Overlapping Climate Policies*

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

July 2022

Abstract

The world is under pressure to deliver on the 2015 Paris Climate Agreement. Individual jurisdictions are enacting policies such as phasing out coal, taxing aviation, and supporting renewable energy. These 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 such "overlapping climate policies" can help combat climate change—depending on their design, location and timing. Some policies are truly complementary while others backfire by raising aggregate emissions. At a conceptual level, our model encompasses most carbon-pricing systems used in practice and a wide range of popular unilateral policies.

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

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

*We thank Jan Abrell, Severin Borenstein, Dallas Burtraw, Jim Bushnell, Meredith Fowlie, Don Fullerton, Reyer Gerlagh, Marten Ovaere, Sebastian Rausch, Mar Reguant, Knut Einar Rosendahl, Herman Vollebergh, Maximilian Willner and Yuting Yang for helpful comments and suggestions. 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, the Toulouse School of Economics, the SWEEEP seminar at ZEW Mannheim, Florence School of Regulation at the European University Institute, and the University of Pennsylvania. 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, 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 (EPRG) 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

The world is under increasing pressure to combat climate change and deliver on the ambition of the 2015 Paris Climate Agreement. Over 60 national and sub-national jurisdictions are putting a price on carbon emissions (World Bank, 2021)—such as the European Union's Emissions Trading System (EU ETS) that spans 30 countries and covers power generation, industrial sectors and domestic aviation; the Regional Greenhouse Gas Initiative (RGGI) for power generation across eleven states in the northeastern US; and China's national emissions trading system that began operating in 2021.

Moreover, individual jurisdictions—countries and states—are pursuing a plethora of other climate policies: supporting renewable energy, phasing out coal-fired power, levying extra carbon taxes on air travel, and so on. A notable example is the UK's Carbon Price Support which from 2013 to 2020 added a unilateral carbon fee of £18/tCO_2 to the allowance price faced by its power generators under the 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).

These examples share a common feature: they are policies enacted by an individual jurisdiction for an individual sector, like electricity or aviation, that operate alongside a multi-jurisdictional carbon-pricing system like the EU ETS or RGGI. In this paper, we refer to these as "overlapping climate policies" and ask a simple question: what is the climate benefit of an overlapping policy? As it is a global public good, any mitigation of climate change will be driven solely by the policy's impact on aggregate emissions.

For a textbook cap-and-trade system with a fixed emissions cap (Montgomery, 1972; Baumol and Oates, 1988), the answer is clear: if an overlapping policy reduces aggregate allowance demand from power generation by 1 ton of CO_2 (for a given system-wide carbon price), then this is—in equilibrium—precisely offset by increased demand of 1 tCO_2 as the system's carbon price adjusts downwards. So the "waterbed effect" in the carbon market is 100% (Fankhauser et al., 2010; Böhringer, 2014) and the policy has no climate benefit.[1] By contrast, a simple carbon tax does not cap emissions so there is no waterbed effect.

Real-world carbon markets are more complicated as they now often involve hybrid designs that combine elements of both price and quantity regulation. North American carbon markets like RGGI use price floors and ceilings as a flexibility mechanism to contain the variability of the allowance price. Since its 2018 reform, the EU ETS features a complex mechanism—the Market Stability Reserve—that cancels allowances under cer-

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

tain market circumstances—and thereby "punctures" the waterbed (Perino, 2018). This enables overlapping policies to reduce aggregate emissions.

We derive a new 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 encompasses price-based 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; Burtraw et al., 2020; Karp and Traeger, 2021) and quantity-based flexibility mechanisms based on allowance banking (Perino, 2018; Gerlagh et al., 2021).

Yet this carbon-market reasoning still has a missing link in the product market which we term "internal carbon leakage". Suppose that a unilateral carbon price on power generation by an individual jurisdiction reduces its domestic emissions demand by 1 tCO_2 but that, within an integrated electricity market, this leads to an increase in its electricity imports which in turn raises emissions demand by 1 tCO_2 in other jurisdictions within the ETS (for a given system-wide carbon price). This overlapping policy has no climate benefit either: its rate of internal carbon leakage is 100%. Hence, an overlapping policy's climate benefit is driven by a combination of the waterbed effect and internal leakage.

We show, more generally, that "supply-side" policies that unilaterally raise the carbon price or directly limit emissions-intensive production have positive internal leakage—sometimes in excess of 100%—as they raise emissions demand in other jurisdictions that "fill the gap" due to lower domestic production.[2] By contrast, "demand-side" policies that reduce the demand for emissions-intensive production, e.g., by promoting renewables or energy efficiency, have negative internal leakage as they displace imported emissions. While some recent empirical work has estimated internal leakage for specific overlapping policies (Vollebergh, 2018; Abrell et al., 2019a; Gerarden et al., 2020), we provide new theoretical insight into its economics across a range of popular policies.[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)

[2]Our use of the term "supply-side" policy differs from the literature which focuses on the market for fossil resources (Sinn, 2008; Harstad, 2012)—our reference point is the market for goods produced by a polluting industry. We discuss broader connections with this strand of research in the conclusion.

[3]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.

it has received much less policy attention, despite likely being more important than its external cousin for sectors such as airlines and electricity.[4]

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 "local" emissions reduction an overlapping policy achieves to its equilibrium impact on aggregate emissions. Section 3 presents a theory of internal carbon leakage that quantifies an overlapping policy's net impact on aggregate emissions demand. Section 4 presents a theory of the waterbed effect that endogenises the system's carbon price and unifies price- and quantity-based flexibility mechanisms. Section 5 derives a multi-period waterbed effect for the reformed EU ETS.

Section 6 illustrates the empirical usefulness of the framework with examples of overlapping policies 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 (see especially Figure 2). Section 7 concludes and suggests future research avenues.

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. The introduction of flexibility mechanisms in cap-and-trade systems has, in part, been motivated by a desire to make overlapping policies more climate effective. For example, in designing the EU ETS's Market Stability Reserve, the European Union noted that "the reserve will also enhance synergy with other climate and energy policies" (European Parliament and Council, 2015)—thus alluding to what are often termed "complementary" policies.

Our results highlight greater subtlety: with a punctured waterbed, demand-side overlapping policies such as renewables support can indeed be truly complementary in that they induce further emissions reductions across the system. Yet, while the UK's Carbon Price Support was very successful at reducing domestic emissions from power generation (Abrell et al., 2022), supply-side overlapping policies are never complementary—and those with high internal carbon leakage can backfire by raising aggregate emissions.

In sum, while we do not attempt to quantify policy cost-effectiveness or broader welfare impacts, our results show how the ability of an overlapping policy to help combat climate change varies enormously depending on its design, location and timing.[5]

[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] We also abstract from other motivations for overlapping policies that may also be important to cost-benefit analysis and policy practice, including those related to distributional goals, local economic development, and the long-term dynamics associated with technological change and other market failures such as innovation externalities (Newbery et al., 2019).

2 Conceptual framework

We begin by setting out a conceptual framework that encompasses a wide range of carbon-market designs and delineates internal carbon leakage and the waterbed effect.

2.1 Internal carbon leakage and the waterbed effect

Consider a multi-jurisdiction—e.g., multi-country—carbon-pricing system that covers multiple sectors, like the EU ETS. (A single-sector cap-and-trade 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. Our main interest is in policies enacted by an individual jurisdiction for an individual sector, like electricity or aviation. For example, jurisdiction i but not other jurisdictions, denoted by j, may have a policy to phase out coal-fired power generation. We assume that, holding fixed the system-wide carbon price τ, the overlapping policy is successful at reducing i's domestic demand for emissions, $\Delta e_i < 0$.

Our main question is, what is the overlapping policy's impact on aggregate equilibrium emissions $\Delta e \equiv \Delta e_i + \Delta e_j$ across all jurisdictions with an equilibrium carbon price? This is the critical issue for combating climate change.

Our framework answers this question using two concepts. First, internal carbon leakage in the product market captures emissions displacement for a given system-wide carbon price. We define the rate of internal carbon leakage associated with i's policy as:

$$L_i \equiv -\Delta e_j / \Delta e_i \text{ (fixed } \tau\text{)}, \tag{1}$$

where Δe_j is the change induced by i's policy in the emissions demand of other jurisdictions j that are part of the carbon-pricing system.[6] For example, this reflects how a coal phase-out by i may induce higher emissions from power generation in j due to its impact on electricity prices—even for a fixed τ. Therefore $\Delta e \equiv [1 - L_i]\Delta e_i$ represents the (net) system-wide change in emissions demand due to the policy for a fixed τ.[7]

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

$$W \equiv 1 - \Delta e / \Delta e \text{ (endogenous } \tau\text{)}. \tag{2}$$

[6] Notice that this is akin to the standard definition of external carbon leakage (e.g., IPCC, 2007) that relates to shifting of emissions to jurisdictions *outside* the system.

[7] Another form of carbon leakage that is outside of our scope occurs when, in the same jurisdiction, some sectors are not covered by the carbon-pricing system (Baylis et al., 2013; Jarke and Perino, 2017); an example is leakage from covered EU ETS sectors like electricity to uncovered sectors such as transport.

This translates the system-wide change in emissions demand due to i's policy into an equilibrium change in emissions (Böhringer and Rosendahl, 2022; Eichner and Pethig, 2019; Osorio et al., 2020).[8] For example, the waterbed effect captures how a coal phase-out may induce a change in the system's carbon price that induces further emissions changes in the electricity sector—and in any other sectors covered by the carbon-pricing system. 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 hybrid carbon-market designs 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:

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

$$\Delta e \;\; = [1 - L_i][1 - W]\Delta e_i. \tag{3}$$

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 $R_i \equiv [1 - L_i][1 - W]$, we can think of policies for which leakage and waterbed effects are such that $R_i \geq 1$ as complementary (or super-additive) policies while those for which $R_i < 1$ 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 in the remainder of the paper to proceed sequentially. First, we derive the rate of internal carbon leakage L_i between different jurisdictions in the same sector for a range of overlapping policies: this maps Δe_i to Δe. Second, given the policy's induced change to aggregate emissions demand, we derive the extent of the waterbed effect W under different carbon-market designs: this maps Δe to Δe .

2.2 Alternative concepts of internal carbon leakage

While other emissions decompositions are possible, we propose Lemma 1 as the simplest and analytically clearest framework. First, some empirical papers estimate the "meta" version of internal leakage $L_i^{\mathsf{M}} = 1 - R_i$ that bundles our L_i with the waterbed effect. A key point is that this obscures two distinct phenomena that should be clearly delineated—partly because of their different policy implications. For example, if $L_i^{\mathsf{M}} = 1$ because

[8] Like other literature, we address the impact of an individual overlapping policy on aggregate emissions *given* the rules of the wider carbon-pricing system. We acknowledge that overlapping policies might, over time, also affect the future design and stringency of the system via the political process.

[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 the equilibrium aggregate impact Δe^* of a given policy-driven local impact Δe_i.

$W = 1$ then no overlapping policy can combat climate change—thus helping explain the EU's 2018 ETS reform; by contrast, if $L_i^{\mathsf{M}} = 1$ because $L_i = 1$ then this as such reveals only the limitations of a single policy.

Second, internal leakage could instead be defined in a "total" form that incorporates any induced change to the system-wide carbon price, $L_i^{\mathsf{T}} \equiv -\Delta e_j / \Delta e_i$ (endogenous τ). For cap-and-trade market designs, $\Delta\tau \neq 0$ so there can be a wedge between L_i^{T} and L_i.

	Fixed carbon price	**Equilibrium carbon price**
ETS sector targeted by overlapping policy	A	B
Any other ETS sectors	zero	C

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

Table 1 clarifies the rationale for our approach by decomposing the impact on equilibrium emissions of i's policy into three parts. First, A captures our notion of internal carbon leakage in the sector targeted by the overlapping policy for a fixed carbon price: $A = [1 - L_i]\Delta e_i$. Second, $B + C$ reflects the waterbed effect and the system-wide adjustment with an endogenous carbon price: $B + C = -W[1 - L_i]\Delta e_i$. Overall this corresponds to Lemma 1 in that $A + B + C = [1 - L_i][1 - W]\Delta e_i$. Note especially that the waterbed effect captures the overlapping policy's indirect impact on *other* sectors—the C.[10]

The concept of "total" internal leakage L_i^{T}, by contrast, captures only the impacts $A + B$ that occur in the sector of the overlapping policy itself. We conclude that an equivalent result to Lemma 1 based on L_i^{T} is necessarily more complex and less intuitive; it would require the waterbed effect W as well as a third term that adjusts for the "missing C"—so it would also no longer be possible to proceed sequentially as in our analysis.[11]

For a *single-sector* ETS like power generation in RGGI, our framework and an approach based on L_i^{T} become equivalent—precisely because then $C \equiv 0$. In this special case, $A + B = [1 - L_i^{\mathsf{T}}]\Delta e_i = [1 - L_i][1 - W]\Delta e_i$ so now also $L_i^{\mathsf{T}} = L_i^{\mathsf{M}} = 1 - R_i$. That is, an analysis based on L_i^{T} is then just as correct as ours—but is still subject to the first critique that it obscures the different policy implications, e.g., of $W = 1$ versus $L_i = 1$.

In sum, we therefore think that Lemma 1 offers the clearest way to conceptualise overlapping climate policies.

[10]For simplicity, Table 1, assumes that the overlapping policy targets a single sector within an ETS; this is consistent with real-world policies aimed at electricity generation or air travel. It also abstracts from *direct* impacts of the overlapping policy on other sectors within the carbon-pricing system (for a fixed τ) that are not targeted by the policy; these are affected only indirectly as captured by the waterbed effect. (For example, an additional carbon fee on electricity generation in i may lead to internal leakage to j in steel production which uses electricity as an input.) This is consistent with prior empirical work on internal carbon leakage, as discussed further in Section 6.

[11]In particular, $C = \left([1 - L_i][1 - W] - [1 - L_i^{\mathsf{T}}]\right)\Delta e_i$ showing the complexity of the third term.

3 A model of internal carbon leakage

We next present a theory of internal carbon leakage at the sectoral level. We consider two groups of overlapping climate policies. First, "supply-side" policies that unilaterally raise the carbon price for emissions-intensive production or directly reduce it as in a coal phase-out. Second, "demand-side" policies that reduce the (residual) demand for emissions-intensive production, e.g., by promoting renewables or energy efficiency.

We derive intuitive formulae for the equilibrium rate of internal carbon leakage L_i. We show that the economics of internal carbon leakage is similar for policies within each group but differs markedly between the two groups: it is always positive—and can exceed 100%—for the former group but is negative for the latter.

3.1 Model setup

We begin with a very simple model to make our main points. A representative firm in jurisdiction k produces output x_k $(k = i, j)$. Firm k's emissions are $e_k = e_k^0 - a_k$ where a_k is abatement and $e_k^0 = \theta_k x_k$ is baseline emissions with an emissions intensity θ_k.

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 case in which firms' cost functions are separable, $G_k(x_k, a_k) = C_k(x_k) + \phi_k(a_k)$.[12] This will deliver easily interpretable leakage formulae. 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 $\phi_k''(a_k) > 0$.

It will be useful to have a metric for the extent of abatement opportunity for firm k. We can also think of k's cost function in terms of output and emissions, $G_k(x_k, e_k) \equiv [C_k(x_k) + \phi_k(\theta_k x_k - e_k)]$ and define the following:

$$A_k \equiv \left(1 - \frac{G_k^{xe}}{G_k^{xx}} \frac{G_k^{ex}}{G_k^{ee}}\right) = \frac{C_k''}{[C_k'' + \theta_k^2 \phi_k'']} \in [0, 1). \tag{4}$$

The limiting case with $A_k \to 1$ corresponds to abatement costs becoming linear ($\phi_k'' \to 0$) so emissions can be reduced without resorting to production cuts. The case in which $A_k \to 0$ corresponds to a Leontief technology: emissions are proportional to output, as abatement is infeasible ($\phi_k'' \to \infty$) so emissions are at their baseline, $e_k = \theta_k x_k$.[13,14]

[12] 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 (CCS) and the purchase of carbon offsets.

[13] The analysis can easily accommodate a fixed cost of abatement, $\phi_k(0) > 0$. If this fixed cost makes abatement unprofitable, then $a_k = 0$, and the following results apply with $A_k = 0$. If instead abatement occurs despite the fixed cost, then the following results apply with the corresponding value of $A_k > 0$.

[14] The stability condition $G_k^{xx} G_k^{ee} - G_k^{xe} G_k^{ex} > 0$ is equivalent to $A_k < 1$. Firm k's cost function satisfies the following standard assumptions. Written in terms of output and emissions, it increases in output, $G_k^x(x_k, e_k) = C_k' + \theta_k \phi_k' > 0$, decreases in emissions, $G_k^e(x_k, e_k) = -\phi_k' < 0$, and is convex in both output

The firms face a demand function $p(X)$ for their product, where $X \equiv x_i + x_j$ is total output. We interpret demand $p(X)$ 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. We assume perfect competition in the product market with the objective to establish a first set of benchmark results that apply across a range of overlapping policies.

Firm k faces a carbon price τ_k on each unit of emissions, which depends on the carbon price τ that is common to both jurisdictions as part of a wider carbon-pricing system. As per our conceptual framework of Lemma 1, τ is held fixed in this analysis.

To maximise profits, firm k solves $\max_{x_k;a_k} \Pi_k = px_k - G_k(x_k, a_k) - \tau_k e_k$. Note it is equivalent for a firm to choose its emissions or abatement. With perfect competition, the two first-order conditions for profit-maximisation are:

$$p = C'_k(x_k) + \theta_k \phi'_k(a_k) \text{ and } \tau_k = \phi'_k(a_k). \tag{5}$$

The product price equals the firm's total marginal cost of output, and the carbon price equals the marginal abatement cost. Putting these together yields a combined condition:

$$p = C'_k(x_k) + \tau_k \theta_k \tag{6}$$

so the product price is equal to marginal cost plus per-unit carbon costs based on its baseline emissions intensity of output.[15] Due to cost separability, the extent of abatement does not affect the product-market outcome. The abatement incentive rises with the domestic carbon price, $da_k/d\tau_k = 1/\phi''_k(\cdot) > 0$ which, in turn, is independent of output.

Our main interest is the rate of internal carbon leakage for different kinds of overlapping policy by jurisdiction i, denoted as λ_i. These reduce i's domestic emissions, $de_i/d\lambda_i < 0$, but may also induce a change in j's emissions. To obtain simple formulae, we focus on a "marginal" policy change, for which internal carbon leakage is given by $L_i = (-de_j/d\lambda_i)/(de_i/d\lambda_i)$. In the benchmark case without abatement, $L_i = (\theta_j/\theta_i)(-dx_j/d\lambda_i)/(dx_i/d\lambda_i)$, where the first term is jurisdictions' "relative dirtiness" and the second term is output leakage $L_i^O \equiv (-dx_j/dx_i)$.

Some equilibrium definitions will prove useful to cast our formulae in familiar terms. First, let $\varepsilon^D \equiv -p(\cdot)/Xp'(\cdot) > 0$ be the price elasticity of demand. Second, let $\sigma_k \equiv x_k/X \in (0, 1)$ be the market share of jurisdiction k's firm (so $\sigma_i + \sigma_j = 1$). Third, let $\widehat{C}'_k(x_k) \equiv [C'_k(x_k) + \tau_k\theta_k]$ be k's total marginal cost of output and define $\eta_k^S \equiv$

and emissions, with $G_k^{xx}(x_k, e_k) = C''_k + \theta_k^2\phi''_k > 0$ and $G_k^{ee}(x_k, e_k) = \phi''_k > 0$.

[15] To guarantee an interior solution for outputs, assume $p(0) > \max_k\{C'_k(0) + \tau_k\theta_k\}$.

$x_k\widehat{C}''_k(x_k)/\widehat{C}'_k(x_k) > 0$ as its elasticity, also noting that $\widehat{C}''_k(x_k) \equiv C''_k(x_k)$. By k's first-order condition, $x'_k(p) = 1/C''_k(x_k) > 0$, i.e., its supply curve is upward-sloping. So $\varepsilon^S_k \equiv px'_k(p)/x_k(p) > 0$ is k's price elasticity of supply and, at the optimum, $\eta^S_k = 1/\varepsilon^S_k$.[16]

3.2 Supply-side overlapping climate policies

We begin with two "supply-side" policies that unilaterally raise the carbon price for emissions-intensive production or directly reduce production, e.g., via 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 $\frac{d}{d\tau}\tau_i(\tau, \lambda_i), \frac{d}{d\lambda_i}\tau_i(\tau, \lambda_i) > 0$. 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). Another possibility is a policy that lifts i's carbon price towards a higher target level $\widehat{\tau}_i$, say with $\tau_i = \tau + \lambda_i(\widehat{\tau}_i - \tau)$ where $\lambda_i \in [0, 1)$. Firm j continues to be subject 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.

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

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

where $\gamma \in \{0, 1\}$ equals zero (one) for a unilateral reduction in carbon-intensive production (unilateral carbon price), and $Z_i \equiv \frac{A_i}{(1-A_i)}\left(1 + \frac{(1-\sigma_j)\varepsilon^S_i/\varepsilon^S_j}{(\sigma_j + \varepsilon^D/\varepsilon^S_j)}\right) \geq 0$ is an abatement effect.

Proposition 1 provides a simple formula to quantify internal carbon leakage. For both supply-side policies, carbon leakage is always positive as the underlying output leakage

[16] These expressions are all evaluated at the output levels of the initial equilibrium.

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. While output leakage is always less than 100%—as i's policy raises the product market price—carbon leakage can exceed 100% if j's firm is sufficiently dirtier, that is, θ_j/θ_i is sufficiently large.

To understand the result, consider the unilateral cut in carbon-intensive production ($\gamma = 0$). The comparative statics are intuitive: output leakage $L_i^O = \sigma_j/(\sigma_j + \varepsilon^D/\varepsilon_j^S)$ 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 the induced abatement A_i. Abatement 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. With near-costless additional abatement, carbon leakage tends to zero, $L_i \to 0$ as $Z_i \to \infty$ ($A_i \to 1$ as $\phi_i''(\cdot) \to 0$). 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 (Newbery et al., 2019): this combines greater market share than single-country action and thereby contains internal leakage.

To illustrate, suppose that the demand elasticity $\varepsilon^D = .5$ and that j has market share $\sigma_j = 20\%$ with a supply-responsiveness $\eta_j^S = .2 \Leftrightarrow \varepsilon_j^S = 5$. With identical emissions intensities $\theta_i = \theta_j$ and no abatement ($A_i = 0$ or $\gamma = 0$), $L_i = 67\%$ is driven by output leakage. If instead j's technology is less responsive with $\eta_j^S = 1 \Leftrightarrow \varepsilon_j^S = 1$ or demand is more elastic with $\varepsilon^D = 2.5$, 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 $A_i = .25$, this yields $L_i = 60\%$, illustrating how abatement can help bring forth an aggregate emissions reduction.

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 production 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 of (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 has internal carbon leakage to jurisdiction j of:*

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

Internal carbon leakage is always negative: j's firm is now directly affected by the policy and responds by also cutting output and emissions. This means that the aggregate emissions reduction here is more pronounced than the local reduction. Akin to Proposition 1, 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 carbon price remains fixed, 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 $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i)$. To first order, this is also true 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 than 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.

In sum, we can already conclude that supply-side policies are necessarily sub-additive with $R_i < 1$ (Lemma 1) for *any* waterbed effect $W \in [0, 1]$ while demand-side policies may be complementary $R_i > 1$ as long as the waterbed is not too pronounced. Propositions 1 and 2 are also consistent with empirical estimates, respectively, of positive internal leakage within the EU ETS from a Dutch unilateral carbon price floor on electricity generation as well as negative internal leakage from German and Spanish renewable support programs—which we discuss in more detail in Section 6.

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

4 A model of the waterbed effect

We now turn to the second building block of our conceptual framework: the waterbed effect $W = 1 - \Delta e \ /\Delta e$, for which the carbon price τ is now derived endogenously. Again we represent the overlapping policy by λ_i and focus on a "marginal" policy change so the waterbed effect is $W = 1 - (de \ /d\lambda_i)/(de/d\lambda_i)$.

We consider a stylised model of an allowance market. By design, the allowance market is geographically blind. We assume *inverse* aggregate demand function for allowances $\rho(e, \lambda_i)$, where e are aggregate emissions and $\partial\rho/\partial e < 0$ and $\partial\rho/\partial\lambda_i \leq 0$.

We first analyse how an anticipated shift in allowance demand affects total emissions and the equilibrium price of allowances when the carbon-pricing system features a (weakly) increasing allowance supply function. We connect its waterbed effect to basic economic principles on pass-through from the classic literature on tax incidence.

Then we consider a two-period model of a market design where a *cumulative* emissions cap is adjusted based on banked allowances—as in the EU ETS since its 2018 reform (Perino, 2018)—and where the overlapping policy is perfectly anticipated by all market participants. We conclude with a unifying result that shows how such a (dynamic) quantity-based flexibility mechanism can be represented by a (static) supply function, thus for the first time allowing for a direct comparison to a price-based system.

4.1 Flexibility mechanisms based on allowances prices

Most real-world carbon-pricing designs—such as the California-Québec system, RGGI (Burtraw et al., 2020), the German ETS for fossil fuels outside the EU ETS, and the new UK ETS—feature flexibility mechanisms based on allowance prices. These programs as well as carbon taxes and cap-and-trade systems with a fixed cap can be represented by a weakly upward-sloping aggregate allowance supply given by $s(\tau)$ with $\partial s/\partial \tau \geq 0$. A cap-and-trade system with a fixed cap (such as the pre-2018 EU ETS) 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 the North American, UK and German cap-and-trade systems, are represented by $s(\tau)$ being perfectly price elastic at a particular $\bar{\tau}$. This mechanism captures all strictly upward-sloping allowance supply curves, e.g., those that trace the marginal damage curve (Roberts and Spence, 1976).

The equilibrium conditions of this carbon-market design that jointly determine e and τ are as follows:

$$\rho(e, \lambda_i) - \tau = 0 \tag{7}$$

$$e - s(\tau) = 0, \tag{8}$$

where condition (7) balances marginal costs of abatement with the carbon price while equation (8) is the market-clearing condition for the allowance market. These conditions yield the overlapping policy's impact on the equilibrium carbon price (see Appendix B.1):

$$\frac{\partial \tau}{\partial \lambda_i} = \frac{\frac{de}{d\lambda_i}}{\frac{\partial s}{\partial \tau} - \frac{\partial e}{\partial \tau}} > 0 \tag{9}$$

where $\partial e/\partial \tau < 0$ is the slope of the aggregate allowance demand curve. The change in the equilibrium allowance price is key to the waterbed effect. Adjustments in total equilibrium emissions e are "spatially blind", i.e., independent of how the overlapping policy is spread over space:

$$\frac{de}{d\lambda_i} = \frac{\partial s}{\partial \tau}\frac{\partial \tau}{\partial \lambda_i} = \frac{de}{d\lambda_i}\frac{\frac{\partial s}{\partial \tau}}{\frac{\partial s}{\partial \tau} - \frac{\partial e}{\partial \tau}} = \frac{de}{d\lambda_i}\frac{\kappa^S}{\kappa^S - \kappa^D} \tag{10}$$

where $\kappa^D = \frac{\partial e}{\partial \tau}\frac{\tau}{e} < 0$ and $\kappa^S = \frac{\partial s}{\partial \tau}\frac{\tau}{s} \geq 0$ are the elasticities of allowance demand and supply.[17]

Proposition 3 *The waterbed effect for a marginal policy overlapping a carbon-pricing system with a (weakly) increasing allowance supply and strictly decreasing allowance demand is:*

$$W = \frac{-\frac{\partial e}{\partial \tau}}{\frac{\partial s}{\partial \tau} - \frac{\partial e}{\partial \tau}} = \frac{\kappa^D}{\kappa^S - \kappa^D} \in [0, 1], \tag{11}$$

[17] Note that equation 10 also holds if the elasticity of allowance supply is negative—as will become relevant in the next subsection.

which is independent of the specifics of the overlapping policy (λ_i) and leakage rates (L_i).

Proposition 3 shows that the waterbed effect of marginal overlapping policies depends only on elasticities of total allowance demand and supply—and is independent of the type of overlapping policy, its geographical impacts, and its internal carbon leakage.

We thus uncover a natural connection between the waterbed effect and classic principles on tax incidence under perfect competition (Jenkin, 1872; Weyl and Fabinger, 2013). In particular, note that Proposition 3 corresponds to the rate of cost pass-through from this literature. Since the allowance supply is assumed to be (weakly) monotonically increasing, it mimics a supply curve. The drop of producer prices in response to a tax-induced shift in inverse demand in the tax-incidence literature exactly mimics the impact of an overlapping policy on the carbon price in a carbon market with a weakly upward-sloping allowance supply curve.

Equation (11) has at opposite ends a zero waterbed for a carbon tax ($\partial s/\partial\tau \to \infty$) and a 100% waterbed effect under a plain cap-and-trade system ($\partial s/\partial\tau = \kappa^S = 0$). For marginal changes, i.e., policies inducing relatively small shifts in allowance demand, this conclusion applies also to step-wise allowance supply functions featured in the California-Québec system, RGGI, the German 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). If the probability that the equilibrium is in any of the horizontal sections of the allowance supply curve is π, then $E(W) = 1 - \pi$. 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 B.3, we extend Proposition 3 more formally to non-marginal policies.[18]

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 (generalising Proposition 3) is also "temporally blind". Consider a two-period version of the above model with *cumulative* allowance supply given by $s(\tau) = s_1(\tau_1) + s_(\tau_2)$, where $\frac{\partial s_t}{\partial \tau_t} \geq 0$ for $t = 1, 2$. Given the interest rate r, intertemporal arbitrage implies that

[18]For non-marginal changes, the size of the demand shift (Δe) matters as it determines the extent of the 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 the internal leakage rate L_i (as $\Delta e = (1 - L_i)\Delta e_i$). Nevertheless, the analytical separation of 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.

carbon prices across the two periods satisfy $\tau_2 = (1+r)\tau_1$.

The three equilibrium conditions of this dynamic carbon market with a price-based flexibility mechanism are:

$$\rho_1(e_1, \lambda_i) - \tau_1 = 0 \quad (12)$$

$$\rho_2(e_2, \lambda_i) - (1+r)\tau_1 = 0 \quad (13)$$

$$e_1 + e_2 - s(\tau_1) = 0, \quad (14)$$

where period-specific emissions e_1 and e_2 and the carbon price τ_1 in period 1 are endogenous variables. Proposition 3 extends to this dynamic setting, simply by reinterpreting the price-elasticities for demand and supply as long-run elasticities capturing the cumulative effect over the two periods (see Appendix B.1):

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 equation 11.*

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 $s_1(\tau_1) + s_(\tau_2)$ that yield the same $s(\tau)$ 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 both temporary and permanent overlapping policies in a simple static setting is therefore straightforward and—with respect to the waterbed effect—without loss of generality.[19]

4.2 Flexibility mechanisms based on allowance banking

With the 2018 reform of the EU ETS, namely the introduction of cancellations within the Market Stability Reserve, a new form of flexibility mechanism gained prominence.[20] Here we present a stylised two-period version of such a mechanism. For the remainder of this section, indices represent time periods $t = 1, 2$. 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

[19]We develop a multi-period version of our overall conceptual framework in Section 6.

[20]We note that RGGI also used to have a banking-based design element; in particular, 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). Within the context of the present paper, this can be represented by $\partial s_2/\partial b = -1$.

earlier periods, where $b = s_1 - e_1$ is banking at the end of period 1 and $\partial s_2/\partial b \in [-1, 0]$.[21] A plain cap-and-trade system is again nested as a special case ($\partial s_2/\partial b = 0$).

The allowance market is assumed to feature perfect intertemporal arbitrage where banking and borrowing constraints do not bind. Hence, allowance prices in the two periods are directly linked by $\tau_2 = (1 + r)\tau_1$. Any overlapping policy is announced at the beginning of period 1 and hence perfectly anticipated by market participants.

Analogous to the dynamic price-based flexibility mechanism, the equilibrium conditions are:

$$
\begin{aligned}
\rho_1(e_1, \lambda_i) - \tau_1 &= 0 && (15)\\
\rho_2(e_2, \lambda_i) - (1 + r)\tau_1 &= 0 && (16)\\
e_1 + e_2 - s_1 - s_2(s_1 - e_1) &= 0. && (17)
\end{aligned}
$$

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

$$
\frac{\partial e_1}{\partial \lambda_i} = -\frac{\frac{de}{d\lambda_i}}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \cdot \left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}} - \frac{\frac{de_1}{d\lambda_i}}{\frac{de}{d\lambda_i}} \right]. \qquad (18)
$$

Given the impact of the overlapping policy on total allowance demand $de/d\lambda_i = de_1/d\lambda_i + de_2/d\lambda_i$, the direction of the policy's impact on equilibrium emissions in period 1 (e_1) depends on the relative size of the two terms in brackets in equation (18). Both have an intuitive economic interpretation. The first term is the ratio of the slopes of first-period and total allowance demand. It captures how much of the total change in emissions induced by the price response materialises in period 1. The second term is the percentage of the shift in the total demand curve occurring in the first period. Shifting the allowance demand curve to the left in period 1 ($de_1/d\lambda_i$), 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 (Perino, 2018) and the indirect price-mediated effect (Rosendahl, 2019) are hence antagonistic (see Figure 1).

Whether an overlapping policy increases or decreases first-period emissions in equilibrium depends on the timing of its impacts. If the policy is front-loaded in that most of the shift in allowance demand occurs early on, then first-period emissions decrease. By contrast, a policy that mainly affects allowance demand in the future raises first-period

[21] Restricting $\partial s_2/\partial b \geq -1$ is somewhat arbitrary as one could imagine systems with more responsive rules. However, imposing this lower bound simplifies the analysis and includes the entire range of values relevant for both the EU ETS and RGGI. For details see Section 5 below.

emissions. In terms of short and long-run elasticities, the term in brackets is more likely to be positive the smaller the difference between the price elasticity of short-run (κ_1^D) and total (κ^D) allowance demand.

This dependence on timing directly carries over to the change in total equilibrium emissions e via adjustment of the cumulative cap, where the banking of allowances $b = s_1 - e_1$ mirrors the change in first-period emissions as the first-period cap s_1 is fixed:

$$\frac{de}{d\lambda_i} = \frac{ds_2}{d\lambda_i} = \frac{\partial s_2}{\partial b}\frac{\partial b}{\partial e_1}\frac{\partial e_1}{\partial \lambda_i} = \frac{\frac{de}{d\lambda_i}\frac{\partial s_2}{\partial b}}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \cdot \left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}} - \frac{\frac{de_1}{d\lambda_i}}{\frac{de}{d\lambda_i}}\right]. \tag{19}$$

Policies that mainly reduce allowance demand early on reduce the cumulative cap as firms respond to the shift in the first-period demand curve by emitting less and banking more. The increase in the bank induces additional reductions in allowance supply. 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 bank results in a smaller reduction in the cap compared to the reference point without the anticipated demand reduction induced by the policy.

Proposition 4 *The waterbed effect* $W = 1 - (de_1/d\lambda_i + de_2/d\lambda_i)/(de_1/d\lambda_i + de_2/d\lambda_i)$ *for a marginal, anticipated policy overlapping a carbon-pricing system with a flexibility mechanism based on allowance banking is:*

$$W = \frac{1 + \frac{\partial s_2}{\partial b}\frac{\frac{de_1}{d\lambda_i}}{\frac{de}{d\lambda_i}}}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \tag{20}$$

where the numerator captures the direct impact of the overlapping policy and the denominator captures the indirect effect mediated through the price response. Defining $\beta = (de_1/d\lambda_i)/(de/d\lambda_i)$ *as the share of the shift in total allowance demand that occurs in period 1,*[22] *the following holds:*

(i) An overlapping policy effective only in period 1 ($\beta = 1$) has a waterbed effect weakly smaller than 1 (because $\partial s_2/\partial b \in [-1, 0]$ and $(\partial e_1/\partial \tau_1)/(\partial e/\partial \tau_1) \in (0, 1)$)

$$W = \frac{1 + \frac{\partial s_2}{\partial b}}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \in [0, 1]. \tag{21}$$

[22] The timing of domestic impacts of an overlapping policy $\beta_i = de_{i1}/(de_{i1} + de_{i2})$ and period-specific rates of internal leakage L_{it} determine β as follows: $\beta = \frac{\beta_i(1-L_{i1})}{\beta_i(1-L_{i1})+(1-\beta_i)(1-L_{i2})}$.

(ii) An overlapping policy effective only in period 2 ($\beta = 0$) has a waterbed effect weakly larger than 1 (because $\partial s_2/\partial b \in [-1, 0]$ and $(\partial e_1/\partial\tau_1)/(\partial e/\partial\tau_1) \in (0, 1)$)

$$W = \frac{1}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial\tau_1}}{\frac{\partial e}{\partial\tau_1}}} \geq 1. \tag{22}$$

(iii) For any given change in total allowance demand ($de/d\lambda_i$) < 0, there exists a threshold $\underline{\beta} = (\partial e_1/\partial\tau_1)/(\partial e/\partial\tau_1) \in (0, 1)$ for which $W = 1$. For all $\beta < \underline{\beta}$, $W > 1$ and vice versa. The larger the difference between short-run and cumulative responsiveness (price elasticity) of allowance demand, the lower $\underline{\beta}$.

(iv) An overlapping policy features a negative waterbed effect, $W < 0$, if it reduces aggregate allowance demand in period 1 and across both periods but sufficiently increases it in period 2 ($\beta > 1, de/d\lambda_i < 0, de_1/d\lambda_i < 0, de_2/d\lambda_i > 0$) according to:

$$\beta > \bar{\beta} = -\frac{1}{\frac{\partial s_2}{\partial b}} \geq 1. \tag{23}$$

In sum, there are three regimes for the waterbed effect:

$$W = \begin{cases} < 0 & \text{if} \quad \beta > \bar{\beta} \geq 1 \\ \in [0, 1] & \text{if} \quad \underline{\beta} \leq \beta \leq \bar{\beta} \\ > 1 & \text{if} \quad \beta < \underline{\beta} \in (0, 1) \end{cases}.$$

The special cases presented in parts (i) and (ii) highlight the direct and the price-mediated indirect effect on cumulative emissions. For policies affecting aggregate demand early on (see Equation (21)), the price effect is always of second order and hence cumulative emissions decrease. However, policies affecting aggregate demand only in the far future (Equation (22)) such as an anticipated coal phase-out have no immediate emissions-demand impact—and hence the price-driven effect dominates. Anticipation of a future reduction in relative scarcity leads to lower carbon-prices in both periods. First-period emissions increase which in combination with the fixed cap in period 1 induces a drop in the bank. This in turn increases the cumulative cap and thus increases aggregate emissions.[23] While the Market Stability Reserve in the EU ETS technically can only cancel allowances, canceling fewer than without the impact of the overlapping policy reflects an increase in the cap relative to the reference point. An example for a policy featuring a negative waterbed effect, as in case (iv) is an amendment of a previously enacted coal phase-out plan that

[23] This effect was first described by Rosendahl (2019) and confirmed by Bruninx and Ovaere (2022) and Gerlagh et al. (2021).

shuts down old inefficient plants earlier but grants new, highly-efficient plants a longer grace period.

In sum, for a given quantity-based flexibility mechanism ($\partial s_2/\partial b$) and given market characteristics ($\rho_t(e_t, \lambda_i)$), an overlapping policy with a given impact on total allowance demand ($de/d\lambda_i$), can increase total emissions ($W > 1$), leave them unaffected ($W = 1$), decrease them ($W < 1$), and even decrease them by more than the initial shift in aggregate demand ($W < 0$)—all driven exclusively by its timing (β). These findings contrast markedly with a price-based flexibility mechanism.

4.3 Linking price- and banking-based flexibility mechanisms

We next show that—despite these differences between price- and quantity-based flexibility mechanisms—there is an economically intuitive link between Propositions 3 and 4.

While the cap adjustment in the flexibility mechanisms based on banking does not explicitly refer to prices, we can construct an "equilibrium expansion path" for overlapping policies of different stringency but identical timing of impacts that mimics an effective supply curve for allowances. Hence, the time-sensitive waterbed effect under a banking-based flexibility mechanism can be represented in a static supply-demand framework; this facilitates intuition and communication of the effects arising. For the purpose of deriving the effective supply curve implied by the banking-based mechanism, we denote the present value of the allowance price path by $\tau = \tau_1 = \tau_2/(1+r)$.

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

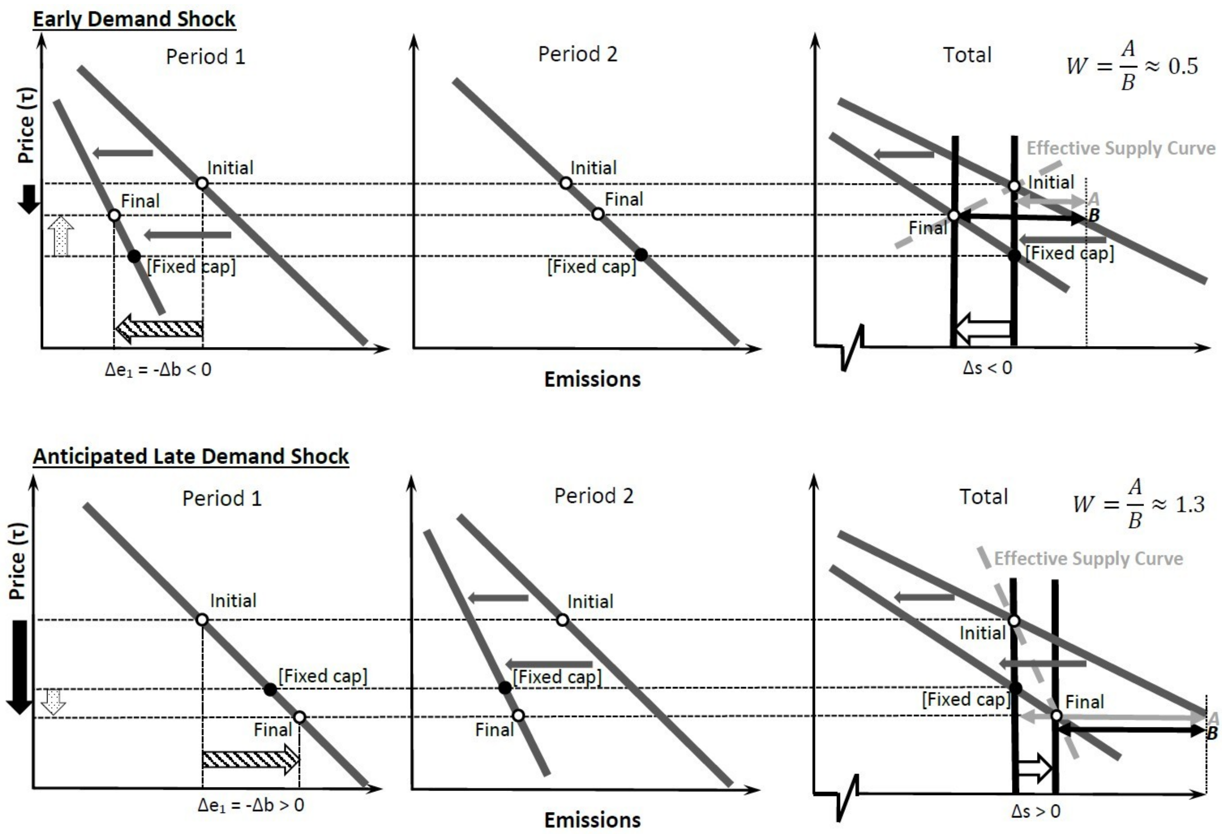
$$\left.\frac{ds}{d\tau}\right|_{equilibrium} = \frac{\frac{ds_2}{d\lambda_i}}{\frac{\partial\tau}{\partial\lambda_i}} = \frac{\frac{\partial s_2}{\partial b}\frac{\partial e}{\partial\tau}}{1+\frac{\partial s_2}{\partial b}\beta}\cdot\left[\beta-\frac{\frac{\partial e_1}{\partial\tau}}{\frac{\partial e}{\partial\tau}}\right]. \tag{24}$$

The equilibrium expansion path (24) is highly instrument-specific. In contrast to the allowance supply function of a price-based flexibility mechanism, it is not a defining feature of the carbon-pricing system but specific to the overlapping policy under consideration. (See Appendix B.2 for a proof: Plugging equation (24) into equation (11) yields the same W as using (20).) Effective supply curves are strictly downward-sloping whenever the waterbed effect of the policy is either above 100% or negative.[24]

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

[24]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 effective supply curve does not induce multiplicity of equilibria given that $\partial s_2/\partial b \geq -1$.

Figure 1: The timing of overlapping policies, waterbed effects and effective supply curves



Notes: An identical shift in total allowance demand ($\Delta e(\tau) = de/d\lambda_i$, 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. Effective supply curve (dashed light grey line, right figures) represents response to a continuum of demand shifts of the same timing. Total demand and effective 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).

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.

The decomposition of Figure 1 is new to the literature, as is the connection between the waterbed effect under price- and banking-based flexibility mechanisms identified by Corollary 1. This 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, 2021a, p. 6) contributions had to resort to ad-hoc assumptions.

More broadly, transforming the supply response of a banking-based flexibility mechanism to a demand-side shock into an allowance supply function allows to use textbook economics to build intuition for the impact of overlapping policies (Gerlagh et al., 2021) on carbon markets—but also of technological change (Bruninx et al., 2020), pandemics (Bruninx and Ovaere, 2022; Gerlagh et al., 2020), stimulus packages (Bruninx and Ovaere, 2022), business cycles (Kollenberg and Taschini, 2016) and, last but not least, reforms of the flexibility mechanism itself (Perino et al., 2022).

5 The waterbed effect in the reformed EU ETS

To further address subtleties originating from the timing of overlapping policies, we take a closer look at the reformed EU ETS in a multi-period context. In particular, we show how to compute $\partial s/\partial b$ when a policy impacts the bank in more than one period.

The EU ETS's Market Stability Reserve (MSR) works as follows.[25] 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 following year is reduced by a certain percentage (intake rate) of the size of the bank (see Table 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. 31st)	Intake rate (%)
2017	16
2018 to 2022	24
2022 to $t_{B=833}$	12

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

Starting in 2023, the maximum number of allowances held in the MSR is limited to the number auctioned in the previous year.[26] Allowances in excess of this upper bound are permanently cancelled. Given that the MSR is seeded with a large quantity of allowances and that the threshold for cancellations is decreasing along with the number of auctioned allowances, any additional allowance drawn into the MSR due to overlapping policies is eventually cancelled.[27]

[25] The rules are in European Parliament and Council (2018) and summarized by Perino (2018).

[26] The target share of auctioning in Phase 4 is 57% (European Parliament and Council, 2018) with the remaining allowances being freely allocated.

[27] At the end of 2020 the MSR contained 1.9 billion allowances with a further 379 million being added

Computing the waterbed effect for the EU ETS faces several challenges. First, the MSR's intake rate changes over time (Table 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, $t_{B=833}$ is itself determined by market outcomes and hence by the overlapping policy. Fourth, the price-mediated Rosendahl effect of anticipated future changes in allowance demand depends on the same dynamics. These complexities mean that W can only be estimated by numerical simulation.[28]

Next, we derive the sensitivity of the cumulative cap to changes in the bank $\partial s/\partial b$ as an explicit function of time. Based on this we derive an instantaneous waterbed effect that captures the first two complexities: the MSR's time-varying intake rate and its multi-period nature. An instantaneous change in the number of banked allowances triggers a sequence of transfers to the MSR.[29] This implies (see Appendix B.4) 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 current rules[30] is given by:

$$\begin{aligned} \frac{\partial}{\partial b} s\left(t, t_{B=833}\right) \quad = \quad & -1 + (1-.16)^{\max[0;\min[2018;t_{B=833}]-\max[2017;t]]} \\ & \times (1-.24)^{\max[0;\min[2023;t_{B=833}]-\max[2018;t]]} \\ & \times (1-.12)^{\max[0;\max[2023;t_{B=833}]-\max[2023;t]]}. \end{aligned} \tag{25}$$

The *instantaneous waterbed effect* $\hat{W}\left(t_a, t, t_{B=833}\right)$ in response to a one-off reduction in aggregate allowance demand in year t that is announced in year $t_a \leq t$ is thus:

$$\hat{W}\left(t_a, t, t_{B=833}\right) = \frac{1 + \frac{\partial}{\partial b} s\left(t, t_{B=833}\right)}{1 + \frac{\partial}{\partial b} s\left(t_a, t_{B=833}\right) \frac{\frac{\partial e_{t_a}}{\partial \tau_{t_a}}}{\frac{\partial e}{\partial \tau_{t_a}}}}. \tag{26}$$

Abstracting from changes in the carbon price induced by the overlapping policy as in

before September 2022 (European Commission, 2021b). The cancellation threshold in 2023 will be below 1 billion.

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

[29] 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.

[30] The European Commission's Fit-for-55 proposal of July 2021 includes a provision that, if implemented, would change 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%.

Perino (2018), this simplifies to:

$$\hat{W}\left(t, t_{B=833}\right)\big|_{\ fixed} = 1 + \frac{\partial}{\partial b} s\left(t, t_{B=833}\right). \tag{27}$$

Equation (26) 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_{833}, jointly determine the size of the instantaneous waterbed effect. (Note that this still ignores the endogeneity of t_{833}.)

6 Empirical illustrations

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 the "effective emissions reduction" rate $R_i \equiv [1 - L_i][1 - W]$. To compute R_i for a policy, we use a combination of different sources to quantify its internal carbon leakage and waterbed effect. The objective of these numerical illustrations is to leverage our theory and prior literature to obtain "ballpark" estimates of likely climate-effectiveness of overlapping policies in Europe and North America; we do not attempt any original empirical work.

Figure 2 is a visual summary of this section. 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 ($R < 0$).

6.1 Moving from theory to empirics

Moving from theory to empirics faces several challenges. First, prior literature relies on different concepts of internal carbon leakage (as discussed in Section 2.2). Some papers estimate internal leakage holding fixed the system-wide carbon price—and hence corresponding 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 $L_i^{\mathsf{M}} = 1 - R_i$. We indicate below which leakage concept is estimated in each study.

For a multi-sector cap-and-trade system, like the EU ETS, we expect that the former two concepts give similar results, $L_i^{\mathsf{T}} \simeq L_i$, as long as the policy is small relative to the

ETS, so that $d\tau/d\lambda_i \simeq 0$, and rely on this in our calibrations (unless stated otherwise).[31] For a single-sector ETS, like RGGI, the latter two concepts coincide, $L_i^{\mathsf{M}} = L_i^{\mathsf{T}} = 1 - R_i$, but do not separately disentangle the waterbed effect.

Second, the application to real-world policies requires extending our single-period conceptual framework from Section 2 to $T > 2$ periods—also to formally accommodate the multi-period EU ETS waterbed effect (Section 5). Denote the system's carbon-price path by $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_T)$, and suppose that i's (marginal) overlapping policy reduces its emissions demand by $(\frac{de_{i1}}{d\lambda_i}, \frac{de_{i2}}{d\lambda_i}, ..., \frac{de_{iT}}{d\lambda_i}) < 0$ in each period (fixed $\boldsymbol{\tau}$) and by $\frac{de_i}{d\lambda_i} = \sum_{t=1}^{T} \frac{de_{it}}{d\lambda_i} < 0$ over time. Also define $\beta_{it} \equiv \frac{de_{it}}{d\lambda_i} \Big/ \frac{de_i}{d\lambda_i} \in (0, 1)$ as the fraction of the change in i's cumulative emissions demand that occurs in period t (where $\sum_{t=1}^{T} \beta_{it} \equiv 1$).

The key metric for climate change is the impact on cumulative equilibrium emissions $\frac{de}{d\lambda_i} = \sum_{t=1}^{T} \frac{de_t}{d\lambda_i}$ (endogenous $\boldsymbol{\tau}$). Internal carbon leakage in period t is $L_{it} \equiv \frac{de_{jt}}{d\lambda_i} \Big/ \frac{de_{it}}{d\lambda_i}$ (fixed $\boldsymbol{\tau}$) so the (net) system-wide change in emissions demand at t due to i's policy is $\frac{de_t}{d\lambda_i} \equiv [1 - L_{it}]\frac{de_{it}}{d\lambda_i}$ which aggregates to $\frac{de}{d\lambda_i} \equiv \sum_{t=1}^{T} \frac{de_t}{d\lambda_i}$. Define $L_i \equiv \sum_{t=1}^{T} \beta_{it} L_{it}$ as the average leakage rate across periods. The multi-period waterbed effect $W = 1 - \frac{de}{d\lambda_i} \Big/ \frac{de}{d\lambda_i}$ is defined as before. Note that for quantity-based mechanisms the temporal distribution of the demand shift is captured by the set of β_{it} affects $de\ /d\lambda_i$ and hence, using (26), the waterbed effect for a policy announced in $t = 1$ can be expressed as $W(\beta_{i1}, ..., \beta_{iT}, t_{833}) \equiv \sum_{t=1}^{T} \beta_{it} \hat{W}(t, t_{833})$ for the reformed EU ETS. This yields a clean generalization of Lemma 1: the change in cumulative equilibrium emissions due to i's policy satisfies $\frac{de}{d\lambda_i} = [1 - L_i][1 - W]\frac{de_i}{d\lambda_i}$.[32] As neither prior empirical work nor our theory of Section 3 provide time-granular estimates of internal leakage, we drop t subscripts in the remainder of this section to work directly with L_i—and usually also drop the policy's i subscript for ease of exposition.

[31] To see this, note that total internal carbon leakage can be written as:

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

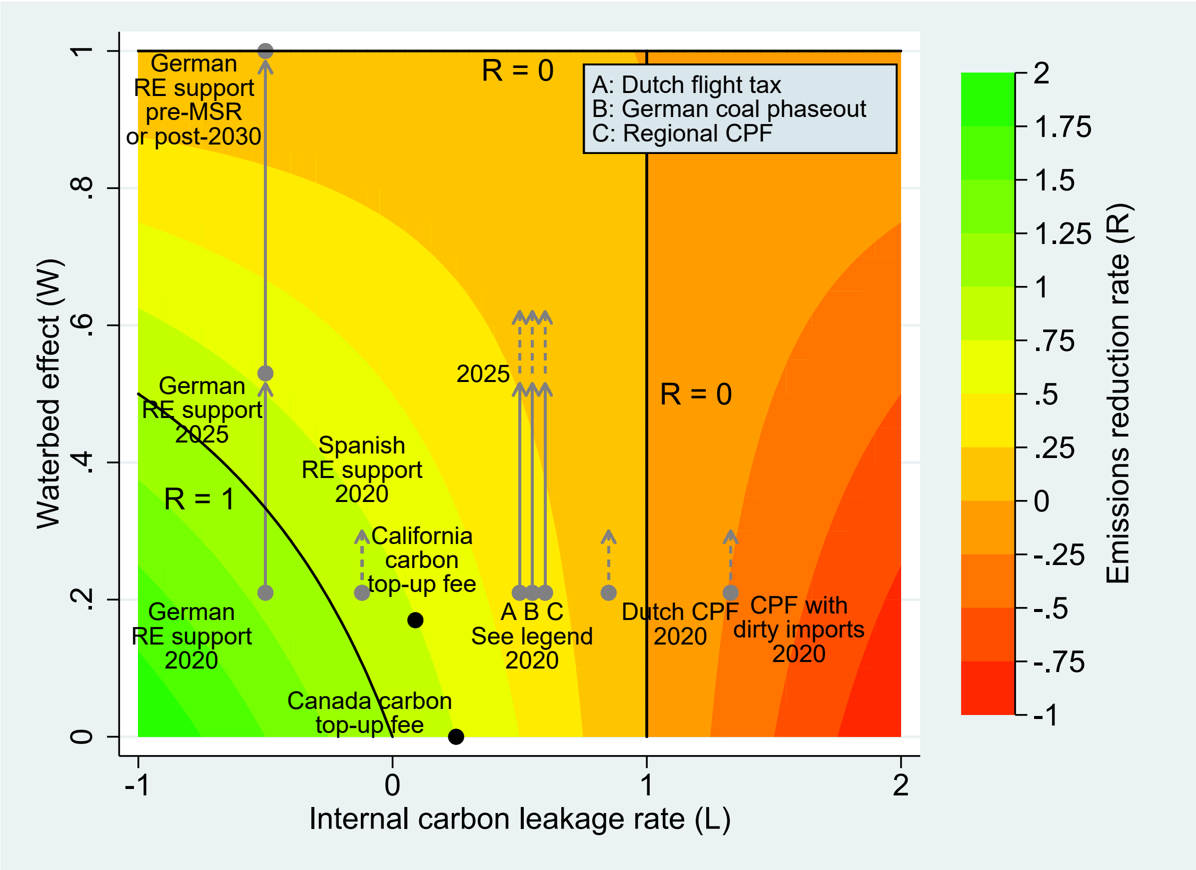
where typically $de_k/d\tau < 0$ $(k = i, j)$. For cap-and-trade systems, $d\tau/d\lambda_i \neq 0$, so there may be a wedge between L_i^{T} and L_i. However, if i's policy is marginal, we expect $d\tau/d\lambda_i \simeq 0$ so that $L_i^{\mathrm{T}} \simeq L_i$.

[32] To see this, note that equilibrium emissions satisfy $\frac{de^*}{d\lambda_i} = [1 - W]\frac{de}{d\lambda_i}$ by construction while also:

$$\frac{de}{d\lambda_i} = \sum_{t=1}^{T} \frac{de_t}{d\lambda_i} = \sum_{t=1}^{T} [1 - L_{it}]\frac{de_{it}}{d\lambda_i} = \frac{de_i}{d\lambda_i}\sum_{t=1}^{T} [1 - L_{it}]\beta_{it} = [1 - L_i]\frac{de_i}{d\lambda_i},$$

from which the result is immediate.

Figure 2: 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).

6.2 Overlapping policies in the EU ETS

We first consider policies for electricity and aviation that overlap the reformed EU ETS. As discussed in Section 5, the eventual impact of a marginal change in the allowance bank in year t on overall EU ETS emissions—and thus the "instantaneous waterbed effect" $\hat{W}_t$ for a fixed carbon price path in Equation (27)—changes over time. 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 2. (The figure abstracts from the Rosendahl effect and other effects driven by adjustments in the carbon price path for anticipated policies as described in Proposition 4 and Equation (26); below we show in sensitivity analysis that $\hat{W}_t$ increases for anticipated policies, especially for the years close to $t_{B=833}$.)

As given by Equation (27), $\hat{W}_t$ depends on the year t in which the policy takes effect and the number of years until the bank drops below 833 million allowances, $t_{B=833}$. We use $t_{B=833} = 2030$ as a lower-end mid-range value,[33] and contrast policies acting in years t = 2020, 2025 and 2030. In a sensitivity analysis, we also consider $t_{B=833} = 2048$, as estimated in Gerlagh et al. (2021), and policies that are announced before they take effect (see Appendix C). As time moves on, $\hat{W}_t$ increases from 0.21 to 0.53 to 1 and all European policies in Figure 2 move north, as indicated by the dotted grey arrows. The values for $\hat{W}_t$ can be calculated using Equation (25) evaluated at $t_{B=833} = 2030$ and $t = 2020, 2025, 2030$. We note here for completeness that, as shown in Appendix C, $\hat{W}_t$ can increase above 1 over time for pre-announced policies.

"Supply-side" overlapping policies

We begin with "supply-side" overlapping policies such as national carbon price floors, coal phase-outs and aviation taxes.

Electricity

We first consider overlapping cost-raising policies such as a national carbon price floor (CPF). 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 CPF for the electricity sector in 2022.[34] It increases from EUR 14.90/tCO_2 in 2022 to EUR 31.90/tCO_2 in 2030—as of June 2022, however, the policy is not binding as the EU ETS carbon price

[33] 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).

[34] 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`.

exceeds the Dutch CPF.

Proposition 1 shows that such 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).[35] Such CPFs in small interconnected countries 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 2).[36] 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.

Cost-raising policies can backfire if imports are substantially dirtier than domestic production (see Proposition 1). We plot a hypothetical "CPF with dirty imports" for which $L = 1.33$ such that EU-wide emissions *increase*, $R < 0$.[37] 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.

Mandates to reduce carbon-intensive production in the electricity sector are also supply-side policies (Proposition 1). 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.[38] 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

[35] 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.

[36] 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.

[37] We assume $\theta_j/\theta_i = 2$, $\varepsilon_j^S = 5 \Leftrightarrow \eta_j^S = 0.2$, $\sigma_j = 0.2$, $\varepsilon^D = 0.5$ and $A_i = 0$.

[38] 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).

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 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).[39] This intra-EU leakage rate of 50%—which we interpret "in spirit" as 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 some 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 United Kingdom (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). Austria is an exception given the proximity of Vienna to Bratislava. Greece, Croatia and Latvia—countries that also have aviation taxes—are relatively small, though their geographies are such that leakage may be less severe than for the Netherlands.

"Demand-side" overlapping policies

We now look at overlapping "demand-side" policies such as renewables support. 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. Consistent with Proposition 2, Abrell et al. (2019a) estimate *negative* carbon (and output) leakage (holding carbon prices fixed by controlling for EUA prices in their regressions) as additional zero-carbon energy depresses wholesale electricity prices and offsets imported gas- and coal-fired electricity in Germany ($L = -0.50$) and Spain ($L = -0.12$). Their leakage estimates correspond to

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

our L in Lemma 1.[40] Similarly, a German government report finds $L = -0.54$ (Klobasa and Sensfuss, 2016), also holding carbon prices fixed.[41] Figure 2 shows that, at least in the year 2020, the renewable support scheme in Germany reduces system-wide emissions considerably ($\hat{W}_{2020} = 0.21, L = -0.50, \hat{R}_{2020} = 1.19$); in fact, by *more* than the domestic emissions reduction in Germany. 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 expect negative internal leakage also for these policies but are not aware of any empirical estimates, so do not include them in Figure 2.

Sensitivity analysis

In Appendix C, we show the sensitivity of our results to $t_{B=833}$, the year in which the MSR will stop taking in allowances. The instantaneous waterbed effect decreases substantially when $t_{B=833}$ lies further in the future. We also show how the performance of two key policies—renewable energy support and a coal phase-out in Germany—changes when we allow for the Rosendahl effect (see Equation (26)). This increases $\hat{W}_t$, 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—consistent with Proposition 4. This highlights the potential unintended consequences of announcing policies that reduce emissions demand far into the future.

6.3 Overlapping policies in North America

We now turn to discussing examples of overlapping carbon policies in North America, two of which are plotted in Figure 2. Recall from Section 4.1 that the waterbed can also be punctured due to its stochastic nature. This feature is relevant for the two carbon markets in the United States.

[40] 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 2. 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 intensity of these marginal units are about equal. Abrell et al. (2019a) show that the German power mix is indeed dirtier than Spain's.

[41] 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.

California-Québec carbon market

California and Québec have a joint carbon market with a hybrid design. There is an auction price floor ($17.71 in 2021)[42] and a price ceiling ($65 in 2021) (Politico, 2018). Before the hard price cap is reached, there are two soft price caps that create horizontal segments in the allowance supply function: up to some limit, allowances will be offered at $41.40 and at $53.20 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, the expected waterbed effect $W = 1 - \pi = 0.17$.[43]

The California-Québec carbon market is known to cause *external* leakage to neighbouring 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.[44] 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 $L =$ 0.65-0.75. Caron et al. (2015) provide a relevant leakage estimate of $L = 0.09$ (holding carbon prices fixed[45]) 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.[46] Figure 2 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

[42]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.

[43]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.

[44]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.

[45]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.

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

($W = 0.17, L = 0.09$). We note, however, that these values are illustrative only given the profound changes in the Western power grid since the Caron et al. (2015) article was written.

Regional Greenhouse Gas Initiative

The 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;[47] 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 the discussion in Section 4.1. Once larger, non-marginal interventions are considered—such that allowance demand moves across one or several steps in the supply schedule—the effective waterbed effect lies between zero and one.

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 as well as the RGGI allowance supply curve, and estimate L^M to other RGGI states that results from New York's policy—the combined effect of internal leakage and RGGI's waterbed effect akin to our $1-R$—at $R = 0.42$.[48] We do not plot New York's carbon fee in Figure 2 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. Note that L^T and L^M (and, thus, $1-R$) are the same in this analysis, since it considers a single sector (electricity). $R = 0.42$ would place the policy in the light-orange region of Figure 2, so "medium climate-effective."

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

[48]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.

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. Such unilateral carbon taxes face no waterbed effect (Proposition 3) but 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 2 plots this policy assuming $L = 0.25$ and $W = 0$, leaving a higher carbon tax in British Columbia reasonably climate-effective ($R = 0.75$).

7 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 policy types have very different internal-leakage properties. Space matters as leakages can differ substantially across industries and jurisdictions. Time matters as it affects 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 the ongoing design of China's national emissions trading system. Future research on hybrid carbon-market designs should pay close attention to internal carbon leakage—more empirical estimates could help improve policy-making substantially.

Our goal in this paper has been to obtain a set of benchmark results that incorporate—and clearly delineate—internal carbon leakage and the waterbed effect. Future research could refine our work in several directions. First, we have assumed a competitive product market; future theory work could examine the roles of market power, economies of scale and product differentiation—these may mitigate internal carbon leakage but also seem unlikely to change its sign. Second, future empirical work could use the variation triggered by an existing overlapping policy to estimate the various demand and supply elasticities shown by our above theory to be crucial to evaluate its climate-effectiveness. Third, while we have focused sharply on climate impacts (via the change in aggregate emissions), future policy work should pursue a full welfare analysis of overlapping policies that incorporates cost-effectiveness as well as wider distributional impacts. Fourth, to highlight the economics of internal carbon leakage, we have abstracted from the potential for external carbon leakage to jurisdictions outside the carbon-pricing system; in some cases both leakages may play an important role.

Finally, our framework has deeper connections to the wider literature on environmental economics. A fixed resource stock that will be exhausted sooner or later (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.[49] Moreover, there is (external) carbon leakage if non-coalition countries increase emissions in response to unilateral action. This has 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 emissions trading system with a fixed cap (like the pre-2018 EU ETS) pursuing overlapping policies that involve cancelling allowances.

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.

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.

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.

[49]The green paradox is concerned with carbon entering the economy (with the burning of fossil resources) while the waterbed effect is concerned with carbon leaving it (due to carbon pricing).

Borenstein, Severin, 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.

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, Charlie Holt, Karen Palmer, and William Shobe, "Quantities with Prices: Price-Responsive Allowance Supply in Environmental Markets," *Resources for the Future Working Paper*, 2020, pp. 20–17. Available at `https://www.rff.org/publications/working-papers/quantities-prices-price-responsive-allowance-supply-environmental-markets/`.

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.

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 2020 for the Purposes of the Market Stability Reserve under the EU Emissions Trading System Established by Directive 2003/87/EC," Technical Report C(2021) 3266 final, European Commission 2021.

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,” *AEA 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.

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.

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.

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, *forthcoming.*

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.

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.

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.

World Bank, "State and Trends of Carbon Pricing 2021," 2021. Washington, DC: World Bank. Available at https://openknowledge.worldbank.org/handle/10986/35620. License: CC BY 3.0 IGO.

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: Proofs of results and robustness discussion for internal carbon leakage

First, we derive two general results, Propositions 1A–2A, on internal carbon leakage for supply-side and demand-side overlapping policies using a general cost function $G_k(x_k, a_k)$ for $k = i, j$. Second, we then obtain Propositions 1–2 from the main text, which use the separable cost function $G_k(x_k, a_k) \equiv [C_k(x_k) + \phi_k(a_k)]$, as simple corollaries (with $G_k^{xa}(x_k, a_k = 0)$ and discuss how the key insights from this simplified model are robust. Third, we extend the baseline model to multi-market settings with demand from consumers in both i and j. Fourth, we discuss robustness for overlapping policies that are not marginal.

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

As in the main text, firm k's emissions are $e_k = e_k^0 - a_k$ where a_k is abatement and $e_k^0 = \theta_k x_k$ is baseline emissions. Standing assumptions are $G_k^x, G_k^a > 0$ and $G_k^{xx}, G_k^{aa} > 0$ so $G_k^{aa} \to \infty$ means that additional abatement is infeasible (corresponding to $A_k = 0$ in our simple model). To maximise profits, firm k solves $\max_{x_k;a_k} \Pi_k = px_k - G_k(x_k, a_k) - \tau_k e_k$. The two first-order conditions are:

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

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 that this optimised cost increases with abatement, $M_k^a(x_k; a_k) \equiv [G_k^{xa} + \theta_k G_k^{aa}] > 0$, 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 is nested where $\psi_k = 0$ while $\psi_k < 1$ again follows by stability. Finally, a key metric to characterise output responses in the general model will be:

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

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.

Supply-side overlapping policies

Proposition 1A. *With general cost functions, a supply-side overlapping policy by jurisdiction i has internal carbon leakage to jurisdiction j of:*

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

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 $Z_i^G \equiv \frac{G_i^{aa}}{M_i^a}\frac{G_i^{xx}}{M_i^a}\left[(1-\psi_i) + \mu_j(1-\psi_j)\frac{G_j^{xx}}{G_i^{xx}}\right] \geq 0$ is an abatement effect.

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

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

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

As j's carbon price remains fixed, $\tau_j = \tau$, differentiating j's first-order conditions yields:

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

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

We proceed in two main steps, first deriving equilibrium changes in output levels, and then deriving changes in emissions—and hence internal carbon leakage. First, combining j's first-order conditions shows that the firms' output changes are related according to:

$$p^0\frac{dx_i}{d\lambda_i} = \left[-p^0 + G_j^{xx}(1-\psi_j)\right]\frac{dx_j}{d\lambda_i}.$$

The same approach for i yields:

$$p^0 \frac{dx_j}{d\lambda_i} = \theta_i \delta_i \frac{d\tau_i}{d\lambda_i} + \left[-p^0 + G_i^{xx}\left(1-\psi_i\right)\right] \frac{dx_i}{d\lambda_i}.$$

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:

$$-\mu_j \frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i} \text{ and } -\mu_i \frac{dx_j}{d\lambda_i} = \mu_i \frac{\theta_i \delta_i}{\left(-p^0\right)} \frac{d\tau_i}{d\lambda_i} + \frac{dx_i}{d\lambda_i}.$$

Hence solving for the equilibrium output responses yields:

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

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. 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 and its first-order condition for abatement we obtain:

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

We similarly obtain for i:

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

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

$$L_i = \frac{\theta_j}{\theta_i} \mu_j \frac{\frac{\mu_i}{\left(1-\mu_i\mu_j\right)} \frac{\delta_i \delta_j}{\left(-p^0\right)}}{\left[\frac{\mu_i}{\left(1-\mu_i\mu_j\right)} \frac{\delta_i^2}{\left(-p^0\right)} + \frac{1}{\theta_i^2 G_i^{aa}}\right]} = \frac{\theta_j}{\theta_i} \mu_j \frac{\delta_j}{\delta_i} \frac{1}{\left[1 + \frac{\left(-p^0\right)}{\delta_i^2 \theta_i^2 G_i^{aa}} \left[\frac{1}{\mu_i} - \mu_j\right]\right]}.$$

It will be useful to rewrite the last term as follows. Recalling the definition $\mu_k \equiv \left(-p^0\right) / \left[-p^0 + G_k^{xx}\left(1-\psi_k\right)\right]$, observe that:

$$\left(-p^0\right)\left[\frac{1}{\mu_i} - \mu_j\right] = \left[\frac{G_i^{xx}\left(1-\psi_i\right)}{-p^0} + \frac{G_j^{xx}\left(1-\psi_j\right)}{\left[-p^0 + G_j^{xx}\left(1-\psi_j\right)\right]}\right]\left(-p^0\right) = G_i^{xx}\left[\left(1-\psi_i\right) + \mu_j\left(1-\psi_j\right)\frac{G_j^{xx}}{G_i^{xx}}\right].$$

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

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

Using these terms in the expression for internal carbon leakage yields the result as claimed for the unilateral carbon prices.

Now consider the unilateral reduction in carbon-intensive production by jurisdiction i, represented as $dx_i/d\lambda_i < 0$, where the common carbon price remains unchanged, $\tau_i = \tau_j = \tau$. This problem has the same structure as before—except that i's output change $dx_i/d\lambda_i < 0$ is determined by policy directly rather than induced in equilibrium by a unilateral carbon price. The remaining choices—abatement by i and output and abatement by j—remain optimal by the respective first-order conditions.

Hence differentiating i's first-order condition for abatement 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 first-order conditions yields:

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

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

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:

$$\frac{de_k}{d\lambda_i} = \theta_k \frac{dx_k}{d\lambda_i} - \frac{da_k}{d\lambda_i}.$$

Using firms' equilibrium output responses and first-order conditions for abatement, we obtain:

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

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 of (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 has internal carbon leakage to jurisdiction j of:*

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

where the rate of output leakage is $L_i^O = -\frac{\mu_j/(1-\mu_j)}{\mu_i/(1-\mu_i)} < 0$.

Proof of Proposition 2A. As explained in the main text, all three of these 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 common carbon price remains unchanged, $\tau_i = \tau_j = \tau$. Thus differentiating i's first-order conditions for the impact of the overlapping policy yields:

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

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

Differentiating j's first-order conditions 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 deriving equilibrium output responses, and then emissions responses. First, combining j's first-order conditions 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:

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

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

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

Solving for equilibrium output responses yields:

$$\frac{dx_i}{d\lambda_i} = \frac{\mu_i(1-\mu_j)}{(1-\mu_i\mu_j)} \frac{1}{(-p^{\prime})} \frac{\partial}{\partial \lambda_i} p(X;\lambda_i) < 0$$

$$\frac{dx_j}{d\lambda_i} = \frac{\mu_j(1-\mu_i)}{(1-\mu_i\mu_j)} \frac{1}{(-p^{\prime})} \frac{\partial}{\partial \lambda_i} p(X;\lambda_i) < 0.$$

So the rate of internal output leakage is:

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

which is always negative. 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 response, its first-order condition for abatement, and the definition of δ_k, we obtain:

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

We similarly obtain for i:

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

Therefore the equilibrium rate of internal carbon leakage is:

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

as claimed in the result.

A.2. Robustness of results with separable cost functions

We now derive Propositions 1–2 from the main text, for separable cost functions, as direct corollaries of Propositions 1A–2A. The key difference is that, in the separable model, output and abatement decisions are independent while, in the general model, abatement can also be induced by changes to output. In comparing the two sets of results, we discuss how the key insights from the simplified model in the main text are nonetheless 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 $G_k^{xa}(x_k, a_k) = 0$. 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 $k = i, j$.

To present leakage formulae in terms of simple demand and supply elasticities, we begin by recording two preliminary results. First, using the price elasticity of demand $\varepsilon^D \equiv -p(\cdot)/Xp'(\cdot) > 0$ 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 the 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 of k's market share, $\sigma_k \equiv x_k/X \in (0, 1)$, and $\eta_k^S = 1/\varepsilon_k^S$ by its first-order condition. Second, using the same ingredients, we also obtain that:

$$\mu_k \equiv \frac{-p'}{(-p' + C_k'')} = \frac{\sigma_k}{(\sigma_k + \varepsilon^D/\varepsilon_k^S)} > 0,$$

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

Supply-side overlapping policies

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

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

where $\gamma = 0$ for a unilateral reduction in carbon-intensive production, $\gamma = 1$ for a unilateral carbon price, and $Z_i \equiv \frac{A_i}{(1-A_i)} \left(1 + \frac{(1-\mu_j)\varepsilon_i^S/\varepsilon_j^S}{(\sigma_j + \varepsilon^D/\varepsilon_j^S)}\right) \geq 0$ is an abatement effect.

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

$$L_i = \frac{\theta_j}{\theta_i} \mu_j \frac{1}{\left[1 + \frac{G_i^{aa}}{M_i^a} \frac{G_i^{xx}}{M_i^a} \left[1 + \mu_j \frac{G_j^{xx}}{G_i^{xx}}\right]\right]} = \frac{\theta_j}{\theta_i} \mu_j \frac{1}{\left[1 + \frac{C_i''}{\theta_i^2 \phi_i''} \left[1 + \mu_j \frac{C_j''}{C_i''}\right]\right]}.$$

Using the two preliminary results, including that $C_j^{00}/C_i^{00} = \sigma_i \varepsilon_i^S / \sigma_j \varepsilon_j^S$, and the definition of k's abatement opportunity from the main text, $A_k = C_k^{00}/[C_k^{00} + \theta_k^2 \phi_k^{00}]$, yields the result. For the unilateral reduction in carbon-intensive production ($\phi = 0$), Proposition 1A simplifies directly to:

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

Comparing this with the general result from Proposition 1A, an obvious difference is the absence of the term δ_j/δ_i, where $\delta_k = (1 + G_k^{ax}/\theta_k G_k^{aa}) > 0$ 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 $\delta_k < 1 \Leftrightarrow G_k^{xa} < 0$ ($k = i, j$), 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 $\delta_k > 1 \Leftrightarrow G_k^{xa} > 0$.

A second difference between the two results arises via the rate of output leakage. In particular, recall that $L_i^O = \mu_j \equiv (-p^0)/\left[-p^0 + G_j^{xx}(1 - \psi_j)\right] \in (0, 1)$ 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 output leakage L_i^O is higher for $G_j^{xa} \neq 0$, all else equal, 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 (Z_i^G) and separable (Z_i) models for the unilateral carbon price—but is irrelevant for the unilateral production cut.

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 of (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 has 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 $C_k''(x_k) = \frac{p(X)}{X}\frac{1}{\varepsilon_k^S}$ 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 simplification 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 $L_i^O = -[\mu_j/(1-\mu_j)]/[\mu_i/(1-\mu_i)] = -G_i^{xx}(1-\psi_i)/G_j^{xx}(1-\psi_j)$ 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.

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

The main text considers a model with a single demand curve that we interpreted to represent consumers in jurisdiction i. We here show that the main insights from this analysis—supply-side (demand-side) overlapping policies have positive (negative) internal carbon leakage—extend to multi-market settings with demand from both consumers in i and j.

Firm i sells x_i to its home market and exports y_i to j's market while firm j sells y_j to 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 $L_i \equiv -\frac{de_j}{de_i} = -\frac{_i}{_j}\frac{dQ_j}{dQ_i}$, 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:

$$L_i = \frac{dx_i}{dQ_i} L_{ii} + \frac{dy_i}{dQ_i} L_{ij}$$

where $L_{ii} \equiv \frac{_i}{_j}\left(-\frac{dx_j}{dx_i}\right)$ is the "single-market" leakage rate arising in i's domestic market while $L_{ij} \equiv \frac{_i}{_j}\left(-\frac{dy_j}{dy_i}\right)$ is the additional "cross-market" leakage rate arising from i's exports to j's home market. Overall leakage is a weighted average of these individual leakages.

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:

$$\frac{\partial \Pi_i}{\partial x_i} = 0 = p_i - C_i^0(x_i) - \tau_i\theta_i \text{ and } \frac{\partial \Pi_i}{\partial y_i} = 0 = p_j - H_i^0(y_i) - \tau_i\theta_i$$

so that the product price, net of non-carbon costs, is equalized across the two markets, with $p_i - C_i^0(x_i) = p_j - H_i^0(y_i) = \tau_i\theta_i$. Similarly, firm j's first-order conditions are:

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

Note that market-level and firm-level responses to an overlapping policy λ_i are related according to $\frac{dX}{d\ _i} + \frac{dY}{d\ _i} = \frac{dQ_i}{d\ _i} + \frac{dQ_j}{d\ _i}$

Generalising the single-market case, define $\mu_{ii} \equiv \frac{p_i^0(X)}{\left[\ p_i^0(X)+C_i^{00}(x_i)\right]} \in (0,1)$, $\mu_{ij} \equiv \frac{p_j^0(Y)}{\left[\ p_j^0(Y)+H_i^{00}(y_i)\right]} \in (0,1)$, $\mu_{ji} \equiv \frac{p_i^0(X)}{\left[\ p_i^0(X)+C_j^{00}(x_j)\right]} \in (0,1)$, and $\mu_{jj} \equiv \frac{p_j^0(Y)}{\left[\ p_j^0(Y)+H_j^{00}(y_j)\right]} \in (0,1)$.

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 $\frac{d\ _i}{d\ _i} > 0$, and (ii) a unilateral policy that requires a cut $\frac{dQ_i}{d\ _i} < 0$ in i's overall production (e.g., a coal phase-out).

Proposition 1M. A supply-side overlapping policy by jurisdiction i has 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 offers a straightforward generalization of our results to a multi-market setting: internal carbon leakage is always positive—and may again exceed 100% where θ_j/θ_i is sufficiently large. 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 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:

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

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 first-order condition for market i gives:

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

Solving these two equations simultaneously yields:

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

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:

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

For a unilateral reduction in fossil production, $\frac{dQ_i}{d\ _i} < 0$, 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 $L_{ii} = \frac{\ _j}{\ _i}\mu_{ji} > 0$ and $L_{ij} = \frac{\ _j}{\ _i}\mu_{jj} > 0$. Moreover, i's profit-maximizing strategy translates $\frac{dQ_i}{d\ _i} < 0$ into cutbacks in both its home-destined and export-destined production, $\frac{dx_i}{d\ _i}, \frac{dy_i}{d\ _i} < 0$ with $\frac{dx_i}{d\ _i} + \frac{dy_i}{d\ _i} = \frac{dQ_i}{d\ _i} < 0$, where it equalizes the marginal profit from each market:

$$p_i^0(X)\left(\frac{dx_i}{d\lambda_i} + \frac{dx_j}{d\lambda_i}\right) - C_i^{00}(x_i)\frac{dx_i}{d\lambda_i} = p_j^0(Y)\left(\frac{dy_j}{d\lambda_i} + \frac{dy_i}{d\lambda_i}\right) - H_i^{00}(y_i)\frac{dy_i}{d\lambda_i}$$

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

$$[p_i^0(X)(1-\mu_{ji}) - C_i^{00}(x_i)]\frac{dx_i}{d\lambda_i} = [p_j^0(Y)(1-\mu_{jj}) - H_i^{00}(y_i)]\left(\frac{dQ_i}{d\lambda_i} - \frac{dx_i}{d\lambda_i}\right)$$

and so i's output responses satisfy:

$$\frac{\frac{dx_i}{d\ _i}}{\frac{dQ_i}{d\ _i}} = \frac{\left[(1-\mu_{jj}) + \frac{H_i^{00}(y_i)}{[\ p_j^0(Y)]}\right]}{\frac{[\ p_i^0(X)]}{[\ p_j^0(Y)]}\left[(1-\mu_{ji}) + \frac{C_i^{00}(x_i)}{[\ p_i^0(X)]}\right] + \left[(1-\mu_{jj}) + \frac{H_i^{00}(y_i)}{[\ p_j^0(Y)]}\right]}$$

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

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

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 overlapping 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 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 a negative rate of internal carbon leakage also for a multi-market setting. Indeed, the leakage rate is identical 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 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:

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

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

Our results so far have focused on marginal overlapping policies, with $d\lambda_i > 0$, that shift equilibrium emissions by small amounts, de_i and de_j. This yields a rate of internal carbon leakage $L_i = \frac{de_j}{de_i}$ that can be seen as an approximation to a non-marginal rate $L_i = \frac{\triangle e_j}{\triangle e_i}$. More generally, an overlapping 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 the results from the first-order approximation.

The first point is that the insight that supply-side policies have positive internal leakage while it is negative for demand-side policies also holds for non-marginal policy. To see why, write the non-marginal leakage rate as:

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

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 to marginal leakage 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 $L_i \equiv \frac{\Delta e_j}{\Delta e_i} = \frac{de_j}{de_i}$ 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, recalling that output leakage in the general case $L_i^O = -[\mu_j/(1-\mu_j)]/[\mu_i/(1-\mu_i)] = -G_i^{xx}(1-\psi_i)/G_j^{xx}(1-\psi_j)$ does not depend on any demand-side properties. Moreover, the simple marginal formulae contain no obvious bias: they could be an over- or underestimate depending on the precise higher-order properties of cost functions, and on whether demand is convex or concave.

In sum, the main insights on internal leakage from marginal analysis also hold for potentially much more complex non-marginal policies.

Appendix B: Proofs of results and robustness discussion 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 discuss the robustness of our results to overlapping policies that are non-marginal. Fourth, we derive the instantaneous waterbed effect in the reformed EU ETS from Section 5.

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

Derivation of Equation (9)

Application of Cramer's rule to conditions (7) and (8) yields:

$$\frac{\partial \tau}{\partial \lambda_i} = \frac{-\frac{\partial \rho}{\partial \lambda_i}}{\frac{\partial \rho}{\partial e}\frac{\partial s}{\partial \tau} - 1}. \tag{A.1}$$

Total differentiation of $\tau = \rho(e, \lambda_i)$ (see first-order condition (7)) yields

$$\frac{de}{d\lambda_i} = -\frac{\frac{\partial \rho}{\partial \lambda_i}}{\frac{\partial \rho}{\partial e}}. \tag{A.2}$$

Extending A.1 with $\left(\frac{\partial \rho}{\partial e}\right)^{-1}$ using (A.2), and substituting the slope of the allowance demand curve $\frac{\partial e}{\partial \rho}$ for the inverse of the slope of the inverse allowance demand curve $(\partial \rho/\partial e)^{-1}$

yields Equation (9).

Independence of timing of allowance demand shifts

Consider the two-period version of the model with a price-based flexibility mechanism where the equilibrium conditions (12) and (13) balance marginal costs of abatement with the carbon price for periods 1 and 2, respectively, while (14) is the market-clearing condition for the allowance market.

Application of Cramer's rule to conditions (12)-(14) yields:

$$\frac{\partial \tau_1}{\partial \lambda_i} = \frac{\frac{\partial \rho_2}{\partial e_2}\frac{\partial \rho_1}{\partial \lambda_i} + \frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial \lambda_i}}{\frac{\partial \rho_2}{\partial e_2} + (1+r)\frac{\partial \rho_1}{\partial e_1} - \frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}\frac{\partial s}{\partial \tau_1}}. \tag{A.3}$$

Total differentiation of $\tau_t = \rho_t(e_t, \lambda_i)$ (see first-order conditions (12) and (13)) yields

$$\frac{de_t}{d\lambda_i} = -\frac{\frac{\partial \rho_t}{\partial \lambda_i}}{\frac{\partial \rho_t}{\partial e_t}}. \tag{A.4}$$

Cancelling $-\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$ from (A.3), using (A.4), 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 yields:

$$\frac{\partial \tau_1}{\partial \lambda_i} = \frac{\frac{de_1}{d\lambda_i} + \frac{de_2}{d\lambda_i}}{\frac{\partial s}{\partial \tau_1} - \frac{\partial e_1}{\partial \tau_1} - (1+r)\frac{\partial e_2}{\partial \tau_2}}.$$

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 the impact of the overlapping policy on the system-wide equilibrium carbon price:

$$\frac{\partial \tau_1}{\partial \lambda_i} = \frac{\frac{de}{d\lambda_i}}{\frac{\partial s}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}} > 0 \tag{A.5}$$

where $\partial e/\partial \tau_1 < 0$ 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:

$$\frac{de}{d\lambda_i} = \frac{\partial s}{\partial \tau_1}\frac{\partial \tau_1}{\partial \lambda_i} = \frac{de}{d\lambda_i}\frac{\frac{\partial s}{\partial \tau_1}}{\frac{\partial s}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}} = \frac{de}{d\lambda_i}\frac{\kappa^S}{\kappa^S + \kappa^D} \tag{A.6}$$

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

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

Derivation of Equation (18)

Application of Cramer's rule to conditions (15)-(17) yields:

$$\frac{\partial e_1}{\partial \lambda_i} \quad = \quad \frac{\frac{\partial \rho_2}{\partial \lambda_i} - (1+r)\frac{\partial \rho_1}{\partial \lambda_i}}{(1+r)\frac{\partial \rho_1}{\partial e_1} - \frac{\partial \rho_2}{\partial e_2}\left(1 + \frac{\partial s_2}{\partial b}\right)}.$$

Cancelling $-\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$, using (A.2), and substituting the slope of the allowance demand curve $\frac{\partial e_t}{\partial \rho_t}$ for the inverse of the slope of the inverse allowance demand curve $(\partial \rho_t/\partial e_t)^{-1}$ yields:

$$\frac{\partial e_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{de_2}{d\lambda_i}\frac{\partial e_1}{\partial \rho_1} - (1+r)\frac{de_1}{d\lambda_i}\frac{\partial e_2}{\partial \rho_2}}{(1+r)\frac{\partial e_2}{\partial \rho_2} - \frac{\partial e_1}{\partial \rho_1}\left(1 + \frac{\partial s_2}{\partial b}\right)}.$$

Using $de/d\lambda_i = de_1/d\lambda_i + de_2/d\lambda_i$, $\frac{\partial e}{\partial \rho_1} = \frac{\partial e_1}{\partial \rho_1} + \frac{\partial e_2}{\partial \rho_1}$ and $\frac{\partial e_2}{\partial \rho_1} = \frac{1+r}{1+r}\frac{\partial e_2}{\partial \rho_1} = (1+r)\frac{\partial e_2}{\partial \rho_2}$ yields

$$\frac{\partial e_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{de}{d\lambda_i}\frac{\partial e}{\partial \rho_1} - \frac{de_1}{d\lambda_i}\frac{\partial e_2}{\partial \rho_2}}{\frac{\partial e}{\partial \rho_1} + \frac{\partial e_1}{\partial \rho_1}\frac{\partial s_2}{\partial b}}.$$

Factoring out $de/d\lambda_i$ in the numerator and dividing both numerator and denominator by $\partial e/\partial \tau_1$ yields (18).

Proof of Corollary 1

First use conditions (15)-(17) and Cramer's rule to compute:

$$\frac{\partial \tau_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{\partial \rho_1}{\partial \lambda_i}\frac{\partial \rho_2}{\partial e_2}\left(1 + \frac{\partial s_2}{\partial b}\right) + \frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial \lambda_i}}{(1+r)\frac{\partial \rho_1}{\partial e_1} - \frac{\partial \rho_2}{\partial e_2}\left(1 + \frac{\partial s_2}{\partial b}\right)}.$$

Cancelling $-\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$, using (A.2), and substituting the slope of the allowance demand curve $\frac{\partial e_t}{\partial \rho_t}$ for the inverse of the slope of the inverse allowance demand curve $(\partial \rho_t/\partial e_t)^{-1}$ yields:

$$\frac{\partial \tau_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{de_1}{d\lambda_i}\left(1 + \frac{\partial s_2}{\partial b}\right) + \frac{de_2}{d\lambda_i}}{(1+r)\frac{\partial e_2}{\partial \rho_2} - \frac{\partial e_1}{\partial \rho_1}\left(1 + \frac{\partial s_2}{\partial b}\right)}.$$

Using $de/d\lambda_i = de_1/d\lambda_i + de_2/d\lambda_i$, $\frac{\partial e}{\partial \rho_1} = \frac{\partial e_1}{\partial \rho_1} + \frac{\partial e_2}{\partial \rho_1}$ and $\frac{\partial e_2}{\partial \rho_1} = \frac{1+r}{1+r}\frac{\partial e_2}{\partial \rho_1} = (1+r)\frac{\partial e_2}{\partial \rho_2}$ yields

$$\frac{\partial \tau_1}{\partial \lambda_i} \quad = \quad -\frac{\frac{de}{d\lambda_i} + \frac{de_1}{d\lambda_i}\frac{\partial s_2}{\partial b}}{\frac{\partial e}{\partial \rho_1} + \frac{\partial e_1}{\partial \rho_1}\frac{\partial s_2}{\partial b}}. \tag{A.7}$$

Next 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:

$$\left.\frac{ds}{d\tau_1}\right|_{equ} = \frac{\frac{\partial s_2}{\partial \tau_i}}{\frac{\partial \tau_1}{\partial \tau_i}} = \frac{\frac{\partial s_2}{\partial b}}{1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}}\cdot\left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}-\beta\right]\cdot(-1)\frac{\frac{\partial e}{\partial \tau_1}\left[1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}\right]}{\left[1+\frac{\partial s_2}{\partial b}\beta\right]}.$$

Cancelling $1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial E_1}{\partial \tau_1}}{\frac{\partial E}{\partial \tau_1}}$ and using $\tau=\tau_1$ yields Equation (24).

It now remains to be shown that Propositions 3 and 4 are equivalent. To see this substitute (24) into (11) to get:

$$W = \frac{-\frac{\partial e}{\partial \tau}}{\left[\frac{\frac{\partial s_2}{\partial b}\frac{\partial e}{\partial \tau}}{1+\frac{\partial s_2}{\partial b}}\cdot\left(\beta-\frac{\frac{\partial e_1}{\partial \tau}}{\frac{\partial e}{\partial \tau}}\right)\right]-\frac{\partial e}{\partial \tau}}.$$

Note that $W\in[0,1]$ only holds for weakly upward-sloping allowance supply curves. This is no longer guaranteed once we substitute in Equation (24). Both values below 0 and above 1 are now possible. Dividing the above equation by $-\frac{\partial e}{\partial \tau}$ and multiplying it by $1+\frac{\partial s_2}{\partial b}\beta$ obtains:

$$W = \frac{1+\frac{\partial s_2}{\partial b}\beta}{\frac{\partial s_2}{\partial b}\cdot\left[\frac{\frac{\partial e_1}{\partial \tau}}{\frac{\partial e}{\partial \tau}}-\beta\right]+1+\frac{\partial s_2}{\partial b}\beta}.$$

Cancel $\frac{\partial s_2}{\partial b}\beta$ in the denominator and substitute $\tau=\tau_1$ to obtain Equation (20).

B.3. Robustness of results with non-marginal overlapping 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 $W=1-\frac{de}{de}$ 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 Proposition 3 and 4 extend to non-marginal changes in policy. To see why, write the non-marginal waterbed effect as:

$$W = 1-\frac{\triangle e}{\triangle e} = 1-\frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de}{d\lambda_i}+\frac{de}{d\tau}\frac{d\tau}{d\lambda_i}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de}{d\lambda_i}\right)d\lambda_i} = -\frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de}{d\tau}\frac{d\tau}{d\lambda_i}\right)d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\triangle\lambda_i}\left(\frac{de}{d\lambda_i}\right)d\lambda_i}. \tag{A.8}$$

Plugging equation (9) into (A.8) we get the non-marginal version of (11):

$$W = \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\Delta\lambda_i} \left(\frac{de}{d\lambda_i} \frac{\frac{\partial e}{\partial \tau}}{\frac{\partial s}{\partial \tau} - \frac{\partial e}{\partial \tau}} \right) d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\Delta\lambda_i} \left(\frac{de}{d\lambda_i} \right) d\lambda_i}$$

Since the slope of the allowances supply curve is assumed to be weakly positive and that of the cumulative allowance demand curve to be strictly negative this expression is always between zero and one. Hence, Proposition 3 extends to non-marginal changes in overlapping policies. Note that for non-marginal overlapping policies, W depends on $de/d\lambda_i$ and hence on de_i and L_i as they determine the size of the change in long-run demand.

Moving to quantity-based flexibility mechanisms, we plug (A.7) into (A.8) and simplify:

$$W = \frac{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\Delta\lambda_i} \left(\frac{de}{d\lambda_i} \frac{1+\frac{\partial s_2}{\partial b}\frac{\frac{de_1}{d\lambda_i}}{\frac{de}{d\lambda_i}}}{1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \right) d\lambda_i}{\int_{\underline{\lambda}_i}^{\underline{\lambda}_i+\Delta\lambda_i} \left(\frac{de}{d\lambda_i} \right) d\lambda_i}$$

If the non-marginal policy change does not affect the demand for allowances in period 2, i.e., if $de_1/d\lambda_i = de/d\lambda_i$ 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 $(\partial s_2/\partial b)(de_1/d\lambda_i)/(de/d\lambda_i) < 0$ 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 quantitatively equivalent to the non-marginal analysis if allowance demand in all periods and allowance supply adjustments (either in prices or in banks) are linear. In this case $\partial e_1/\partial \tau_1$, $\partial e/\partial \tau_1$, $\partial s/\partial \tau_1$ and $\partial s_2/\partial b$ are all constants.

B.4. Proof of equations (25) and (26) for the EU ETS

Equation (25) follows directly from the parameters presented in Table 2 and the explanation given in the main text. See also Perino (2018).

The instantaneous waterbed effect $\hat{W}(t_a, t, t_{B=833})$ measures the waterbed effect of a reduction in allowance demand in a single year (t). Hence, the β measuring the temporal distribution of a policy's impact that appears in Equation (20) is equal to 1. Since we now consider a setting with more than two periods and the time of announcement of the overlapping policy is no longer fixed, we need to explicitly take this into account. In a market with perfect intertemporal arbitrage prices will respond to the announcement of a policy t_a. The denominator of equation (20) capturing the price effect is therefore adjusted accordingly resulting in equation 26.

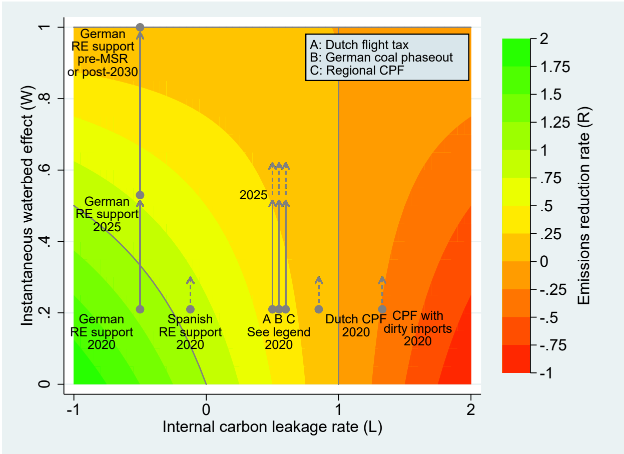
Appendix C: Sensitivity to $t_{B=833}$ and the Rosendahl effect

In the main text, 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 (see Equation (26)). 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.

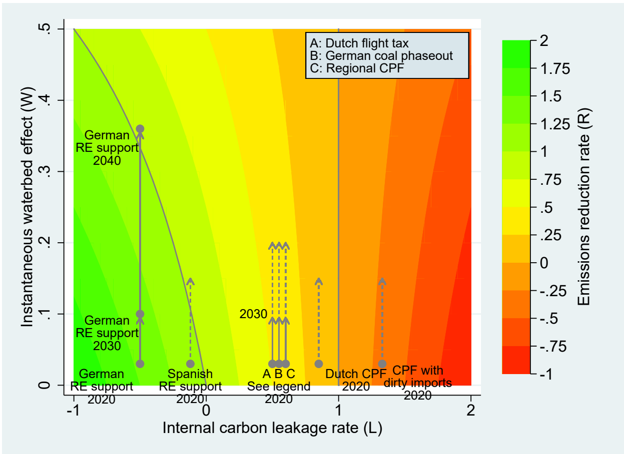
Panel (b) of Figure A1 shows that, compared to our original estimates in Panel (a), the instantaneous waterbed effect decreases substantially when $t_{B=833}$ lies further in the future. The waterbed effect can only go below 100% if the MSR takes in allowances; if allowances still flow into the MSR in 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 $\hat{W}_t$ substantially, 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.

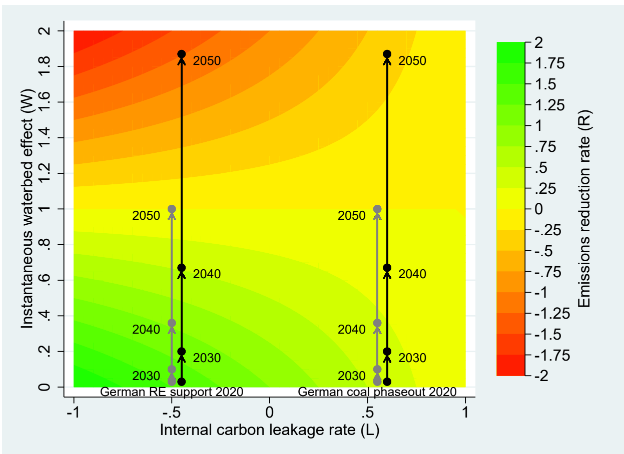
Figure A1: Leakage and waterbed effects in the EU ETS under varying assumptions



(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 2 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 $t_{B=833} = 2048$.