead of $t_{B=833} = 2030$. Panel (c) adds the Rosendahl effect as estimated in Gerlagh et al. (2021), together with their estimate of $t_{B=833} = 2048$.