UNIVERSITY OF CAMBRIDGE | Energy Policy Research Group

Overlapping Climate Policies

EPRG Working Paper 2034

Cambridge Working Paper in Economics 20111

Grischa Perino, Robert A. Ritz and Arthur A. van Benthem

Abstract Major carbon-pricing systems in Europe and North America involve multiple jurisdictions (countries or states). Individual jurisdictions often pursue additional Initiatives – such as unilateral carbon price oors, legislation to phase out coal, aviation taxes or support programs for renewable energy – that overlap with the wider carbon-pricing system. We develop a general framework to study how the climate benefit of such overlapping policies depends on their design, location and timing. Some policies leverage additional climate benefits elsewhere in the system while others backfire by raising aggregate emissions. Our model encompasses almost every type of carbon-pricing system used in practice.

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

JEL Classification H23 (externalities), Q54 (climate)

Contact Robert Ritz – rar36@cam.ac.uk
Publication November 2020

www.eprg.group.cam.ac.uk

Overlapping Climate Policies*

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

November 19, 2020

Abstract

Major carbon-pricing systems in Europe and North America involve multiple jurisdictions (countries or states). Individual jurisdictions often pursue additional initiatives—such as unilateral carbon price floors, legislation to phase out coal, aviation taxes or support programs for renewable energy—that overlap with the wider carbon-pricing system. We develop a general framework to study how the climate benefit of such overlapping policies depends on their design, location and timing. Some policies leverage additional climate benefits elsewhere in the system while others backfire by raising aggregate emissions. Our model encompasses almost every type of carbon-pricing system used in practice.

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

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

1 Introduction

The world is under increasing pressure to deliver on the ambition of the 2015 Paris Climate Agreement, and over 60 national and sub-national jurisdictions are putting a price

*We thank Severin Borenstein, Dallas Burtraw, Jim Bushnell, Reyer Gerlach, Marten Ovaere, Sebastian Rausch, Mar Reguant, Knut Einar Rosendahl, Herman Vollebergh and Maximilian Willner for helpful comments and suggestions. 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, and the Analytics at Wharton Data Science and Business Analytics Fund 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.

on carbon emissions (World Bank, 2020). Two features of the carbon-pricing landscape are striking. First, by using hybrid designs that combine elements of price and quantity regulation, practice has run far ahead of the simple carbon tax and cap-and-trade policies emphasised by textbook economics. North American carbon markets—such as the Regional Greenhouse Gas Initiative (RGGI)—use price floors and ceilings to contain the variability of the allowance price. Since its 2018 reform, the European Union's Emissions Trading System (EU ETS) features a complex allowance cancellation mechanism. Second, major carbon-pricing systems involve multiple jurisdictions: the EU ETS covers 27 member states plus linked countries like Norway and the UK while RGGI involves ten states in the northeastern United States.

Individual jurisdictions, in turn, often pursue unilateral climate initiatives that overlap with the wider carbon-pricing system. The EU is a classic example, with individual countries "doing more" than what is centrally provided by the EU ETS. The UK in 2013 introduced a carbon fee that adds £18/tCO_2 to the allowance price faced by its power generators under the EU ETS; the Netherlands are committed to introducing a similar unilateral carbon price floor for electricity and industrial sectors.[1] There is a plethora of national policies to support renewable energy (notably solar and wind), and an increasing number of countries are legislating to phase out coal-fired power and impose additional carbon taxes on air travel.[2] These examples share a common feature: they are policies by an individual jurisdiction that operate alongside a wider carbon-pricing system.

Our question in this paper is simple: What is the climate benefit of such overlapping policies? As it is a global public good, any mitigation of climate change will be driven solely by changes in aggregate emissions. For a cap-and-trade system with a fixed emissions cap, like the pre-2018 EU ETS, the answer is clear: if an overlapping policy reduces EU-wide emissions demand (say, from power generation) by 1 ton of CO_2, this will be precisely offset by increased demand of 1 tCO_2 elsewhere in the system—the "waterbed effect" is 100%. At the opposite end, a simple carbon tax does not have an emissions cap and so the waterbed effect is zero. Our main interest, therefore, is in real-world hybrid carbon-market designs which typically feature dynamic "punctured" waterbeds that lie between these two extremes. A punctured waterbed enables overlapping policies to have a global climate benefit.

[1]The EU ETS includes power generation, industrial sectors, and domestic aviation and is the world's largest carbon-pricing system. The UK's Carbon Price Support 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).

[2]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. The Powering Past Coal Alliance currently includes 34 national and 33 sub-national governments including twelve EU member states committed to phasing out coal. Motivations for overlapping policies range from climate benefits to concerns about low or volatile carbon prices to other market failures such as innovation externalities (Newbery et al., 2019).

Yet this chain of reasoning still has a missing link which we refer to as "internal carbon leakage". Suppose that a unilateral Dutch carbon price on power generation reduces its domestic emissions demand by 1 tCO_2 but, within an integrated European electricity market, this leads to an increase in Dutch electricity imports which in turn raises emissions demand by 1 tCO_2 in other EU ETS countries. This overlapping policy has no climate benefit either: its rate of internal carbon leakage is 100%. This conclusion, in turn, applies irrespective of the extent of the waterbed effect. In sum, the answer to our question must be driven by a combination of the waterbed effect and internal leakage.

This paper provides a novel integrated approach through which to understand and quantify the overall emissions impact of an overlapping policy that applies only to part of a multi-jurisdiction carbon-pricing system. Section 2 presents a model-independent conceptual framework that provides a mapping from the "local" emissions reduction the overlapping policy achieves to its "global" impact which includes any knock-on effects elsewhere in the system. Internal carbon leakage captures emissions displacement within the system (e.g., greater product imports from a neighbouring country) for a given system-wide carbon price. The waterbed effect endogenises the policy's interaction with the system's carbon price (and any emissions cap).[3]

Section 3 introduces a theory of internal carbon leakage that focuses on emissions displacement between different jurisdictions in the same sector. We consider two groups of overlapping policies. First, "supply-side" policies that unilaterally raise the carbon price or directly limit emissions-intensive production. Second, "demand-side" policies that reduce the (residual) demand for emissions-intensive production, e.g., by promoting renewables or introducing a carbon-consumption tax. We show that supply-side policies feature positive internal carbon leakage—sometimes in excess of 100%—while demand-side policies have negative internal leakage. While some recent empirical work has estimated internal leakage for specific policies (Vollebergh, 2018; Abrell et al., 2019; Gerarden et al., 2020), our first contribution is to provide new theoretical insight into the economics of internal carbon leakage that applies across a range of commonly-used overlapping policies.[4]

Section 4 presents a general two-period analysis of the waterbed effect. While the literature has studied the waterbed effect in specific circumstances, notably the EU ETS (Fankhauser et al., 2010; Böhringer, 2014; Perino, 2018; Rosendahl, 2019a), there is still very limited understanding of its operation across different types of hybrid carbon-market designs. Our model encompasses price-based flexibility mechanisms based on past al-

[3]The strength and timing of the overlapping policy are exogenous in our framework; we are interested in the extent to which a given local reduction in emissions demand translates into system-wide emissions impacts for different overlapping policies and carbon-market designs.

[4]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.

lowance prices (including price ceilings and floors) (Roberts and Spence, 1976; Pizer, 2002; Newell et al., 2005; Borenstein et al., 2019; Burtraw et al., 2020), quantity-based flexibility mechanisms based on past allowance banking as well as a simple carbon tax and cap-and-trade.[5] We uncover a natural connection between the extent of the waterbed and classic principles from the literature on tax incidence (Jenkin, 1872; Weyl and Fabinger, 2013). Our second contribution, therefore, is to bring together results from prior literature in a unifying framework that covers almost every type of carbon-pricing system used in practice, and connect them to simple economic principles.

Section 5 illustrates the empirical usefulness of the modelling framework. It derives values for internal leakage and the waterbed effect using a combination of simple formulae from our theory results and prior empirical work. We cover overlapping policies in Europe and in North American carbon-pricing systems such as RGGI, the California-Québec carbon market, and Canada's new federal minimum carbon price (see Figure 4). Consistent with our theory, we find that supply-side (demand-side) overlapping policies have positive (negative) internal leakage. Our findings illustrate how a policy's overall climate benefit varies widely depending on its design, location and timing. Section 6 concludes.[6]

We hope that our analysis, by providing practical guidance on the climate benefits of 25 different combinations of overlapping policy instruments (see Figure 1) and types of carbon-pricing designs (see Figure 3), will be of value to policymakers trying "in real time" to gauge the attractiveness of domestic climate initiatives. A hybrid carbon-market design raises the stakes for what are often termed "complementary" policies: some are truly complementary in the sense that they induce further emissions reductions elsewhere while others can backfire by raising aggregate emissions.[7]

Finally, while some of the economic issues are similar, our focus in this paper on internal leakage differs from the external carbon leakage to jurisdictions *outside* a unilateral carbon-pricing system that has been the main concern of the literature (Fowlie, 2009; Hanna, 2010; Kahn and Mansur, 2013; Martin et al., 2014; Aldy and Pizer, 2015; Caron et al., 2015; Fowlie et al., 2016; Fischer et al., 2017; Fowlie and Reguant, 2018). This research has examined situations where the scope of the product market is wider than that of a carbon-pricing system; by contrast, we are interested in leakage among different

[5]A two-period model is necessary to be able to incorporate banking of allowances in a cap-and-trade system which, in turn, can interact with the extent of the waterbed. We also derive an analytical waterbed effect in a many-period representation of the EU ETS's Market Stability Reserve.

[6]Formal proofs are in Appendix A (internal carbon leakage) and Appendix B (waterbed effect).

[7]Our analysis in this paper focuses on the emissions impacts of overlapping policies which is only one—albeit important—component of a broader welfare analysis. Another salient aspect is the fiscal cost of different policies; a unilateral carbon tax raises additional government revenue while a renewables support program may be very costly. While we show some policies have much more favourable leakage and waterbed properties than others, we do not formally rank them. Welfare analysis can build on the results in our paper and is an important topics for future research on the economics of the environment.

jurisdictions *inside* the system.[8] The cement and steel industries are classic examples of the former; our illustrations of the latter are drawn from airline and electricity markets.

2 Conceptual framework

We begin by setting out a simple conceptual framework that encompasses a wide range of carbon-market designs and highlights the dual role of internal carbon leakage and the waterbed effect in determining the climate benefits of different overlapping policies.

Consider a multi-jurisdiction carbon-pricing system that may cover a single sector (like RGGI) or multiple sectors (like the EU ETS). An "overlapping policy", in general, is any unilateral policy that applies only to part of the system; our leading example is a policy by a single jurisdiction that applies only to a subset of competing firms in a sector. For simplicity, we consider two time periods, $t = 1, 2$, and think of the first period as the short run and the second period as the long run.[9] Denote by $\boldsymbol{\tau} = (\tau_1, \tau_2)$ the system-wide carbon price at each time, which is determined by the carbon-market design.

We are interested in unilateral policies by country (jurisdiction) i that, holding fixed the carbon price path $\boldsymbol{\tau}$, are successful at reducing i's domestic demand for emissions in each period, $\Delta e_{it} < 0$, and hence also $\Delta e_i \equiv \Delta e_{i1} + \Delta e_{i2} < 0$ over time. Let Δe^*_t denote the policy's impact on aggregate emissions across all countries at time t at equilibrium carbon prices (relative to a baseline without the unilateral policy). Our main question is, what is the policy's impact on long-run equilibrium emissions, $\Delta e^* \equiv \Delta e^*_1 + \Delta e^*_2$? This is the critical issue for the policy's effectiveness in combating climate change.

Our framework answers this question using two concepts: internal carbon leakage and the waterbed effect. Internal leakage captures emissions displacement within the system (e.g., greater product imports from a neighbouring country) for a given system-wide carbon price. We define the rate of internal carbon leakage associated with i's policy at time t as:

$$L_{it} \equiv -\Delta e_{-it}/\Delta e_{it}, \tag{1}$$

where Δe_{-it} is the change induced by i's policy in the emissions demand of other countries that are part of the carbon-pricing system.[10] Therefore $\Delta e_t \equiv [1 - L_{it}]\Delta e_{it}$ represents

[8] Another form of carbon leakage 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.

[9] We use a two-period model to incorporate banking of allowances in a cap-and-trade system. The logic of our framework otherwise also applies to a single-period setting.

[10] 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.

the (net) system-wide change in emissions demand at time t so $\Delta e \equiv \Delta e_1 + \Delta e_2$ is the long-run system-wide change in emissions demand due to the policy (for fixed $\boldsymbol{\tau}$).

The waterbed effect then captures the system-wide impacts arising from any induced changes to the equilibrium path of the system-wide carbon price. In particular, the extent of the waterbed effect is defined according to the following expression:

$$W \equiv 1 - \Delta e^*/\Delta e. \tag{2}$$

This translates the system-wide change in emissions demand due to i's policy into an equilibrium change in overall emissions that incorporates any induced changes to the carbon price path. A textbook cap-and-trade system with a fixed emissions cap has $W = 1$ while a carbon tax has $W = 0$; real-world hybrid carbon-market designs typically feature punctured waterbeds, $W \in (0, 1)$.[11]

We can now state the central equation of our conceptual framework. Let $\beta_i \equiv \Delta e_{i1}/\Delta e_i \in [0, 1]$ denote the fraction of the policy's domestic impact that occurs in the first period, and define $L_i \equiv (\beta_i[1 - L_{i1}] + (1 - \beta_i)[1 - L_{i2}])$ as the policy's weighted-average rate of internal carbon leakage across the two periods. We can then write the equilibrium change in long-run emissions as:

$$\Delta e^* = [1 - L_i][1 - W]\Delta e_i, \tag{3}$$

which incorporates the equilibrium carbon price path via the waterbed effect. This shows how internal carbon leakage and the waterbed effect together drive the sign and magnitude of the overlapping policy's impact on long-run equilibrium emissions, Δe^*. 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 "global" emissions rise ($\Delta e^* > 0$) even though "local" emissions fall ($\Delta e_i < 0$).

We do not attempt to quantify the policy's impact on i's domestic emissions demand, $\Delta e_i < 0$; rather we are interested in the mapping from a given policy-driven local impact Δe_i to the equilibrium global impact Δe^*. We also do not attempt to endogenise the extent to which the policy operates in the short run or the long run, β_i; rather we will explore later the impact of β_i on the extent of the waterbed effect. In short, we take the unilateral policy's size (Δe_i) and its period-by-period timing (β_i) as given, and ask to what extent it translates into a long-run global climate gain.

Leveraging this conceptual framework, the remainder of the paper proceeds in three

[11] The latter include caps adjusted based on prices (Burtraw et al., 2020), taxes adjusted based on (cumulative) emissions (Metcalf, 2020) and caps adjusted based on unused allowances (Newell et al., 2005; Perino, 2018).

steps. First, we derive the rate of internal carbon leakage L_{it} between different jurisdictions in the same sector for a range of "supply-side" and "demand-side" overlapping policies. Given the policy-timing parameter β_i, these can be aggregated to give the average leakage rate L_i. Second, given the policy's induced changes to aggregate emissions demand $(\Delta e_1, \Delta e_2)$, we derive the extent of the waterbed effect W under different carbon-market designs. Third, we illustrate the empirical usefulness of the framework by deriving values for (L_i, W) for real-world overlapping policies in Europe and North America.

3 A model of internal carbon leakage

We next present a new theory of internal carbon leakage arising at the sectoral level from unilateral policies within a wider carbon-pricing system. We consider two groups of overlapping policies. First, "supply-side" policies that unilaterally raise the carbon price for emissions-intensive production or directly reduce production, for example, by way of a coal phase-out. Second, "demand-side" policies that reduce the (residual) demand for emissions-intensive production, for example, by promoting renewables or energy efficiency or introducing a carbon consumption tax. We show that the economics of internal carbon leakage is similar for policies within each group but differs markedly across the two groups.

In the model, emissions-intensive firms in each of two countries decide on their production volume and abatement effort. Countries can be heterogeneous in their technologies along three dimensions: production costs, emissions intensities, and abatement opportunities. The benchmark with fixed emissions intensities of production is nested as a special case of the model, allowing us to clearly delineate the effects of abatement on internal leakage. The unilateral policies overlap with a carbon price that is common across countries (jurisdictions) within the wider carbon-pricing system and—as motivated by the conceptual framework—is held fixed.

For expositional convenience, we focus in the main text on a model in which production costs and abatement costs are separable. This yields simple intuitive formulae for the equilibrium rate of period-by-period internal carbon leakage L_{it}. The model is static so time subscripts are omitted to simplify notation. In Appendix A, we solve the model with general cost functions that allow non-separability between production and abatement; Section 3.4 provides a summary of the additional effects that arise when abatement decisions have an impact on the product-market equilibrium (and vice versa).

3.1 Model setup

There are two countries, i and j, where the latter can be interpreted as an aggregate of all countries except i. A representative firm in each country k produces output x_k

($k = i, j$) and the firms face a demand function $p(X)$ for their product, where $X \equiv x_i + x_j$ is total output and $p'(\cdot) < 0$. 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 in the absence of abatement for which its emissions intensity of output is θ_k, which can be thought of as the "dirtiness" of the marginal plant in country k.

Firm k has a cost function $G_k(x_k, a_k) \equiv [C_k(x_k) + \phi_k(a_k)]$, where $C_k(x_k)$ is its production cost function and $\phi_k(a_k)$ is its abatement cost function. For a well-behaved solution, we make standard assumptions about cost convexity: $C_k(0) = C_k'(0) = 0$, $C_k'(x_k) > 0$ for $x_k > 0$, and $C_k''(x_k) > 0$ as well as $\phi_k'(a_k) > 0$ for $a_k > 0$ and $\phi_k''(a_k) > 0$.[12] It will be useful to have a simple metric for the extent of abatement opportunity for firm k. To do so, we can think of its 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 additional incremental abatement being very cheap in that cost convexity $\phi_k'' \to 0$. This means that output and emissions become strongly detached: due to abatement, it is possible to reduce emissions without resorting to output reductions. Conversely, the case in which $A_k \to 0$ corresponds to a Leontief technology: emissions are proportional to output (perfect complements) due to abatement being infeasible, with $\phi_k'' \to \infty$ and hence emissions remain at their baseline level, $e_k = \theta_k x_k$. We can therefore use $A_k \in [0, 1)$ as a metric of firm k's abatement opportunities.[13]

Firm k faces a carbon price $\tau_k(\tau)$ on each unit of emissions, which depends on the carbon price τ that is common to both countries as part of a wider carbon-pricing system. As we shall see, a unilateral policy by country i that raises its domestic carbon price leads to $\tau_i(\tau) > \tau$; otherwise $\tau_i(\tau) = \tau$. As per our conceptual framework, the system-wide carbon price τ itself is held fixed in this leakage analysis (and endogenised in our subsequent waterbed analysis).

To maximise profits, firm k therefore solves $\max_{x_k, a_k} \Pi_k = p x_k - G_k(x_k, a_k) - \tau_k e_k$. Note it is equivalent for a firm to choose its emissions or abatement. With perfect competition

[12] This kind of separable cost function can be interpreted as featuring an end-of-pipe technology which cleans up production *ex post*. Examples include carbon capture and storage (CCS) and the purchase of carbon offsets. Appendix A solves the model for general cost functions $G_k(x_k, a_k)$ with a non-zero cross partial, $G_k^{xa}(x_k, a_k) \neq 0$.

[13] Note that $A_k < 1$ is equivalent to the stability condition $G_k^{xx} G_k^{ee} - G_k^{xe} G_k^{ex} > 0$. Firm k's cost function satisfies other standard assumptions made in environmental economics. 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 and emissions, with $G_k^{xx}(x_k, e_k) = C_k'' + \theta_k^2 \phi_k'' > 0$ and $G_k^{ee}(x_k, e_k) = \phi_k'' > 0$.

in the product market, the two first-order conditions for profit-maximisation are:

$$p(X) - C_k'(x_k) - \theta_k \phi_k'(a_k) = 0 \tag{5}$$

$$-\tau_k + \phi_k'(a_k) = 0. \tag{6}$$

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 first-order condition:

$$p(X) - C_k'(x_k) - \tau_k \theta_k = 0 \tag{7}$$

so product price is equal to the marginal cost of production plus per-unit carbon costs based on its baseline emissions intensity of output. It does not depend on the extent of abatement which, due to cost separability, does not affect the product-market outcome. At an interior solution, k's abatement incentive rises with its domestic carbon price, $da_k/d\tau_k = 1/\phi_k''(\cdot) > 0$ which, in turn, is independent of output. To guarantee an interior solution in the product market, we assume $p(0) > \max_k\{C_k'(0) + \tau_k\theta_k\}$.

Our main interest lies in characterising the internal carbon leakage associated with a unilateral policy introduced by country i alone. We assume, and later verify, that the unilateral policy is successful at reducing i's domestic emissions, $\Delta e_i < 0$. However, through the product market, i's unilateral policy may also induce changes in the behaviour of the firm in country j—and hence also lead to a change in its emissions, Δe_j. (Given that the system-wide carbon price is held fixed, changes in emissions and changes in emissions demand are equivalent in our analysis of internal leakage.)

We define the rate of internal carbon leakage as $L_i \equiv (-\Delta e_j/\Delta e_i)$. In the benchmark case without abatement, firms' emissions intensities are fixed so carbon leakage is $L_i = (\theta_j/\theta_i)\,(-\Delta x_j/\Delta x_i)$, where the first term is the "relative dirtiness" of firms and the second term is *output* leakage, so also let $L_i^{\mathrm{O}} \equiv (-\Delta x_j/\Delta x_i)$. To simplify the analysis, as is common practice, we focus on "marginal" unilateral policies for which the carbon-leakage rate is approximated by $L_i \simeq (-de_j/de_i)$.[14] That is, we consider a small deviation from the initial equilibrium in which there is no unilateral policy.

Some equilibrium definitions will prove useful to cast our leakage 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 country 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 $\eta_k^S \equiv x_k\widehat{C}_k''(x_k)/\widehat{C}_k'(x_k) > 0$ as its elasticity, also noting that $\widehat{C}_k''(x_k) \equiv C_k''(x_k)$.[15] By k's first-order condition, $x_k'(p) = 1/C_k''(x_k) > 0$, i.e., its supply curve is upward-sloping. So $\varepsilon_k^S \equiv px_k'(p)/x_k(p) > 0$ is k's

[14] In Section 3.4, we discuss the general case with non-marginal policy changes.

[15] Note that $\widehat{C}_k'(x_k) \equiv G_k^x(x_k, a_k)$ where abatement is optimally chosen according to $\tau_k = \phi_k'(a_k)$.

price elasticity of supply and, at the firm's optimum, $\eta_k^S = 1/\varepsilon_k^S$. These expressions are all evaluated at the output levels of the initial equilibrium without any unilateral policy.

3.2 "Supply-side" unilateral policies

We begin by considering two kinds of "supply-side" policies that unilaterally raise the carbon price for emissions-intensive production or directly reduce production, for example, by way of a coal phase-out. We will see that such policies have positive internal leakage which, under some circumstances, can exceed 100%.

For concreteness, we can think of the demand curve $p(X)$ as that of consumers in country i who are served partly by domestic production and partly by imports from j. This is a natural interpretation, for example, in the context of an integrated market for electricity or in the context of aviation where i's consumers have a choice of whether to use an airport in i or j. Internal carbon leakage then captures the extent to which i's consumers are increasingly served by j's producers. As will become clear, this interpretation also facilitates the comparison between different types of unilateral policies.[16]

3.2.1 Unilateral carbon pricing

Our first overlapping policy λ_i imposes an additional carbon price only in country 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 designed to "top up" the system-wide carbon price, $\tau_i = \tau + \lambda_i$, like for Great Britain's Carbon Price Support for power generation that runs alongside the EU ETS. Another possibility could be a unilateral policy that seeks to lift i's carbon price towards a higher target level $\widehat{\tau}_i$, say with $\tau_i = \tau + \lambda_i(\widehat{\tau}_i - \tau)$. Our setup is general in that it allows the top-up to be non-uniform. 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 j's carbon price remains unchanged, its 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 is signed by j's output response.

Proposition 1 *A unilateral carbon price by country i has internal carbon leakage to country j of:*

$$L_i = \frac{\theta_j}{\theta_i} \frac{\frac{\sigma_j}{(\sigma_j+\varepsilon^D/\varepsilon_j^S)}}{\left[1 + \frac{A_i}{(1-A_i)}\left(1 + \frac{(1-\sigma_j)\varepsilon_i^S/\varepsilon_j^S}{(\sigma_j+\varepsilon^D/\varepsilon_j^S)}\right)\right]} > 0.$$

[16] An alternative interpretation is that $p(X)$ reflects aggregate consumer demand across both countries.

Proposition 1 provides a simple formula to quantify internal carbon leakage. Carbon leakage is always positive as the underlying output leakage is positive—due to the asymmetric cost shock, i's firm loses market share to j's. Output leakage is always less than 100% as i's policy raises the market price, i.e., there is positive carbon cost pass-through. Yet carbon leakage can exceed 100% if j's firm is sufficiently dirtier than i's.

To understand the expression for L_i, consider the benchmark with zero abatement by i's firm, $A_i = 0$. The comparative statics are intuitive: output leakage 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). Output leakage then maps into carbon leakage by way of the relative emissions intensity θ_j/θ_i, so is more pronounced if j has relatively dirtier production.

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.

Greater abatement opportunities by country i, as measured by a higher value of $A_i > 0$, mitigate internal carbon leakage. 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. In the limit, with near-costless additional abatement, carbon leakage tends to zero—even if it may have been very high without any abatement (i.e., $L_i \to 0$ as $A_i \to 1$ as $\phi_i''(\cdot) \to 0$).

Finally, observe that the formula for L_i does not depend on the precise functional form of the carbon price in country i $\tau_i = \tau_i(\tau, \lambda_i)$; at the margin, this matters for the *absolute* output and emissions impacts but not for the *relative* effects—which is what our leakage rate for a marginal policy captures.

To get a sense for magnitudes, it is helpful to consider a numerical example. Suppose that the demand elasticity $\varepsilon^D = \frac{1}{2}$ and that j has have market share $\sigma_j = 20\%$ with a supply-responsiveness determined by a modest cost elasticity $\eta_k^S = .2 \Leftrightarrow \varepsilon_j^S = 5$. These parameter values might be plausible in the context of electricity markets in which country i imports some of its consumption from generators located in j. As a baseline, countries have identical emissions intensities, $\theta_i = \theta_j$, and there is no abatement, $A_i = 0$. Then we find internal carbon leakage $L_i = 67\%$, driven solely by output leakage. Now suppose instead that j's firm is twice as dirty (which approximates emissions from coal vs gas-fired generation); then internal leakage also doubles to $L_i = 133\%$. Finally, let country i have significant abatement opportunity as implied by $A_i = \frac{1}{4}$, with the same supply elasticity $\varepsilon_i^S = 5$. This yields a large drop in internal leakage back down to $L_i = 60\%$, showing how abatement can help bring forth an overall emissions reduction.

3.2.2 Unilateral reduction in carbon-intensive production

Our second policy has country i institute a unilateral reduction in carbon-intensive production. A topical example of such a policy is a phase-out of coal-fired power generation, which a number of European countries have individually committed to—again alongside these plants being covered by the EU ETS. Formally, we suppose that i's policy λ_i imposes a (marginal) reduction in i's output, $dx_i < 0$. In contrast to the previous policy, this policy does not affect the carbon price faced by i's firm so $\tau_k = \tau$ for $k = i, j$.

The economics of internal carbon leakage is similar to the first policy: the unilateral reduction in i's production leads to a degree of output leakage, as j's firm "fills the gap" in market supply, which then translates into carbon leakage. An important difference, however, is that the reduction in carbon-intensive production is directly mandated by policy rather than induced in equilibrium by i's unilateral carbon price—and therefore does not incentivise any abatement.

Proposition 2 *A unilateral reduction in carbon-intensive production by country i has internal carbon leakage to country $j$of:*

$$L_i = \frac{\theta_j}{\theta_i} \frac{\sigma_j}{(\sigma_j + \varepsilon^D / \varepsilon_j^S)} > 0.$$

The formula of Proposition 2 is a special case of Proposition 1 in which country i's abatement is zero. Like before, the rate of output leakage satisfies $L_i^{\mathrm{O}} = \sigma_j/(\sigma_j + \varepsilon^D/\varepsilon_j^S) \in (0, 1)$, with intuitive properties, and this translates into carbon leakage by way of the relative emissions intensity θ_j/θ_i. It is also clear that, overall, a unilateral reduction in carbon-intensive production has less attractive internal leakage than a unilateral carbon price: the absence of induced abatement means that a key factor that can mitigate carbon leakage is not present.

With our earlier parameter values, $\sigma_j = 20\%$, $\varepsilon^D = \frac{1}{2}$, $\varepsilon_j^S = 5$ and $\theta_i = \theta_j$, we again find $L_i = 67\%$. If j's technology is less responsive with $\eta_j^S = 1 \Leftrightarrow \varepsilon_j^S = 1$ or market demand is more elastic with $\varepsilon^D = 2\frac{1}{2}$, then leakage falls considerably to $L_i = 28\%$.[17]

3.3 "Demand-side" unilateral 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

[17] Our exposition here is in terms of unilateral action by one country within a multi-jurisdiction carbon-pricing system. An alternative interpretation is that i and j are players within a single jurisdiction. Then Proposition 2 derives within-country internal carbon leakage from coal-fired generation to, say, gas-fired generation in response to a (policy-induced) cut in the former.

program, and a carbon-consumption tax. We will see that these have markedly different properties from supply-side policies: their internal carbon leakage is negative.

We retain the interpretation that the demand curve reflects that of consumers in country i who are served partly by domestic production and partly by imports from j. Formally, we model a unilateral policy λ_i by country i and write the demand curve as $p(X; \lambda_i)$ where $\frac{\partial}{\partial \lambda_i} p(X; \lambda_i) < 0$ so the unilateral policy reduces demand, and the residual demand of i's firms—but also that of j's firms. Both firms continue to face the common carbon price $\tau_i = \tau_j = \tau$.

The policies fit into this setup as follows. First, for i's renewables support program, we write demand as $p(X; \lambda_i) = p(X + \lambda_i)$ where λ_i is the volume of zero-carbon renewables production supported by the policy. Second, for the energy-efficiency program, write direct demand as $D(p; \lambda_i) = (1 - \lambda_i)D(p)$ so it reduces demand by a fraction λ_i (for a given p) and hence $p(X; \lambda_i) = D^{-1}(X/(1 - \lambda_i))$. Third, for the carbon-consumption tax, write demand as $p(X; \lambda_i) = [p(X) - \lambda_i \theta_i]$ 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 3 *The unilateral policies by country i of (i) a renewables support program that brings in additional zero-carbon production, (ii) an energy-efficiency program that reduces demand for carbon-intensive production, and (iii) a carbon consumption tax have identical internal carbon leakage to country j of:*

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

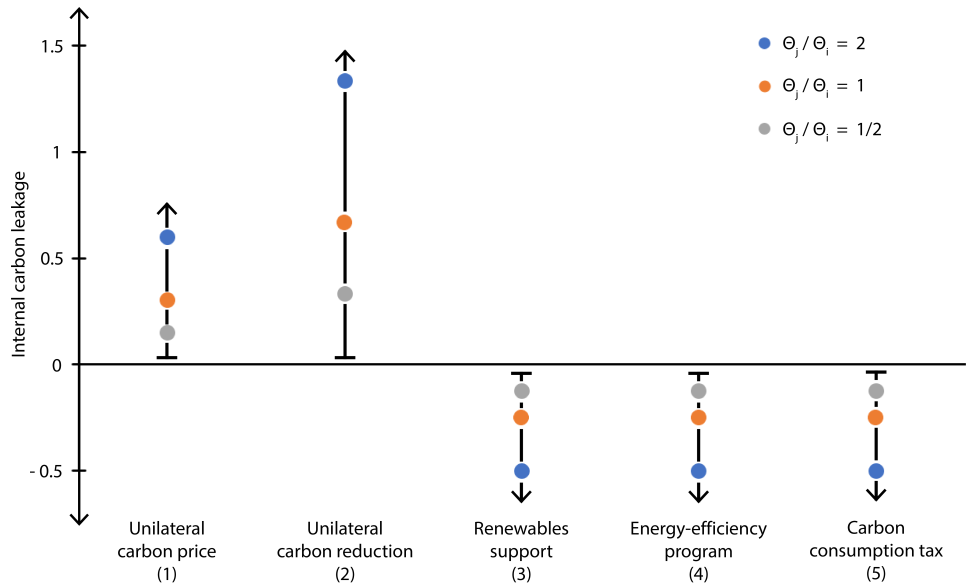
Internal carbon leakage is always negative: j's firm is now directly affected by the policy and responds by also cutting output and emissions. This means that the "global" emissions reduction here is more pronounced than the "local" reduction. Akin to Propositions 1 and 2, leakage is more strongly negative 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 for both countries, unilateral action here brings no extra abatement incentive (i.e., $da_k/d\lambda_i = 0$ for $k = i, j$).

Proposition 3'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 unilateral policies have identical leakage properties.

To illustrate magnitudes, 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 "global" reduction in emissions demand is now twice the size of the "local" reduction.

Summarising our analysis thus far, Figure 1 graphically illustrates the key differences in rates of internal carbon leakage across different overlapping policies.

Figure 1: Internal carbon leakage for different types of unilateral climate policies



Notes: Based on Propositions 1-3 and parameter values $\sigma_j = 20\%$, $\varepsilon^D = \frac{1}{2}$, $\varepsilon_i^S = \varepsilon_j^S = 5$, $A_i = \frac{1}{4}$, and using three different values of the relative emissions intensity θ_j/θ_i.

3.4 Extensions and robustness

Different types of unilateral action can yield very different outcomes in terms of internal carbon leakage. Three overarching drivers of leakage are (i) the countries' relative baseline emissions intensities, θ_j/θ_i (ii) output leakage as driven by market shares and demand/supply elasticities, L_i^O and (iii) the potential for emissions abatement, A_i, A_j.[18]

We next develop some general intuition for internal leakage, and then discuss the implications of non-separability of cost functions and of our focus on marginal policies as well as an alternative definition of internal carbon leakage.

[18] A unifying feature across all our unilateral policies is that internal carbon leakage tends to zero where either j's market share is tiny, $\sigma_j \to 0$, j is supply-unresponsive, $\varepsilon_j^S \to 0$, or j is already zero-carbon, $\theta_j/\theta_i \to 0$. In the first two cases, output leakage is zero; in the third, the change in j's emissions is zero.

General intuition. We begin with a model-independent decomposition that confirms how the drivers of internal leakage from our simple equilibrium model are more generally applicable. Consider any (marginal) unilateral policy $d\lambda_i$ that is successful at reducing country i's domestic emissions, $de_i/d\lambda_i < 0$. Recalling that firm k's emissions $e_k = e_k^0 - a_k$, where baseline emissions $e_k^0 = \theta_k x_k$, allows us to write the first-order rate of internal carbon leakage as:

$$L_i \equiv \frac{de_j/d\lambda_i}{-de_i/d\lambda_i} = \frac{\theta_j}{\theta_i} L_i^{\mathrm{O}} \frac{(1-\alpha_j)}{(1+\alpha_i)} \tag{8}$$

where $L_i^{\mathrm{O}} = (dx_j/d\lambda_i)/(-dx_i/d\lambda_i)$ is output leakage and $\alpha_i \equiv (da_i/d\lambda_i)/(-de_i^0/d\lambda_i) \geq 0$ and $\alpha_j \equiv (da_j/d\lambda_i)/(de_j^0/d\lambda_i)$ are market-based measures of the extent of abatement, relative to the change in baseline emissions, by the two countries.[19]

This expression shows how internal carbon leakage is generally driven by the same three factors identified in our simple model: relative emissions intensities, output leakage, and abatement effects. Moreover, it remains true that (i) output leakage typically determines the sign of carbon leakage, (ii) a positive leakage rate is mitigated by firms' abatement opportunities, and (iii) the relative baseline emissions intensity acts as a scaling factor. So these qualitative properties from our simple model are robust in a more general context.

Non-separable technology. A simplifying assumption is that firm k's cost function $G_k(x_k, a_k) \equiv [C_k(x_k) + \phi_k(a_k)]$ has a zero cross-partial, $G_k^{xa}(x_k, a_k) = 0$, so output and abatement decisions are independent. Separability is a common assumption but also restrictive. For a general cost function $G_k(x_k, a_k)$ the first-order conditions are $p = G_k^x + \tau_k\theta_k$ and $\tau_k = G_k^a$. An immediate implication is that abatement can now be induced by changes to output (via $G_k^{xa} \neq 0$) so is no longer solely tied to the carbon price.

Consider our supply-side policy of a unilateral reduction in emissions-intensive production. In Appendix A, we show that the rate of internal carbon leakage now becomes:

$$L_i = \frac{\theta_j}{\theta_i} L_i^{\mathrm{O}} \frac{\delta_j}{\delta_i} > 0, \tag{9}$$

generalising Proposition 2, where $\delta_k = (1 + G_k^{ax}/\theta_k G_k^{aa}) > 0$ captures the extent of non-separability.[20] If $\delta_k < 1 \Leftrightarrow G_k^{xa} < 0$ $(k = i, j)$, given the rate of output leakage L_i^{O}, j tends to abate more given its output increase—which pushes internal leakage down; at the same time, however, i's output reduction undermines its abatement incentive—which pushes internal leakage up. Output leakage L_i^{O} itself is more pronounced than for a separable technology. If $G_j^{xa} < 0$, then abatement raises the marginal return to output, and vice

[19]Note that $\alpha_k = 0$ if and only if $A_k = 0$ in terms of our earlier measure of abatement.

[20]The property $\delta_k > 0$ corresponds to abatement increasing k's optimised marginal cost of production.

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, the non-separability effect 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 $G_j^{xa} = 0$.

Internal carbon leakage for the three demand-side policies (Proposition 3) is affected by non-separability in a similar way. For these, the impact on the rate of output leakage L_i^{O} is ambiguous as both countries now experience a direct change on their marginal return to output. Output leakage, all else equal, will be exacerbated—that is, become more negative—if the non-separability effect is stronger for j. Given output leakage, the same two contrasting effects as above are again at play (via δ_j/δ_i). The other supply-side policy of a unilateral carbon price (Proposition 1), in addition to these effects, features the familiar carbon-price-induced abatement by i.

In sum, the overarching insight from our simple model with separability is robust: demand-side policies have negative internal leakage while supply-side policies do not.

Non-marginal policy. Our analysis has focused on marginal unilateral policies that shift the equilibrium by a small amount, via the approximation $L_i \equiv \frac{\triangle e_j}{-\triangle e_i} \simeq \frac{de_j}{-de_i}$. Clearly, this is a good approximation if changes in policy are relatively modest; how good is it more generally? As we have seen, marginal rates of internal leakage depend on first-order derivatives of demand (via the demand elasticity) and second-order derivatives of cost functions (via supply elasticities and abatement opportunities). This is true in our simple model but also more generally with additional second-order cost terms due to non-separability. The key point is that the non-marginal leakage rate will be similar to marginal leakage as long as either second-order demand terms and third-order cost terms are small. This obtains exactly if the demand curve is linear ($p'(X)$ is constant) and the cost functions are quadratic in output and abatement ($G_k^{xx}, G_k^{aa}, G_k^{xa}$ all constant). In such cases, we have that $L_i \equiv \frac{\triangle e_j}{-\triangle e_i} = \left[\int_0^{\lambda_i} \left(\frac{de_j}{-de_i}\right)\left(\frac{de_i}{d\lambda_i}\right) d\lambda_i\right] \Big/ \left[\int_0^{\lambda_i} \left(\frac{de_i}{d\lambda_i}\right) d\lambda_i\right] = \frac{de_j}{-de_i}$ since the marginal leakage rate is constant with respect to λ_i. Given the ubiquity of linear-quadratic models in economics, this suggests that our simple formulae should be acceptable approximations to more complex non-marginal leakage rates. The simple formulae contain no obvious bias; they could be an over- or underestimate depending on whether demand is convex or concave and on the precise higher-order properties of cost functions.

Finally, we should point out that the equivalence of our three demand-side policies in terms of internal leakage (Proposition 3) is tied to marginal changes; this, in effect, implies that the three policies are identically-sized and therefore directly comparable. If these unilateral policies had different "sizes" in terms of the non-marginal policy change $\triangle\lambda_i \equiv \int_0^{\lambda_i} d\lambda_i > 0$, then their leakage properties are no longer necessarily identical. However, they would remain identical if, as per the previous argument, the rate of internal carbon leakage is itself invariant as in a linear-quadratic model.

"Total" internal carbon leakage with an endogenous allowance price. As motivated by our conceptual framework, our leakage analysis has taken the system-wide carbon price τ as given and fixed; our subsequent waterbed analysis will endogenise the allowance price for a cap-and-trade system. An alternative definition of "total" internal carbon leakage directly features any induced change in the common carbon price:

$$L_i^{\mathrm{T}} = \frac{\frac{de_j}{d\lambda_i} + \frac{de_j}{d\tau}\frac{d\tau}{d\lambda_i}}{-\left(\frac{de_i}{d\lambda_i} + \frac{de_i}{d\tau}\frac{d\tau}{d\lambda_i}\right)}. \tag{10}$$

where typically $de_k/d\tau < 0$ for $k = i, j$. For a carbon tax, the two leakage concepts are equivalent, $L_i^{\mathrm{T}} = L_i$, as then $d\tau/d\lambda_i = 0$ (no waterbed effect). For cap-and-trade systems, depending on the details of the unilateral policy, we instead have $d\tau/d\lambda_i \neq 0$ so there can be a wedge between L_i^{T} and L_i.

Consider a demand-side unilateral policy and suppose that countries have identical emissions intensities $\theta_i = \theta_j$. From the viewpoint of firm k, it is equivalent for its demand to decline by a small amount or its cost to increase by a small amount; the demand-side policies, as we have seen, have a symmetric impact on both countries and, since $\theta_i = \theta_j$, this is also true for the system-wide change in τ—so we have that $L_i^{\mathrm{T}} = L_i < 0$. By contrast, if $\theta_i > \theta_j$ then also $L_i^{\mathrm{T}} < L_i$ (and vice versa). For supply-side unilateral policies, we typically have $L_i > 0$ so expect that total leakage is lower than our measure, $L_i^{\mathrm{T}} < L_i$.

In sum, supply-side unilateral policies will typically, if not always, look better in terms of total internal carbon leakage than with a fixed carbon price while the wedge between the measures is more detail-dependent for demand-side policies. In general, as long as $d\tau/d\lambda_i \simeq 0$ for i's policy, we expect that the two concepts give similar results, $L_i^{\mathrm{T}} \simeq L_i$.

4 A model of the waterbed effect

We now derive the second building block of our conceptual framework: the waterbed effect $W = 1 - \Delta e^*/\Delta e$ of overlapping policies under a wide range of real-world carbon-pricing schemes. The carbon price path $\boldsymbol{\tau}$ that was treated as given in Section 3 is now derived endogenously.

Consider a stylised two-period model of an intertemporal allowance market. By design, the allowance market is geographically blind. The policy's impact is thus represented by Δe and β, where $\Delta e = [1 - L_i]\Delta e_i$ is the total change in allowance demand at a given carbon price and $\beta = \Delta e_1/\Delta e$ is the fraction of this impact that is effective in period 1.[21]

[21] We can write $\beta = \frac{\beta_i[1-L_{i1}]}{\beta_i[1-L_{i1}]+(1-\beta_i)[1-L_{i2}]}$ to express it in terms of previous timing and leakage

We assume a baseline inverse aggregate demand function for allowances $\rho_t(e_t)$ where e_t are aggregate emissions in period $t = 1, 2$ and so $\rho_1(e_1 - \beta\Delta e)$ and $\rho_2(e_2 - (1-\beta)\Delta e)$ are the post-intervention inverse demands. Again, we focus on marginal unilateral policies overlapping the carbon-pricing scheme. We assume that any overlapping policy is announced at the beginning of period 1, regardless of the timing of its impacts. Hence, policy impacts are perfectly anticipated by all market participants. We restrict attention to markets with perfect intertemporal arbitrage in which any borrowing and banking constraints do not bind, i.e., that $\tau_2 = (1+r)\tau_1$. For a carbon tax or a binding price corridor (i.e., a combination of a price floor and a price ceiling), the interest rate r reflects an exogenous increase in the carbon price.

We first analyse how an anticipated shift in allowance demand affects total emissions and the equilibrium price of allowances when the carbon-pricing scheme features a (weakly) increasing allowance supply function. Then we consider a design where the long-run cap is adjusted based on banked allowances as has been the case in the EU ETS since the 2018 reform (Perino, 2018). Due to peculiar design features that cannot be fully captured by a two-period model, we conclude this section by taking a closer look at the EU ETS's Market Stability Reserve.[22]

4.1 Flexibility mechanisms based on past allowances prices

We begin with flexibility mechanisms based on past allowance prices that specify variations of (weakly) upward-sloping allowance supply curves.[23] A carbon tax, a plain cap-and-trade system with a fixed cap as well as carbon price floors and ceilings are all special cases. Hence, these mechanisms include most carbon-pricing designs observed in the real world such as the California-Québec scheme and RGGI (Burtraw et al., 2020), the pre-2018 EU ETS (Perino and Willner, 2016), and all carbon taxes.[24]

variables. Note that $\beta = 0$ if $\beta_i = 0$ and $\beta = 1$ if $\beta_i = 1$. Depending on the rates of internal carbon leakage in each period, β can be negative or exceed 100%. For all demand side policies and for supply-side policies in relatively dirty countries (i.e., $L_{it} < 1$ for both periods) the country-specific and the aggregate variables are positively related. A constant leakage rate ($L_{i1} = L_{i2} = L_i$) implies $\beta = \beta_i$ and $\Delta e = \Delta e_{i1}(1 - L_i)/\beta_i$.

[22]In this section we restrict attention to cap adjustments brought about by formally codified flexibility mechanisms. On top of that unilateral policies might impact future caps by changing the outcome of future political negotiations about the stringency of the carbon-pricing system.Newell et al. (2005); Kuusela and Lintunen (2020); Pizer and Prest (2020) present models of policy updating but do not explicitly discuss the impact of unilateral policies on the political economy of re-negotiations.

[23]Such mechanisms have been described, among others, by Roberts and Spence (1976); Unold and Requate (2001); Pizer (2002); Newell et al. (2005); Hepburn (2006); Burtraw et al. (2020) and Traeger et al. (2020). In principle our model can also capture downward-sloping allowance supply curves. However, both from an environmental and an economic perspective, we are not aware of any arguments in their favour; nonetheless, the next subsection shows that—at least implicitly—they do exist.

[24]The Emissions Assurance Mechanism proposed by Metcalf (2020) adjusts tax rates based on cumulative emissions. It therefore specifies prices as a function of quantities rather than quantities as a function

Allowance supply is given by a fixed number of allowances s_1 issued in period 1 and a flexible number of allowances, $s_2(\tau_1)$, issued in period 2 with $\partial s_2/\partial\tau_1 \geq 0$. A cap-and-trade scheme with a fixed cap or any vertical section of an allowance supply curve are represented by $\partial s_2/\partial\tau_1 = 0$. A carbon tax or a horizontal section of an allowance supply curve, such as a price floor, are represented by $\partial s_2/\partial\tau_1$ being perfectly price elastic at a particular $\bar{\tau}_1$.[25]

The three equilibrium conditions of this carbon-market design are:

$$\begin{aligned} \rho_1(e_1 - \beta\Delta e_1) - \tau_1 &= 0 \quad (11) \\ \rho_2(e_2 - (1-\beta)\Delta e_2) - (1+r)\tau_1 &= 0 \quad (12) \\ e_1 + e_2 - s_1 - s_2(\tau_1) &= 0, \quad (13) \end{aligned}$$

where (11) and (12) balance marginal costs of abatement with the carbon price for periods 1 and 2, respectively, while (13) is the market-clearing condition for the allowance market.

Using Cramer's rule and the implicit function theorem (see Appendix B), the equilibrium conditions yield the impact of the unilateral policy on the system-wide equilibrium carbon price:

$$\begin{aligned} \frac{\partial\tau_1}{\partial\Delta e} &= \frac{1}{\frac{\partial s_2}{\partial\tau_1} - \frac{\partial e}{\partial\tau_1}} > 0 \quad (14) \\ \frac{\partial\tau_1}{\partial\beta} &= 0 \quad (15) \end{aligned}$$

where $\partial e/\partial\tau_1 < 0$ is the slope of the total allowance demand curve. The change in the equilibrium allowance price is key to identify the waterbed effect. (It also allows to compute total internal carbon leakage that incorporates the induced price change; see Section 3.4). Temporal and geographical distributions of the change in allowance demand are irrelevant. Note that the former is in stark contrast to what we find for flexibility mechanisms based on past banking in subsection 4.2. Adjustments in total equilibrium emissions e^* are also independent of how the unilateral policy is spread over time and space:

$$\frac{\partial\Delta e^*}{\partial\Delta e} = \frac{\partial s_2}{\partial\Delta e} = \frac{\partial s_2}{\partial\tau_1}\frac{\partial\tau_1}{\partial\Delta e} = \frac{\frac{\partial s_2}{\partial\tau_1}}{\frac{\partial s_2}{\partial\tau_1} - \frac{\partial e}{\partial\tau_1}} = \frac{\kappa^S}{\kappa^S + \kappa^D} \quad (16)$$

of prices. While the formal representation would slightly differ from the setup used below, qualitatively the mechanisms mimics a flexibility mechanism based on past allowances prices.

[25]Then the mechanism fixes prices at $\tau_1 = \bar{\tau}_1$ and $\tau_2 = (1+r)\bar{\tau}_1$. Furthermore, if we interpret r not as the rate of time preference but as a policy parameter, i.e. the tax rate escalator, it captures all cases of period-specific linear tax schemes.

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

Proposition 4 *The waterbed effect for a marginal policy overlapping a carbon-pricing scheme with a (weakly) increasing allowance supply and strictly decreasing allowance demand is:*

$$W \;=\; 1 - \frac{\partial \Delta e^*}{\partial \Delta e} = \frac{-\frac{\partial e}{\partial \tau_1}}{\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}} = \frac{\kappa^D}{\kappa^S + \kappa^D} \in [0, 1], \tag{17}$$

which is independent of the specifics of the overlapping policy ($\Delta e_{it}, \beta_i$) and leakage rates (L_{it}).

Proposition 4 shows that the waterbed effect depends only on characteristics of total allowance demand and supply—and is independent of the type of overlapping policy, its temporal and geographical impacts, and its internal carbon leakage.

Equation (17) has at opposite ends of the spectrum a zero waterbed for a carbon tax ($\partial s_2/\partial \tau_1 \to \infty$) and a 100% waterbed effect under a plain cap-and-trade system ($\partial s_2/\partial \tau_1 = \kappa^S = 0$). In a deterministic setting and for marginal changes, i.e., policies inducing relatively small shifts in allowance demand, this conclusion applies also to stepwise allowance supply functions featured in the California-Québec scheme and RGGI. If the initial equilibrium is in a vertical (horizontal) section of the supply curve, the waterbed effect is 100% (zero).

Once larger interventions are considered—such that allowance demand moves across one or several kinks in the 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.[26] 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.[27] If the probability that any of the price bounds binds is π, then $E(W) = \pi$. Ex-post the waterbed effect is either zero or 100%. As in the case with strictly upward-sloping allowance supply curves, the waterbed effect then increases in the elasticity of allowance demand (κ^D) and decreases in the elasticity of allowance supply (κ^S).

Proposition 4 also uncovers a natural connection between the waterbed effect and classic principles from the literature on tax incidence (Jenkin, 1872; Weyl and Fabinger, 2013). In particular, note that condition (17) corresponds to the cost pass-through rate from the tax-incidence literature. Since the allowance supply is assumed to be (weakly)

[26] The impact of a small change in the ambition of a large policy is again accurately described by marginal analysis.

[27] See Borenstein et al. (2019) for an assessment of such uncertainties in the California carbon market.

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.[28]

4.2 Flexibility mechanisms based on past allowance banking

With the 2018 reform of the EU ETS, namely the introduction of cancellations within the Market Stability Reserve, an entirely new form of flexibility mechanism entered the scene. Here we present a stylised two-period version of such a mechanism. It adjusts the long-run cap based on the number of allowances banked for future use in early periods, $s_2(b)$ where $b = s_1 - e_1$ is banking at the end of period 1 and $\partial s_2/\partial b \in [-1, 0]$.[29] A plain cap-and-trade scheme is again included as a special case ($\partial s_2/\partial b = 0$).

The responsiveness of the adjustment mechanism, given by $\partial s_2/\partial b$, allows the waterbed effect to deviate from 100%. However, in contrast to the price-based flexibility mechanism, the waterbed effect here is not fully determined by the mechanism and the slope of allowance demand. Given both pre-intervention allowance demand functions and $\partial s_2/\partial b$, the waterbed effect can still take virtually any value, driven by the timing of the impact of the overlapping policy. Depending on whether it occurs in period 1 or 2, a given shift in total allowance demand has very different impacts on first-period allowance banking. As banking balances temporal differences in allowance scarcity, it is thus intuitive that the waterbed effect depends on the timing of demand shifts.

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

$$
\begin{aligned}
\rho_1(e_1 - \beta\Delta e) - \tau_1 &= 0 && (18)\\
\rho_2(e_2 - (1-\beta)\Delta e) - (1+r)\tau_1 &= 0 && (19)\\
e_1 + e_2 - s_1 - s_2(s_1 - e_1) &= 0. && (20)
\end{aligned}
$$

Cramer's rule and the implicit function theorem (see Appendix B) yield the response of

[28] The trade-off between the price and quantity response due to an exogenous shift in allowance demand is also at the heart of instrument choice (Weitzman, 1974). Phrased in the terms of our model, Weitzman's famous criterion states that a waterbed effect of zero (plain tax) is preferred over a waterbed effect of 100% (plain cap-and-trade) if and only if the slope of marginal damages from emissions is larger than the slope of firms' allowance demand. This serves as a reminder that a waterbed effect of zero is not necessarily preferred on welfare grounds.

[29] Restricting $\partial s_2/\partial b > -1$ is somewhat arbitrary as one could imagine schemes with more responsive rules. However, imposing this lower bound simplifies the analysis and includes the entire range of values relevant for the EU ETS. For details see Lemma 1.

short-run equilibrium emissions to the overall change in allowance demand:

$$\frac{\partial e_1^*}{\partial \Delta e} = -\frac{1}{1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \cdot \left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}} - \beta\right]. \tag{21}$$

The sign of the first term is unambiguously positive (noting that $\partial s_2/\partial b \in [-1, 0]$ and $(\partial e_1/\partial \tau_1)/(\partial e/\partial \tau_1) \in (0, 1)$). The sign of the term in brackets, however, depends on the relative size of the fraction of the demand reduction impacting period 1 (β) and the indirect effect caused by the adjustment in the equilibrium price $(\partial e_1/\partial \tau_1)/(\partial e/\partial \tau_1)$. The latter determines how emissions changes triggered by a price change are distributed across periods in a market with perfect intertemporal arbitrage. Phrased in terms of short and long-run elasticities, the term in brackets is more likely to be positive the smaller the difference between the short (κ_1^D) and long-run (κ^D) price elasticity of allowance demand. Put in terms of features of the overlapping policy, if the reduction in allowance demand is sufficiently frontloaded, equilibrium emissions in period 1 decrease. Recall that $\Delta e < 0$, i.e., a positive sign of (21) implies that e_1^* decreases if the policy's impact is more pronounced. Hence, in contrast to price-based flexibility mechanisms, the timing of policy impacts matters.

A similar argument also holds for the change in total equilibrium emissions e^* via adjustment in the long-run cap:

$$\frac{\partial e^*}{\partial \Delta e} = \frac{\partial s_2}{\partial \Delta e} = \frac{\partial s_2}{\partial b}\frac{\partial b}{\partial e_1^*}\frac{\partial e_1^*}{\partial \Delta e} = \frac{\frac{\partial s_2}{\partial b}}{1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \cdot \left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}} - \beta\right]. \tag{22}$$

Again, policies mainly reducing allowance demand early on reduce the long-run cap. Those reducing allowance demand in the distant future tend to increase it.[30]

Proposition 5 *The waterbed effect for a marginal, anticipated policy overlapping a carbon-pricing scheme with a flexibility mechanism based on past allowance banking is:*

$$W = 1 - \frac{\partial \Delta e^*}{\partial \Delta e} = \frac{1+\frac{\partial s_2}{\partial b}\beta}{1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \tag{23}$$

where the numerator captures the direct impact of the overlapping policy (Perino, 2018) and the denominator the indirect effect mediated through the price response (Rosendahl, 2019b). Recalling that $\beta = [1 + (1 - L_{i2})/(1 - L_{i1}) \cdot (1 - \beta_i)/\beta_i]^{-1}$ *makes apparent that*

[30] Note that the increase in allowances triggered by the overlapping policy does not require that the flexibility mechanism actually creates additional allowances. It might merely reflect a reduction in the number of allowances cancelled compared to the reference scenario without the overlapping policy.

the waterbed effect depends on the temporal structure of the overlapping policy (β_i) and its internal carbon leakage (L_{i1}, L_{i2}). Several special cases are worth highlighting:

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

$$W = \frac{1 + \frac{\partial s_2}{\partial b}}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} \in (0, 1). \tag{24}$$

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

$$W = \frac{1}{1 + \frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}} > 1. \tag{25}$$

(iii) For any given change in total allowance demand $\Delta e < 0$, 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 and long-run 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 ($\beta > 1, \Delta e < 0, \Delta e_1 < 0, \Delta e_2 > 0$) according to:

$$\beta > \bar{\beta} = -\frac{1}{\frac{\partial s_2}{\partial b}} \geq 1. \tag{26}$$

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

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

Proposition 5's equation (23) captures two opposing effects: the direct and the price-mediated indirect effect on long-run emissions. The former is represented by the second term in the numerator, the latter by the second term in the denominator. The two special cases presented in parts (i) and (ii) highlight the two effects. For policies affecting aggregate demand early on (see Equation (24)), the price effect is always of second order and the waterbed is punctured (Perino, 2018). However, policies affecting aggregate demand only in the far future (Equation (25)) have no direct impact—and hence the price-driven effect in the denominator of Equation (23) dominates. Ceteris paribus, such policies increase the supply of allowances, i.e., they refill the waterbed—the "Rosendahl effect".[31] Anticipation of a future reduction in relative scarcity reduces the incentives to bank allowances, and the drop in the bank reduces the number of allowances cancelled by the flexibility mechanism.

In sum, for a given quantity-based flexibility mechanism ($\partial s_2/\partial b$) and given market characteristics ($\rho_1(e_1), \rho_2(e_2)$), an overlapping policy with a given impact on total allowance demand (Δe), 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 the timing of the policy's impact on aggregate allowance demand (β). This in turn implies that the waterbed effect is also a function of changes over time in internal carbon leakage[32] and that for any given change in total allowance demand, there is a variant of the unilateral policy distinguished only by the temporal distribution of demand shifts that features a waterbed effect of 100% (case (iii) in Proposition 5). Both points are in stark contrast to price-based flexibility mechanisms where policy timing and internal leakage were irrelevant (Proposition 4). An example for a policy featuring a negative waterbed (case (iv) in Proposition 5) is an amendment of a previously enacted coal phase-out plan that shuts down old inefficient plants earlier but grants new, highly-efficient plants a longer grace period.

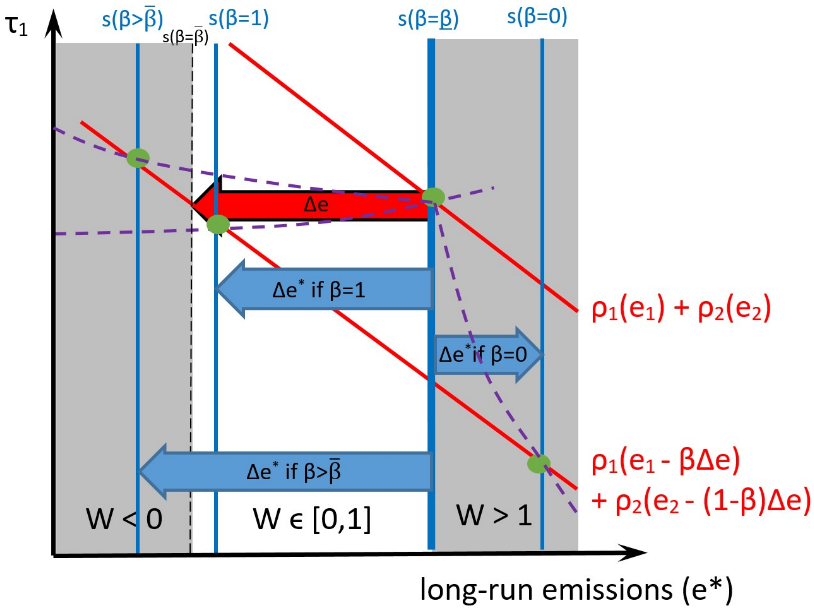
Figure 2 illustrates Proposition 5. The horizontal distance between inverse demand curves (red) represents the total change in the aggregate demand for allowances at a given carbon price (Δe). Three variants of the overlapping policy are presented that differ only in the timing of impacts (β). They occur either only in period 1 ($\beta = 1$), only in period 2 ($\beta = 0$), or a substantial decrease in demand is followed by a smaller increase ($\beta > \bar{\beta}$). Supply responses (blue) are starkly different: for $\beta = 1$ the waterbed effect is around 0.1, for $\beta = 0$ around 1.4, and for $\beta > \bar{\beta}$ it is $-.2$. The three dashed purple curves connect market equilibria for a continuum of demand shocks. Hence, they can be interpreted as effective supply curves indicating how equilibrium prices and quantities respond to variations in the inverse aggregate emissions demand curve.

Based on these effective supply curves, Corollary 1 shows formally how the results of our earlier Proposition 4—and hence of the literature on tax incidence—can be transferred to banking-based flexibility mechanisms. There are two qualifications. First, one needs to compute the effective supply curve for each and every overlapping policy separately

[31] This effect was first described by Rosendahl (2019b) and confirmed by Bruninx et al. (2019); Gerlagh et al. (2019); Pahle et al. (2019); Rosendahl (2019a). For a discussion see Perino (2019). Rosendahl (2019a) and Gerlagh et al. (2019) refer to it as a "green paradox".

[32] Numerical simulations of different policies using calibrated models can be found in Bruninx et al. (2019); Gerlagh et al. (2019); Osorio et al. (2020); Pahle et al. (2019) and Rosendahl (2019a).

Figure 2: Waterbed effect for a banking-based flexibility mechanism



Notes: Shift in total allowance demand (Δe, red) induced by overlapping policy; supply response (blue); temporal distribution of the impact (β); response to continuum of demand shifts (dotted purple) with β fixed.

and, second, these curves might be strictly downward-sloping which allows for waterbed effects above 100% and below zero.

Corollary 1 *Propositions 4 and 5 are equivalent when considering the equilibrium expansion path as an instrument-specific effective allowance supply function*

$$\left.\frac{\partial s_2}{\partial \tau_1}\right|_{equilibrium} = \frac{\frac{\partial s_2}{\partial \Delta e}}{\frac{\partial \tau_1}{\partial \Delta e}} = \frac{\frac{\partial s_2}{\partial b}\frac{\partial e}{\partial \tau_1}}{1+\frac{\partial s_2}{\partial b}\beta} \cdot \left[\beta - \frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}\right]. \tag{27}$$

See Appendix B for a proof.

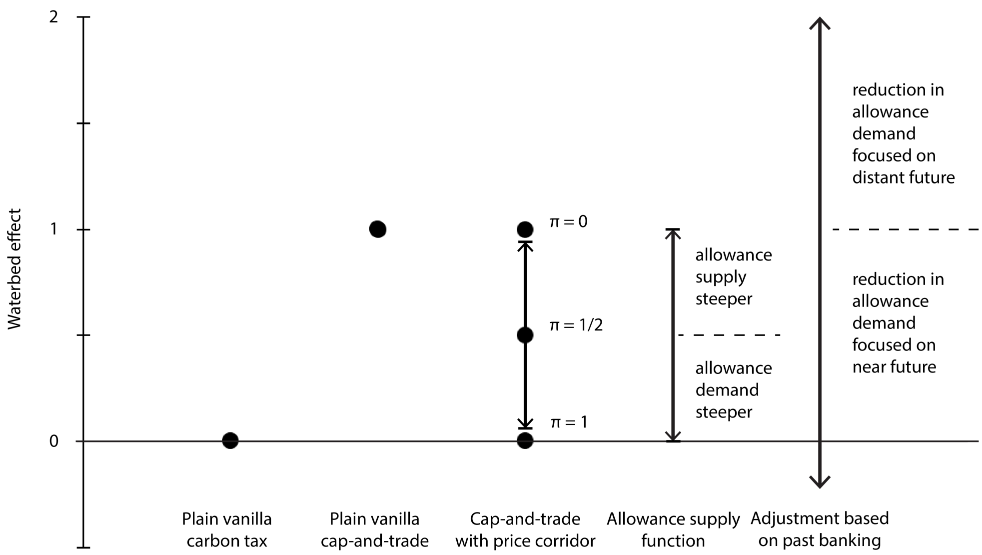
The equilibrium expansion path (27) is highly instrument-specific which highlights that the interpretation as an effective allowance supply function is illustrative at best. In contrast to the allowance supply function specified in Section 4.1 it is not a common and defining feature of the carbon-pricing scheme but specific to the overlapping policy under consideration. Hence, if several such policies are considered simultaneously, either by the same or different jurisdictions, each faces its own equilibrium expansion path. Plugging (27) into Equation (17) yields the same W as using (23).

To advise policymakers on the climate benefit of an overlapping policy within a carbon market that adjusts the cap based on past allowance banking such as the EU ETS, one needs to compute a separate effective supply curve for each and every variant of the policy—and they differ in fundamental ways based on seemingly minor differences such as the exact time profile of reductions in aggregate allowance demand.

The allowance supply function interpretation also highlights an inherent instability induced by combining overlapping policies with a carbon-market design like the reformed EU ETS. They have the potential to generate downward-sloping effective supply curves and hence multiplicity of equilibria and essentially erratic responses. The latter are most obvious for cases when the slopes of supply and demand curves are similar and neither strictly convex nor strictly concave.[33]

Summarising our analysis, Figure 3 presents the possible ranges of the waterbed effect for typical carbon-market designs captured by Propositions 4 and 5.[34] While the extent of the waterbed is unique for a carbon tax ($W = 0$) and cap-and-trade ($W = 1$), all hybrid policies and flexibility mechanisms yield ranges that depend on the specifics of the carbon-market design, the probability that any price bounds are binding (π), the scale and timing of the overlapping policy, or the long and short-run elasticity of emissions demand.

Figure 3: Waterbed effects for typical carbon-pricing policies



Notes: The expected (but not the ex-post) waterbed effect of a marginal overlapping policy in a cap-and-trade scheme with a price corridor depends on the probability that price bounds are binding (π).

[33]Existence of equilibria is typically not an issue. For any history of emissions there is a unique cap. The allowance supply in period 2 is always perfectly price-inelastic. The "effective supply curve" is an equilibrium expansion path, i.e., its mere existence requires that an equilibrium exists.

[34]Note that Propositions 4 and 5 apply to policies announced at the beginning of period 1. Due to the design of both types of flexibility mechanisms—both respond to market outcomes in period 1—an unanticipated policy implemented in period 2 has a waterbed effect of $W = 1$. It simply escapes the radar of the flexibility mechanism.

4.3 The reformed EU ETS

To further address subtleties with respect to the timing of overlapping policies, we now take a closer look in a many-period context at the reformed EU ETS. We also show how to compute $\partial s/\partial b$ and an instantaneous waterbed effect in this context—where $\partial s/\partial b$ is not explicitly specified by the market rules but rather is the result of an interaction between the rules and (future) market outcomes.

The EU ETS's flexibility mechanism, the Market Stability Reserve (MSR), works as follows.[35] 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 October of the following year (but not before January 2019) is reduced by a certain percentage of the size of the bank (see Table 1). Allowances withheld are placed in the MSR and released in installments of 100 million/year once the bank has dropped below 400 million. We label $t_{B=833}$ the year in which the bank drops below the 833 million threshold and the MSR hence stops taking in allowances.

Year (if bank > 833 million on Dec. 31^{st})	Intake rate (%)
2017	16*
2018 - 2021	24
2021 - $t_{B=833}$	12

Table 1: Intake rates for the EU ETS Market Stability Reserve (MSR)

* Two-thirds of 24 percent because the withdrawals that would be due in Oct.-Dec. 2018 do not materialise as the MSR is instituted at the beginning of 2019 (European Commission, 2018).

Starting in 2023, the maximum number of allowances held in the MSR is limited to the number auctioned in the previous year.[36] 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 is eventually cancelled.[37]

Computation of the waterbed effect for the EU ETS faces several challenges. First, the MSR's intake rate changes over time (Table 1). 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 bank in all future periods up to the point at which the TNAC drops

[35] The rules are laid down in European Parliament and Council (2018) and discussed by Perino (2018).

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

[37] At the end of 2019 the MSR contained 1.3 billion allowances with a further 332.5 million being added before September 2021 (European Commission, 2020). The cancellation threshold in 2023 will be below 1 billion.

below 833 million. Third, the point in time at which this threshold is reached is itself determined by market outcomes and hence by the overlapping policy itself.[38] Fourth, the size of the indirect price-mediated Rosendahl effect of anticipated future changes in allowance demand depends on the same dynamics. These complexities imply that W can only be estimated by numerical simulation.[39]

To illustrate some of these intricacies, we compute the sensitivity of the long-run cap to changes in the bank $\partial s/\partial b$. Based on this we derive an instantaneous waterbed effect that captures only the first two complexities: the MSR's time-varying intake rate and the uncertain multi-period nature of its activity. An instantaneous change in the number of banked allowances triggers a sequence of transfers to the MSR. Only 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. This implies (see Appendix B for proof):

Lemma 1 *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 long-run cap $(\partial s/\partial b)$ in the EU ETS is given by:*

$$\begin{aligned}\frac{\partial}{\partial b}s\left(t,t_{B=833}\right) = &-(1-.16)^{\max[0,\min[2018,t_{B=833}]-\max[2017,t]]} \\ &\times(1-.24)^{\max[0,\min[2022,t_{B=833}]-\max[2018,t]]} \\ &\times(1-.12)^{\max[0,\max[2022,t_{B=833}]-\max[2022,t]]}.\end{aligned} \tag{28}$$

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:

$$\hat{W}\left(t_a,t,t_{B=833}\right) = \frac{1+\frac{\partial}{\partial b}s\left(t,t_{B=833}\right)}{1+\frac{\partial}{\partial b}s\left(t_a,t_{B=833}\right)\frac{\frac{\partial e_{t_a}}{\partial \tau_{t_a}}}{\frac{\partial e}{\partial \tau_{t_a}}}}. \tag{29}$$

Abstracting from changes in the carbon price $((\partial e_{t_a}/\partial \tau_{t_a})/(\partial e/\partial \tau_{t_a})=0)$, this simplifies to:

$$\hat{W}\left(t,t_{B=833}\right)|_{\tau fixed} = 1+\frac{\partial}{\partial b}s\left(t,t_{B=833}\right). \tag{30}$$

Lemma 1 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

[38] Further increasing complexity, Gerlagh et al. (2019) show that the mechanism creates multiple equilibria.

[39] See Bruninx et al. (2019); Gerlagh et al. (2019); Pahle et al. (2019); Rosendahl (2019a) for simulation results and Rosendahl (2019b); Perino (2019) for informal discussions.

scheme stops responding to past market outcomes, t_{833}, jointly determine the size of the instantaneous waterbed effect. Note that this still ignores the endogeneity of t_{833} and that most overlapping policies shift allowance demand in several years (see Section 4.1).

5 Illustrations of unilateral overlapping policies

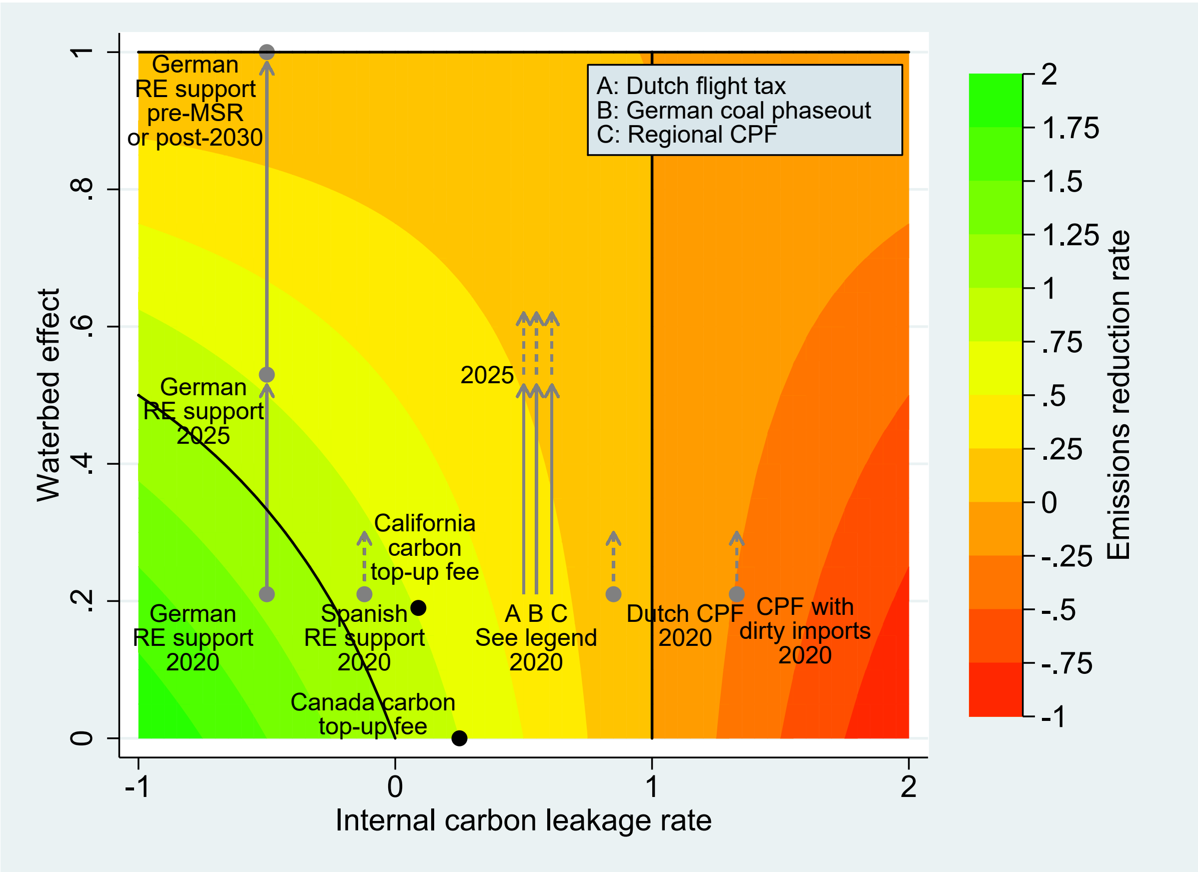
There are many real-world unilateral policies that overlap with wider carbon-pricing systems, leading to different degrees of waterbed effects and internal carbon leakage. We now illustrate how several such overlapping policies fit into our model's conceptual framework from Section 2. The equilibrium change in long-run emissions is $\Delta e^* = [1-L_i][1-W]\Delta e_i$, and our main outcome of interest here is the effective emissions reduction rate $R_i \equiv [1-L_i][1-W]$. We use a combination of sources to quantify leakage and waterbed effects, which allows us to compute R_i for a range of policies. (A limitation is that our sources do not provide time-varying estimates of internal carbon leakage, hence we leave out the t subscript for L_i (and we sometimes leave out subscripts altogether for ease of exposition).)

Figure 4 is the visual summary of this section. It plots the contour lines of R in (L, W)-space along with various policy examples for which we have found estimates of L and W using existing literature. 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; policies in the darker orange regions have little effect, or worse, increase aggregate emissions. The evidence is consistent with the predictions from our theory of internal carbon leakage: a unilateral carbon price floor, aviation tax, and coal phase-out have positive leakage (Propositions 1 and 2) while renewables support has negative leakage (Proposition 3). The following subsections explain the various overlapping policies in more detail, first for those overlapping the EU ETS and then for those in North America.

5.1 Overlapping policies in the EU ETS

We first consider policies overlapping the reformed EU ETS for which the waterbed effect depends on the timing of the policies—a result of the flexibility mechanism based on past allowance banking (the MSR). As discussed in Section 4.3, 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 referred to in Equation (30) in Lemma 1—changes over time. Therefore, the effective emissions reduction rate for policies in the EU ETS changes over time, and we refer to it as $\hat{R}_{it} = (1-L_i)(1-\hat{W}_t)$.

Figure 4: Unilateral 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. 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 schemes and for a proposed regional carbon price floor.

As given by Equation (28), $\hat{W}_t$ 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[40] 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. (2019). As time moves on, $\hat{W}_t$ increases from 0.21 to 0.53 to 1 and all European policies in Figure 4 move north, as indicated by the dotted lines. The values for $\hat{W}_t$ can be calculated using Equation (28) evaluated at $t_{B=833} = 2030$ and $t = 2020, 2025, 2030$. The internal leakage rate L_i is policy specific and we discuss empirical estimates for various policies below. Note that Figure 4 shows a sequence of emissions reduction rates for policies operating in different years. The overall performance of a policy that is in effect for multiple years can be summarised by the emissions-reduction weighted average over the values of $\hat{R}_{it}$ along the grey lines for the relevant time period.

We finally note that the policies that we highlight below (e.g., a carbon price floor, a coal phase-out, an aviation tax, or renewables support) likely have negligible external carbon leakage to regions outside the EU ETS, justifying our focus on internal leakage.

"Supply-side" unilateral policies

Following the structure of Section 3.1, we now discuss "supply-side" unilateral policies such as national carbon price floors, aviation taxes, and low-carbon mandates.

Electricity

We first consider unilateral cost-raising policies such as a national carbon price floor (CPF) for electricity generation. For example, the Dutch government announced a national CPF for the electricity sector in 2018 and is awaiting a final vote in parliament as of the fall of 2020. It is slated to increase from EUR 12.30/tCO_2 in 2020 to EUR 31.90/tCO_2 in 2030. In 2013, Great Britain introduced a carbon fee for its power sector. Proposition 1 shows that such policies, if binding, suffer from intra-EU leakage as domestic 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, recent estimates find $L \simeq 0.85$ for the Dutch CPF, while a regional CPF including the Benelux, France and Germany faces $L = 0.61$ (Frontier Economics, 2018; Vollebergh, 2018).[41] Such CPFs in small interconnected countries are un-

[40] This date is subject to substantial uncertainty, with estimates ranging from 2020 (Perino, 2018) to the second half of the 2030s (Quemin and Trotignon, 2018), and $t_B = 833 = 2030$ as a mid-range value (Vollebergh, 2018).

[41] 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 $L = 0.85$. Vollebergh (2018) estimates internal carbon leakage to be 85% for the Dutch price floor and 61% for a regional CPF including the Benelux, France and Germany.

likely 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 4).[42] 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*, $R < 0$.[43] 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 2). 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.[44] This would lead to reduced demand for allowances both before and after this date, relative to the counterfactual.[45] 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.

Post-2030, $\hat{W}_t = 1$, so all unilateral policies within the EU ETS end up at $R = 0$.

Aviation

As another example policy, 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).[46] This intra-EU leakage rate of 50% is in line

[42] We expect internal carbon leakage to be lower for Great Britain's carbon fee as import supply is more inelastic due to interconnection constraints, but we are not aware of any empirical estimates.

[43] 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 $A_i = 0$.

[44] 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).

[45] The law also contains a provision for cancelling allowances. It is supposed to create a waterbed effect of zero taking the MSR into account. However, the law assumes that the MSR induces a waterbed effect strictly below 1 and ignores leakage effects. The numbers calculated in Equation (28) are before cancellations are taken into account and hence can be used as an indication of how many allowances would need to be cancelled in order to achieve $W = 0$.

[46] 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.

with Proposition 1. As a result, the Dutch government abolished the tax in July 2009. The Netherlands will reintroduce a modest ticket tax of EUR 7 on all flights starting in January 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" unilateral policies

We now look at unilateral "demand-side" policies such as renewables support. 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 3, Abrell et al. (2019) estimate *negative* carbon (and output) leakage as additional zero-carbon energy depresses wholesale electricity prices and offsets imported gas- and coal-fired electricity in Germany ($L = -0.50$) and Spain ($L = -0.12$).[47] Similarly, a German government report finds $L = -0.65$ (Klobasa and Sensfuss, 2016). Figure 4 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 3 showed 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 4.

[47] In their Table 3, Abrell et al. (2019) 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 4. Schnaars (2019) provides an even more negative carbon leakage rate of -73%, 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. (2019) show that the German power mix is indeed dirtier than Spain's.

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

So far, we have assumed the MSR will stop taking in allowances in 2030 ($t_{B=833} = 2030$). In Figure 5, Panel (b), we investigate how the effective emissions reduction rate changes when we assume $t_{B=833} = 2048$ (following Gerlagh et al. (2019)). 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 (29) in Lemma 1). We use Gerlagh et al. (2019)'s estimates of the Rosendahl effect but note that estimates in the literature differ and this is a highly active area of research.

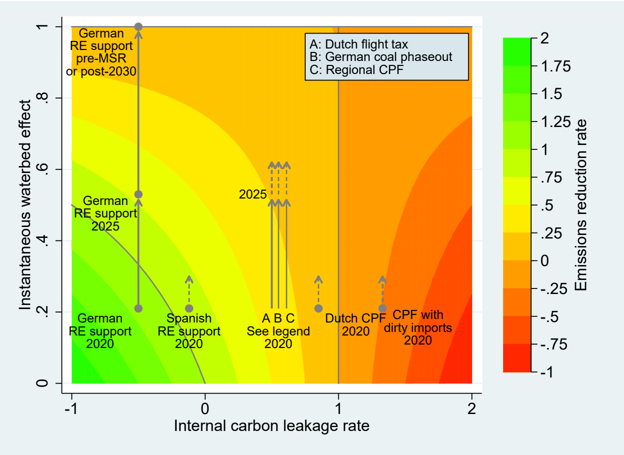
Panel (b) of Figure 5 shows that, compared to our original estimates in Panel (a), the instantaneous waterbed effect decreases substantially when $t_{B=833}$ lies further in the future. The waterbed effect can only go below 100% if the MSR takes in allowances; if allowances still flow into the MSR in 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 5 and highlights the potential unintended consequences of announcing policies that reduce emissions demand far into the future.

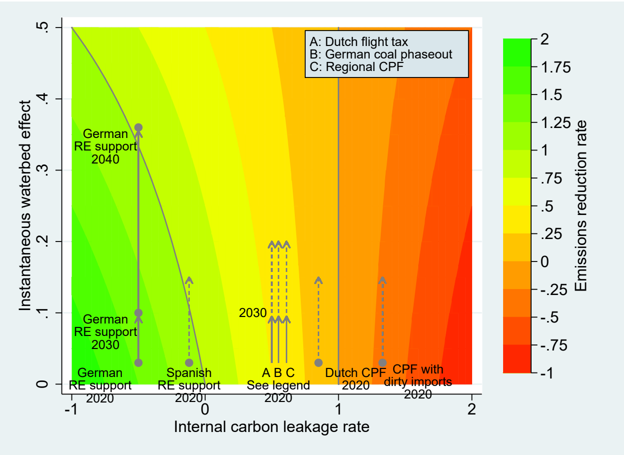
5.2 Overlapping policies in North America

We now turn to discussing examples of unilateral carbon policies in North America, two of which are plotted in Figure 4. Recall from Section 4.1 that the waterbed can also be punctured due to the stochastic nature of when a carbon price corridor is binding. A cap-and-trade system in which the carbon price trades at an auction price floor or cap has $W = 0$ while in the intermediate price range $W = 1$. The expected waterbed effect that applies to an overlapping policy thus depends on the probability that the auction price floor or cap is binding in a given year. The higher the probability that the system will trade at the price floor or cap, the more effective the puncture (Figure 3, policy 3). This feature is relevant for the two carbon markets in the United States.

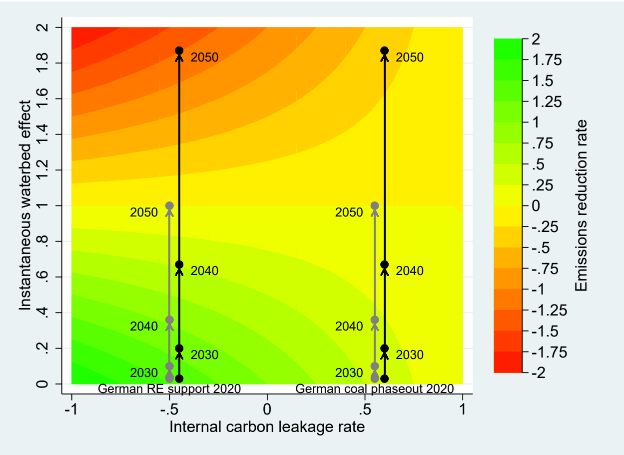
Figure 5: Leakage and waterbed effects in the EU ETS under varying assumptions



(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 4 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. (2019), together with their estimate of $t_{B=833} = 2048$.

California-Québec carbon trading

California and Québec have a joint carbon market with an auction price floor (\$16.68 in 2020) and, in a proposal to take effect in 2021, a price ceiling of \$61.25 (Politico, 2018). During periods when the auction price floor binds—which it did in various auctions in the year 2016[48]—the unsold allowances are first placed in a holding account, from which they are re-introduced after two consecutive sold-out auctions (subject to a 25% volume limit per auction). If unsold for 24 months, they are moved to the Allowance Price Containment Reserve (APCR). In the case of the proposed price floor, the APCR will be practically infinite (i.e., the threshold for releasing allowances is very high) so moving allowances into the APCR is essentially the same as retiring them. Borenstein et al. (2017) report a post-reform estimate of 47% (34%) that the price floor (ceiling) binds. Thus, in expected terms, $W = 1 - 0.47 - 0.34 = 0.19$.

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.[49] 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 $L = 0.09$ 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.[50] Figure 4 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 $R = 0.74$ tons ($W = 0.19, L = 0.09$).

[48]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.

[49]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 recently resurfaced in discussions among states.

[50]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.

Regional Greenhouse Gas Initiative

The Regional Greenhouse Gas Initiative (RGGI) that caps CO_2 emissions from electricity in ten Northeastern states has a price floor which was binding during 2010-2012;[51] the states decided to retire unsold allowances in such cases. Allowances currently trade at higher prices and the policy also has a price ceiling. 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, which would apply in-state and to imported electricity from other RGGI states (Forbes, 2020b). This policy is somewhat different from the ones we have examined in our theory—there is a border tax and external leakage to non-RGGI states may be nontrivial due to the interconnectedness of the power grid.

Shawhan et al. (2019) estimate the emissions leakage to other RGGI states and to non-RGGI states that results from New York's policy and find that this depends heavily on whether the cost of renewables is relatively low or high vs. natural gas—the former leading to lower RGGI allowance prices. They find emissions leakage to other RGGI states of $L = 0.58$ under the low-cost renewables scenario, but a negative leakage rate of $L = -0.42$ when the relative cost of renewables is high. They also estimate the external carbon leakage from New York to non-RGGI states at $L = -1.55$ and $L = 0.04$ for the low and high renewable cost scenarios, respectively.[52] This underscores that external and internal leakage are distinct phenomena that can even have different signs, and that careful empirical estimates of internal carbon leakage are essential for assessing the effectiveness of unilateral policies.[53] The border tax moves the carbon-pricing policy away from the setting of Proposition 1 towards Proposition 3 where it reduces emissions both inside and outside jurisdiction i—hence leakage can be negative. We do not plot New York's carbon fee in Figure 4 as we are not aware of an empirical estimate of the fraction of the time that the system is expected to trade at the price floor or ceiling, so W is missing.

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

[52] The reason for the difference is that the carbon fee increases non-emitting generation in New York. In the low-cost scenario this increase is quite substantial and reduces dirty imports from non-RGGI to RGGI states, causing negative *external* leakage. With the high-cost assumptions the carbon fee reduces emitting generation in other RGGI states, causing negative *internal* leakage (but there is little effect on imports to RGGI from the rest of the Eastern Interconnection).

[53] We further note that Fell and Maniloff (2018) estimate carbon leakage from the introduction of RGGI as a whole. They find substantial positive *external* leakage ($L = 0.51$) from RGGI to non-RGGI states. As this is a very different policy than New York's proposed carbon price—it operates in a larger region and does not have a border-tax adjustment—we have no a priori reason to expect that external leakage rates for these policies 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 4(ii)) 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 4 plots this policy assuming $L = 0.25$ and $W = 0$, leaving a higher carbon tax in British Columbia reasonably climate-effective ($R = 0.75$).

6 Conclusion

This paper has presented a new modelling framework—based on internal carbon leakage and waterbed effects—to understand the impacts of overlapping climate policies within a wider carbon-pricing system. Design matters in that different policy types have very different leakage properties. Space matters as internal leakage rates can differ substantially across industries and countries. Time matters as it affects the magnitude of the waterbed effect. Our results provide policy-relevant guidance on the climate benefits of 25 different combinations of unilateral policies and types of carbon-pricing systems.

The issues we have highlighted are critical for the design of new climate policies and extend beyond policy-making in Europe and North America. The ongoing design of China's national cap-and-trade system—and, in particular, the extent to which it will feature a waterbed effect—will greatly affect how effective additional province-level climate action will be. A carbon price floor would strengthen the market design by preventing an inefficiently-low price and may puncture the waterbed; subnational governments then ought to think carefully about selecting unilateral policies with limited internal leakage. More empirical estimates of internal leakage could substantially improve policy-making.

Over time, as carbon pricing is adopted by additional jurisdictions and international carbon markets become increasingly linked, the issue of internal carbon leakage may become even more salient. Imagine a future with a global carbon-pricing system. Suppose that one country or region, say the EU, wishes to push harder on decarbonisation of an individual sector, say steel. Depending on the nature of competition in steel, there may be "internal" leakage to non-EU jurisdictions. This unilateral EU policy will also experience a global waterbed effect: zero under a global carbon tax but potentially large under global cap-and-trade. With appropriate reinterpretation, our results can help understand the impacts of overlapping policies in a future with a more global carbon-pricing system.

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.

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.

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.

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, Marten Ovaere, Kenneth Gillingham, and Erik Delarue, "The Unintended Consequences of the EU ETS Cancellation Policy," 2019. MPRA Paper No. 96437. Available at `https://mpra.ub.uni-muenchen.de/96437/`.

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.

European Commission, "Publication of the Total Number of Allowances in Circulation in 2017 for the Purposes of the Market Stability Reserve under the EU Emissions Trading System Established by Directive 2003/87/EC," Technical Report C(2018) 2801 final, European Commission 2018.

_ , "Publication of the Total Number of Allowances in Circulation in 2019 for the Purposes of the Market Stability Reserve under the EU Emissions Trading System Established by Directive 2003/87/EC," Technical Report C(2020) 2835 final, European Commission 2020.

European Parliament and Council, "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.

Fischer, Carolyn, Mads Greaker, and Knut Einar Rosendahl, "Robust Technology Policy against Emission Leakage: The Case of Upstream Subsidies," *Journal of Environmental Economics and Management*, 2017, *84*, 44–61.

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, "Endogenous Emission Caps Always Induce a Green Paradox," 2019. CESIfo Working Paper No. 7862. Available at `https://www.cesifo.org/en/publikationen/2019/working-paper/endogenous-emission-caps-always-induce-green-paradox`.

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.

Hanna, Rema, "US Environmental Regulation and FDI: Evidence from a Panel of US-Based Multinational Firms," *American Economic Journal: Applied Economics*, 2010, *2* (3), 158–189.

Hepburn, Cameron, "Regulation by Prices, Quantities, or Both: A Review of Instrument Choice," *Oxford Review of Economic Policy*, 2006, *22* (2), 226–247.

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.

Kahn, Matthew E. and Erin T. Mansur, "Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?," *Journal of Public Economics*, 2013, *101*, 105–114.

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`.

Kuusela, Olli-Pekka and Jussi Lintunen, "A Cap-and-Trade Commitment Policy with Allowance Banking," *Environmental and Resource Economics*, 2020, *75* (3), 421–455.

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.

Metcalf, Gilbert E, "An emissions assurance mechanism: adding environmental certainty to a US carbon tax," *Review of Environmental Economics and Policy*, 2020, *14* (1), 114–130.

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**, "Procrastinating Reform: The Impact of the Market Stability Reserve on the EU ETS," *Journal of Environmental Economics and Management*, 2016, *80*, 37–52.

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

Pizer, William A and Brian C Prest, "Prices versus quantities with policy updating," *Journal of the Association of Environmental and Resource Economists*, 2020, *7* (3), 483–518.

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, "Intertemporal Emissions Trading and Market Design: An Application to the EU-ETS," 2018. Centre for Climate Change Economics and Policy Working Paper No. 348; available at http://www.lse.ac.uk/GranthamInstitute/wp-content/uploads/2019/01/working-paper-316-Quemin-Trotignon.pdf.

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 New Green Paradox," 2019. CESIfo Working Paper No. 7645. Available at https://www.cesifo.org/en/publikationen/2019/working-paper/eu-ets-and-new-green-paradox.

_ , "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," 2019. Working Paper, University of Hamburg. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3411782.

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/.

Traeger, Christian, Grischa Perino, Karen Pittel, Till Requate, and Alex Schmitt, "The Flexcap—An Innovative CO2 Pricing Method for Germany," *DICE Report*, 2020, *18* (1), 42–48.

Unold, Wolfram and Till Requate, "Pollution Control by Options Trading," *Economics Letters*, 2001, *73* (3), 353–358.

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.

Weitzman, Martin L., "Prices vs. Quantities," *Review of Economic Studies*, 1974, *41* (4), 477–491.

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 2020," 2020. Washington, DC: World Bank. Available at https://openknowledge.worldbank.org/handle/10986/33809. 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 on internal carbon leakage

General results (non-separable cost functions)

In this appendix, we derive three results, Propositions 1A–3A, on internal carbon leakage using a general cost function $G_k(x_k, a_k)$ for $k = i, j$. The separable cost function $G_k(x_k, a_k) \equiv [C_k(x_k) + \phi_k(a_k)]$ from the main text is nested where $G_k^{xa}(x_k, a_k) = 0$ so that Propositions 1–3 will follow as simple corollaries. 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).

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. 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 first-order conditions are:

$$p = G_k^x + \tau_k \theta_k$$

$$\tau_k = G_k^a.$$

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 (with $G_k^{xa} = 0$) and, more generally, is satisfied if either $G_k^{xa} \geq 0$ or $G_k^{xa} < 0$ but not too negative. Intuitively, the condition limits the degree of cost complementarity between output and abatement such that 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 from the main text 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:

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

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

Proposition 1A. *With non-separable cost functions, a unilateral carbon price by country i has internal carbon leakage to country j of:*

$$L_i = \frac{\theta_j}{\theta_i}\mu_j \frac{\delta_j}{\delta_i}\frac{1}{\left[1+\frac{G_i^{aa}}{M_i^a}\frac{G_i^{xx}}{M_i^a}\left[(1-\psi_i)+\mu_j\left(1-\psi_j\right)\frac{G_j^{xx}}{G_i^{xx}}\right]\right]} > 0,$$

where the rate of output leakage is $L_i^O = \mu_j \in (0,1)$.

Proof of Proposition 1A. Since i's carbon price $\tau_i = \tau_i(\tau, \lambda_i)$ under a unilateral carbon price, differentiating i's first-order conditions yields:

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

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

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

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

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

We now proceed in two main steps. First, we derive equilibrium changes in output levels. Second, we derive changes in emissions—and hence the rate of internal carbon leakage.

Combining j's first-order conditions shows that firms' output changes are related according to:

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

The same approach for i yields:

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

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:

$$-\mu_j\frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i}$$

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

Solving for equilibrium output responses yields:

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

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

Therefore the rate of internal output leakage is:

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

which is always positive but less than 100% by stability. 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:

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

We similarly obtain for i:

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

Therefore the rate of internal carbon leakage due to the unilateral policy is:

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

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

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

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

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

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

Proposition 2A. *A unilateral reduction in carbon-intensive production by country i has internal carbon leakage to country j of:*

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

where the rate of output leakage is $L_i^O = \mu_j \in (0, 1)$.

Proof of Proposition 2A. A unilateral reduction in carbon-intensive production by country i is represented as $dx_i/d\lambda_i < 0$. The common carbon price remains unchanged, $\tau_i = \tau_j = \tau$. This problem has the same structure as that underlying Proposition 1A—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:

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

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

Writing j's first-order conditions in more compact form, using the definitions of ψ_j and μ_j, gives:

$$-\mu_j \frac{dx_i}{d\lambda_i} = \frac{dx_j}{d\lambda_i} > 0.$$

So the rate of internal output leakage is:

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

which is always positive but less than 100% by stability, exactly as in Proposition 1A.

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:

$$\frac{de_j}{d\lambda_i} = \theta_j \delta_j \frac{dx_j}{d\lambda_i} = -\theta_j \delta_j \mu_j \frac{dx_i}{d\lambda_i} > 0.$$

We similarly obtain for i:

$$\frac{de_i}{d\lambda_i} = \theta_i \delta_i \frac{dx_i}{d\lambda_i} < 0.$$

Therefore the equilibrium rate of internal carbon leakage due to the unilateral policy is:

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

which is a special case of the expression in Proposition 1A, as claimed.

Proposition 3A. *The unilateral policies by country i of (i) a renewables support program that brings in additional zero-carbon production, (ii) an energy-efficiency program that reduces demand for carbon-intensive production, and (iii) a carbon-consumption tax have identical internal carbon leakage to country 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 3A. As explained in the main text, all three of these demand-side unilateral 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 unilateral 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, we derive equilibrium changes in output levels. Second, we derive changes in emissions—and hence the rate of internal carbon leakage.

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:

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

using the definition of ψ_k.

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

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

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

Solving for equilibrium output responses yields:

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

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

So the rate of internal output leakage is:

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

which is always negative.

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:

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

We similarly obtain for i:

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

Therefore the equilibrium rate of internal carbon leakage due to the unilateral policy 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.

Specific results (separable cost functions)

We here state Propositions 1–3 as simple corollaries of Propositions 1A–3A. 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 expressions for internal carbon leakage in terms of simple demand and supply elasticities, we begin with 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:

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

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 approach, we also obtain that:

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

Proposition 1 (Separable cost functions). *A unilateral carbon price by country i has internal carbon leakage to country j of:*

$$L_i = \frac{\theta_j}{\theta_i}\frac{\frac{\sigma_j}{(\sigma_j+\varepsilon^D/\varepsilon_j^S)}}{\left[1 + \frac{A_i}{(1-A_i)}\left(1 + \frac{(1-\sigma_j)\varepsilon_i^S/\varepsilon_j^S}{(\sigma_j+\varepsilon^D/\varepsilon_j^S)}\right)\right]} > 0.$$

Proof of Proposition 1. The expression for internal carbon leakage from Proposition

1A simplifies directly 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 and recalling the definition of k's abatement opportunity from the main text, $A_k = C_k''/[C_k'' + \theta_k^2\phi_k'']$, yields the result as claimed.

Proposition 2 (Separable cost functions). *A unilateral reduction in carbon-intensive production by country i has internal carbon leakage to country j of:*

$$L_i = \frac{\theta_j}{\theta_i}\frac{\sigma_j}{(\sigma_j + \varepsilon^D/\varepsilon_j^S)}.$$

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

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

Using the relationship $\mu_j = \sigma_j/(\sigma_j + \varepsilon^D/\varepsilon_j^S)$ yields the result as claimed.

Proposition 3 (Separable cost functions). *The unilateral policies by country i of (i) a renewables support program that brings in additional zero-carbon production, (ii) an energy-efficiency program that reduces demand for carbon-intensive production, and (iii) a carbon-consumption tax have identical internal carbon leakage to country 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 3. The expression for internal carbon leakage from Proposition 3A simplifies as:

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

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

Appendix B: Proof of results on waterbed effects

Derivation of Equation (14)

Application of Cramer's rule to conditions (11)-(13) yields:

$$\frac{\partial \tau_1}{\partial \Delta e} = \frac{-\frac{\partial\rho_1}{\partial e_1}\frac{\partial\rho_2}{\partial e_2}[\beta + (1-\beta)]}{-\frac{\partial\rho_1}{\partial e_1}\frac{\partial\rho_2}{\partial e_2}\frac{\partial s_2}{\partial\tau_1} + \frac{\partial\rho_2}{\partial e_2} + (1+r)\frac{\partial\rho_1}{\partial e_1}}.$$

Cancelling $-\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$ 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 $\frac{1}{\frac{\partial \rho_t}{\partial e_t}}$ yields:

$$\frac{\partial \tau_1}{\partial \lambda} = \frac{1}{\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e_1}{\partial \tau_1} - (1+r)\frac{\partial e_2}{\partial \tau_2}}.$$

Using $\frac{\partial e}{\partial \tau_1} = \frac{\partial e_1}{\partial \tau_1} + \frac{\partial e_2}{\partial \tau_1}$ 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 Equation (14).

Proof of part iii) of Proposition 4

The waterbed effect is given by:

$$W = \frac{-\frac{\partial e}{\partial \tau_1}}{\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}}.$$

Differentiating w.r.t. the slope of the allowance supply function $\partial s_2/\partial \tau_1$ yields:

$$\frac{\partial W}{\partial\left(\frac{\partial s_2}{\partial \tau_1}\right)} = \frac{\frac{\partial e}{\partial \tau_1}}{\left(\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}\right)^2} < 0,$$

while differentiating w.r.t. the slope of the total allowance demand function $\partial e/\partial \tau_1$ yields:

$$\frac{\partial W}{\partial\left(\frac{\partial e}{\partial \tau_1}\right)} = \frac{\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1} + 1}{\left(\frac{\partial s_2}{\partial \tau_1} - \frac{\partial e}{\partial \tau_1}\right)^2} > 0.$$

Derivation of Equation (21)

Application of Cramer's rule to conditions (18)-(20) yields:

$$\frac{\partial e_1^*}{\partial \Delta e} = \frac{(1+r)\beta\frac{\partial \rho_1}{\partial e_1} - (1-\beta)\frac{\partial \rho_2}{\partial e_2}}{(1+r)\frac{\partial \rho_1}{\partial e_1} + \left(1 + \frac{\partial s_2}{\partial b}\right)\frac{\partial \rho_2}{\partial e_2}}.$$

Analogous to the derivation of Equation (14), cancelling $\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$, using $\frac{1}{\frac{\partial \rho_t}{\partial e_t}} = \frac{\partial e_t}{\partial \tau_t}$ and $(1+r)\frac{\partial e_2}{\partial \tau_2} = \frac{\partial e_2}{\partial \tau_1}$, yields:

$$\frac{\partial e_1^*}{\partial \Delta e} = \frac{\beta\frac{\partial e_2}{\partial \tau_1} - (1-\beta)\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e_2}{\partial \tau_1} + \left(1 + \frac{\partial s_2}{\partial b}\right)\frac{\partial e_1}{\partial \tau_1}}.$$

Using $\frac{\partial e}{\partial \tau_1} = \frac{\partial e_1}{\partial \tau_1} + \frac{\partial e_2}{\partial \tau_1}$ and canceling $\frac{\partial e}{\partial \tau_1}$ yields Equation (21).

Proof of Corollary 1

First use conditions (18)-(20) and Cramer's rule to compute:

$$\frac{\partial \tau_1}{\partial \Delta e} = -\frac{\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}\left[\beta + (1-\beta)\left(1+\frac{\partial s_2}{\partial b}\right)\right]}{(1+r)\frac{\partial \rho_1}{\partial e_1} + \left(1+\frac{\partial s_2}{\partial b}\right)\frac{\partial \rho_2}{\partial e_2}}.$$

Again, cancelling $\frac{\partial \rho_1}{\partial e_1}\frac{\partial \rho_2}{\partial e_2}$, using $\frac{1}{\frac{\partial \rho_t}{\partial e_t}} = \frac{\partial e_t}{\partial \tau_t}$ and $(1+r)\frac{\partial e_2}{\partial \tau_2} = \frac{\partial e_2}{\partial \tau_1}$ and $\frac{\partial e}{\partial \tau_1} = \frac{\partial e_1}{\partial \tau_1} + \frac{\partial e_2}{\partial \tau_1}$, yields:

$$\frac{\partial \tau_1}{\partial \Delta e} = \frac{1+\frac{\partial s_2}{\partial b}\beta}{\frac{\partial e}{\partial \tau_1}\left(1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}\right)}.$$

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:

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

Cancelling $1+\frac{\partial s_2}{\partial b}\frac{\frac{\partial E_1}{\partial \tau_1}}{\frac{\partial E}{\partial \tau_1}}$ yields Equation (27).

It now remains to be shown that Propositions 4 and 5 are equivalent. To see this substitute (27) into (17) to get:

$$W = \frac{-\frac{\partial e}{\partial \tau_1}}{\left[\frac{\frac{\partial s_2}{\partial b}\frac{\partial e}{\partial \tau_1}}{1+\frac{\partial s_2}{\partial b}\beta} \cdot \left(\beta - \frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}}\right)\right] - \frac{\partial e}{\partial \tau_1}}.$$

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 (27). Both values below 0 and above 1 are now possible. Dividing the above equation by $-\frac{\partial e}{\partial \tau_1}$ and multiplying it by $1+\frac{\partial s_2}{\partial b}\beta$ obtains:

$$W = \frac{1+\frac{\partial s_2}{\partial b}\beta}{\frac{\partial s_2}{\partial b}\cdot\left[\frac{\frac{\partial e_1}{\partial \tau_1}}{\frac{\partial e}{\partial \tau_1}} - \beta\right] + 1 + \frac{\partial s_2}{\partial b}\beta}.$$

Cancel $\frac{\partial s_2}{\partial b}\beta$ in the denominator to obtain Equation (23).

Proof of Lemma 1

Equation (28) follows directly from the parameters presented in Table 1 and the explanation given in the paragraph preceding the Lemma. See also Perino (2018).

The instantaneous waterbed effect $\hat{W}\left(t_a, t, t_{B=833}\right)$ 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 (23) 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 (23) capturing the price effect is therefore adjusted accordingly.