Intertemporal Price Dynamics and Linking of Carbon Markets*

Jan Abrell Dallas Burtraw Michael Pahle Simon Quemin Matt Woerman

This version: 30 January 2025

Abstract

The efficiency of carbon pricing hinges on markets broad enough to encompass heterogeneous economies and sectors. Linking carbon markets is gaining increasing attention, in particular for advanced systems in which caps are rapidly decreasing and prices are rising. Paradoxically, realizing linking surfaces economic and political threats that tend to be more severe the more advanced a system is. One is an immediate convergence of prices due to intertemporal arbitrage opportunities that emerge already when firms anticipate linking in the future. This paper shows that ensuring gradual price convergence across linked markets requires two instruments—one to establish the exchange rate at which permits trade in each period, another to align intertemporal arbitrage incentives with the targeted sequence of exchange rates—and characterizes how they should be calibrated jointly. These tools may enable policy makers to control price convergence while initiating linking, and offer more flexibility to manage the distributional and environmental impacts of linking, thus overcoming a main barrier to higher stringency. A calibrated numerical application to the EU ETS1 and ETS2 illustrates the different price paths and emission levels under various gradual linking policies.

Keywords Linking, banking, EU ETS, ...

JEL classification H23, Q58, ...

*Abrell: Chair of Energy Economics, University of Basel, Faculty of Business and Economics, Peter-Merian-Weg 6, CH-4002 Basel, jan.abrell@unibas.ch Butraw: Resources for the Future (RFF), 1616 P Street NW, Washington DC 20036; burtraw@rff.org Pahle: Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A31, D-14473 Potsdam; michael.pahle@pik-potsdam.de Quemin: Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A31, D-14473 Potsdam; Climate Economics Chair, Paris-Dauphine University PSL, Place du Maréchal de Lattre de Tassigny, F-75016 Paris; Électricité de France R&D, SYSTEME Department, EDF Lab Paris-Saclay, 7 Boulevard Gaspard Monge, F-91120 Palaiseau; simon.quemin@pik-potsdam.de. Woerman: Colorado State University, Department of Agricultural and Resource Economics, 301 University Avenue, Fort Collins, CO 80523-1172; matt.woerman@colostate.edu.

1 Introduction

Economic costs in carbon pricing programs pose a crucial barrier to realizing international commitments to reduce greenhouse gas emissions. There are currently 75 different carbon pricing programs in place around the world covering 24% of emissions, mostly through carbon markets, but with individual program characteristics and widely varying prices (World Bank, 2024). The discrepancy in abatement costs across markets raises aggregate costs, setting back progress globally. Economic advice is to harmonize and link markets to broadly identify least-cost abatement opportunities; however, this brings an associated harmonization (equalization) of marginal costs, presenting severe distributional challenges.

This work develops a theoretical and numerical model to identify policy levers for regulators to gradually link carbon markets while managing distributional impacts that can otherwise make the energy transition politically unsustainable. Various challenges coming along with linking have long been recognized in the literature and, paradoxically, the higher the efficiency gains the more severe these challenges may be. Regulatory disadvantages may arise from the linked system's inconsistency with original domestic policy objectives, and from the partial de facto cession of discretionary control over the domestic emissions trading system (Flachsland et al., 2009; Burtraw et al., 2013).[1] Distributional challenges may also arise through effects on consumers or fiscal revenues. In principle, these implications could be dealt with by suitable emission permit allocations or financial transfers so that all jurisdictions gain from linking. Yet when such a redistribution proves to be infeasible, uniform pricing may be inefficient (Abrell et al., 2018; Bataille et al., 2018; Stiglitz, 2019). Challenges for linking may be further aggravated in the case of linking an advanced system featuring a relatively high price with a recent system featuring a relatively low price. This is because linking may lead to sizable price shifts in both systems, which might incur additional adjustment costs or erode the value of investments. Finally, in advanced systems deeply entrenched design paths may exist that further complicate a necessary alignment of the rules between systems (Metcalf & Weisbach, 2012; Mehling et al., 2018, 2024). All this leads to the paradoxical situation in which larger efficiency gains from linking come along with greater challenges too.

The tradeoff between costs and distributional outcomes is especially salient in the EU, which maintains the world's oldest carbon emissions trading system (ETS1) covering the electricity and industrial sectors, with a carbon price of approximately 75 EUR per ton. Beginning in

[1] Conversely, linking also may insulate policy from capture by special interests within each jurisdiction. See Section 2 for a detailed literature review.

2027, the EU will launch a second emissions trading system (ETS2) covering transportation and buildings with prices that might initially be half that of the ETS1, at least according to the Impact Assessment by the EU Commission. But over time the relationship between marginal costs and prices is expected to flip. These programs will initially be maintained separately because of distributional concerns about impacts on households and businesses, but this price discrepancy requires some higher-cost abatement, while lower-cost opportunities remain unutilized, generating inefficiencies.

While the expected intersection of prices in the two systems may facilitate linking over time, the anticipation of linking may actually lead to early and politically problematic convergence. According to current law, the EU Commission will analyze the feasibility of linking only in 2031. In other words, it has indirectly committed to not link the ETS1 and ETS2 before then. Presumably, the reasons for doing so are that the ETS2 needs time to find equilibrium first, as well as distributional implications. That is, phasing in the ETS2 may be problematic if prices in 2027 turn out to be high already, which especially for Eastern Europe countries with no national carbon pricing in place could be considered a policy shock. Indeed, in anticipation of future linking firms may buy and bank allowances in the ETS2 for later use (Edenhofer et al., 2021). This in turn could lead to a rapid convergence of prices, and in particular a relatively high price in the ETS2 already in its first year of operation (2027). While this type of banking is speculative (based on future linking), the fact that all main enabling factors—harmonized market rules, same jurisdiction, converging prices—are conducive to linking makes the risk seem relatively low and such banking behavior likely.

Globally one observes a similar challenge is replicated across sectors and countries generally, with carbon prices that are even more disparate. Rapidly equating carbon prices by linking markets would result in distributional outcomes that are politically untenable. Decarbonization policies should be not only cost efficient but also feasible, fair, effective, and credible in the long term. Fabra & Reguant (2024) provide an overview of the various dimensions of this broader notion of efficiency and of the various associated tradeoffs. Gradual linking policies as considered in this paper should be viewed in this context, as a way to balance and deal with cost efficiency, distributional, environmental, and credibility impacts. A pressing policy problem is how to design carbon markets to capture gains from trade while governing the distributional impacts and incremental convergence in prices over the long run.

This paper shows that a linking policy ensuring gradual price convergence across markets over time needs two instruments—one to establish the exchange rate at which permits trade in each period, another to align intertemporal arbitrage incentives with the targeted sequence

of exchange rates. If exchange rates are used in isolation, anticipation of the future trading terms makes the current ones mute. That is, arbitrage entails that price ratios are governed by the exchange rate that will prevail in the future, making the phase-in of the link obsolete. Moreover, the paper shows how banking conversion rates should be set to sustain the targeted sequence of exchange rates in equilbrium, and that such a "two-legged" policy gives policy makers more flexibility to manage the distributional and environmental impacts of linking. Because these impacts are nontrivial, the paper provides a calibrated numerical application to the EU ETS1 and ETS2 to illustrate the different price paths and emission levels that may result under various gradual linking policies. Finally, the paper discusses several policy implementation issues and compares the performance of price-based policies (such as the ones analyzed here) and quantity-based policies (e.g., volume limits on the number of permits that can be traded across markets or vintage restrictions on banked permits).

The paper is structured as follows. Section 2 reviews the literature and describes our contributions. Section 3 provides a simple theoretical model to study gradual linking policies. Section 4 describes the numerical implementation of the model and how it is parametrized to the EU ETS1 and ETS2. Section 5 illustrates the impacts of various gradual linking policies between the ETS1 and ETS2 using the calibrated numerical model. Section 6 discusses policy implementation issues and alternative instruments. Section 7 concludes. An Appendix collects the proofs and analytical examples.

2 Literature

Our paper connects and contributes to two strands of literature that we review below.

2.1 Conversion rates in permit markets

In a permit market with unrestricted banking and borrowing, intertemporal arbitrage implies that the permit price grows at the rate of interest (Rubin, 1996). There is a finite supply of permits, which are optimally used as any exhaustible resources according to the Hotelling depletion rule. Under uncertainty and when borrowing is prohibited, the price grows at a rate as high as the interest rate in expectation (Schennach, 2000). Yet such a path is generally not socially optimal as the private interest rate is different from (typically higher than) the social interest rate.[2] More formally, the decentralized problem of minimizing abatement costs

[2]In general, capital market interest rates tend to be higher than social interest rates.

over time for firms is generally not congruent with the centralized problem of maximizing the consumer surplus net of abatement costs and environmental damages, both with flow and stock pollutants (Kling & Rubin, 1997; Leiby & Rubin, 2001).[3]

Against this background, the literature has essentially focused on how to address uncertainty about damage and abatement costs as well as asymmetric information about costs in permit markets with banking. Some flexibility in arbitraged price paths can be introduced with the use of non-unitary intertemporal trading ratios (ITRs) that adjust—typically discount—the emission or compliance value of banked permits in subsequent periods. For instance, Yates & Cronshaw (2001) and Feng & Zhao (2006) compare the efficiency of banking regimes with different ITRs. In essence, ITRs enable a decoupling between discounted marginal abatement costs (and prices) across periods through adjusted intertemporal arbitrage. Such a decoupling is a prerequisite when regulators seek to adjust permit supply over time as new information arrives in order to achieve first best (Newell et al., 2005; Pizer & Prest, 2020).[4]

In the case of heterogeneous or localized pollution damages (e.g., non-uniformly mixed pollution), trading ratios across sources are generally warranted to operate a trade-off between minimizing abatement costs and minimizing damages (Mendelsohn, 1986; Muller & Mendelsohn, 2009; Holland & Yates, 2015; Fowlie & Muller, 2019). Trading ratios allow the flexible control of marginal costs across sources, which reduces cost effectiveness but enables to discriminate emissions based on their location and damages caused. Because the focus of these papers is on spatial heterogeneity, granular but static frameworks are used. As a result, the interaction between trading ratios and ITRs, which are also aimed at controlling marginal costs across periods, is not considered.

The contribution of our paper lies in the combination of the spatial and temporal dimensions. When both intra- and inter-period trade is allowed, we show that both types of ratios—which for disambiguation we call trading and banking conversion rates—are needed to control the marginal costs and prices across space and time, and we characterize how they must be set in tandem to achieve a given objective.

Finally, note that trading and banking conversion rates are not just theoretical concepts, but are actually part of the existing policy toolkit. For instance, Pizer & Prest (2020) describe

[3]With a flow poluutant, for instance, the former problem equates discounted marginal costs over time, whereas the latter equates marginal cost and marginal damage in every period (Kling & Rubin, 1997). In general, both solutions need not coincide. See also Leard (2013) for a discussion of the environmental impacts of banking and associated emission reshuffling over time.

[4]Decoupling can also be achieved by the "conversion function" in Karp & Traeger (2024), which modifies the compliance value of permits over time without the direct use of ITRs.

examples of ITRs used or proposed in US programs such as the Acid Rain Trading Program and the "Waxman-Markey Bill". In addition, the NOx Budget Program in the eastern US established a 2-for-1 surrender rate for the use of banked allowances over a specified quantity (EPA, 2003). More recently, the US EPA's Good Neighbor Plan annually recalibrates the compliance value of the allowance bank for emitting sources with higher emissions rates (EPA, 2023). Similarly, exchange rates for emission allowances and discount rates for offsets have also been used or proposed in a variety of carbon markets, see notably Quemin & de Perthuis (2019) and Woerman (2023) for examples and discussions.

2.2 Linking permit markets

The benefits and drawbacks of linking permit markets are well researched and long established (e.g., Flachsland et al., 2009; Tuerk et al., 2009). The main economic benefit is an increase in allocative efficiency that results from the equalization of (or a reduced dispersion in) permit prices between markets. Linking allows for a (more) cost-effective split of abatement efforts across markets as permits flow where they reduce compliance cost the most, and all markets are better off from linking in aggregate. Intuitively, the higher the initial price differential across markets, the higher the efficiency gains from such "effort sharing". Yet at the same time, the induced distributional and price impacts are larger, making the acceptability of linking less obvious and the decision to link more difficult to make politically (e.g., Burtraw et al., 2013; Quemin & de Perthuis, 2019). In addition, linking brings about efficiency gains from "risk sharing" as shocks can be pooled and spread over a larger market with larger buffer capacity (Doda et al., 2019). While price volatility decreases on average, it may increase in some linked markets when external shocks are "imported" through the link.

While in principle linking-induced efficiency gains could facilitate setting more ambitious emission targets and increase environmental benefits, the literature finds ambiguous results regarding the cap-setting dynamics under linking (e.g., Doda et al., 2019; Holtsmark & Weitzman, 2020). For instance, total emissions tend to increase as a result of linking when emission caps are set non-cooperatively (Helm, 2003), but this result is reversed in a dynamic setting where investment in abatement technology is endogenous (Holtsmark & Midttømme, 2021). Emissions may also increase as a result of linking when caps are intensity-based, as in tradable performance standards programs (Long et al., 2025). In addition, there are economic reasons why permit prices could differ across markets, including differences in labor or capital markets and energy structures (Bataille et al., 2018), the public good nature of emission

reductions (Chichilnisky & Heal, 1994), and pre-exisiting inequality or fairness issues (Abrell et al., 2018). Other drawbacks of linking include reduced autonomy in terms of governance, market design, and price developments.

Against this background, linking with restrictions or "controls" that may be gradually lifted over time has the potential to capture efficiency gains all the while managing distributional impacts and initial differences across markets. The literature has studied the effects of such restrictions to trade, such as exchange rates or quantitative trading limits, but only in static frameworks (Quemin & de Perthuis, 2019; Woerman, 2023). Conversely, dynamic models of linking assume unrestricted banking and borrowing, and do not feature exchange rates, trade limits, or intertemporal trading ratios (Doda et al., 2019; Holtsmark & Midttømme, 2021).

Our paper characterizes how effective gradual linking policies should be designed when banking is allowed. Specifically, we show how a policy that performs well in a static setting (e.g., exchange rates to allow for the flexible control of the price ratios between markets) may be insufficient and lead to detrimental outcomes in a dynamic setting. Intuitively, because current permit prices reflect their expected value in the future, intertemporal arbitrage entails that the future exchange rate (provided that it is known) is effective before it is in place while the current exchange rate is mute. Thus, implementing gradual linking requires driving a wedge between markets' Hotelling price paths, which can only occur with the use of intertemporal trading ratios that are set to sustain the targeted sequence of exchange rates. In other words, the joint use of trading and banking conversion rates allow policy makers to accommodate the need to communicate early about future policy to avoid credibility issues (e.g., Salant, 2016; Dubois et al., 2025) with a smooth transition to the future policy regime that would otherwise materialize and be priced in "too early" through arbitrage.

Our premise in this paper is that the policy focus is on ensuring gradual price convergence across linked markets. Such endeavor may result from various political economy constraints that reach beyond economic efficiency (e.g., Fabra & Reguant, 2024), which we take as given and abstract from. These include adjustment costs and some form of systemic inertia that may call for a phase-in of the price and distributional implications of full linking in the long run.[5] In other words, we do not consider the socially optimal gradual linking policy (if there is one), but we aim to assess the joint impacts of a targeted sequence of trading and banking conversion rates on cost and environmental effectiveness. This allows us to delineate first-pass policy recommendations for policy selection—e.g., we identify value ranges where rates are welfare improving. Because trading and banking conversion rates may have opposite effects

[5]See notably Williams III (2010) and Vogt-Schilb et al. (2018) for related analyses.

on emissions depending on how they are set, we utilize calibrated numerical simulations to determine which effect is more likely to dominate. More generally, the numerical simulations allow us to explore the feasible policy parameter space in detail.

3 A simple model of gradual linking

3.1 Modeling framework

The modeling framework borrows from and adjusts that in Pizer & Yates (2015) to represent gradual linking in a simple, parsimonious manner.[6] The key novelty of the model is that it incorporates both exchange rates or trading ratios—as in Holland & Yates (2015), Quemin & de Perthuis (2019) and Woerman (2023)—and intertemporal trading ratios—as in Yates & Cronshaw (2001), Feng & Zhao (2006) and Pizer & Prest (2020). Specifically, consider a two-period, two-market model in which firms are regulated under separate or linked emission permit markets. Periods are indexed by $t \in \{1, 2\}$ and markets are indexed by $m \in \{H, L\}$. Without loss of generality, H denotes the high abatement cost or high permit price market under autarchy, whereas L denotes the low cost or low price market. Regulators distribute permits denominated by their origin and vintage, i.e., a volume of ω_{mt} is allocated in market m in period t. Firms are required to surrender (the equivalent of) one period-t m-permit for each unit of emissions generated in market m and period t.

Linking and banking give firms some flexibility to shift emissions between markets and time periods, respectively. We consider linking and banking policies that use permit conversion rates. Linking allows permits originating from one market to be used for compliance in the other market. The linking policy we consider adjusts and governs the compliance value of traded permits through a *linking conversion rate* ρ_t that is specific to period t, whereby ρ_t L-permits are equivalent to or fungible with one H-permit in period t—and vice versa. Yet bear in mind that one m-permit always gives the right to emit one unit of emissions in market m. By convention, let Δ_t denote the amount of permits—denominated in H-permits or tons of emissions in market H—transferred from market L to market H in period t.[7]

Banking allows permits allocated in period 1 to be saved for compliance in period 2. The banking policy we consider adjusts and governs the compliance value of banked permits

[6] By contrast, the model in Pizer & Yates (2015) is designed to study "delinking".

[7] In some policy configurations below, permits may flow in the "unnatural" direction from H to L (i.e., $\Delta_t < 0$) despite the fact that H (resp. L) is the high (resp. low) price market under autarchy.

through a *banking conversion rate* δ_m that is specific to market m, whereby δ_m period-1 permits banked in market m are equivalent to or fungible with one period-2 m-permit. Yet bear in mind that one period-t permit always gives the right to emit one unit of emissions in period t. Let B_m denote the amount of banked period-1 permits carried over into period 2 in market m, with the convention that borrowing is not allowed, i.e., $B_m \geq 0$.[8] Importantly, as will become clear, the use of non-unitary linking and/or banking conversion rates makes the volume of overall emissions variable (i.e., no longer equal to $\sum_{mt} \omega_{mt}$ units of emissions) and a function of the policy design.

Another facet of the policy we consider that lies at the intersection of linking and banking concerns the fate, or the process of conversion, of permits when trade across markets. Two broad approaches can be delineated, with or without retirement and re-issuance. In the first case, when a firm in H buys a permit from L, that L-permit is retired in L and ρ^{-1} H-permits are re-issued in H in its place. In the second case, no retirement and re-issuance occurs, and the firm still holds the L-permit, which can be surrendered against ρ^{-1} units of emissions in H. The second approach ensures the denomination of permits is immutable—so that one can always keep track of permits by their origin and vintage—but begs the question of the compliance value of banked foreign permits in the next period (i.e., whether δ_L, δ_H, or no banking conversion rate applies to L-permits banked in H).

To capture these two approaches in the model, we represent them through different specifications of the upper constraint on banking. In the first case, domestic banking constraints are $B_H \leq \omega_{H1} + \rho_1^{-1}\omega_{L1}$ and $B_L \leq \omega_{L1} + \rho_1\omega_{H1}$. But both these constraints always hold with strict inequality as they are equivalent to having positive emissions.[9] In the second case, domestic banking constraints are $B_H \leq \omega_{H1}$ and $B_L \leq \omega_{L1}$, i.e., only domestic permits can be banked into the next period. This specification allows us to set aside (for now) the issue of the treatment of banked foreign permits. It is also convenient to carefully disentangle the effects of and interactions between linking and banking in the first place. Later in the paper, we will discuss and relax this specification.

For ease of exposition, we characterize the market equilibrium under different linking and banking policies with an optimization problem. That is, we consider a representative agent seeking to minimize aggregate compliance costs across markets and time periods. Let e_{mt}

[8] A few permit markets allow for limited borrowing, such as the EU ETS, but changing the constraint from $B_m \geq 0$ to $B_m \geq -\overline{B}_m$ with $\overline{B}_m > 0$ the borrowing limit does not affect our results.

[9] Specifically, $B_H = \omega_{H1} + \rho_1^{-1}\omega_{L1}$ entails that all period-1 L-permits are transferred to H: $\Delta_1 = \rho_1^{-1}\omega_{L1}$. This implies $e_{L1} = -B_L \leq 0$ (e.g., as per (1c)), a contradiction. Hence, $B_H < \omega_{H1} + \rho_1^{-1}\omega_{L1}$ always holds. By the same token, $B_L < \omega_{L1} + \rho_1\omega_{H1}$ ensures that $e_{H1} > 0$.

and $C_{mt}(e_{mt})$ denote the aggregate emissions and abatement cost functions in market m and period t. The functions C_{mt} are defined with respect to emissions (i.e., $-C'_{mt}$ denote marginal abatement costs) and satisfy standard assumptions (i.e., $C'_{mt} < 0$, $C''_{mt} > 0$). For simplicity, we assume the discount factor is one (i.e., the opportunity cost of capital is zero) and ignore positivity constraints on emissions.[10] This leads to following optimization problem

$$\min_{e_{mt},\Delta_t,B_m} \sum_{t\in\{1,2\}} \sum_{m\in\{H,L\}} C_{rt}(e_{rt}) \tag{1a}$$

$$\text{subject to} \quad e_{H1} \leq \omega_{H1} + \Delta_1 - B_H \quad (p_{H1}) \tag{1b}$$

$$e_{L1} \leq \omega_{L1} - \rho_1\Delta_1 - B_L \quad (p_{L1}) \tag{1c}$$

$$e_{H2} \leq \omega_{H2} + \Delta_2 + \delta_H B_H \quad (p_{H2}) \tag{1d}$$

$$e_{L2} \leq \omega_{L2} - \rho_2\Delta_2 + \delta_L B_L \quad (p_{L2}) \tag{1e}$$

$$-B_m \leq 0 \quad (\lambda_m) \quad \text{for } m \in \{H, L\} \tag{1f}$$

$$\Delta_1 \leq e_{H1} \quad (\gamma_H) \tag{1g}$$

$$-\rho_1\Delta_1 \leq e_{L1} \quad (\gamma_L) \tag{1h}$$

The shadow values or Lagrange multipliers associated with all constraints are given within parentheses in (1b–1h). The constraints (1g) and (1h) are the equivalents of $B_H \leq \omega_{H1}$ and $B_L \leq \omega_{L1}$ when permits that trade across markets are not retired and re-issued, respectively. With retirement and re-issuance of traded permits, these two constraints are irrelevant.

The standard assumptions about abatement costs ensure that the solution to the optimization problem (1) is equivalent to the competitive market equilibrium outcome with many firms in each market. That is, the competitive equilibrium permit price in market m and period t is equal to the shadow price p_{mt}. The first-order conditions are

$$-C'_{m1}(e_{m1}) = p_{m1} - \gamma_m \quad \text{for } m \in \{H, L\} \tag{2a}$$

$$-C'_{m2}(e_{m2}) = p_{m2} \quad \text{for } m \in \{H, L\} \tag{2b}$$

$$p_{H1} - \gamma_H = \rho_1(p_{L1} - \gamma_L) \tag{2c}$$

$$p_{H2} = \rho_2 p_{L2} \tag{2d}$$

$$p_{m1} = \delta_m p_{m2} + \lambda_m \quad \text{for } m \in \{H, L\} \tag{2e}$$

With retirement and re-issuance of traded permits, $\gamma_H = \gamma_L = 0$ always. Without retirement and re-issuance, when $\gamma_m > 0$, note that a wedge is introduced between the price and marginal

[10] Alternatively, assuming $-C'_{mt}(e_{mt}) \to \infty$ in the limit as $e_{mt} \to 0$ ensures $e_{mt} > 0$ in equilibrium.

abatement cost in market m in period 1. For instance, this occurs in H when $e_{H1} = \Delta_1$ or equivalently $B_H = \omega_{H1}$, that is when H banks the entirety of its period-1 permits, buys some period-1 permits from L for compliance in period 1, and would like to buy more to bank more but cannot do so. In this case, the period-1 price in H is equal to marginal abatement costs plus a positive price adder reflecting the binding constraint on trade or equivalently banking. The different cases that can occur are further studied in the following.

Before turning to the analysis of various gradual linking policies—i.e., different combinations of linking conversion rates (ρ_1, ρ_2) and banking conversion rates (δ_H, δ_L)—note that, as shown in Section 4.1, the framework can be extended to represent any number of periods, multiple individual firms within each market, as well as a mix of auctioning and free allocation. This only makes exposition more tedious, but does not alter the nature of our main results. See Section 6 for a description of how results carry over to the case with $T > 2$ periods.

In the following, four types of illustrative linking policies are considered, as shown in Table 1. The first type features two fully linked markets with unitary linking and banking conversion rates (Section 3.2). This is the reference case (first best). The second type shuts down linking in period 1—linking is only effective in period 2 but is antiticipated in period 1 (Section 3.3). This case illustrates that the first best may still obtain through banking (i.e., arbitrage). The third and fourth types seek to implement gradual linking, respectively with linking conversion rates only (Section 3.4) and with both linking and banking conversion rates (Section 3.5). Our analysis will show that the use of appropriate banking conversion rates is needed to implement the targeted exchange rates in each period. When the temporal dimension is not adressed with a specific instrument, banking entails that the exchange rate in period 2 backpropagates to period 1, by a similar arbitrage argument as with anticipated linking in the second policy type. In addition, using both types of conversion rates not only implements gradual linking as desired, but it also provides policy makers with more flexibility to address the distributional, environmental and price impacts of their linking policy.

Table 1: Main linking policies considered

Policy description	ρ_1	ρ_2	δ_H	δ_L
First best (1:1 linking in periods 1 and 2, 1:1 banking)	1	1	1	1
Anticipated linking (1:1 linking in period 2, 1:1 banking)	n.a.	1	1	1
Naïve or partial gradual linking policy	var.>1	1	1	1
Successful or flexible gradual linking policy	var.	var.	var.	var.

Note: n.a.: not applicable. var.: variable.

3.2 Full 1:1 linking with 1:1 banking

We start with the reference case—the first best—which features 1:1 linking in both periods and 1:1 banking in both markets. The first-order conditions in (2) rewrite

$$\begin{aligned} -C'_{rt}(e^{fl\star}_{mt}) = p^{fl\star}_{mt} \text{ for } m \in \{H, L\} \text{ and } t \in \{1, 2\}, \\ p^{fl\star}_{Ht} = p^{fl\star}_{Lt} \equiv p^{fl\star}_{t} \text{ for } t \in \{1, 2\}, \\ \text{and } \lambda^{fl\star}_{m} = p^{fl\star}_{m1} - p^{fl\star}_{m2} = p^{fl\star}_{1} - p^{fl\star}_{2} \equiv \lambda^{fl\star} \text{ for } m \in \{H, L\}, \end{aligned}$$

which characterize the full linking equilibrium, denoted by the superscript fl.

Result 1. *1:1 linking in both periods equates prices (and marginal abatement costs) across markets in both periods, and with 1:1 banking the single price is either constant or decreasing across periods when the no-borrowing constraint is binding. This holds irrespective of whether traded permits are retired and re-issued or not.*

Proof. See Appendix A.1. □

Result 1 shows that the first best coincides with the well-known Rubin-Schennach result for a single market with 1:1 banking. This is because 1:1 linking is effective in both periods, so the two markets are de facto merged into one. That is, there is no need to shift "extra" period-1 L-permits to H for compliance in period 2 through increased trading in period 1, higher banking in L, and lower banking in H, as trading is also possible in period 2.[11] The possible price differential between the two periods relfects the cost of the no-borrowing constraint, i.e., the cost saving from borrowing one permit if this was allowed (whether this constraint is binding or not depends on the permit allocation profile in both markets). See Appendix B for additional insights based on a stylized analytical example.

3.3 Anticipated 1:1 linking with 1:1 banking

Another important aspect of the first best is that there are multiple solutions to (1). That is, different combinations of the flexibility levers provided by linking (i.e., Δ_t) and banking (i.e., B_m) can implement the emission and price paths defined by $e^{fl\star}_{mt}$ and $p^{fl\star}_{t}$. Formally, there is always some $\varepsilon > 0$ that is sufficiently small so that the tuple $(\Delta^{fl\star}_1 - \varepsilon, B^{fl\star}_H - \varepsilon, \Delta^{fl\star}_2 + \varepsilon, B^{fl\star}_L + \varepsilon)$ sustains the first-best equilibrium. Specifically, feasibility requires $\varepsilon < B^{fl\star}_H$ with retirement

[11] Formally, $\Delta^{fl\star}_1 < e^{fl\star}_{H1}$ and $\gamma^{fl\star}_H = 0$.

and re-issuance of permits when they trade across markets, and $\varepsilon < \min\{B_H^{fl\star}, \omega_{L1} - B_L^{fl\star}\}$ without it.[12] See Appendix B for analytical examples. In words, some linking-induced inter-market flexibility can be replaced by some banking-induced inter-temporal flexibility.

Pushing the above logic one step further, if a possibly not so small $\varepsilon = \Delta_1^{fl\star}$ is feasible, the first best (and 1:1 price convergence in period 1) can be attained without linking in period 1 (i.e., only with linking in period 2 and banking across periods).

Result 2. *If 1:1 linking in period 2 is anticipated in period 1, banking may provide sufficient flexibility to shift the overall first-best volume of permit trade to period 2. This leads to the same emission and price paths as in the first best, notably with 1:1 price convergence already in period 1 despite the fact that linking is not yet effective.*

Result 2 requires banking to remain positive in both markets to maintain the full connection between markets in both periods through intertemporal arbitrage in each market. Namely, when $p_{m1} = p_{m2}$ holds, $p_{H2} = p_{L2}$ through 1:1 linking in period 2 implies $p_{H1} = p_{L1}$ even without 1:1 linking in period 1. Formally, this is equivalent to determining the conditions for the solution to (1) to be unaltered by the inclusion of the additional constraint $\Delta_1 = 0$. This ultimately depends on the feasibility of the first-best emission path relative to allocation profile in each market.

Result 2 shows that the mere announcement of linking at a future date can cause prices to converge before the link is effective through arbitrage. Convergence may be partial or may not occur when (*i*) banking—and in turn arbitrage—is constrained in one or both markets; (*ii*) there is uncertainty—or a lack of credibility—about the implementation of the link. For instance, when banking is positive in market m, $p_{m1} = p_{m2} = \phi p_2^{2l} + (1-\phi)p_{m2}^{a}$, where ϕ is the subjective probability of linking in period 2, as viewed by market participants in period 1, and where the superscripts $2l$ and a refer to period-2 linking and autarchy.[13] The above mechanism of "early" price convergence through banking and arbitrage is the same as that which is at play in the case of gradual linking considered below.

[12] If, additionally, a vintage restriction applies in period 2 whereby period-1 L-permits cannot be surrendered against compliance in region H, ε must also satisfy $\varepsilon < \omega_{L2} - \Delta_2^{fl\star}$. See Section 6.

[13] See Pizer & Yates (2015) for a proof in the symmetric case of uncertainty about future "delinking". See also Salant (2016) for a similar argument in the presence of regulatory uncertainty.

3.4 Naïve or partial gradual linking policy

As discussed in the Introduction, there are various reasons why policy makers may seek to implement linking gradually and have some control over the rate of price convergence across markets. In this section, we assume linking is effective in both periods, but permits trade at parity no sooner than in period 2. Specifically, we consider the case where policy makers set $\rho_1 > \rho_2 = 1$, but do not implement a specific banking policy in parallel (i.e., $\delta_H = \delta_L = 1$). As Results 3 and 4 show, introducing a linking policy without a companion banking policy in markets that allow banking is likely to result in a failed attempt at implementing gradual linking, and to lead to higher total emissions than provided for by emission caps.

Result 3. *When banking is positive in both markets, permits trade at parity and prices are equal across markets already in period 1 (i.e., ρ_1 is mute). This cannot occur with retirement and re-issuance of traded permits.*

When banking is zero in at least one market, permits trade at parity only in period 2, and the price ratio in period 1 is set by ρ_1. Without permit retirement and re-issuance, the price ratio in period 1 may in some cases be ambiguous and strictly comprised between 1 and ρ_1.

Proof. See Appendix A.2. □

Result 3 shows that 1:1 price convergence occurs in period 1 when both markets have positive banking, even though the current exchange rate is above one. Because prices are expected to have the same price in period 2 due to 1:1 linking, prices immediately converge on identical Hotelling paths by arbitrage, which makes ρ_1 obsolete. Only in "corner" cases where banking is zero in one or both markets may ρ_1 be effective in period 1, as the temporal connection between markets breaks down. This connection is more likely to hold in full when both markets have strong incentives to bank,[14] which is more likely to happen when permit allocation is frontloaded relative to the cost-effective emission path. Thus, it follows that gradual linking requires a two-pronged policy, with both linking and banking conversion rates. Indeed, only the use of different banking conversion rates can drive a wedge between markets' Hotelling price paths and thereby make gradual price convergence possible.

A striking feature of setting linking conversion rates such that $\rho_1 > \rho_2$ is that every period-1 permit traded from H to L effectively increases the quantity of emissions allowed, despite the fixed number of permits. As a result, firms have a strong incentive to maximize the number of period-1 permits traded in this way. For example, rather than H banking a period-1

[14] The connection—and early price convergence—would always hold if borrowing were allowed.

permit, total compliance costs are reduced if H instead trades that permit to L, L banks a period-1 permit into period 2, and finally L trades a period-2 permit back to H. Firms in both markets will prefer to have L's permits banked, and H will only bank once L has banked all of its period-1 permits. Thus, the outcomes under this naïve attempt depends on the demand for banking and, hence, the allocation of permits across the two periods.

Result 4. *Total emissions increase when the trade direction is reversed in period 1, which is more likely to occur the higher ρ_1 is, the more markets have room to adjust banking and shift trade to period 2, and with retirement and re-issuance of traded permits.*

In a static setting, linking conversion rates may also lead to increased emissions, but for that to happen, they must be set in an arguably poor manner ex ante, namely when ρ is above the ratio of autarchy prices p_H/p_L or below 1 (Quemin & de Perthuis, 2019; Woerman, 2023).[15] In a dynamic setting with banking, Result 4 shows that total emissions may increase even when ρ is set adequately, namely when it lies between 1 and p_H/p_L and partially closes the differential in autarchy prices. Total emissions are more likely to increase and will tend to be higher the more trade incentives are distorted across periods (i.e., the higher ρ_1 is) and the more L (resp. H) can increase (resp. decrease) banking. This is more likely to be feasible with permit retirement and re-issuance (the upper constraint on B_l is relaxed) and the more permit allocations tend to be frontloaded relative to cost-effective emission paths (L has more room to bank, while H has more room to sell to L in period 1).

More generally, the above "one-legged" gradual linking policy—beyond failing at implementing gradual price convergence in the presence of banking in both markets—does not provide policy makers with agility in dealing with its distributional and environmental impacts and the slopes of the market's price paths to 1:1 linking in period 2. However, distributional and price impacts of linking are one of the main reasons to introduce gradual linking, and total emissions become an integral part of the policy once exchange rates are used. Thus, policy makers need to have a handle on those. For instance, in the most likely case where the price ratio in period 1 is effectively set by ρ_1,[16] total emissions increase and H's price path decreases over time to reach L's flat price path from above.

[15] When $\rho > p_H/p_L$, the initial ranking of marginal abatement costs and prices between markets is reversed, permits flow from H to L, and abatement occurs where it is most expensive. When $\rho < 1$, the dispersion of marginal abatement costs and prices is greater than under autarchy. See also Section 5.2.

[16] That is, in case (*iii*) or if $e^n_{H1} > \omega_{H1}$ in Appendix A.2.

3.5 Successful or flexible gradual linking policy

Let us now turn to the case where both linking and banking policies are used. While we are primarily interested in gradual linking between markets—i.e., with $\rho_1 > \rho_2 \geq 1$—notice that the two results established below hold for any pairs (ρ_1, ρ_2) and (δ_H, δ_L).

Result 5. *With retirement and re-issuance of traded permits, the rate of price convergence across linked markets is uniquely pinned down and determined by*

$$\frac{p_{H1}}{p_{L1}} \Big/ \frac{p_{H2}}{p_{L2}} = \frac{\delta_H}{\delta_L} \text{ when banking is positive in both markets,}$$
$$\text{or } = \frac{\rho_1}{\rho_2} \text{ when banking is zero in one or both markets.}$$

Without retirement and re-issuance of traded permits, the rate of price convergence is determined as in the above, except when one market banks the entirety of its period-1 permits while the other market does not bank any of its period-1 permits. In those cases, the rate of price convergence is ambiguous and can be above or below ρ_1/ρ_2, namely

$$\frac{p_{H1}}{p_{L1}} \Big/ \frac{p_{H2}}{p_{L2}} < \frac{\rho_1}{\rho_2} \text{ when } \rho_1 > \rho_2,\ B_H = 0 \text{ and } B_L = \omega_{L1},$$
$$\text{or } > \frac{\rho_1}{\rho_2} \text{ when } \rho_1 < \rho_2,\ B_H = \omega_{H1} \text{ and } B_L = 0.$$

Proof. See Appendix A.3. □

Result 5 shows that only in the "corner" cases where banking is zero in one or both markets do we have that price ratios in periods 1 and 2 are governed by the targeted exchange rates. In the "interior" cases where banking is positive in both markets, the rate of price convergence across markets subject to gradual linking is governed by the ratio of banking conversion rates, rather than the ratio of targeted exchange rates. Therefore, adequate gradual linking implementation policy requires that banking conversion rates be set in accordance with the targeted linking conversion rates as per the "general rule" given in Result 6.

Result 6. *The targeted price ratios of ρ_1 in period 1 and ρ_2 in period 2 are always sustained in equilibrium for all banking conversion rates that satisfy $\delta_H/\delta_L = \rho_1/\rho_2$. This holds irrespective of whether traded permits are retired and re-issued or not.*

Proof. See Appendix A.4. □

Result 6 admits two interpretations. The first, with which we are by now familiar, is that gradual price convergence across linked markets requires the use of banking conversion rates to drive a wedge between Hotelling paths that are otherwise only governed by the opportunity cost of capital (and the constraints on banking). Accordingly, banking discount rates must be set in line with the targeted price ratios in each period as per $\delta_H/\delta_L = \rho_1/\rho_2$. This ensures that the exchange rates (ρ_1, ρ_2) apply when banking is positive in both markets.

The second interpretation is that "well-set" banking conversion rates adjust banking incentives to control for the distortion in trade incentives between periods induced by using different exchange rates. Recall that setting $\rho_1 > \rho_2 \geq 1$ has two effects. First, cost-effective trade from L to H is more profitable (less costly) in period 2 than in period 1, as this relaxes the overall constraint on emissions. This incentivizes higher in banking L to increase sales in period 2 and lower banking in H. Second, reverse trade from H to L in period 1 may even become more profitable, as this further relaxes the emission constraint. Indeed, having L bank the equivalent of ρ_1 L-permits for each H-permit traded in this way, and then sell only the equivalent of ρ_2 L-permits to return that H-permit to H in period 2 increases emissions by $\rho_1 - \rho_2$. Now, introducing banking conversion rates such that $\delta_H/\delta_L = \rho_1/\rho_2 > 1$ mitigates these incentives by making banking in L less attractive than banking H, and vice versa. For instance, when $\rho_1 = 2\rho_2$, setting $\delta_H = 2\delta_L$ eliminates the gain, in terms of relaxation of the overall emission constraint, from strategies (1) and (2) above. Namely, an L-permit is twice as much valuable for compliance in H in period 2, but because the compliance value in L of an L-permit is halved in period 2, the two effects cancel each other out.

With this interpretation in mind, Result 5 shows that, with positive banking in both markets, setting $\delta_H/\delta_L > \rho_1/\rho_2$ entails lower (slower) price convergence than targeted. This is because banking incentives are "over-adjusted" relative to distorted trading incentives, so that higher banking in H and lower banking in L sustain a price ratio which is higher than ρ_1 in period 1. By the same token, price convergence is greater (faster) than targeted when $\delta_H/\delta_L < \rho_1/\rho_2$, with the price ratio in period 1 comprised between ρ_2 and ρ_1.

Another important policy aspect of the "general rule" in Result 6 is that, not only is it always successful at implementing gradual linking with targeted price ratios, but it further provides flexibility to policy makers to deal with the distributional and environmental impacts of their policy. This flexibility stems from the fact that multiple combinations of banking conversion rates such that $\delta_H/\delta_L = \rho_1/\rho_2$ can be used. Intuitively, since δ_m governs the slope of the Hotelling price path in market m, different pairs (δ_H, δ_L) implement different paths to gradual price convergence. Moreover, banking conversion rates allow for more control over (1) total

emissions that will vary when exchange rates are used, which in turn affects the level of the price paths; (2) the direction of trade, which in turn affects overall cost effectiveness of the policy. The impacts of using different linking and banking conversion rates are studied in more detail with numerical simulations in the following section.

Finally, the astute reader will have noted that only in arguably "rare" instances do we have positive banking in both markets.[17] This is essentially because our stylized frictionless model leads to "polar" equilibria, in which banking is as low as possible in H (ideally zero) and as high as possible in L since $\rho_1 > \rho_2$. In practice, however, a number of frictions or factors may constrain banking in L (e.g., risk aversion or financial constraints) or maintain positive banking in H (e.g., hedging needs or liquidity requirements). In other words, one can expect the situation with positive banking in both markets to be (much) more frequent in practice than in the model, which makes the case for using a "two-legged" gradual linking policy more relevant. We discuss this and related issues further in Section 6.

4 Numerical model

In this section, we first describe how we implement (and extend) the model numerically, and then present how the numerical model is parametrized.

4.1 Model formulation

Outline. There are $i \in \{1, 2, \ldots, I\}$ firms that are regulated in $m \in \{1, 2, \ldots, M\}$ permit markets in T periods indexed by $t \in \{1, \ldots, T\}$. We introduce firms and markets to allow for multiple firms within a given market.

Permit markets are characterized by a sequence of emission limits or caps $\overline{E}_{mt}$. These limits are brought to the market in the form of permits. The initial allocation of permits is done through auctions or free allocation. The amount of permits grandfathered to firm i in market m and period t is given by ω_{imt}. The overall amount of auctioned permits in market m and period t is thus given by $\overline{E}_{mt} - \sum_i \omega_{imt}$.

Markets can be linked in the sense that permits from market m can be transferred to market $\tilde{m}$ and—in the process—transformed into $\tilde{m}$-permits. Linking may be subject to restrictions in the sense that the trading conversion rate $\rho_{m\tilde{m}t}$ specifies the amount of units of permits

[17] As a case in point, this can even never occur with retirement and re-issuance of traded permits.

from market m needed to create one permit in market $\tilde{m}$, and vice versa for $\rho_{\tilde{m}mt}$.

Markets may allow for banking, but banking may be subject to provisions that adjust the compliance value of permits when they are banked from one period to the next. In the model, this is captured by adjusting the stock of banked permits between periods. Specifically, the banking conversion rate, δ_{mt}, is the share of the stock of permits in period $t-1$ that is available in period t. Furthermore, banking might be restricted in terms of the total number of permits max_{mt}^{bank} that can be in the bank at the end of period t in market m.

Besides policy makers that determine the emission caps and the linking and banking rules, there are two main actors: Firms and traders. Firms typically generate emissions in different markets and have to surrender permits of the respective market for emissions caused in that market in each period. Firms can trade only in markets in which they are regulated but are not allowed to trade in other markets. Firms can bank permits of the markets in which they are regulated. Traders do not have compliance obligations and are free to trade permits between all markets. They are, however, not allowed to bank permits. Traders are introduced for simplicity to (1) ensure arbitrage and (2) avoid each firm being active in each market.

Firms. Firms are characterized by a baseline emission path $\overline{e}_{imt}$ (i.e., the emission of firm i in market m in period t in case of no abatement actions). Abatement (or emission reduction) r_{imt} is costly. Abatement cost functions $C_{imt}\left(r_{imt}\right)$ are assumed to be twice differentiable, increasing and strictly convex. Firms receive some permits for free (ω_{imt}) and can buy (a_{imt}^{buy}) or sell (a_{imt}^{sell}) permits in the market. To demonstrate compliance, firms need to surrender permits (a_{imt}^{surr}) in each period to cover their emissions in market m. Finally, a_{imt}^{stock} denotes firm i's stock of banked m-permits at the end of period t.

Firms are assumed to maximize their profits from permit trading net of abatement cost over the whole time horizon. With a discount factor of one, the optimization problem for firm i is

$$\max_{r_{imt},a_{imt}^{buy},a_{imt}^{sell},a_{imt}^{stock},a_{imt}^{surr}\geq 0} \sum_t \sum_m \left[p_{mt}(a_{imt}^{sell}-a_{imt}^{buy}) - C_{imt}\left(r_{imt}\right) - p_{mt}^{bank} a_{imt}^{stock}\right]$$

$$\begin{aligned}
a_{imt}^{stock} &= \delta_{mt} a_{im(t-1)}^{stock} + \omega_{imt} - a_{imt}^{surr} + a_{imt}^{buy} - a_{imt}^{sell} && (p_{imt}^{stock}) && \forall m,t \\
a_{imt}^{surr} &\geq (\overline{e}_{imt} - r_{imt}) && (p_{imt}^{surr}) && \forall m,t \\
a_{im0}^{stock} &\text{ given}
\end{aligned}$$

The last term in the objective function denotes the cost to acquire permits that allow banking in the case of a limit on the total permit bank in market m, with p_{mt}^{bank} the associated shadow

price. The second equation represents the stock-flow relation for permits. The third equation ensures that permits surrendered equal at least emissions caused.

The equilibrium conditions are derived by taking the first-order conditions

$$\frac{\partial C_{imt}}{\partial r_{imt}} \geq p^{surr}_{mt} \quad \perp \quad r_{imt} \geq 0 \qquad \forall i,m,t \qquad (5)$$

$$p^{stock}_{imt} \geq p_{mt} \quad \perp \quad a^{sell}_{imt} \geq 0 \qquad \forall i,m,t \qquad (6)$$

$$p_{mt} \geq p^{stock}_{imt} \quad \perp \quad a^{buy}_{imt} \geq 0 \qquad \forall i,m,t \qquad (7)$$

$$p^{stock}_{imt} + p^{bank}_{mt} \geq \delta_{m(t+1)} p^{stock}_{im(t+1)} \quad \perp \quad a^{stock}_{imt} \geq 0 \qquad \forall i,m,t \qquad (8)$$

$$p^{stock}_{imt} \geq p^{surr}_{imt} \quad \perp \quad a^{surr}_{imt} \geq 0 \qquad \forall i,m,t \qquad (9)$$

$$a^{stock}_{imt} = \delta_{mt} a^{stock}_{im(t-1)} + \omega_{imt} - a^{surr}_{imt} + a^{buy}_{imt} - a^{sell}_{imt} \quad \perp \quad p^{stock}_{imt} \text{ free} \qquad \forall r,t \qquad (10)$$

$$a^{surr}_{imt} \geq (\overline{e}_{imt} - r_{imt}) \quad \perp \quad p^{surr}_{imt} \geq 0 \qquad \forall m,t \qquad (11)$$

$$a^{stock}_{im0} \text{ given}$$

Equations (5–8) are zero-profit conditions that state that the receptive variable must be zero if the left-hand side (unit cost) exceeds the right hand side (unit profit). For instance, (5) states that if the cost of abating one unit of emissions exceeds the value of surrendering a permit, abatement is zero. In contrast, if the marginal abatement cost equals the permit price, abatement is positive. Next, (6) states that firm i will sell permits in market m and period t ($a^{sell}_{imt} \geq 0$) if the cost of selling a permit measured as the opportunity cost of holding a permit (p^{stock}_{imt}) equals the market price received for that permit (p_{mt}). For permit purchase, (7) follows the same logic but with opposite signs. The zero-profit condition for the permit bank in (8) equates the cost of acquiring the bank to the benefit derived from the bank: The cost is given as the price of acquiring a banked permit (p^{stock}_{imt}) plus the price of being allowed to bank (p^{bank}_{mt}), which is positive if the banking restriction in (14) binds. The benefit is defined as the next period value of a banked permit taking into account the banking conversion rate. Next, (9) expresses the tradeoff between banking and surrendering a permit, which connects the respective prices. The value of a banked permit (p^{stock}_{imt}) is determined in conjunction with the stock-flow relation for the permit stock in (10). Finally, (11) ensures compliance (i.e., surrendered permits equal at least baseline emissions net of abatement) and determines the price for surrendering a permit (p^{surr}_{imt}).

Traders. Traders arbitrage between markets but are not allowed to bank permits. In order to create one permit in market $\tilde{m}$ they have to buy $\rho_{m\tilde{m}t}$ permits from market m. Traders

maximize their profits under the assumption of perfect competition choosing the amount of permits traded from market m to market $\tilde{m}$ for each period t, $x_{m\tilde{m}t}$, such that

$$\max_{x_{m\tilde{m}t}\geq 0} \sum_t \sum_{m,\tilde{m}} (p_{\tilde{m}t} - \rho_{m\tilde{m}t} p_{mt}) x_{m\tilde{m}t}.$$

The corresponding equilibrium conditions are

$$\rho_{m\tilde{m}t} p_{mt} \geq p_{\tilde{m}t} \quad \perp \quad x_{m\tilde{m}t} \geq 0 \quad \forall m, \tilde{m}, t. \tag{12}$$

Market clearing. The model is closed with the condition that markets must be in equilibrium in each period, namely

$$\overline{E}_{mt} - \sum_i \omega_{imt} + \sum_i a^{sell}_{imt} + \sum_{\tilde{m}} x_{\tilde{m}mt} \geq \sum_i a^{buy}_{imt} + \sum_{\tilde{m}} x_{m\tilde{m}t} \quad \perp \quad p_{mt} \geq 0 \quad \forall m, t. \tag{13}$$

The supply on the left-hand side is given by the amount of auctioned permits, the sum of permits sold by firms, and the amount of imported permits. The demand on the right-hand side is given by the sum of permits bought by firms and the amount of exported permits.

In the case of limited banking, the market clearing condition for banking permits determines the respective price

$$max^{bank}_{mt} \geq \sum_i a^{stock}_{imt} \quad \perp \quad p^{bank}_{mt} \geq 0 \quad \forall m, t. \tag{14}$$

The full numerical model is given by equations (5) to (14).

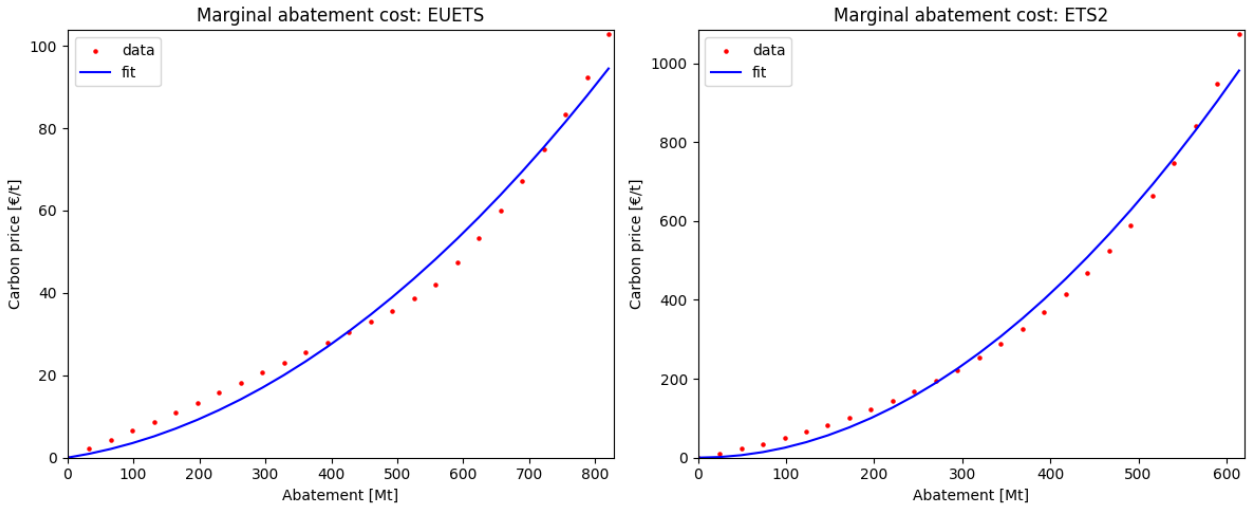
4.2 Model parameterization

The numerical model is parameterized by calibrating the coefficients of the marginal abatement cost (MAC) curves and baseline emissions (i.e., emissions with no abatement actions). We determine these parameters for two markets, the EU Emissions Trading System (ETS1) and the upcoming EU Emissions Trading System for Transport and Heat (ETS2).

We follow the literature deriving the MAC coefficients based on a global Computable General Equilibrium (CGE) model (e.g., Böhringer et al., 2006; Klepper & Peterson, 2006). The methodology is described in detail in Abrell & Rausch (2017). The static multi-sector multi-region model is based on the GTAP 10 database (Aguiar et al., 2019) using 2014 as the

base year. The CGE resolves all EU countries with sufficient sectoral details to represent the coverage of the two ETSs. For each ETS, we impose a sequence of reduction requirements ranging from 0 to 50% in steps of 2%. For each sector and reduction requirement, the CGE model produces the price that clears the permit market and the associated abatement level. To determine the coefficients of the MAC curves, we fit a quadratic function under the restriction that the constant term is zero (Figure 1). Baseline emissions are determined for the base year 2014 and are taken from the CGE model (Table 2).

Figure 1: Marginal Abatement Cost Curves



Note: Marginal abatement cost curves derived from the CGE model. Red dots correspond to price-abatement pairs received from the CGE model and the blue lines are the MAC curves fitted under the assumption of a quadratic function that passes through the origin (the constant term is zero).

The model parameters are assumed to be constant in time. That is, we assume that no technological progress takes place and that baseline emissions stay constant over time. We make these assumptions for lack of good or uniformly agreed parameters charactering technological progress for abatement technologies or the permanence of emission reductions.

Table 2: Numerical model parameters

	Marginal Abatement Cost			**Baseline Emissions**
	Constant [$]	Linear [$/t]	Quadratic [$/t^2]	[10^6t]
ETS1	0	0.025025	0.00011	1'640
ETS2	0	0	0.002601	1'280

Note: Dollar values are denominated in USD_{2014}.

5 Numerical illustration

In this section, we describe the linking scenarios studied and discuss the numerical results.

5.1 Scenario outlines

Our numerical simulations systematically analyze the impact of different linking and banking conversion rates. We keep the ovreall emission reduction targets constant across all scenarios with variations in the conversion rates only. We implement the scenarios for two time periods. The first period corresponds to the year 2027 when ETS2 starts to become functional and 2030 is the final period (in line with the objectives in the current legislation). For simplicity, we assume a zero continuation value for banked permits at the end of period 2. A positive value would increase the incentive to bank and banking levels throughout, which would make the case for using linking and banking conversion rates jointly even more salient.

The absolute emission reduction targets are derived from relative reduction targets provided in the current legislation and from historical emissions for the year 2014 in the ETS1 and ETS2 perimeters (Table 3). The European Environmental Agency reports verified emissions in the ETS1.[18] We use the value for the year 2014 including the correction for scope expansion and excluding non-EU countries. Graichen & Ludig (2024) estimate the historical emissions for ETS2. As they do not provide a value for 2014 we use the 2015 value instead.[19] Günther et al. (2024) and Pahle et al. (2025) provide estimates for the emission caps in the ETS1 and ETS2 for the years 2027 and 2030 based on the current legislation.

Table 3: Reduction targets

	2014	**2027**		**2030**	
	Observed [Mt]	**Target [Mt]**	**Reduction**	**Target [Mt]**	**Reduction**
ETS1	1630	1095	33%	825	49%
ETS2	1350	1036	23%	849	37%
Total	2980	2131	28%	1674	44%

Note: Observed emissions are based on EEA's Emissions Trading viewer for ETS1 and Graichen & Ludig (2024) for ETS2 (2015 value is used). Emission targets (caps) are taken from Günther et al. (2024) and Pahle et al. (2025). The reduction column provides the relative reduction compared to the base year 2014.

[18] The Emissions Trading Viewer accessible at this link.

[19] Historical emissions and baseline emissions in the numerical model are in line. For the ETS1 the observed value amounts to 1'630 Mt in 2014 compared to 1'640 Mt in the CGE model. For the ETS2 the observed value in 2015 is 1'350 Mt compared to 1'228 Mt in the CGE model in 2014.

In all scenarios, all permits are auctioned and no free allocation takes place.[20] The amount of auctioned permits equals the emission target in the respective period; but an alternative case of full frontloading is also considered in which both the emission targets for periods 1 and 2 are aunctioned in period 1. By default, banking is limited to not exceed the amount of auctioned permits in both markets taking into account the linking conversion rate ($max_{m1}^{bank} = \overline{E}_{m1} + \sum_{\tilde{m}} \rho_{\tilde{m}m1}\overline{E}_{\tilde{m}1}$). This corresponds to the case of retirement and re-issuance of permits when they trade across markets. Without retirement and re-issuance, banking is bounded from above by the amount of permits auctioned in the market considered ($max_{m1}^{bank} = \overline{E}_{m1}$).

5.2 Numerical results

Reference scenarios. We start with reference scenarios, some of which include traditional 1:1 linking or banking or both. These results are in the first block of Table 4. Linking (resp. banking) equates prices and marginal abatement costs across markets (resp. periods). When banking is allowed, both markets bank permits, which yields equal emissions in both periods because MAC curves and baseline emissions are time invariant and there is no discounting. Full flexibility with both 1:1 linking and 1:1 banking achieves the first-best outcome for the specified level of emissions by equalizing marginal costs across markets and periods, which provides numerical evidence for Result 1 from our theory model.

Linking and banking each yield efficiency gains, as seen by the lower average abatement costs, compared to \$66 per ton with no flexibility. Efficiency gains from 1:1 linking alone (reducing average cost by \$26 per ton) are higher than those from 1:1 banking alone (reducing by only \$8 per ton) because the price differentials across markets are larger than those across periods. These gains are sub-additive, however, with full flexibility reducing average costs by \$30 per ton, which suggests some "redundancy" in the flexibility provided by linking and banking.

Anticipation of linking. The first best may result even without linking in period 1—so long as the link in period 2 is known in advance—as the second block of Table 4 shows. This occurs when permits are sufficiently frontloaded that both markets bank and all cost-effective trades between markets can be shifted to period 2, which provides numerical evidence for Result 2 from our theory model. In that case, the least-cost price signal is effective in both markets and periods through banking, despite the link not being active in period 1. Without sufficient frontloading, emissions in H are constrained in period 1 relative to first best—H

[20] This is for convenience but without loss of generality for our results under perfect competitition.

does not bank and would reduce compliance costs by borrowing permits from period 2 if that were allowed—and some efficiency gains are left unrealized, yielding an average abatement cost of $2 per ton greater than first best.

Basic effects of linking conversion rates. With uniform linking conversion rates across periods ($\rho_1 = \rho_2 = \rho$) and unitary banking conversion rates ($\delta_L = \delta_H = 1$), our model collapses to the dynamic equivalent of a static model that illustrates the findings of Quemin & de Perthuis (2019) and Woerman (2023). These results are in the third block of Table 4. As described in that literature, the politically plausible range for the linking conversion rate is between 1 and the ratio of autarchic market prices; here we use $\rho = 2$ as an example. This rate yields beneficial outcomes with permits flowing from L to H, reducing both the initial price wedge (efficiency gains) and total emissions (environmental dividend) as every 2 L-permits traded are converted into 1 H-permit.

Outside of that range, linking can have perverse outcomes. When ρ is below 1 (here, $\rho = 0.5$), permits again flow from L to H, but the initial price wedge becomes inverted (L becomes the high-price market, and H the low-price market) and total emissions increase (as every L-permit traded is converted into 2 H-permit). When ρ is above the ratio of autarchic market prices (here, $\rho = 6$), permits instead flow in the unnatural direction from H to L, which causes both total emissions and the initial price wedge to increase.

Basic effects of banking conversion rates. To prepare for the analysis of gradual linking policies, we first consider the effects of uniform banking conversion rates set below or above one ($\delta_L = \delta_H = \delta \lessgtr 1$) with unitary linking conversion rates ($\rho_1 = \rho_2 = 1$). These results are in the fourth block of Table 4. Setting $\delta < 1$ reduces banking incentives as banking leads to less emissions being allowed, and more trade occurs than in the first best as the relative value of flexibility levers is distorted in favor of linking. Banked period-1 permits are δ^{-1} times less valuable than period-2 permits, hence the Hotelling price path is upward sloping. Instead, setting $\delta > 1$ increases banking incentives as banking leads to more emissions being allowed, and less trade occurs than in the first best as the relative value of flexibility levers is distorted in favor of banking. Banked period-1 permits are δ times more valuable than period-2 permits, hence the Hotelling price path is downward sloping.

Table 4: Reference scenarios and basic linking policies

Policy		L = ETS1						H = ETS2						H+L (t1+t2)	
		ω_{Lt}	e_{Lt}	$\rho_t\Delta_t$	$(\delta_L)B_L$	$-C'_{Lt}$	p_{Lt}	ω_{Ht}	e_{Ht}	Δ_t	$(\delta_H)B_H$	$-C'_{Ht}$	p_{Ht}	Emission	Cost/ton
No flexibility	t1	1099	1099	n.a.	n.a.	46	46	946	946	n.a.	n.a.	208	208	3656	66
(autarchy, no banking)	t2	837	837	n.a.	n.a.	91	91	774	774	n.a.	n.a.	537	537		
Spatial flexibility	t1	1099	975	124	n.a.	65	65	946	1070	124	n.a.	65	65	3656	40
(1:1 linking, no banking)	t2	837	615	222	n.a.	141	141	774	996	222	n.a.	141	141		
Temporal flexibility	t1	1099	968	n.a.	131	67	67	946	860	n.a.	86	353	353	3656	58
(autarchy, 1:1 banking)	t2	837	968	n.a.	131	67	67	774	860	n.a.	86	353	353		
Full flexibility	t1	1099	795	190	114	100	100	946	1033	190	103	100	100	3656	36
(1:1 linking, 1:1 banking)	t2	837	795	156	114	100	100	774	1033	156	103	100	100		
1:1 linking in t2 only	t1	1099	834	n.a.	265	92	92	946	946	n.a.	0	208	208	3656	38
with 1:1 banking	t2	837	834	267	265	92	92	774	1041	265	0	92	92		
1:1 linking in t2 only	t1	1936	795	n.a.	1141	100	100	1720	1033	n.a.	687	100	100	3656	36
with 1:1 banking	t2	0	795	346	1141	100	100	0	1033	346	687	100	100		
Linking: ρ_1=ρ_2=2	t1	1099	788	225	85	101	101	946	950	113	109	202	202	3476	45
Banking: δ_H=δ_L=1	t2	837	788	133	85	101	101	774	950	67	109	202	202		
Linking: ρ_1=ρ_2=0.5	t1	1099	849	112	138	89	89	946	1098	224	72	44	44	3894	30
Banking: δ_H=δ_L=1	t2	837	849	126	138	89	89	774	1098	252	72	44	44		
Linking: ρ_1=ρ_2=6	t1	1099	1002	35	62	61	61	946	854	6	98	364	364	3713	59
Banking: δ_H=δ_L=1	t2	837	1002	-103	62	61	61	774	854	-17	98	364	364		
Linking: ρ_1=ρ_2=1	t1	1099	953	146	0	69	69	946	1066	146	27	69	69	3642	40
Banking: δ_H=δ_L=0.5	t2	837	626	210	0	138	138	774	998	210	14	138	138		
Linking: ρ_1=ρ_2=1	t1	1099	759	192	148	107	107	946	1025	192	113	107	107	3916	31
Banking: δ_H=δ_L=2	t2	837	1046	86	295	54	54	774	1085	86	225	54	54		

Note: Allocation, emission, trade, and banking levels are in million tons. Prices and marginal abatement costs are in USD per ton. Banking is given in period-1 units in lines t1, and in period-2 units in lines t2, which differ when $\delta \neq 1$.

Table 5: Naïve or partial gradual linking policies

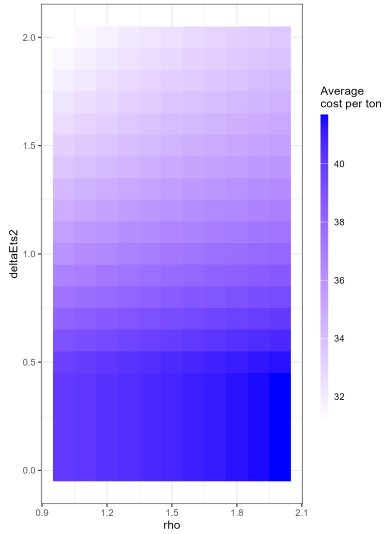
		L = ETS1						H = ETS2						H+L (t1+t2)	
Policy		ω_{Lt}	e_{Lt}	$\rho_t\Delta_t$	$(\delta_L)B_L$	$-C'_{Lt}$	p_{Lt}	ω_{Ht}	e_{Ht}	Δ_t	$(\delta_H)B_H$	$-C'_{Ht}$	p_{Ht}	Emission	Cost/ton
Linking: ρ_1=2; ρ_2=1	t1	1099	822	27	250	94	94	946	960	14	0	188	188	3642	38
Banking: δ_H=δ_L=1	t2	837	822	264	250	94	94	774	1038	264	0	94	94		
Linking: ρ_1=2; ρ_2=1	t1	1936	1174	-1174	1936	36	71	1720	1063	-587	70	71	71	4243	22
Banking: δ_H=δ_L=1	t2	0	942	994	1936	71	71	0	1063	994	70	71	71		
Linking: ρ_1=3; ρ_2=1	t1	1099	873	-86	312	84	84	946	918	-28	0	252	252	3713	46
Banking: δ_H=δ_L=1	t2	837	873	275	312	84	84	774	1049	275	0	84	84		
Linking: ρ_1=3; ρ_2=1	t1	1936	1333	-1333	1936	18	54	1720	1084	-444	192	54	54	4544	17
Banking: δ_H=δ_L=1	t2	0	1043	893	1936	54	54	0	1084	893	192	54	54		
Linking: ρ_1=1.5; ρ_2=1	t1	1099	804	67	228	98	98	946	991	45	0	147	147	3633	37
Banking: δ_H=δ_L=1	t2	837	804	261	228	98	98	774	1035	261	0	98	98		
Linking: ρ_1=1.5; ρ_2=1	t1	1936	1014	-1014	1936	59	82	1720	1044	-676	0	88	88	3994	27
Banking: δ_H=δ_L=1	t2	0	884	1051	1936	82	82	0	1051	1051	0	82	82		
Linking: ρ_1=3; ρ_2=2	t1	1099	839	-123	383	91	91	946	905	-41	0	272	272	3547	45
Banking: δ_H=δ_L=1	t2	837	839	381	383	91	91	774	965	190	0	181	181		
Linking: ρ_1=3; ρ_2=2	t1	1936	1007	-1007	1936	60	90	1720	966	-336	419	180	180	3780	37
Banking: δ_H=δ_L=1	t2	0	842	1093	1936	90	90	0	966	547	419	180	180		

Note: Allocation, emission, trade, and banking levels are in million tons. Prices and marginal abatement costs are in USD per ton. Banking is given in period-1 units in lines t1, and in period-2 units in lines t2, which differ when $\delta \neq 1$. All cases shown assume no retirement and re-issuance of permits when they trade across markets (i.e., $B_L \leq \omega_{L1}$). When this constraint is binding (i.e., when $B_L = \omega_{L1}$), the dual associated with this constraint is positive and drives a wedge between the MAC, $-C'_{L1}$, and the market price, p_{L1}. In particular, $-C'_{H1}/-C'_{L1} = \rho_1$ and $p_{H1}/p_{L1} = \rho_2$.

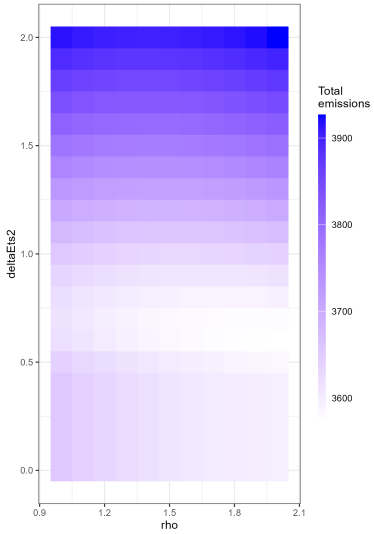
Table 6: Successful or flexible gradual linking policies

		L = ETS1						H = ETS2						H+L (t1+t2)	
Policy		ω_{Lt}	e_{Lt}	$\rho_t\Delta_t$	$(\delta_L)B_L$	$-C'_{Lt}$	p_{Lt}	ω_{Ht}	e_{Ht}	Δ_t	$(\delta_H)B_H$	$-C'_{Ht}$	p_{Ht}	Emission	Cost/ton
Linking: ρ_1=2; ρ_2=1	t1	1099	956	143	0	69	69	946	999	72	19	137	137	3584	41
Banking: δ_H=1; δ_L=0.5	t2	837	631	206	0	137	137	774	999	206	19	137	137		
Linking: ρ_1=2; ρ_2=1	t1	1936	789	451	696	101	101	1720	950	225	996	202	202	3082	60
Banking: δ_H=1; δ_L=0.5	t2	0	394	-46	348	202	202	0	950	-46	996	202	202		
Linking: ρ_1=2; ρ_2=1	t1	1099	822	265	12	94	94	946	960	132	119	188	188	3642	38
Banking: δ_H=2; δ_L=1	t2	837	822	27	12	94	94	774	1038	27	238	94	94		
Linking: ρ_1=2; ρ_2=1	t1	1936	1099	121	716	46	46	1720	1041	60	740	92	92	4335	19
Banking: δ_H=2; δ_L=1	t2	0	1099	-383	716	46	46	0	1041	-383	1479	46	46		
Linking: ρ_1=2; ρ_2=1	t1	1099	864	235	0	86	86	946	972	117	91	171	171	3584	40
Banking: δ_H=1.5; δ_L=0.75	t2	837	729	108	0	114	114	774	1019	108	137	114	114		
Linking: ρ_1=2; ρ_2=1	t1	1936	933	297	705	73	73	1720	992	149	877	145	145	3769	34
Banking: δ_H=1.5; δ_L=0.75	t2	0	808	-279	529	97	97	0	1036	-279	1315	97	97		
Linking: ρ_1=2; ρ_2=1	t1	1099	979	120	0	65	65	946	1006	60	0	129	129	3596	42
Banking: δ_H=0.8; δ_L=0.4	t2	837	615	222	0	141	141	774	996	222	0	141	141		
Linking: ρ_1=2; ρ_2=1	t1	1936	759	492	685	108	108	1720	941	246	1025	215	215	2794	76
Banking: δ_H=0.8; δ_L=0.4	t2	0	187	87	274	269	269	0	907	87	820	269	269		

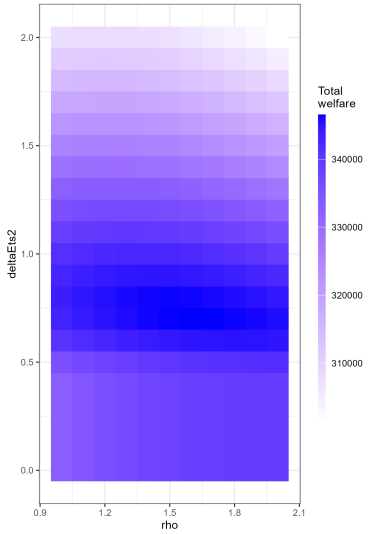
Note: Allocation, emission, trade, and banking levels are in million tons. Prices and marginal abatement costs are in USD per ton. Banking is given in period-1 units in lines t1, and in period-2 units in lines t2, which differ when $\delta \neq 1$. All cases shown are identical with *and* without retirement and re-issuance of traded permits. All cases shown assume linking and banking conversion rates satisfy $\delta_H/\delta_L = \rho_1/\rho_2$.



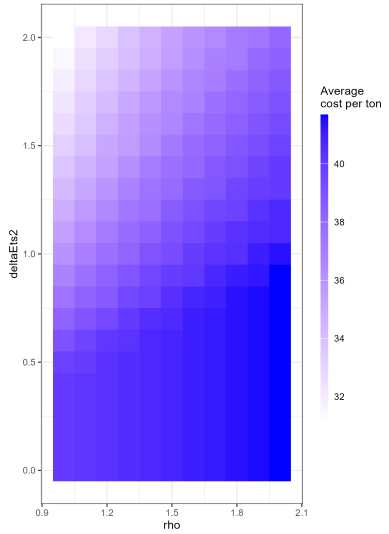
(a) Naive: AvgCost



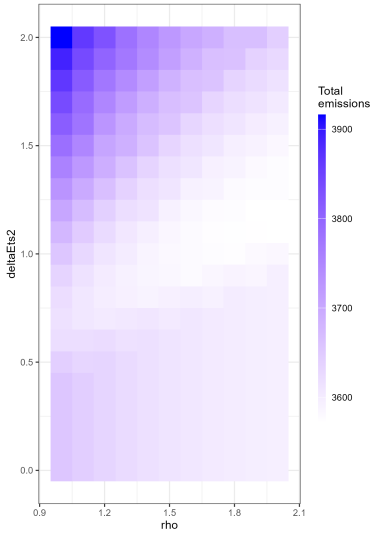
(b) Naive: Emissions



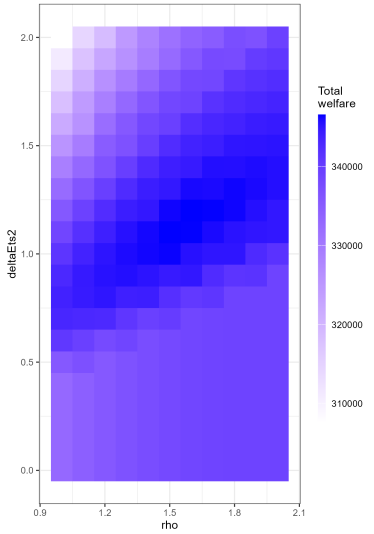
(c) Naive: Welfare



(d) Grad: AvgCost



(e) Grad: Emissions



(f) Grad: Welfare

Naïve gradual linking policies. We now consider potential policies to achieve gradual price convergence in a linked system. We begin by comparing the impacts of different targeted rates of price convergence across markets ($\rho_1 > \rho_2 \geq 1$) with unitary banking conversion rates ($\delta_L = \delta_H = 1$), which our theory model indicates will not be successful at controlling price convergence in all cases. These results are in Table 5. Because $\rho_1 > \rho_2$, the terms of trade from L to H are more profitable in period 2. This leads to as little trade from L to H as possible in period 1—and at times even trade in the unnatural direction from H to L—and as small banking as possible in H (zero, if feasible). Two broad cases can be delineated.

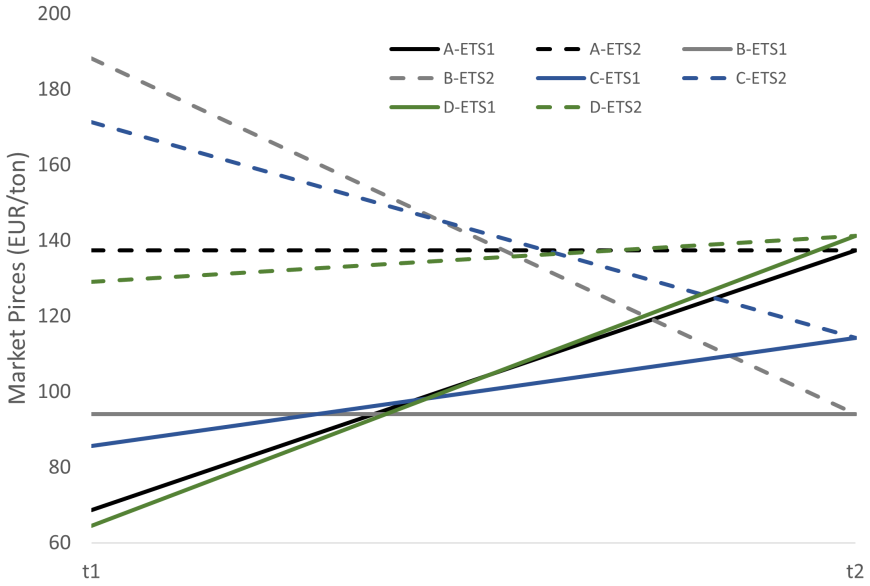
First, when permits are not sufficiently frontloaded and H does not bank. In this case, the smallest amount possible of trade from L to H occurs in period 1 so that $B_H = 0$. Trade may even reverse from H to L when ρ_1 is sufficiently large (here, $\rho_1 = 3$) as this relaxes the overall emission contraint. The absence of banking in H makes it possible that both the price and MAC ratios are governed by the current exchange rate in each period. Such a "corner" case with zero banking is the only situation where the rate of price convergence is "as desired" (i.e., in line with the sequence of exchange rates). That said, the targeted rate of price convergence can only occur with the price path in H decreasing over time, reaching the flat price path in L from above in period 2 (when $\rho_2 = 1$). In addition, total emissions increase when permits flow in the unnatural direction in period 1.

Second, when permits are sufficiently frontloaded and H features positive banking. Permit allocations in period 1 are so great that both markets bank permits and that permits flow from H to L in period 1, which relaxes the overall emissions constraint. As both markets bank permits, ρ_2 defines the price ratio in *both* periods, i.e., the price ratio converges to ρ_2 in period 1 through arbitrage even though the exchange rate ρ_1 applies. Not only does the attempt to phase in price convergence across markets fails, but also total emissions systematically increase as a result of trade in the unnatural direction in period 1.

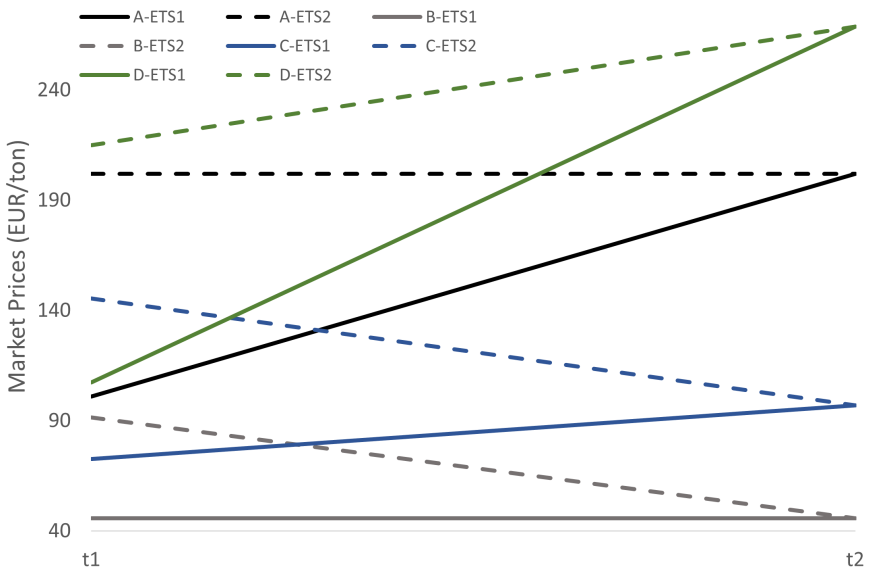
While ρ_1 may be mute when it comes to governing prices, it always governs the MAC ratio in period 1. In the extreme case were all permits are allocated in period 1, observe that the price in L in period 1 is greater than the MAC in that market-period. This occurs as the (discounted) value of a period-1 permit banked for use in period 2, reflecting the MAC in period 2, causes excess permit demand in period 1. In this context, all period-1 permits in L are banked into period 2. Then, L must rely exclusively on H-permits acquired at the linking conversion rate for compliance in period 1. Thus, L will reduce emissions such that their MAC equals their permit cost, which in turn equals the MAC (and the permit price) in H adjusted for the linking conversion rate. As a result, the relative MACs in period 1—but

Figure 2: Price paths in ETS1 and ETS2 when $\delta_H/\delta_L = \rho_1/\rho_2$

(a) Reference permit allocation



(b) Maximum permit frontloading



Note: (a): Permit allocation per period is as calibrated in Section 4.2; (b): All period-2 permits are front-loaded to period-1 allocation. The scenarios A, B, C, and D implement different pairs (δ_H, δ_L) that satisfy $\delta_H/\delta_L = \rho_1/\rho_2$ with $\rho_1 = 2$ and $\rho_2 = 1$, namely A: $\delta_H = 1$, $\delta_L = 0.5$; B: $\delta_H = 2$, $\delta_L = 1$; C: $\delta_H = 1.5$, $\delta_L = 0.75$; D: $\delta_H = 0.8$, $\delta_L = 0.4$. Aggregate emissions vary (w.r.t. to the fixed number of permits initially equivalent to 3.66 billion tons) and are inversely commensurate with the average price in each scenario, namely A(a): 3.58, A(b): 3.08; B(a): 3.64, B(b): 4.34; C(a): 3.58, C(b): 3.77; D(a): 3.60, D(b): 2.79.

not the relative prices—are governed by the linking conversion rate.

These findings of when ρ_1 is effective or not provide numerical evidence for Result 3 from our theory model. The rates of price convergence in these scenarios provide examples of Result 5. As permits flow in the unnatural direction under many of these naïve policies, these results showing an increase in emissions provide numerical evidence for Result 4.

Successful gradual linking policies. Not only is a two-legged gradual linking policy (i.e., leveraging both linking and banking conversion rates) always successful at implementing the targeted rate of price convergence, it also gives more flexibility to policy makers to deal with the distributional, environmental and price impacts of their policy. For instance, total emissions can be reduced, as opposed to increased in general with a one-legged policy (i.e., using only linking conversion rates), and more options for setting the price paths are available. Results from these successful policies are in Table 6 and Figure 2.

In all of these cases, linking and banking conversion rates are aligned as in Result 6. And in every case, the desired rate of price convergence—defined by ρ_1 and ρ_2—is achieved, providing numerical evidence for Result 6. Observe these cases include "corner" cases featuring zero banking in one or both markets. Figure 2 shows that for a particular rate of price convergence (here, $\rho_1 = 2$ and $\rho_2 = 1$), many different price paths can be achieved depending on the choice of banking discount rates and the allocation of permits over time periods.

The choice of linking and banking conversion rates will also affect total emissions. Recall that these rates may have reinforcing or opposing effects on emissions depending on how they are set. In all of the cases here, permits always trade in the natural direction from L to H in period 1, which serves to decrease emissions because $\rho_1 > 1$. Setting δ below (resp. above) one decreases (resp. increases) the incentive to bank, and each banked permit allows for fewer (resp. more) emissions, which tends to decrease (resp. increase) total emissions. Frontloading permit allocation increases banking and exacerbates these effects of δ, so total emissions may increase or decrease, and potentially by substantial amounts relative to first best. In all four cases without frontloading, however, we find that total emissions decrease under successful gradual price convergence.

6 Policy implementation and extensions

In this section, we discuss implementation issues with our gradual linking policy, and assess how it would perform compared to alternative policy instruments. In doing so, we go beyond the confined (simple) problem setting considered above and explore extensions and further considerations, also with a view on future work and model refinement.

Policy implementation. We deem a gradual linking policy "successful" if it results in a Pareto-improvement along three dimensions—allocative efficiency, distributional efficiency, and environmental efficacy. That is, policy "success" means three goals are met: some price convergence across markets (efficiency gains), which is controlled and rises gradually over time (distributional aspects), and no environmental worsening (total emissions cannot increase). Our theoretical model has established the relation between linking and banking conversion rates necessary to ensure the first two goals are met—the "two-legged" policy in Result 6. Our numerical simulations have further illustrated principles and desirable value ranges for the choice of conversion rates to ensure linking leads to an environmental dividend.

In practice, policy implementation is likely to span more than two periods. The necessary relation between linking and banking conversion rates to ensure gradual price convergence in Result 6 naturally extends to an abitrary number T of periods. Using the notations defined in Section 4, and assuming positive banking in both H and L throughout, the intertemporal arbitrage condition yields $\prod_{t=2}^{T} \delta_{Ht}/\delta_{Lt} = \rho_1/\rho_T$. Importantly, this can be expanded into

$$\frac{\delta_{H2}}{\delta_{L2}} \times \cdots \times \frac{\delta_{HT}}{\delta_{LT}} = \frac{\rho_1}{\rho_2} \times \cdots \times \frac{\rho_{T-1}}{\rho_T}, \tag{15}$$

showing that a targeted sequence of price ratios $(\rho_1, \ldots, \rho_T)$ can be sustained in equilibrium with a sequence of banking conversion rates calibrated to the associated changes in linking conversion rates across periods, i.e., $\delta_{Ht}/\delta_{Lt} = \rho_{t-1}/\rho_t$.[21] Note that so long as policy makers commit to a given pace of price convergence $(\rho_1, \ldots, \rho_T)$, the introduction of uncertainty or shocks in the model does not change the definition of the companion sequence of banking conversion rates, as the same intertemporal condition holds in expectations.

However, with the realization of shocks or the arrival of new information, policy makers may want to adjust the pace of price convergence.[22] Because our gradual linking policy rule can

[21] The sequence of banking conversion rates that sustains $(\rho_1, \ldots, \rho_T)$ in equilibrium can also be defined recursively (period across period) starting backwards from period T.

[22] As in the deterministic case, we do not consider how linking conversion rates are set or adjusted.

be broken down by period, as (15) shows, and because it is independent of history,[23] it can accommodate such regulatory changes. For instance, if in period $\tilde{t}$ policy makers revise future price ratios to $(\rho'_{\tilde{t}+1}, \ldots, \rho'_T)$, then it suffices to adjust the banking conversion rates such that $\delta_{H\tilde{t}+1}/\delta_{L\tilde{t}+1} = \rho_{\tilde{t}}/\rho'_{\tilde{t}+1}$ and $\delta_{Ht}/\delta_{Lt} = \rho'_{t-1}/\rho'_t$ for $t \in (\tilde{t}+2, \ldots, T)$.

Another implementation aspect is whether or not a gradual linking policy can be enforced unilaterally, for instance in the case of linking between distinct jurisdictions. Observe that for a given targeted sequence of price ratios (ρ_1, ρ_2), Result 5 shows that only the ratio of banking conversion rates matters. This means that for "success" as we define it, only one market may introduce a non-unitary banking conversion rate—i.e., $\delta_H = \rho_1/\rho_2$ or $\delta_L = \rho_2/\rho_1$—to ensure that price ratios are effectively governed by ρ_1 first and ρ_2 then. In addition, the choice of the linking conversion rates may also be made unilaterally, but the implementing market needs to specify applicable rates for both imports and exports. For instance, if only H introduces linking conversion rates, our results carry over with a rate of ρ_t for imports and a rate of $1/\rho_t$ for exports. Overall, therefore, gradual linking with a targeted rate of price convergence may in principle be implemented unilaterally by one market, say H, through a pair of directional linking conversion rates per period $(\rho_t, 1/\rho_t)$ and a banking conversion rate $\delta_H = \rho_1/\rho_2$.

Finally, two caveats are worth mentioning when it comes to practical implementation of the gradual linking policies considered here. First, banking and linking may not be as substitutable in practice—in terms of flexibility provided—as our stylized model suggests. This is because the inter-market and intertemporal trade dimensions may be subject to frictions, and because leveraging one dimension can partly alleviate frictions that may hinder the other dimension. For instance, trading may be hindered by transaction costs (e.g., Baudry et al., 2021), while banking may be constrained by limited foresight and financial constraints (e.g., Quemin & Trotignon, 2021) or the perceived credibility of the program (e.g., Dolphin et al., 2023).[24] While frictions make the case for using the "two-legged" gradual linking policy more relevant in practice (since banking is more likely to be positive in both markets than in the frictionless model), they may also affect arbitrage and lead to a rate of price convergence across markets that may slightly deviate from δ_H/δ_L or ρ_1/ρ_2.

Second, we assume abatement cost functions and baseline emissions are exogenously given and independent of the permit price. However, our results would be altered in the presence of investment in permanent abatement or in improved abatement technologies, notably because

[23] To see this, note that our results hold independently of the initial stock of permits held by firms.

[24] Banking has the potential to increase economic efficiency in the presence of transaction costs (Toyama, 2024), and the existence of a bank and associated vested interest may contribute to the program's durability.

investment can act as a substitute or complement for banking (e.g., Phaneuf & Requate, 2002; Slechten, 2013; Pommeret & Schubert, 2018). As a result, endogenous investment would alter arbitrage conditions, and in turn, the rate of price convergence. One might conjecture that trading and banking conversion rates may no longer ensure the targeted price paths, and that another policy instrument would be needed to address the investment channel. Arguably, such a "three-legged" gradual linking policy may be difficult to define and parametrize.

Alternative instruments: Limits to trade or fungibility. Other policy tools to control price convergence that may have more traction with policy makers are limits on trade volume or the fungibility of banked permits across markets. This is essentially because they are the most elementary regulatory intervention—but as we will see in this case, simplicity does not mean effectiveness. For instance, the volume of inter-market trade per period may be limited to a given share of the domestic allocation volume, or the compliance value of banked permits in a given market for use in the other market may be restricted and vintage-dependent.[25] Typically, such quantity constraints require that imported permits are rationed among firms, which could be implemented through a constraint on the ratio of surrender of premits for compliance for firms so that only a certain percentage can be from the other market.[26]

On the face of it, these policies seem easier to implement, but various issues arise. Notably, they do not provide an adequate handle on price convergence and may bring about unclear distributional effects. To see this, consider the case of no-linking in period 1 with volume-constrained 1:1 linking in period 2. That is, $\Delta_1 = 0$ and $\Delta_2 \leq \alpha\omega_{L2} + \beta B_L$, where $\alpha \in (0;1)$ is the share of vintage-2 permits L is allowed to transfer to H and $\beta \in (0;1)$ is the share of banked vintage-1 permits L is allowed to transfer to H. With γ denoting the Lagrange

[25] In the previous sections, banking conversion rates apply domestically. Here, conversion would apply to banked permits only when they flow to the other market. Assuming such a "fungibility rate" is less than one, it would only kick in after all possible vintage-2 permit trades have been exhausted: Before then, the fungibility rate has no effect; afterwards, it has the same effect as an exchange rate. It is thus uncertain ex-ante whether or not a fungibility rate will have an effect.

[26] That share is not transferable or bankable. This is the approach taken in the EU and California ETSs for limiting the use of offsets for compliance. With unilateral implementation, this restriction would have to be adopted by the importing (likely high cost) program.

multiplier associated with the constraint on Δ_2, the first-order conditions write

$$\begin{aligned} -C'_{mt}(e_{mt}) &= p_{mt} \quad \text{for } m \in \{H, L\} \text{ and } t \in \{1, 2\}, \\ p_{H1} &= p_{H2} + \lambda_H, \\ p_{L1} &= p_{L2} + \lambda_L + \beta\gamma, \\ \text{and } \; p_{H2} &= p_{L2} + \gamma. \end{aligned}$$

We see that α does not appear in the first-order conditions whereas β enters as an additive term (offset) in the intertemporal arbitrage condition in L, which is relevant only when $\gamma > 0$. By contrast, a banking conversion rate always alters the slope of the Hotelling path.

Assume both markets have a positive bank (i.e., $\lambda_H = \lambda_L = 0$). If the limit is not binding ($\gamma = 0$), then prices are equal in both periods by arbitrage. If the limit is binding ($\gamma > 0$), then only a share of the period-2 price wedge, i.e., $p_{H2} - p_{L2} = \gamma$, prevails in period 1, i.e., $p_{H1} - p_{L1} = \gamma(1 - \beta)$. The period-1 wedge is smaller than the period-2 wedge and it is decreasing with β—both wedges are equal when $\beta = 0$ but a positive β adds an offset term that partially narrows the wedge in period 1.[27] In words, the incentive for L to bank more in period 1 in order to sell more to H in period 2, which increases with β, leads to closer prices in period 1 than in period 2 when the link is effectively in place. This is the opposite outcome that policy makers seek to implement.

In addition, the following issues arise. First, as in a static setting, the price wedge induced by the binding limit create economic rents, the apportionment of which is not clear ex ante and may depend on markets' relative bargaining powers (Quemin & de Perthuis, 2019). Second, it is uncertain ex-ante whether or not the limit will be binding, i.e., whether prices will be equal in period 2 or not. That is, a quantitative limit does not provide a direct handle on the price wedge or ratio in period 2 as an exchange rate does. In addition, this uncertain price wedge backpropagates to period 1—in part or in full—through arbitrage (see Hintermann & Gronwald, 2019, for a similar effect with offsets). In other words, quantity limits can blur the information contained in prices, and the reduction in the informational efficiency of prices is exacerbated through banking and arbitrage. More generally, in the context of partitioned regulation, Abrell & Rausch (2017) show that price bounds are welfare superior to quantity bounds in triggering a shift in abatement effort from one system to the other in the presence of uncertainty about abatement technologies and baseline emissions.[28]

[27] Notice, the wedge cannot be nil in period 1 because the case $\gamma > 0$ is not feasible with $\beta = 1$.

[28] A loose parallel with the design of supply adjustment mechanisms is that using prices to condition supply

7 Conclusion

This paper has shown that a linking policy ensuring gradual price convergence across permit markets needs two instruments—one to establish the exchange rate at which permits trade in each period, another to align intertemporal arbitrage incentives with the targeted sequence of exchange rates. If exchange rates are used in isolation, anticipation of the future trading terms makes the current ones mute. That is, price ratios are governed by the exchange rate that will prevail in the future, making the phase-in of the link obsolete. The paper has also shown how banking conversion rates should be set to sustain the targeted sequence of price ratios in equilbrium, and that such a "two-legged" policy gives policy makers more flexibility to manage the distributional and environmental impacts of linking. In a calibrated numerical application to the EU ETS1 and ETS2, the paper has finally illustrated the different price paths (in each market) and emission levels under various gradual linking policies.

changes effectively implements a hybrid approach (Roberts & Spence, 1976; Burtraw et al., 2022), whereas using quantities to condition supply changes (as with the MSR in the EU ETS) effectively redoubles the reliance of policy on quantities and does not leverage the efficiency of prices, which tends to "destabilize" the market (Heijmans, 2023; Brinker et al., 2025).

References

Abrell, J. & Rausch, S. (2017). Combining Price and Quantity Controls under Partitioned Environmental Regulation. *Journal of Public Economics*, **145**, 226–242.

Abrell, J., Rausch, S. & Schwarz, G. A. (2018). How Robust is the Uniform Emissions Pricing Rule to Social Equity Concerns? *Journal of Environmental Economics and Management*, **92**, 783–814.

Aguiar, A., Chepeliev, M., Corong, E. L., McDougall, R. & van der Mensbrugghe, D. (2019). The GTAP Data Base: Version 10. *Journal of Global Economic Analysis*, **4**(1), 1–27.

Bataille, C., Guivarch, C., Hallegatte, S., Rogelj, J. & Waisman, H. (2018). Carbon Prices across Countries. *Nature Climate Change*, **8**, 648–650.

Baudry, M., Faure, A. & Quemin, S. (2021). Emissions Trading with Transaction Costs. *Journal of Environmental Economics and Management*, **108**, 102468.

Böhringer, C., Hoffmann, T. & Manrique de Lara Peñate, C. (2006). The Efficiency Costs of Separating Carbon Markets under the EU Emissions Trading Scheme: A Quantitative Assessment for Germany. *Energy Economics*, **28**(1), 44–61.

Brinker, T., Quemin, S. & Schoch, N. (2025). Emissions Trading with Supply Adjustment Mechanisms and Market Power. Working Paper.

Burtraw, D., Holt, C., Palmer, K. & Shobe, W. (2022). Price-Responsive Allowance Supply in Emissions Markets. *Journal of the Association of Environmental and Resource Economists*, **9**(5), 851–884.

Burtraw, D., Palmer, K., Munnings, C., Weber, P. & Woerman, M. (2013). Linking by Degrees: Incremental Alignment of Cap-and-Trade Markets. RFF Discussion Paper 13-04.

Chichilnisky, G. & Heal, G. (1994). Who Should Abate Carbon Emissions? An International Viewpoint. *Economics Letters*, **4**(4), 443–449.

Doda, B., Quemin, S. & Taschini, L. (2019). Linking Permit Markets Multilaterally. *Journal of Environmental Economics and Management*, **98**, 102259.

Dolphin, G., Pahle, M., Burtraw, D. & Kosch, M. (2023). A Net-Zero Target Compels a Backward Induction Approach to Climate Policy. *Nature Climate Change*, **13**, 1033–1041.

Dubois, L., Sahuc, J.-G. & Vermandel, G. (2025). A General Equilibrium Approach to Carbon Permit Banking. *Journal of Environmental Economics and Management*, **129**, 103076.

Edenhofer, O., Kosch, M., Pahle, M. & Zachmann, G. (2021). A Whole-Economy Carbon Price for Europe and How to Get There. Bruegel, Policy Contribution 06/2021.

EPA (2003). NOx Budget Trading Program, 2003 Progress and Compliance Report. US Environmental Protection Agency.

EPA (2023). EPA's "Good Neighbor" Plan Cuts Ozone Pollution – Overview Fact Sheet. US Environmental Protection Agency.

Fabra, N. & Reguant, M. (2024). The Energy Transition: A Balancing Act. *Resource and Energy Economics*, **76**, 101408.

Feng, H. & Zhao, J. (2006). Alternative Intertemporal Permit Trading Regimes with Stochastic Abatement Costs. *Resource and Energy Economics*, **28**, 24–40.

Flachsland, C., Marschinski, R. & Edenhofer, O. (2009). To Link or Not to Link: Benefits and Disadvantages of Linking Cap-and-Trade Systems. *Climate Policy*, **9**(4), 358–372.

Fowlie, M. L. & Muller, N. Z. (2019). Market-Based Emissions Regulation When Damages Vary across Sources: What Are the Gains from Differentiation? *Journal of the Association of Environmental and Resource Economists*, **6**(3), 593–632.

Graichen, J. & Ludig, S. (2024). Supply and Demand in the ETS 2. Report by Öko-Institut e.V. for the German Environmental Agency (Umweltbundesamt).

Günther, C., Pahle, M., Govorukha, K., Osorio, S. & Fotiou, T. (2024). Carbon Prices on the Rise? Shedding Light on the Emerging EU ETS2. SSRN Working Paper 4808605.

Heijmans, R. J. R. K. (2023). Adjustable Emissions Caps and the Price of Pollution. *Journal of Environmental Economics and Management*, **118**, 102793.

Helm, C. (2003). International Emissions Trading with Endogenous Allowance Choices. *Journal of Public Economics*, **87**(12), 2737–2747.

Hintermann, B. & Gronwald, M. (2019). Linking with Uncertainty: The Relationship Between EU ETS Pollution Permits and Kyoto Offsets. *Environmental and Resource Economics*, **74**, 761–784.

Holland, S. P. & Yates, A. J. (2015). Optimal Trading Ratios for Pollution Permit Markets. *Journal of Public Economics*, **125**, 16–27.

Holtsmark, B. & Weitzman, M. L. (2020). On the Effects of Linking Cap-and-Trade Systems for CO_2 Emissions. *Environmental and Resource Economics*, **75**(3), 615–630.

Holtsmark, K. & Midttømme, K. (2021). The Dynamics of Linking Permit Markets. *Journal of Public Economics*, **198**, 104406.

Hotelling, H. (1931). The Economics of Exhaustible Resources. *Journal of Political Economy*, **39**(2), 137–175.

Karp, L. & Traeger, C. (2024). Smart Cap. *Journal of the European Economic Association*, in press, jvae030.

Klepper, G. & Peterson, S. (2006). Marginal Abatement Cost Curves in General Equilibrium: The Influence of World Energy Prices. *Resource and Energy Economics*, **28**(1), 1–23.

Kling, C. L. & Rubin, J. D. (1997). Bankable Permits for the Control of Environmental Pollution. *Journal of Public Economics*, **64**(1), 101–115.

Leard, B. (2013). The Welfare Effects of Allowance Banking in Emissions Trading Programs. *Environmental and Resource Economics*, **55**, 175–197.

Leiby, P. N. & Rubin, J. D. (2001). Intertemporal Permit Trading for the Control of Greenhouse Gas Emissions. *Environmental and Resource Economics*, **19**, 229–256.

Long, X., Astier, N. & Zhang, D. (2025). Is Broader Trading Welfare Improving for Emission Trading Systems? *Journal of Environmental Economics and Management*, **130**, 103110.

Mehling, M. A., van Asselt, H., Dröge, S., Das, K. & Hall, C. (2024). Bridging the Divide: Assessing the Viability of International Cooperation on Border Carbon Adjustments. MIT-CEEPR Working Paper 2024-06.

Mehling, M. A., Metcalf, G. E. & Stavins, R. N. (2018). Linking Heterogeneous Climate Policies (Consistent with the Paris Agreement). *Environmental Law*, **48**(4), 647–698.

Mendelsohn, R. (1986). Regulating Heterogeneous Hmissions. *Journal of Environmental Economics and Management*, **13**(4), 301–312.

Metcalf, G. E. & Weisbach, D. (2012). Linking Policies When Tastes Differ: Global Climate Policy in a Heterogeneous World. *Review of Environmental Economics and Policy*, **6**(1), 110–129.

Muller, N. Z. & Mendelsohn, R. (2009). Efficient Pollution Regulation: Getting the Prices Right. *American Economic Review*, **99**(5), 1714–1739.

Newell, R. G., Pizer, W. A. & Zhang, J. (2005). Managing Permit Markets to Stabilize Prices. *Environmental and Resource Economics*, **31**, 133–157.

Pahle, M., Quemin, S., Osorio, S., Günther, C. & Pietzcker, R. (2025). The Emerging Endgame: The EU ETS on the Road Towards Climate Neutrality. *Resource and Energy Economics*, **81**, 101476.

Phaneuf, D. J. & Requate, T. (2002). Incentives for Investment in Advanced Pollution Abatement Technology in Emission Permit Markets with Banking. *Environmental and Resource Economics*, **22**, 369–390.

Pizer, W. A. & Prest, B. C. (2020). Prices versus Quantities with Policy Updating. *Journal of the Association of Environmental and Resource Economists*, **7**(3), 483–518.

Pizer, W. A. & Yates, A. J. (2015). Terminating Links Between Emission Trading Programs. *Journal of Environmental Economics and Management*, **71**, 142–169.

Pommeret, A. & Schubert, K. (2018). Intertemporal Emission Permits Trading under Uncertainty and Irreversibility. *Environmental and Resource Economics*, **71**, 73–97.

Quemin, S. & de Perthuis, C. (2019). Transitional Restricted Linkage Between Emissions Trading Schemes. *Environmental and Resource Economics*, **74**(1), 1–32.

Quemin, S. & Trotignon, R. (2021). Emissions Trading with Rolling Horizons. *Journal of Economic Dynamics and Control*, **125**, 104099.

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

Rubin, J. D. (1996). A Model of Intertemporal Emission Trading, Banking and Borrowing. *Journal of Environmental Economics and Management*, **31**(3), 269–86.

Salant, S. W. (2016). What Ails the European Union's Emissions Trading System? *Journal of Environmental Economics and Management*, **80**, 6–19.

Schennach, S. M. (2000). The Economics of Pollution Permit Banking in the Context of Title IV of the 1990 Clean Air Act Amendments. *Journal of Environmental Economics and Management*, **40**(3), 189–210.

Slechten, A. (2013). Intertemporal Links in Cap-and-Trade Schemes. *Journal of Environmental Economics and Management*, **66**(2), 319–336.

Stiglitz, J. E. (2019). Addressing Climate Change Through Price and Non-Price Interventions. *European Economic Review*, **119**, 594–612.

Toyama, Y. (2024). Dynamic Incentives and Permit Market Equilibrium in Cap-and-Trade Regulation. *American Economic Journal: Microeconomics*, **16**(3), 374–423.

Tuerk, A., Mehling, M., Flachsland, C. & Sterk, W. (2009). Linking Carbon Markets: Concepts, Case Studies and Pathways. *Climate Policy*, **9**(4), 341–357.

Vogt-Schilb, A., Meunier, G. & Hallegatte, S. (2018). When Starting with the Most Expensive Option Makes Sense: Optimal Timing, Cost and Sectoral Allocation of Abatement Investment. *Journal of Environmental Economics and Management*, **88**, 210–233.

Williams III, R. C. (2010). Setting the Initial Time-Profile of Climate Policy: The Economics of Environmental Policy Phase-Ins. NBER Working Paper 16120.

Woerman, M. (2023). Linking Carbon Markets with Different Initial Conditions. *Journal of Environmental Economics and Management*, **119**, 102820.

World Bank (2024). State and Trends of Carbon Pricing 2024. Washington, DC: World Bank.

Yates, A. J. & Cronshaw, M. B. (2001). Pollution Permit Markets with Intertemporal Trading and Asymmetric Information. *Journal of Environmental Economics and Management*, **42**, 104–118.

Appendix

A Proofs

A.1 Proof of Result 1

Fix $\rho_1 = \rho_2 = 1$ and $\delta_H = \delta_L = 1$ in (2). In this case, permits always flow from L to H, i.e., $\Delta_1 > 0$ and $\Delta_2 > 0$ (if not, simply relabel the markets). Without retirement and re-issuance of traded permits, inspection of the first-order conditions yields $\gamma_H = 0$ (and $\gamma_L = 0$ since $\Delta_1 > 0$). Suppose $\gamma_H > 0$. H imports L-permits to cover the entirety of its period-1 emissions ($\Delta_1 = e_{H1}$) and banks the entirety of its period-1 allocation ($B_H = \omega_{H1} > 0$). $B_H > 0$ implies $\lambda_H = 0$, hence $p_{H1} = p_{H2}$. This in turn implies that $p_{H2} = p_{L2} = p_{H1} = p_{L1} + \gamma_H > p_{L1}$. Yet $p_{L2} > p_{L1}$ entails $\lambda_L < 0$, a contradiction. Thus, it must be that $\gamma_H = 0$, and the first best equilibrium is characterized by

$$\begin{aligned} -C'_{rt}(e^{fl\star}_{mt}) = p^{fl\star}_{mt} \text{ for } m \in \{H, L\} \text{ and } t \in \{1, 2\}, \\ p^{fl\star}_{Ht} = p^{fl\star}_{Lt} \equiv p^{fl\star}_{t} \text{ for } t \in \{1, 2\}, \\ \text{and } \lambda^{fl\star}_{m} = p^{fl\star}_{m1} - p^{fl\star}_{m2} = p^{fl\star}_{1} - p^{fl\star}_{2} \equiv \lambda^{fl\star} \text{ for } m \in \{H, L\}. \end{aligned}$$

With retirement and re-issuance, the same result obtains as $\gamma_H = \gamma_L = 0$ always holds.

A.2 Proof of Results 3 and 4

Because $\rho_1 > \rho_2 = 1$, trade incentives are distorted between periods so that firms are better off when (1) as much trade from L to H as possible takes place in period 2; (2) when feasible, as much (reverse) trade as possible from H to L takes place in period 1. Doing so relaxes the overall constraint on emissions, despite the number of permits being fixed. This entails that banking levels in H and L are as low and high as possible, respectively.

Specifically, let the superscript n denote the naïve linking policy considered here. If $e^n_{H1} > \omega_{H1}$, permits flow from L to H in period 1 just as much as needed to cover the difference, and $B_H = 0$. Because ρ_1 L-permits are needed to cover for each "missing" H-permits, total emissions are reduced by $(\rho_1 - 1)(e^n_{H1} - \omega_{H1})$. If $e^n_{H1} < \omega_{H1}$, all "excess" H-permits flow to L in period 1 as long as L can bank the equivalent of $\rho_1(\omega_{H1} - e^n_{H1})$ additional L-permits, and $B_H = 0$. Total emissions are increased by $(\rho_1 - 1)(\omega_{H1} - e^n_{H1})$. With permit retirement and

re-issuance, either of the above cases holds. Without retirement and re-issuance, two other cases can occur when the upper constraint on B_L is binding, which we explore below.

Without retirement and re-issuance, inspection of (2) with $\rho_1 > \rho_2 = 1$ and $\delta_H = \delta_L = 1$ shows that $\gamma_H = 0$ (γ_H cannot be positive because $\delta_H\rho_2/\delta_L\rho_1 = 1/\rho_1 < 1$; see the proof of Result 6) and $\gamma_L \geq 0$. In this policy configuration, permits always flow from L to H in period 2, but in period 1 they may flow in both directions. If $e^n_{H1} < \omega_{H1}$, the first-order conditions admit the four distinct cases below. Which case occurs depends on the magnitude of ρ_1, the demand for banking, and the allocation of permits over time in both markets.

(*i*) $e^n_{H2} + e^n_{L2} > \omega_{H2} + \omega_{L2} + \omega_{L1}$. Full price convergence in period 1: $p_{H1} = p_{L1}$. Aggregate emissions in period 2 are sufficiently great that L banks the entirety of its period-1 permits—and buys some permits from H for compliance in period 1—and H also banks some of its period-1 permits ($\gamma_L > 0$ and $\lambda_L = \lambda_H = 0$). As banking occurs in both markets and prices are equal in period 2, arbitraged Hotelling price paths are identical and prices fully converge in period 1.

(*ii*) $e^n_{H2} + e^n_{L2} = \omega_{H2} + \omega_{L2} + \omega_{L1}$. Partial price convergence in period 1: $\rho_1 p_{L1} > p_{H1} > p_{L1}$. Aggregate emissions in period 2 require L to bank the entirety of its period-1 permits—and buy from H for compliance in period 1—but H does not bank any of its period-1 permits ($\gamma_L > 0$, $\lambda_L = 0$, and $\lambda_H > 0$). With H not banking, its prices do not follow a Hotelling path ($p_{H1} > p_{H2}$ instead) so early price convergence no longer holds. Yet the demand to bank period-1 permits remains sufficiently high to drive up the price of L's period-1 permits above what would be expected by the exchange rate ρ_1.

(*iii*) $\omega_{H2} + \omega_{L2} < e^n_{H2} + e^n_{L2} < \omega_{H2} + \omega_{L2} + \omega_{L1}$. No early price convergence: $p_{H1} = \rho_1 p_{L1}$. Aggregate emissions in period 2 are sufficiently low that L banks some—but not the entirety—of its period-1 permits, and H does not bank any of its period-1 permits ($\gamma_L = 0$, $\lambda_L = 0$, and $\lambda_H > 0$). With this lower demand for period-1 permits, the price ratio in period 1 is set by ρ_1. As L banks some period-1 permits, its prices remains on a Hotelling path ($p_{L1} = p_{L2}$). The absence of banking in H imply that its prices are decreasing across periods, with the rate of decrease fixed by ρ_1 ($p_{H1} = \rho_1 p_{H2} > p_{H2}$).

(*iv*) $e^n_{H2} + e^n_{L2} = \omega_{H2} + \omega_{L2}$. No early price convergence: $p_{H1} = \rho_1 p_{L1}$. Aggregate emissions in period 2 are so low that no banking occurs ($\gamma_L = 0$, $\lambda_L > 0$, and $\lambda_H > 0$). The price ratio is period 1 is set by ρ_1, and neither set of prices is on a Hotelling path.

For clarity, with permit retirement and re-issuance, only cases (iii) and (iv) can hold. Finally, if $e^n_{H1} > \omega_{H1}$, a similar case as (iii) occurs since $\gamma_H = \gamma_L = 0$ and $\lambda_H > \lambda_L = 0$. This holds with and without permit retirement and re-issuance.

A.3 Proof of Result 5

Notice, $p_{H2} = \rho_2 p_{L2}$ always holds. When $B_m > 0$ in each market m, $p_{m1} = \delta_m p_{m2}$, which concludes. When $B_m = 0$ in each market m, $p_{H1} = \rho_1 p_{L1}$ because $\gamma_H = \gamma_L = 0$, which concludes. When $B_m > 0$, $B_{\tilde{m}} = 0$ and $\gamma_m = 0$, $p_{H1} = \rho_1 p_{L1}$ because $\gamma_{\tilde{m}} = 0$, with $p_{m1} = \delta_m p_{m2}$ and $p_{\tilde{m}1} > \delta_{\tilde{m}} p_{\tilde{m}2}$. Without permit retirement and re-issuance, two other cases where $\gamma_m > 0$ must be addressed. When $B_L = \omega_{L1}$ and $\gamma_L > 0$, which can only occur when $\rho_1 > \rho_2$, noting that $p_{H1} = \rho_1(p_{L1} - \gamma_L) < \rho_1 p_{L1}$ concludes. When $B_H = \omega_{H1}$ and $\gamma_H > 0$, which can only occur when $\rho_1 < \rho_2$, noting that $p_{H1} - \gamma_H = \rho_1 p_{L1} < p_{H1}$ concludes.

A.4 Proof of Result 6

The proof only needs to address the two ambiguous corner cases in Result 5 without permit retirement and re-issuance. Note that the first-order conditions in (2) for any pairs (ρ_1, ρ_2) and (δ_H, δ_L) imply that γ_H and γ_L cannot be positive at the same time. If $\gamma_H > 0$, then $\Delta_1 = e_{H1} > 0$. Thus $e_{L1} > -\rho_1 \Delta_1$ always holds, so $\gamma_L = 0$. Symmetrically, $\gamma_L > 0 \Rightarrow \gamma_H = 0$. Intuitively, when permis flow from H to L, the constraint on the maximum permit transfer from L to H is de facto non-binding—and vice versa.

Assume $\gamma_H > 0$. This implies $\gamma_L = 0$ and $B_H = \omega_{H1} > 0$, so $\lambda_H = 0$. Thus, (2) reduces to

$$p_{H1} = \delta_H p_{H2} = \delta_H \rho_2 p_{L2} \leq \frac{\delta_H \rho_2}{\delta_L} p_{L1} \text{ and } p_{H1} > \rho_1 p_{L1},$$

which is feasible only if $\delta_H \rho_2 / \delta_L \rho_1 > 1$. By the same token, now assume $\gamma_L > 0$. This implies $\gamma_H = 0$ and $B_L = \omega_{L1} > 0$, so $\lambda_L = 0$. Thus, (2) reduces to

$$p_{L1} = \delta_L p_{L2} = \frac{\delta_L}{\rho_2} p_{H2} \leq \frac{\delta_L}{\delta_H \rho_2} p_{H1} \text{ and } p_{H1} < \rho_1 p_{L1},$$

which is feasible only if $\delta_L \rho_1 / \delta_H \rho_2 > 1$. Therefore, setting banking conversion rates such that $\delta_H / \delta_L = \rho_1 / \rho_2$ is the only policy configuration where both γ_H and γ_L can be nil, and where

price ratios are equal to current linking conversion rates in both periods. Formally,

$$\frac{\delta_H}{\delta_L} = \frac{\rho_1}{\rho_2} \Rightarrow \gamma_H = \gamma_L = 0 \Rightarrow \begin{cases} p_{Ht} = \rho_t p_{Lt} \text{ for } t \in \{1,2\}, \\ p_{m1} = \delta_m p_{m2} + \lambda_m \text{ for } m \in \{H,L\}. \end{cases}$$

In all other cases, including those with permit retirement and re-issuance, invoking Result 5 with $\delta_H/\delta_L = \rho_1/\rho_2$ concludes.

B Stylized analytical example

Assume quadratic, time-invariant abatement cost functions, i.e., C_{mt} with $C''_{mt}(\cdot) = c_m > 0$, and time-invariant baseline emissions, i.e., $\overline{e}_{mt} = \overline{e}_m$, for market $m \in \{H, L\}$ and time period $t \in \{1, 2\}$. To reduce the number of relevant Lagrange multipliers, assume all permits are allocated in period 1, i.e., $\omega_{m1} = \omega_m$ and $\omega_{m2} = 0$. This ensures that both markets always have an incentive to bank, i.e., $B_m > 0$ and $\lambda_m = 0$, and that only case (i) in Appendix A.2 is relevant. Optimality in market m under autarchy satisfies $c_m(\overline{e}_m - (\omega_m - B_m)) = c_m(\overline{e}_m - B_m)$, which gives

$$B^a_m = \omega_m/2 \text{ and } p^a_{m1} = p^a_{m2} = p^a_m = c_m(\overline{e}_m - \omega_m/2).$$

That is, half of the permits are banked from period 1 into period 2, and prices are equalized across periods. Now consider the case where markets H and L are linked in both periods. We solve for the competitive equilibrium in two steps. First, we derive the optimal permit transfers from H to L in periods 1 and 2, $\Delta^\star_1$ and $\Delta^\star_2$, for given banking decisions B_H and B_L. Second, we derive the optimal banking decisions, $B^\star_H$ and $B^\star_L$, given optimal transfers $\Delta^\star_1$ and $\Delta^\star_2$. Obviously, proceeding the other way around gives the same outcome.

Step 1. In each period, for given banking decisions B_H and B_L, linking equates marginal abatement costs between markets. For instance, in period 2, permit transfer from L to H satisfies $c_H(\overline{e}_H - (B_H + \Delta^\star_2)) = c_L(\overline{e}_L - (B_L - \Delta^\star_2))$, which defines $\Delta^\star_2$ as a function of B_H and B_L

$$\Delta^\star_2(B_H, B_L) = \frac{c_H(u_H - B_H) - c_L(u_L - B_L)}{c_H + c_L}.$$

By the same token,

$$\Delta^\star_1(B_H, B_L) = \frac{c_H(\overline{e}_H - \omega_H + B_H) - c_L(\overline{e}_L - \omega_L + B_L)}{c_H + c_L}.$$

Observe that the sum of periodic permit trades does not depend on banking decisions, i.e., on how permits are shifted over time. Specifically, total trade is proportional to two times the differential in prices under autarchy

$$\Delta_1^\star + \Delta_2^\star = 2\Delta^a \quad \text{with } \Delta^a = \frac{(p_H^a - p_L^a)}{c_H + c_L}.$$

Step 2. In each market, given allowance transfers $\Delta_1^\star$ and $\Delta_2^\star$, banking equates marginal abatement costs across periods. For instance, intertemporal arbitrage in market H requires that $c_H(\overline{e}_H - (\omega_H + \Delta_1^\star(B_H, B_L) - B_H)) = c_H(\overline{e}_H - (B_H + \Delta_2^\star(B_H, B_L)))$, which gives

$$B_H^\star = \frac{\omega_H + \Delta_1^\star(B_H^\star, B_L^\star) - \Delta_2^\star(B_H^\star, B_L^\star)}{2}.$$

By the same token,

$$B_L^\star = \frac{\omega_L - \Delta_1^\star(B_H^\star, B_L^\star) + \Delta_2^\star(B_H^\star, B_L^\star)}{2}.$$

Observe that the sum of banked permits does not depend on permit trades, i.e., on how permits are shifted between regions. Specifically, total banking amounts to half of the total emission permits

$$B_H^\star + B_L^\star = \frac{\omega_H + \omega_L}{2}.$$

Multiplicity of solutions. There are multiple solutions to the above optimality conditions due to "redundancy" in flexibility levers with linking in both periods and banking across periods. Mathematically, there is a continuum of individual values for $(\Delta_1^\star, \Delta_2^\star)$ and $(B_H^\star, B_L^\star)$ that satisfy the optimal values of $\Delta_1^\star + \Delta_2^\star$ and $B_H^\star + B_L^\star$ in aggregate, respectively. For instance

(1) Banking as under autarchy, $B_H^\star = B_H^a$ and $B_L^\star = B_L^a$, with equal partition of total trade across periods, $\Delta_1^\star = \Delta_2^\star = \Delta^a$. This shows that the first best can be attained with linking in both periods but without banking when allocation is split evenly across periods in both markets, i.e., $\omega_{m1} = \omega_{m2} = \omega_m/2$.

(2) No linking in period 2 with all trade occurring in period 1, $\Delta_1^\star = 2\Delta^a$ and $\Delta_2^\star = 0$, with corresponding adjustment through banking, $B_H^\star = B_H^a - \Delta^a$ and $B_L^\star = B_L^a + \Delta^a$. This shows that the first best can be attained without linking in period 2 as long as banking provides sufficient room to shift all trades to period 1, i.e., provided that $\omega_H > 2\Delta^a$.

(3) No linking in period 1 with all trade occurring in period 2, $\Delta_1^\star = 0$ and $\Delta_2^\star = 2\Delta^a$, with corresponding adjustment through banking, $B_H^\star = B_H^a + \Delta^a$ and $B_L^\star = B_L^a - \Delta^a$. This

shows that the first best can be attained without linking in period 1 as long as banking provides sufficient room to shift all trades to period 2, i.e., provided that $\omega_L > 2\Delta^a$.

In effect, any convex combination of solutions (2) and (3) is also a solution. As discussed above, whether or not these solutions to the optimality conditions can implement the fist best depends on the feasibility of the associated banking and emission paths and, hence, ultimately on the allocation profile over time. They also require common knowledge of whether and when linking is in place so that banking can adjust accordingly.

Solution (2) forms the basis of the analysis of delinking in Pizer & Yates (2015), who describe a variety of delinking policies that implement the fist best. Solution (3) echoes the analysis in Section 3.3 showing that the mere anticipation of 1:1 linking in period 2 leads to 1:1 price convergence in period 1 through intertemporal arbitrage—even without linking being effective—provided there is positive, unconstrained banking in both markets. Here, feasibility is ensured by assumption as the allocation profile is such that there is always an incentive to bank in both markets (i.e., case (i) in Appendix A.2 applies). As our analysis in Section 3.5 shows, controlling price convergence over time requires the combined use of linking and banking conversion rates to allow some control over Hotelling price paths in each market.