Resource and Energy Economics 68 (2022) 101300

Contents lists available at ScienceDirect

Resource and Energy Economics

journal homepage: www.elsevier.com/locate/ree

Raising climate ambition in emissions trading systems: The case of the EU ETS and the 2021 review

Simon Quemin[a,b,c,d,*]

[a] Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Germany
[b] Grantham Research Institute, London School of Economics and Political Science, United Kingdom
[c] Climate Economics Chair, Paris-Dauphine University, PSL Research University, France
[d] Électricité de France R&D, EDF Lab Paris-Saclay, SYSTEME Department, France

ARTICLE INFO

Article history:
Received 2 May 2021
Received in revised form 23 January 2022
Accepted 21 March 2022
Available online xxxx

JEL classification:
D47
D81
G41
H32
Q58

Keywords:
Emissions trading
Rolling horizon
Market stability reserve
EU ETS
Climate ambition

ABSTRACT

We provide a quantitative assessment of policy options to inform the 2021 review of the EU Emissions Trading System (ETS) and raise climate ambition. We use a permit trading model in which firms utilize rolling finite planning horizons, which replicates historical price and banking developments well compared to an infinite horizon. When firms have bounded foresight, indirectly raising ambition through the Market Stability Reserve (MSR) is not equivalent to directly raising ambition through the emissions cap trajectory. Leveraging the MSR turns out to be efficiency improving as it compensates for firms' bounded foresight by frontloading abatement efforts. We analyze the MSR interaction with the cap trajectory to exploit synergies and minimize the cost of raising ambition. We also provide a comparative assessment of a complete suite of changes in the MSR parameters. Whatever its parameters, MSR-induced resilience to demand shocks remains limited by design: the MSR acts more as an unconditional price support provider than as a responsive price stabilizer.

© 2022 Published by Elsevier B.V.
CC_BY_4.0

1. Introduction

Strengthening climate change mitigation targets is a critical policy objective worldwide. In the EU, for instance, climate ambition needs to be ramped up to align with the commitments made under the Paris Agreement and the objectives of the Green Deal (e.g. carbon neutrality by 2050). In this context, attributing a stricter emission reduction target to the perimeter covered under the EU emissions trading system (ETS) looks appealing and the ongoing ETS review offers an adequate implementation window. As a case in point, Parry (2020) finds that relying on more robust pricing in the ETS is the policy option that yields the largest welfare gains in raising ambition and is a Pareto improvement for Member States. Yet, Parry's model does not capture the intricacies and specificities of the EU ETS, and specific modeling needs to be carried out to appraise possible levers and explore how best to raise ambition.

---

[*] Correspondence to: Potsdam-Institut für Klimafolgenforschung e.V., Telegrafenberg A 31, 14473 Potsdam, Germany.
E-mail address: Simon.Quemin@pik-potsdam.de.

https://doi.org/10.1016/j.reseneeco.2022.101300
0928-7655/© 2022 Published by Elsevier B.V.
CC_BY_4.0

In any ETS, the most natural (i.e. direct) way to raise ambition is to increase the stringency of the limit (or cap) on emissions or the rate at which it decreases over time. In the EU ETS, annual emissions caps are set far in advance and gradually adjusted downwards at a rate called the linear reduction factor (LRF). Regulated firms receive for free or buy at auctions a corresponding overall number of emissions rights or permits, which they must annually surrender for compliance against their emissions. Firms are allowed to cost optimize their compliance strategies through trading or timing their permit use across years. Since permits do not expire, firms can bank (i.e. stockpile) unused permits for compliance in future years (e.g. Rubin, 1996; Schennach, 2000). To illustrate, banking occurs when permits are relatively more abundant now than they will be in the future – be it due to past oversupply or as a precautionary measure if firms anticipate future permit scarcity relative to their needs.

In the EU ETS, there is also an indirect way to raise ambition through the so-called market stability reserve (MSR) and its add-on cancellation mechanism (CM). The MSR is a permit supply control instrument unique of its kind which started operating in 2019 and adjusts supply based on market-wide banking: the more firms bank permits, the more it reduces supply (e.g. Perino and Willner, 2016).[1] Specifically, when banking is above a given threshold, a given share of banked permits is withdrawn from auctions and placed in the reserve. Conversely, when banking is below another given threshold, a given volume of permits is released from the reserve and added to auctions. On top of this, the CM cancels any permit stored in the reserve in excess of a given volume. This entails that the cumulative emissions cap can be reduced, i.e. the MSR and CM have the appealing potential to raise both short- and long-run ambition levels (e.g. Perino, 2018; Quemin and Trotignon, 2021). In fact, the 2018 price rally is in large part attributable to the introduction of the MSR and CM, which increased and brought permit scarcity back to the market earlier than what was originally anticipated, if at all, by market actors. As we will indeed see, the direct and indirect ways of raising ambition are not equivalent as they ultimately depend on how firms respond to them.

In this paper, we evaluate and compare the impacts of realistic regulatory changes to inform the ongoing ETS review and raise ambition.[2] To do so, we use the competitive intertemporal emissions permit trading model developed in Quemin and Trotignon (2021), henceforth QT21. The model features the MSR and CM and incorporates uncertainty about the future permit demand. A key novelty is that firms can utilize rolling finite horizons as part of their decision-making process, e.g. as a way of dealing with uncertainty (e.g. Goldman, 1968; Spiro, 2014; Grüne et al., 2015; van Veldhuizen and Sonnemans, 2018).[3] Rolling horizons can: (1) reconcile annual banking developments over 2008–17 with implicit discount rates inferred from futures' yield curves where a standard infinite horizon can only do so with discount rates above implicit rates, (2) reproduce average annual prices over 2008–17 with a twice better fit than an infinite horizon, and (3) pick up the 2018 price rally in the wake of the 2018 market reform where an infinite horizon falls short of it. We interpret these results through the lens of Friedman's (1953) black box approach. That is not to say that firms actually utilize rolling horizons but this modeling assumption has the comparative advantage of reproducing historical market outcomes more satisfactorily than an infinite horizon model does. In turn, if firms de facto – or behave as if they – focus more on the short to mid term than on the long term, this has important ramifications for policy design and outcomes (e.g. Fuss et al., 2018).

As a first contribution, we leverage our calibrated model to evaluate various realistic options in revising the main MSR parameters (viz. the intake rate, the intake and release thresholds)[4] to provide some guidance for the ongoing review. We highlight three key results. First, because the MSR is a trigger mechanism, when the intake and release thresholds are constant over time as in the status quo, raising the intake rate creates additional volatility due to stronger oscillations around the thresholds without increasing cancellations and ambition. Second, the position of the intake threshold matters more than that of the release threshold in terms of market outcomes, a lower intake threshold sustaining higher prices and ambition. Third, as a potent avenue for the review, combining MSR thresholds that are declining over time with a higher intake rate can keep induced oscillations in check, thereby leading to higher prices and ambition without unduly destabilizing the market.

As a second contribution, we study the least-cost ways to raise ambition through two levers: a direct increase in the LRF (i.e. the rate at which the emissions cap decreases over time) and indirectly leveraging the MSR through changes in its parameters. Importantly, these levers are not equivalent in terms of increased stringency over time (as perceived by firms) and they also interact (as complements or substitutes) in non-straightforward ways. Indeed, the MSR has the potential to make long-term scarcity embedded in the cap trajectory more tangible in the short to mid term by frontloading abatement efforts (transitional stringency is higher); and coupled with the CM, it further increases long-term ambition (cumulative stringency is higher). By contrast, higher scarcity induced by an LRF increase is more prevalent in the long term than in the short term. These properties are visually depicted in Fig. 1.

Our analysis characterizes how, in a context where firms are – or behave as if – boundedly farsighted, transitional stringency is as important as cumulative stringency for policy design and impacts. Specifically, we analyze the synergies between an LRF increase and changes in MSR parameters to minimize the regulatory cost of an ambition ramp-up. Our analysis of the LRF-MSR

---

[1] By contrast, supply-side control instruments typically introduce price steps in otherwise inelastic supply curves (e.g. Roberts and Spence, 1976; Weitzman, 1978) in order to constrain price variability and regulatory costs in the face of ex-ante uncertainty in demand for emissions permits and abatement costs (e.g. Fell et al., 2012; Borenstein et al., 2019; Burtraw et al., 2020).

[2] On 14 July 2021, the European Commission published its proposal for a revised ETS Directive (Marcu and Cabras, 2021). See (European Commission 2015; European Commission 2021) for details, in particular the impact assessment that uses a modeling approach based on the model developed herein.

[3] See Section 2 in QT21 for more details on theoretical aspects and micro-foundations for rolling horizons as well as for anecdotal evidence that this assumption is relevant in the context of the EU ETS.

[4] Specifically, the MSR adjusts yearly auction volumes a_t based on banking in the previous year b_{t-1}: a_t is reduced by $b_{t-1} \times \textit{intake rate}$ if $b_{t-1} > \textit{intake threshold}$; else a_t is increased by a fixed release quantity if $b_{t-1} < \textit{release threshold}$; else a_t is unchanged. See Section 2.1 for details.

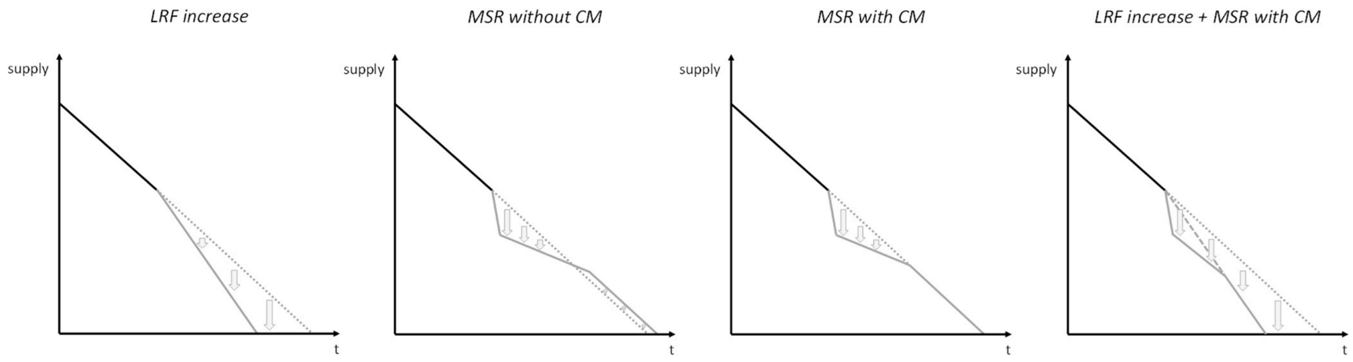


Fig. 1. Stylized representation of LRF vs. MSR impacts on supply over time.

interaction shows that declining thresholds are always less costly than constant thresholds for a given ambition target, especially when coupled with an increase in the intake rate for ambitious targets. In general, leaving more traction to an enhanced MSR than to an LRF increase turns out to be efficiency-improving as the MSR partially compensates for bounded foresight by frontloading abatement efforts.

As a third contribution, we assess the ability of the MSR to improve the market resilience to demand shocks. We find that whatever the changes in its parameters, MSR-induced resilience remains limited and one-sided by design – the MSR is geared towards supply contraction and can limitedly respond to shocks in the short and longer terms alike. In essence, the MSR acts more as an unconditional price support mechanism than as a price stabilizer. Contrary to what its name suggests, the MSR (and CM) should therefore rather be thought of as a potent indirect means of raising ambition rather than as a more standard supply-side instrument à la Roberts and Spence (1976) and Weitzman (1978).

Since the early contributions by Fell (2016) and Perino and Willner (2016), the literature has recently witnessed a flurry of papers on the MSR and related aspects, see inter alia Perino and Willner (2017), Chaton et al. (2018), Bocklet et al. (2019), Carlén et al. (2019), Kollenberg and Taschini (2019), Mauer et al. (2019), Beck and Kruse-Andersen (2020), Bruninx et al. (2020), Osorio et al. (2020), Bruninx and Ovaere (2021), Gerlagh et al. (2021), Quemin and Trotignon (2021) and Tietjen et al. (2021) – see Perino et al. (2021) for a recent review and discussion. A finding that emerges from scanning this burgeoning literature is that market outcomes and MSR impacts are sensitive to model parametrization and calibration as well as underlying assumptions about the achievement of complementary policies.

Within this literature, the first key differentiator of our paper is that we provide a thorough quantitative assessment of realistic regulatory changes to inform the 2021 review. Other works typically evaluate the impacts of the 2018 reform, keeping the MSR parameters unchanged. To the best of our knowledge, Osorio et al. (2020) is the only paper that provides a similar analysis, which is conducted based on detailed electricity sector and industry model using an infinite horizon. This brings us to the second key differentiator of our paper: though we use a more stylized framework, our model is finely calibrated to reproduce historical price and banking developments and thereby capture the resultant of observed firms' behavior, notably through rolling planning horizons. These two works thus complement each other well.

The remainder is structured as follows. Section 2 sets forth our model of emissions trading with rolling horizons and its calibration to historical data. Section 3 leverages our calibrated model to inform the 2021 review with a quantitative assessment of realistic changes in relevant MSR parameters, explore the nature of the interaction between the LRF and MSR, and assess the extent of MSR-induced resilience to demand shocks. Section 4 concludes. An Appendix provides complementary simulation results and details on the model calibration.

2. Model

In this section, we first describe our emissions trading model in the presence of the MSR and CM, and then its calibration to past annual market price and banking outcomes. We borrow the modeling framework developed in QT21 where firms utilize rolling finite horizons as part of their planning and optimization processes. This section is as parsimonious as can be and only provides key building blocks and insights – the reader is referred to QT21 for details.

2.1. Description

2.1.1. Economic environment

We consider an emissions trading system in discrete time where compliance is required in each year t. Annual caps on system-wide emissions consist of freely allocated and auctioned permits, f_t and a_t, and o_t denotes the total amount of offset credits surrendered in year t. As is standard, we assume that regulated firms acquit their compliance obligations in full by remitting enough permits or offsets to cover their yearly emissions.

Permits are tradable across firms (spatial flexibility) and years (temporal flexibility) but there are some restrictions on the temporal dimension. While banking (i.e. storing past or current vintage permits for future compliance) is fully authorized, borrowing (i.e. frontloading future vintage permits for current compliance) is upper bounded. Specifically, borrowing is tacitly allowed on a year-on-year basis as free allocation in year $t+1$ typically takes place two to four months before year-t compliance is due, and no permit vintage restriction applies (European Parliament and Council, 2003). Letting b_t denote the total volume of banked permits in year t, the following constraint must hold for all t: $b_t + f_{t+1} \geq 0$.

Future demand for permits is uncertain in nature as firms' counterfactual baseline emissions are affected by external factors (i.e. independently of the permit price) such as business cycle fluctuations (e.g. Chèze et al., 2020) and the variable performances of complementary climate and energy policies (e.g. Borenstein et al., 2019). Uncertainty prevails on the supply side as well since future cap trajectories $\{f_t\}_t$ and $\{a_t\}_t$ are subject to regulatory changes and $\{o_t\}_t$ depends on external offset market conditions (e.g. Ellerman et al., 2016).

2.1.2. Firms' behavior

As firms cost minimize over time, limited intertemporal trading opportunities imply a no-arbitrage condition whereby two conditions must hold with complementary slackness

$$b_t + f_{t+1} \geq 0 \quad \perp \quad p_t - \beta \mathbb{E}_t\{p_{t+1}\} \geq 0, \tag{1}$$

where $\mathbb{E}_t\{\cdot\}$ denotes expectation conditional on all information available to the firms in year t, $\beta = (1+r)^{-1}$ is the discount factor with r the discount rate, and p_t the market price in year t. Condition (1) specifies a quasi Hotelling's rule whereby the expected price can rise at a rate at most as high as the discount rate in the competitive equilibrium (Hotelling, 1931). The price cannot rise at a rate greater than r, for otherwise firms would buy and bank more permits to sell them later on – until they break even and the market price coincides with the cost-of-carry price (e.g. Ellerman and Montero, 2007). The converse may not hold due to limited borrowing, i.e. arbitrage cannot prevent the price from increasing at a rate below r when the borrowing constraint is binding, or expected to be binding (e.g. Schennach, 2000).

Since our aim is to analyze the temporal dimension of the system in the presence of supply-side control, we abstract from its spatial trading component and take the perspective of the regulated perimeter as a whole. We use a representative firm approach which is well-documented and widely employed in the literature (e.g. Fell et al., 2012; Kollenberg and Taschini, 2019) since the decentralized competitive market equilibrium can be characterized indirectly as the solution to joint cost minimization among all firms (e.g. Montgomery, 1972; Rubin, 1996).

We let e_t and u_t denote the representative firm's levels of realized and baseline emissions in year t, respectively. Abatement $u_t - e_t \geq 0$ is costly and we let C_t denote its abatement cost function in year t, with C' and $C'' > 0$. In year t, u_t, f_t, a_t and o_t[5] as well as the state variable b_{t-1} are given to the firm. The firm selects its emission e_t and implied bank b_t by minimizing its expected net present value of compliance costs

$$\min_{\{e_\tau\}_{\tau \geq t}} \mathbb{E}_t\Big\{\sum_{\tau \geq t} \beta^{\tau - t} C_\tau(u_\tau - e_\tau)\Big\} \tag{2a}$$

$$\text{subject to } 0 \leq e_\tau \leq u_\tau, \tag{2b}$$

$$\text{and } b_\tau = b_{\tau-1} + f_\tau + a_\tau + o_\tau - e_\tau \geq -f_{\tau+1}, \tag{2c}$$

where (2b) contains feasibility constraints for the emission path and (2c) describes the law of motion for the state variable, i.e. annual market clearing. In (2c), the constraint on the bank ensures that cumulative supply equals cumulative emissions, i.e. overall market clearing.

2.1.3. Rolling horizon

In a context of uncertainty and limited information, firms typically plan and optimize over rolling finite horizons (e.g. Goldman, 1968; Sahin et al., 2013; Spiro, 2014; Grüne et al., 2015; van Veldhuizen and Sonnemans, 2018). Specifically, with some finite horizon $h \geq 0$, the firm selects year-t emission e_t and bank b_t by solving

$$\min_{\{e_\tau\}_{\tau=t}^{t+h}} \sum_{\tau=t}^{t+h} \beta^{\tau-t} C_\tau(\hat{u}_\tau^t - e_\tau) \tag{3a}$$

$$\text{subject to } 0 \leq e_\tau \leq \hat{u}_\tau^t, \quad \text{and } b_\tau = b_{\tau-1} + \hat{f}_\tau^t + \hat{a}_\tau^t + \hat{o}_\tau^t - e_\tau \geq -\hat{f}_{\tau+1}^t. \tag{3b}$$

where $\hat{x}_\tau^t$ denotes the year-t forecast for $x = \{u, f, a, o\}$ in year $\tau \geq t$. With a rolling horizon, the firm solves for the equilibrium path from year t to $t+h$ given its current forecasts $\{\hat{x}_\tau^t\}_\tau$, but only implements the first year of the plan, which pins down the state variable for next year. In year $t+1$, the firm revises its forecasts and initiates a new planning cycle from year $t+1$ to $t+h+1$, taking the state variable b_t as given (see Section 2.2 for a description of forecast heuristics and their updates). This year-on-year solving-and-updating of finite plans with sequential execution of first-year-only decisions unfolds over time.

As h grows, one would like to have the solution paths generated by a rolling horizon converge to those of the infinite horizon. As shown in QT21, two assumptions are needed to arrive at this convergence property. First, we derive the expected equilibrium paths under the infinite horizon invoking a first-order approximation along certainty-equivalent paths for the exogenous variables x. This approach, suggested by Schennach (2000), reduces the dimensionality of the infinite horizon problem to that of a rolling horizon. Second, we impose that certainty equivalent paths coincide with the corresponding forecasts, i.e. $\hat{x}_\tau^t = \mathbb{E}_t\{x_\tau\}$. In this case, the solution paths generated by a rolling finite horizon and a given discount rate, or by the infinite horizon and a higher discount rate, are observationally equivalent. Crucially, however, this equivalence breaks down in the presence of supply-side control.

2.1.4. Supply-side control

The market stability reserve (MSR) is a banking corridor consisting of a reserve of permits, with stock s_t in year t, and a set of parameters: an intake rate IR_t, a release quantity RQ_t, and intake-release thresholds IT_t and RT_t. It is a rule-based mechanism adjusting yearly auctions a_t based on a past bank index $B_t = \frac{1}{3}b_{t-1} + \frac{2}{3}b_{t-2}$ whereby[6]

---

[5] Offset usage is assumed exogenous to the firm's problem. In Section 2.2, we explain how we tackle offset usage for ex-post model calibration and why this assumption is innocuous for our analysis in Section 3.

[6] Because of a mismatch between the compliance and auction calendars (i.e. the official figure for b_{t-1} can only be used from September of year t onward) total MSR operations over year t are de facto based on B_t.

- if $B_t > IT_t$: a_t is reduced by $\min\{IR_t \times B_t;\, a_t\}$ and $s_{t+1} = s_t + \min\{IR_t \times B_t;\, a_t\}$, or
- if $B_t < RT_t$: a_t is increased by $\min\{RQ_t;\, s_t\}$ and $s_{t+1} = \max\{s_t - RQ_t;\, 0\}$, or
- else: the MSR is inactive and no change occurs.

The MSR endogenizes the auction schedule $\{a_t\}_t$ (i.e. rearranges annual auctions over time) based on the number of banked permits, which reflects both past market outcomes (a legacy) and anticipated future permit need (through intertemporal compliance cost minimization). In principle, it leaves cumulative supply as set by the cap trajectory unchanged: all permits withdrawn from the market should ultimately be released (Perino and Willner, 2016).

Additionally, the MSR is equipped with a bolt-on cancellation mechanism (CM), which may further adjust the MSR stock as follows

- if $s_t > L_t$: s_t is reduced by $s_t - L_t$, or
- else: the CM is inactive and no change occurs.

The CM shaves off all reserve permits in excess of a predefined upper bound from the MSR stock (here, L_t in year t) and permanently invalidates them. As a result, cumulative emissions allowed under the system are now endogenously determined and the amount by which they will be reduced has become a market outcome, hence ex ante uncertain (Perino, 2018).

Given our indirect planning approach, we solve for the intertemporal competitive equilibrium as a fixed point of a mapping between the firm's beliefs about the MSR-driven supply impact stream and the equilibrium stream in the spirit of rational expectations equilibria (Lucas and Prescott, 1971). At each step in this recursive procedure, the firm holds beliefs about future control impacts and supply schedules, optimizes based on its beliefs, and updates them based on the resulting control impact stream. The equilibrium is obtained when the firm's beliefs coincide with the actual law of motion for control impacts generated by the optimal choices induced by these beliefs. In equilibrium, the firm fully understands the interplay between its decisions and associated control impacts over time, and has no incentive to deviate.[7]

2.1.5. Current parametrization

In the status quo, i.e. pursuant to the 2018 reform (European Parliament and Council, 2018), the MSR is parametrized as follows

- $IT_t = 833$ MtCO$_2$ for all $t \geq 2019$,
- $RT_t = 400$ MtCO$_2$ for all $t \geq 2019$,
- $IR_t = 24\%$ for $t \in [2019;\, 2023]$ and 12% for all $t \geq 2024$,
- $RQ_t = 100$ MtCO$_2$ for all $t \geq 2019$.

In addition, the MSR is exogenously seeded with 1.55 GtCO$_2$: 0.9 billion ad-hoc backloaded permits (European Commission, 2014) plus an estimated 0.65 billion unallocated Phase-III permits.[8] As per CM rules, the reserve is capped by realized auctions in the previous year and 'excess permits' are cancelled from 2023 onward, that is

- $L_t = \infty$ for $t \in [2019;\, 2022]$ and a_{t-1} for all $t \geq 2023$.

Notice the degree of intricacy of the supply-side control. The CM is controlled by a_{t-1}, which is determined by: (1) current MSR actions and (2) anticipated future MSR and CM actions through intertemporal banking decisions. Given current MSR and CM parameters and initial system conditions (i.e. b_{2017} and b_{2018} in the order of 1.6 GtCO$_2$) the MSR is bound to begin by absorbing an endogenously-determined amount of permits before possibly releasing them back in later years, and the CM is set to cancel an endogenously-determined share of reserve permits. As shown in QT21 and Appendix A, these amounts depend on firm's behavior (e.g. sophistication level, planning horizon). In Section 3, we further quantify how they hinge on MSR design and parameters.

2.2. Calibration

2.2.1. Demand and supply

Permit demand is driven by unconstrained emissions. To construct counterfactual baseline CO$_2$ emissions for the EU ETS perimeter (i.e. emissions as they would be absent the scheme, but accounting for industrial production growth and complementary energy and climate policies), we decompose them into three Kaya indexes: production, energy intensity and carbon

---

[7] It is important to notice that, although firms could possibly be better off collectively if the market-wide bank was below the intake threshold – this would mitigate MSR-induced supply contraction – in a competitive equilibrium they cannot coordinate individual banking decisions to 'game' the MSR, i.e. the equalization of marginal abatement costs across firms and time periods must hold. See Section 3.5 in QT21 for details.

[8] These permits are gradually seeded into the MSR between 2019 and 2021, but the timeline is irrelevant for our results. Our 0.65 billion estimate falls within the expected range (European Commission 2015).

intensity. To compute these indexes ex post, we consider that the permit price has had negligible impacts on production, energy efficiency and renewables deployment and we use various databases from Eurostat and the IEA.[9] To compute these indexes ex ante, we assume that: (1) annual production growth is 1%, (2) the current 2030 energy efficiency and renewables targets are attained linearly, and (3) the implied trends continue to be valid afterward. Fig. 2a depicts the reconstructed and projected paths for these three indexes between 1990 and 2050, and Fig. 2b shows that resulting baseline emissions path – in line with the historical trend, it is downward sloping and reaches zero in 2096.

Regarding demand forecasts, we assume that the firm uses a simple heuristic congruent with the deterministic part of an AR (1) process. It is slightly tweaked to accommodate for growth and varying trend. That is, the year-t forecast for baseline emissions in year $t+1$ is defined by

$$\hat{u}_{t+1}^{t} = \varphi(1+\gamma_t)u_t + (1-\varphi)\bar{u}_{t+1}^{t}, \tag{4}$$

where u_t is the realized baseline in year t, $\varphi \in [0;1]$ the persistence parameter and γ_t the annual growth rate as expected in year t. The trend $\bar{u}$ can vary over time, viz. the trend in year t for some future year $t' > t$ ($\bar{u}_{t'}^{t}$) is in line with the achievement of the renewable energy and energy efficiency targets as set out in Climate Energy Package that prevails in year t. We set $\varphi = 0.9$ as in Fell (2016) and γ_t aligns with GDP growth forecasts by the European Commission over 2008–2020 with a 1% p.a. growth rate afterward. It is important to notice that the firm makes imperfect demand forecasts, i.e. realized baselines u_t (as shown in Fig. 2b) typically differ from their forecasts $\hat{u}_t^{\tau<t}$ (defined as per (4)). The firm hence adjusts its demand forecasts and associated intertemporal decisions on a yearly basis.

Parametrizing supply is an easier task. In year t, the firm observes permit supply $f_t + a_t$ and forecasts future supply to coincide with the cap path as given in prevailing regulatory texts (e.g. EU Directives or Decisions). As soon as regulation is amended or upon release of actual supply data, the firm corrects its forecasts. For instance, the firm considers a post-2020 cap path based on the effective linear reduction factor (LRF), viz. 1.74% before and 2.2% after the 2018 reform (from 2021 on, 57% of the cap is auctioned off). The grey line in Fig. 2b depicts total annual supply $\{f_t + a_t + o_t\}_t$.[10] The peak in 2011–12 is due to a massive usage of offsets, totalling about 1 GtCO_2. The subsequent dip is due to the backloading of 0.9 GtCO_2 over 2014–16 and to non-issued Phase-III permits, totalling about 0.6 GtCO_2 over 2013–17.

2.2.2. In-sample calibration

We restrain our calibration sample to 2008–2017 for two reasons. First, we leave aside the trial Phase I (2005–2007) since banking and borrowing across Phases I and II was not allowed, de facto restricting the firm's horizon with regard to Phase-I permits usage. Second, we aim at exploiting the regulatory regime shift induced by the 2018 market reform. Our calibration methodology is described in details in Appendix A.

In sample, we utilize a two-step procedure in the spirit of standard least squares maximum likelihood estimations with one free parameter. In a fist step, we select the couple (r, h) to minimize annual deviations between the simulated and observed bank levels over 2008–2017. In a second step, given the selected (r, h), we fit the abatement cost function by minimizing the distance between the simulated and yearly-averaged prices over 2008–2017. Specifically, assuming an infinite horizon $h = \infty$, a discount rate $r \approx 8\%$ is found to replicate past banking best. While this aligns with standard rates of return on risky assets (e.g. Jordà et al., 2019), it lies in the higher range of the rates inferred from futures' yield curves (see Appendix A). Alternatively, assuming a central value for such rates, i.e. $r = 3\%$, a rolling finite horizon of $h = 11$ years is found to yield a similar banking fit. That is, the rolling horizon can reconcile past bank dynamics with implicit discount rates whereas the infinite horizon falls short of it. Moreover, the price fit obtained with the rolling horizon in the second step of the procedure is more than twice as good (both in size and sign) than with the infinite horizon.[11]

2.2.3. Out-of-sample calibration

We now further compare how in-sample calibrated infinite vs. rolling horizons fare out of sample, i.e. their ability to pick up the observed price quadrupling in 2018 (from 7 to 25 € per tCO_2). As illustrated in Appendix A, the rolling horizon captures the new pricing regime better than the infinite horizon. Specifically, although a price increase occurs in 2018 with both horizons in response to the reform, it is four times bigger with the rolling horizon and closer to actual price developments.[12] As we

---

[9] For production, our assumption is tenable as there is no evidence of carbon leakage (Naegele and Zaklan, 2019). For energy efficiency, it is reasonable as the index declines less over 2005–15 with the ETS in place than before without it (Fig. 2a). For carbon intensity, it is supported by evidence that permit-price equivalents of renewable subsidies were significantly higher than market prices (Marcantonini and Ellerman, 2015; Abrell et al., 2019). The databases we use for Kaya index computation are referenced in QT21. This assumption will become less tenable when prices reach higher levels than those that prevailed in the past – but note that baselines are also often exogenously given (i.e. price-unresponsive) in related modeling approaches.

[10] Kyoto offsets are authorized over 2008–20 up to an overall limit $O \approx 1.6$ GtCO_2. Offset usage is assumed to be exogenous to the firm: in year t, o_t is given and equal to realized usage, and the firm forecasts that the remainder (i.e. $O - \sum_{\tau=2008}^{\tau=t} o_\tau$) is equally split across the remaining years. From 2021 on, supply coincides with the announced cap which is set to decline at a 2.2% LRF, implying that it becomes nil in 2058.

[11] With both the rolling and infinite horizons, the initial one-off measure to backload 900 million permits does not affect firms' decisions (and thus has no impact on prices) as it simply redistributes annual auction volumes (and thus does affect the overall auction volume) over the firms' planning horizon.

[12] Notice that our calibrated rolling horizon model does not explain the 2018 price rally in its entirety and the better price fit remains qualitative. Indeed, the post-2018 pricing regime reflects a variety of other factors that are beyond the scope of our stylized model, including increased speculative market activity on the part of financial entities (e.g. Friedrich et al., 2020; Quemin and Pahle, 2021), transaction costs (e.g. Baudry et al., 2021) or other bullish factors such as the ratcheting up of climate ambition targets.

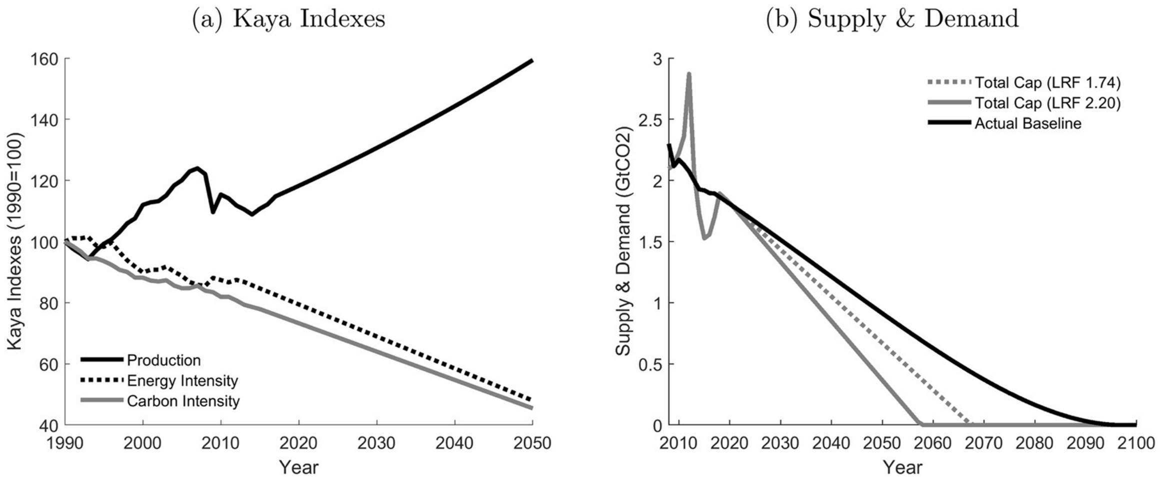


Fig. 2. Kaya indexes, baseline emissions and total cap on emissions. *Note:* The amounts by which the cap declines yearly correspond to the LRF multiplied by the 2010 emissions of the covered perimeter in Phase III: 38.3 and 48.4 million under a LRF of 1.74% and 2.2%, respectively.

elaborate further in Section 3, this is because with the rolling horizon yearly MSR-driven supply cuts have a larger relative impact on the firm's perceived overall abatement effort, and more of it is abated early on given the lower discount rate. These two effects concur to yield higher price (and banking) levels than with the infinite horizon.

We finally interpret our calibration results in the spirit of Friedman (1953) black box model approach. That is, we compare the merits of two alternative assumptions for firms' behavior (infinite vs. rolling horizons) in their ability to reproduce past market outcomes. Our results lend more empirical support to the latter – both in and out of sample. That is not to say that all firms actually use rolling finite horizons, nor a fortiori the same horizon and discount rate. Rather, this representation has the comparative advantage of replicating past market outcomes more satisfactorily than the infinite horizon and, in the first order, may be thought of as capturing the resultant of individual firms' behaviors.

3. Simulations

In this section, we utilize our calibrated model to inform and feed into policy debates on three interrelated issues: (1) a quantitative assessment of realistic changes in the MSR parameters to provide some guidance on regulatory amendments for the 2021 market review, (2) an evaluation of possible ways of ramping up ambition through the MSR by exploring its interaction with the cap trajectory (LRF), and (3) an appraisal of the MSR-induced resilience to future demand shocks. Based on the comparative merits of a rolling finite horizon both in and out of sample (see Section 2.2), we focus on this case to limit the number of scenarios.[13]

3.1. Assessing the market impacts of the 2018 reform (status quo)

We begin by describing market outcomes in the status quo, i.e. after the market reform passed in 2018, which comprised an LRF increase from 1.74% to 2.2% from 2021 on, the reinforcement of the MSR and the introduction of the CM. Fig. 3 depicts the equilibrium price and bank paths with the calibrated rolling horizon.[14] Relative to no reform ('LRF 1.74 no reform'), the sole LRF increase ('LRF 2.2 only') leads to higher prices and a shorter banking period, albeit with higher banking volumes initially. Intuitively, this is because less permits are minted and put into circulation. Adding the MSR and CM on top of it ('LRF 2.2 + MSR CM') further hikes the price and reduces the bank. We expound on the underlying mechanism below. The rest of the exposition leaves aside the second half of the century because permits are quickly exhausted (yearly supply dries out in 2058 and the permit bank empties a few years later) and both the MSR and CM have become inactive (see Appendix B).[15] That is, this period is essentially inconsequential for the purposes of our analysis.

Transitional stringency is as important as cumulative stringency.

---

[13] See QT21 for a quantitative assessment of market outcomes under infinite vs. rolling finite horizons.

[14] Prices are in current Euros, using observed inflation rates over 2008–2018 and 1.5% p.a. afterward.

[15] In this period with no permits left in circulation, there are no permits left to trade (i.e. the market has terminated de facto) or to surrender for compliance against emissions. Hence, in order to meet compliance, the firm has no choice but to emit zero, i.e. abate its entire baseline emissions. The price thus reflects the firm's marginal cost of abating its annual baseline emissions – and because baselines are by construction set to decline over time and reach zero by the end of the century, so does the price.

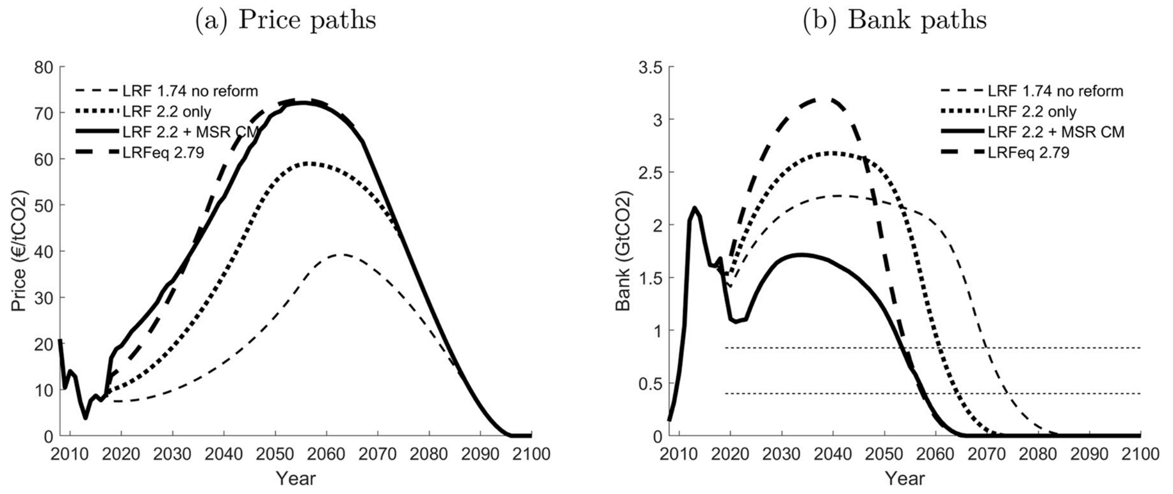


Fig. 3. Market impacts of the 2018 reform (status quo). *Note:* Release and intake thresholds are respectively set at 0.400 and 0.833 $GtCO_2$ and remain constant over time. Intake rate of 24% over 2019–2023 and 12% afterward. LRF_{eq} is the LRF which, on its own, generates the same 2008–2100 cumulative emissions as those generated by a 2.2% LRF with the MSR+CM.

The MSR+CM has the potential to increase both the transitional and cumulative stringency of the ETS' emissions targets. These two properties are visually depicted in Fig. 1. The former property occurs as the MSR temporarily withdraws some permits from circulation by postponing some auctions, i.e. it de facto frontloads abatement efforts and makes long-term scarcity more tangible earlier on. This is especially relevant if market participants tend to focus more on the short to mid term than on the long term – in fact, in this case, this property leads to gains in cost-effectiveness (see QT21).[16] The latter property occurs through the CM, i.e. cumulative stringency increases as the bulk of MSR-withdrawn permits are cancelled and never re-enter the market. In comparison, the impact of an LRF increase is more salient in the long term than in the short term, and as such might not be proportionally reflected in boundedly farsighted agents' early abatement decisions.

Specifically, with the MSR in place and banking needs initially above the intake threshold, the firm foresees a supply tightening over its horizon due to MSR intakes (and reinjections may also be too far off into the future to be relevant for the firm's current planning). This drives up current abatement and banking, which in turn inflates future MSR intakes. This further raises the firm's overall abatement forecast, which leads to higher banking and future MSR intakes, and so forth.[17] As a result, annual MSR intakes last over three decades, which translates into total cancellations of 8.7 $GtCO_2$ (see Appendix B).[18]

Finally, to further get a sense of why transitional stringency is key when firms are boundedly farsighted, we illustrate these properties with two examples.

Example 1. We simulate what the 2018 price response would be if the 2018 market reform only comprised the LRF increase from 1.74% to 2.2% (no MSR). While the induced reduction in cumulative supply is substantial (9 $GtCO_2$), the price response is less than commensurate with the rolling horizon as the impact on transitional scarcity is less marked. Specifically, in our calibrated model, this represents only 13% of the price rise witnessed in conjunction with the MSR and CM. Note that this is despite the fact that the CM's impacts on cumulative stringency is of similar size (8.7 $GtCO_2$). Rather, this is because the MSR brings into present times the long-term scarcity implied by the emissions cap trajectory.[19]

Example 2. We simulate the market outcomes obtained under a sole LRF increase yielding the same 2008–2100 cumulative emissions as under a 2.2% LRF with MSR and CM. With the rolling horizon, this equivalent LRF, or LRF_{eq}, is of 2.79%. As Fig. 3a shows, prices are initially lower with the LRF_{eq} before catching up and surpassing prices with the 2.2% LRF and MSR+CM, as firms gradually factor in the cap's actual long-term stringency and realize they had underestimated it. Specifically, the 2018 price jump is twice as small with the LRF_{eq} as with the 2.2% LRF and MSR+CM.[20]

---

[16] In addition to gains in cost-effectiveness, environmental benefits also result from earlier abatement (e.g. Leard, 2013). We ignore this welfare dividend in our analysis to focus on cost-effectiveness aspects exclusively (see notably Section 3.3), but note that both channels are independent and additive.

[17] This self-reinforcing effect gradually subsides and stops because $IR < 1$. Relatedly, recall that although firms would collectively be better off stockpiling less permits to reduce MSR intakes, they cannot collude to coordinate individual banking decisions in a competitive equilibrium (see footnote [7])

[18] By contrast, with an infinite horizon model, MSR-induced backloading has less of an impact of the firm's planning (MSR intakes are spread over a longer horizon, their impacts on current decisions are also lower due to the higher discount rate, and possible reinjections are accounted for). As a result, MSR intakes are lower (due to lower banking) and end sooner (the intake cut-off year occurs two decades earlier), which translates into smaller cancellations of 4.2 $GtCO_2$ (see Appendix A).

[19] By contrast, with an infinite horizon model, the LRF increase captures 80% of the simulated 2018 price rise obtained in conjunction with the MSR and CM (see Appendix A).

[20] By contrast, with an infinite horizon model: (1) the LRF_{eq} is lower (2.38%) because total cancellations are lower (4.2 $GtCO_2$); and (2) price paths are almost identical because the MSR-driven auction backloading is essentially irrelevant in this case (see Appendix A).

3.2. Assessing changes in the MSR parameters for the 2021 review

We take the MSR framework as given and vary each of its main parameters in isolation (the intake rate *IR*, the intake and release thresholds *IT* and *RT*) relative to their values in the status quo (Section 2.1) in order to single out their respective impacts on market outcomes. While the review package may comprise changes in a combination of MSR parameters, we do not quantify their interactions to limit the number of scenarios. Yet, the way parameters interact with one another will be clear from the analysis of individual changes.

We also take the CM framework as given and its trigger parameter *L* as currently set, though we wish to underline that it needs to be enshrined in law as part of the review.[21] We assume that agreement on the 2021 review package takes time. Specifically, following the 2015–2018 reform timeline, we consider that regulatory amendments are voted in – and thus anticipated from – 2023 for implementation in 2024 and maintained unchanged thereafter. In Appendix D, we also evaluate the impacts of a free-allocation phase-out in the context of a transition to border carbon adjustment mechanisms as mentioned in the EU Green Deal.

With constant thresholds, a higher intake rate raises volatility but not ambition.

As a threshold-based trigger mechanism, the MSR is subject to discontinuities.[22] Therefore, raising the intake rate increases the magnitude and frequency of the induced threshold effects, which materialize as banking oscillations around the intake threshold. These are the resultant of a conflict between an MSR-driven downward dragging force[23] and an upward restoring force as long as firms have an incentive to accumulate a bank. This conflict and induced banking oscillations are more salient the higher the intake rate (Fig. 4b), in turn leading to more erratic streams of yearly auctions and MSR intakes (see Appendix B).

As oscillations in the bank and annual supply are transmitted to prices, a higher intake rate is conducive to larger price variability with a negligible price increase on average (Fig. 4a). This contrasts with the second objective of the MSR to improve the resilience of the market and ultimately of the price signal. The negligible increases in average price levels result from slightly larger cancellations (i.e. lower cumulative supply), ranging from 8.7 to 9.2 $GtCO_2$ with an intake rate of 12% and 48% respectively. Importantly, even though intake rates vary by a factor of 4, cancellations are almost identical. This is because in cumulative terms, large but irregular yearly intakes generated by a high intake rate roughly tally with smaller but steadier yearly intakes generated by a lower intake rate (see Appendix B).

With the current thresholds (400–833 $MtCO_2$ constant over time), a higher intake rate than in the status quo (12% from 2024 on) does not increase ambition and tends to destabilize the market by making annual supply bumpier and prices more volatile. In practice, as a result of more pronounced threshold effects, a higher intake rate makes future supply conditions harder to gauge for market actors. One may argue that this could make the system at large susceptible to manipulation and speculative gaming not related to fundamentals, especially as the permit bank nears the MSR thresholds (e.g. Pahle and Quemin, 2020). This may further amplify price volatility and be detrimental for market functioning.

Given an intake rate, a lower intake threshold yields higher prices and ambition.

Given an intake rate, the lower the intake threshold, the longer the intake period and thus the larger the cumulative intakes and cancellations. Price paths are hence ordered by decreasing intake threshold height, with an average wedge of 6€/tCO_2 between intake thresholds set at 433 and 1233 $MtCO_2$ (Fig. 5a), which is reflected in and driven by cancellations of 9.3 and 6.9 $GtCO_2$, respectively. Banking paths are ordered similarly, i.e. banking levels are higher when thresholds are lower.[24] What may seem counterintuitive on the face of it results from an anticipation effect. That is, with a lower intake threshold, as forward-looking firms foresee a longer intake period and thus a larger overall supply cutback, they stockpile more permits, which itself leads to larger yearly and cumulative intakes (see Section 3.1).

Lowering the intake threshold allows an increase in price and ambition levels without inducing volatility but there are decreasing returns. For instance, a lowering from 1233 to 1033 $MtCO_2$ raises cancellations by 1 $GtCO_2$ compared to 0.4 (0.1) $GtCO_2$ for a lowering from 833 to 633 (633−433) $MtCO_2$. Decreasing returns occur because: (1) the anticipation effect driving higher banking levels and MSR intakes saturates for low intake thresholds and (2) lowering the intake threshold prolongs the intake period at the end of the banking period when the bank is relatively low and decreases sharply.

Together with the intake rate, the position of the intake threshold is a pivotal policy handle. In comparison, the position of the release threshold has a negligible bearing on market and ambition outcomes.[25] This is because reinjections may only occur when the release threshold is passed (in the 50′s at the earliest) which is irrelevant for market outcomes since: (1) this is beyond the firms' planning horizon until the 40′s and (2) the MSR has been depleted by the CM and is already empty when it could release permits, so no reinjections take place.

With declining thresholds, a higher intake rate raises ambition but not volatility.

---

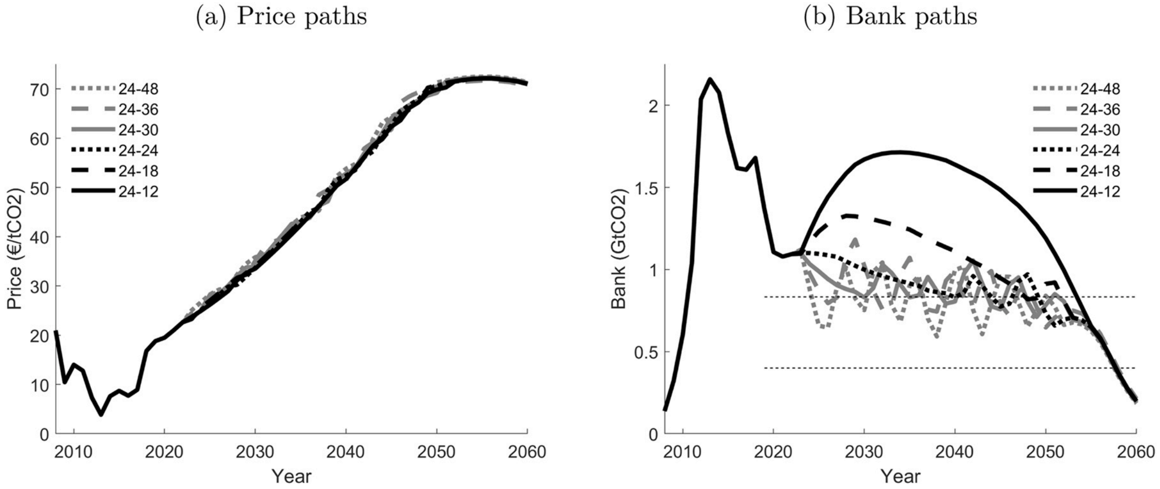
[21] See QT21 for a quantitative assessment of the MSR impacts with and without the CM.

[22] That is, annual supply is highly sensitive to when the MSR is active or inactive. For instance, when the bank in year t is 834 $MtCO_2$ auctions in year $t+1$ are curtailed by 100 up to 400 $MtCO_2$ with an intake rate of 12% and 48% respectively, while they are unaltered when the bank in year t is 832 $MtCO_2$.

[23] More precisely, as the MSR takes in permits and cuts back on yearly auctions, it forces firms to tap into their private bank of permits to compensate for reduced contemporaneous supply.

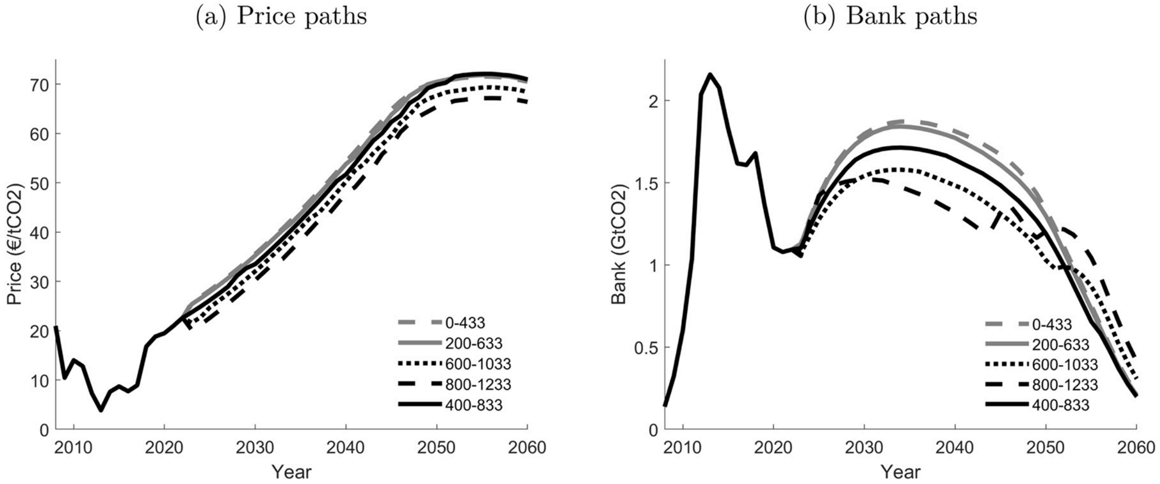
[24] The ordering is altered by threshold effects in the 40′s and 50′s and does not hold with 800–1233 $MtCO_2$ thresholds until 2029 since the MSR is initially inactive and does not eat away at the bank in this case.

[25] We obtain similar results by varying the breadth between thresholds, which we do not report for brevity.



(a) Price paths (b) Bank paths

Fig. 4. Different intake rates from 2024 on with thresholds constant over time. *Note:* Release and intake thresholds are respectively set at 0.400 and 0.833 $GtCO_2$ and remain constant over time. Intake rate of 24% over 2019–2023 and 12, 18, 24, 30, 36 or 48% afterward.



(a) Price paths (b) Bank paths

Fig. 5. Different threshold positions from 2024 on with given intake rates. *Note:* Release and intake thresholds are respectively set at 0.400 and 0.833 $GtCO_2$ over 2019–2023; afterward, their position varies keeping a constant breadth of 0.433 $GtCO_2$ (not depicted to reduce visual clutter). Intake rate of 24% over 2019–2023 and 12% afterward.

One may argue that intake-release thresholds should be aligned with evolving banking needs. Because banking will eventually decrease in the course of time as a result of firms' optimizing behavior under a decreasing cap path (see e.g. Appendix F), this implies that the thresholds should be gradually adjusted downward. One may conceive of various ways of implementing declining thresholds but one practical regulatory approach could be to align them with the LRF. We follow this approach and assume that both thresholds are decreasing linearly over time to become nil in the same year as the cap (in 2058 with a 2.2% LRF).[26]

Relative to constant thresholds, one may intuit that when thresholds are declining over time: (1) the intake period is longer, which leads to larger total intakes and thus higher ambition and price levels and (2) threshold effects are less frequent, which mitigates induced oscillatory behavior. This is readily apparent in comparing Figs. 4 and 6: for a given intake rate with declining thresholds, the price is higher (because cancellations are larger, ranging from +0.5 to +2 $GtCO_2$ with a 12% and 48% intake rate, respectively) and more stable (because intake and auction streams are more regular, see Appendix B). Oscillations may still

---

[26] A similar argument is that thresholds should be adjusted to reflect market size. Since the LRF dictates how the annual supply volume evolves over time, it could readily be used to adjust thresholds.

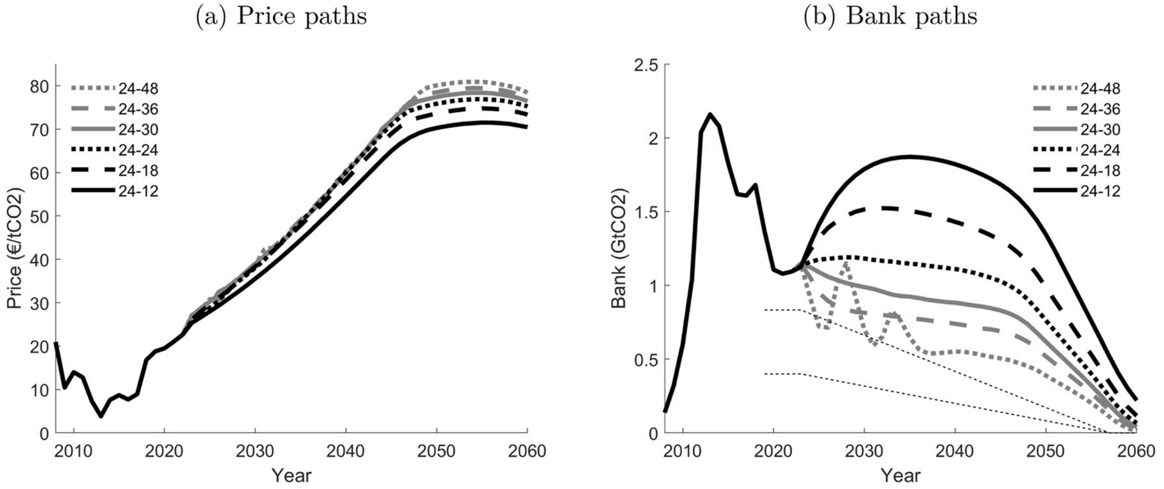


Fig. 6. Linearly declining thresholds with different intake rates from 2024 on. *Note:* Release and intake thresholds are respectively set at 0.400 and 0.833 $GtCO_2$ over 2019–2023; afterward, they linearly decline to reach zero in the same year as the cap (2058). Intake rate of 24% over 2019–2023 and 12, 18, 24, 30, 36 or 48% afterward.

materialize if the intake rate is large enough, which happens only with a 48% intake rate over 2025–2035. Otherwise, price and banking levels are monotonically increasing with the intake rate.

Combining an increase in the intake rate with declining thresholds is a promising option for the review as this allows raising ambition without inducing more volatility. The increase in the intake rate need not be significant because of decreasing returns: an increase from 12% to 18% already reaps the bulk of the higher ambition potential (1.3 out of the 2 $GtCO_2$ additional cancellations obtained with a 48% intake rate). This can readily be seen by comparing price paths in Fig. 6a: those with a 18% intake rate or more are similar until the late 40′s and grouped together above that with a 12% intake rate (see also Fig. 7).

3.3. Raising ambition through the LRF and the MSR

There are two ways of increasing climate ambition within the ETS perimeter: increasing the LRF and leveraging the MSR and CM. Because they are not equivalent in terms of perceived supply impacts and associated abatement decisions when firms have bounded foresight, synergies can be exploited by utilizing these two options hand in hand to ensure complementarity and minimize regulatory costs. Indeed, the LRF and MSR interact. For instance, changing the LRF changes banking incentives and thus MSR intakes, and ultimately both transitional and cumulative stringency. This underlines the need to understand the

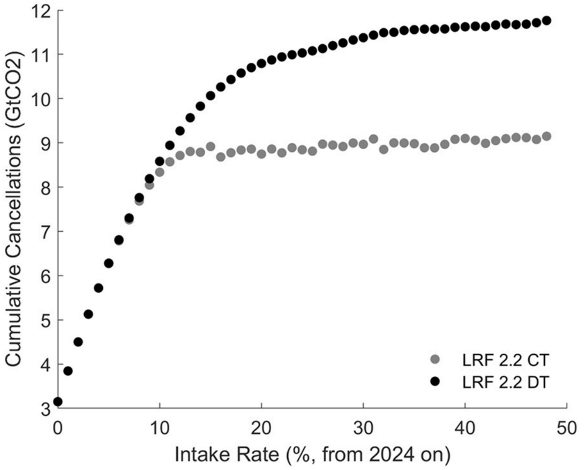


Fig. 7. LRF-MSR interaction as a function of the intake rate. *Note:* Intake rate of 24% over 2019–2023 and varying afterward along the x-axis. LRF is fixed at 2.2%. CT (DT) indicates that release and intake thresholds set at 0.400 and 0.833 $GtCO_2$ are constant over time (over 2019–2023; then linearly declining to reach 0 in 2058, the year the cap becomes nil given the 2.2% LRF).

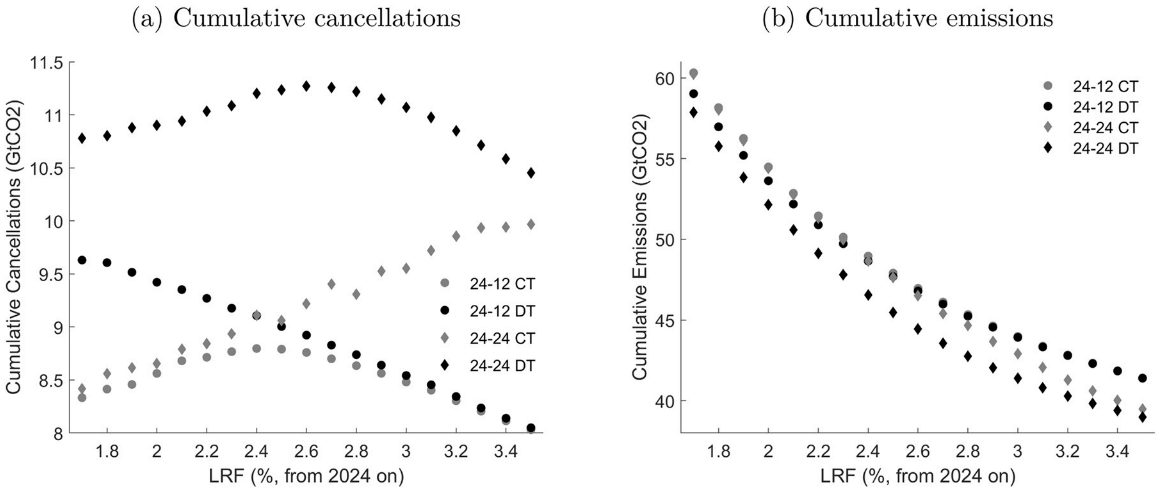


Fig. 8. LRF-MSR interaction as a function of the LRF. *Note:* Intake rate of 24% over 2019–2023 and 12 or 24% afterward. Post-2023 LRF is varied along the x-axis. CT (DT) indicates that release and intake thresholds set at 0.400 and 0.833 $GtCO_2$ are constant over time (over 2019–2023; then linearly declining to reach 0 in the same year as the cap given the prevailing LRF).

nature of the LRF-MSR interaction, which we explore below. In Appendix E, we also provide a more concrete case study of various LRF-MSR settings in the context of a 2030 ambition ramp-up.

The LRF-MSR interaction: complements or substitutes in raising ambition?

We first analyze how total cancellations vary as a function of the intake rate with constant or declining thresholds with a fixed 2.2% LRF. Because the LRF is fixed, cumulative emissions vary in symmetrical quantities, i.e. one more tCO_2 cancelled implies one less tCO_2 emitted. With constant thresholds, cancellations sharply increase with the intake rate below 10% but a saturation effect occurs at larger rates (Fig. 7). The current post-2023 intake rate (12%) is located at the kink before the saturation plateau: raising it would increase volatility but not cancellations (see Section 3.2) which is visible here as the plateau wobbles.

By contrast, the saturation effect is less marked with declining thresholds and occurs only at higher intake rates. This is conducive to larger cancellations for a given intake rate, although there are still decreasing returns to increasing the intake rate. For instance, raising the post-2023 intake rate from 12% to 24% leads to + 1.8 $GtCO_2$ cancellations with declining thresholds compared to + 0.1 $GtCO_2$ with constant thresholds. Additionally, oscillatory behavior and volatility are mitigated (see Section 3.2 and Appendix B).

Not only do MSR impacts depend on its parameter values, but also on the LRF. Hence, we now analyze how total cancellations vary as function of the LRF for given intake rates and types of thresholds. As the LRF varies, a purely symmetrical relationship between cumulative cancellations and emissions no longer holds. Specifically, as the LRF rises, cumulative supply and thus emissions are reduced (*direct effect*). On top of that, changing the LRF also affects banking strategies and thus MSR intakes, cancellations and cumulative emissions (*indirect effect*). The symmetrical relationship breaks down due to the indirect effect, hence Fig. 8 displays both total cancellations and emissions. Fig. 8a confirms what we already know: given an LRF, cancellations are always larger the higher the intake rate with given thresholds, or with declining thresholds for given intake rates, or both.

We now want to know under which conditions increasing the LRF raises or reduces cancellations ceteris paribus, i.e. the extent to which an LRF increase and MSR-driven cancellations are complements or substitutes in curbing cumulative emissions (the indirect effect). All else contant, the higher the LRF the shorter the banking period, but the higher the banking levels early on when the bank is accumulating. With the MSR in place, this implies that MSR intakes are larger early on (*short-term effect*) as banking levels are higher, but smaller later on (*long-term effect*) as the intake period is shorter. Depending on which effect dominates, increasing the LRF has an ambiguous impact on cancellations, i.e. it can either reinforce or undermine the MSR ability to raise ambition. Intuitively, the larger the intake rate, the larger annual MSR intakes early on and the shorter the intake period. One may hence expect that the short-term effect is more likely to dominate with a higher intake rate.

In Fig. 8a, the LRF and the MSR are complements (substitutes) when curves are upward (downward) sloping. Even though no general results emerge, they can be explained through the lens of the conflict between the short-term (positive) and long-term (negative) effects of an LRF increase on intakes. With constant thresholds, the short-term effect always dominates for the higher intake rate (24%) while this only holds for small LRFs with the lower intake rate (12%). With declining thresholds, the long-term effect always dominates for the lower intake rate while this only holds for large LRFs with the higher intake rate. This is because of an exact (a less than) one-to-one mapping between the reduction in the banking period and that in the intake period with declining (constant) thresholds due to a higher LRF.

Table 1
Relative costs of various LRF-MSR combinations for given ambition targets.

		Cumulative emissions ($GtCO_2$)						
Thresholds	IR	44	45	46	47	48	49	Ref
Constant	24–12	0.47% (2.998)	0.43% (2.847)	0.33% (2.714)	0.29% (2.596)	0.28% (2.491)	0.18% (2.396)	51.5
	24–24	0.48% (2.866)	0.31% (2.752)	0.45% (2.646)	0.23% (2.553)	0.23% (2.473)	0.39% (2.371)	51.3
Declining	24–12	0.31% (2.986)	0.28% (2.833)	0.26% (2.699)	0.14% (2.577)	$\star$ (2.466)	$\star$ (2.366)	50.9
	24–24	$\star$ (2.647)	$\star$ (2.539)	$\star$ (2.450)	$\star$ (2.364)	0.08% (2.285)	0.11% (2.209)	49.1

Note: Intake rate of 24% over 2019–2023 and 12 or 24% afterward. Release and intake thresholds are fixed at 0.400 and 0.833 $GtCO_2$, either: permanently constant, or constant over 2019–2023 and then linearly declining to reach 0 in the same year as the cap given the LRF. Numbers within parentheses give the LRF required to attain the various ambition targets (specified in cumulative emissions, one target per column) given different MSR parametrizations. The $\star$ indicates the lowest-cost combination to attain a given target, and percentages measure the additional net present value of compliance costs for the other possible LRF-MSR combinations. 'Ref' indicates cumulative emissions with a 2.2% LRF and the given MSR parametrizations.

Finally, Fig. 8b shows the implications of the LRF-MSR interaction for cumulative emissions. As a result of the direct effect of increasing the LRF, cumulative emissions are reduced, but at a decreasing rate. Indeed, cumulative emissions are given by the area below the supply curve as shown in Fig. 2b, so as the LRF becomes larger, the lower the amount by which the integral is reduced – hence the decreasing convex trend. On top of that, the indirect effect generates second-order deviations around the trend (30% in relative magnitude at most) due to the LRF-MSR interaction and varying cancellations. This notwithstanding, the LRF-MSR interaction is an important aspect of the policy which need be assessed as part of the review, especially as it has potential to lower the costs of an ambition ramp-up.

Exploiting synergies between the LRF and the MSR to raise ambition.

Multiple LRF-MSR combinations can achieve the same ambition target. With a sole efficiency criterion in mind, one seeks the setting which minimizes the net present value of compliance costs. On the face of it, one might argue that this is the one which leaves most of the traction to the LRF, as the MSR can be thought of as distorting firms' intertemporal decision making, hence increasing regulatory costs. This reasoning would hold with fully farsighted firms. Yet, when firms have bounded foresight, the MSR coupled with the CM has potential to improve upon efficiency relative to a sole LRF generating identical cumulative emissions (see QT21). In essence, by frontloading future scarcity, the MSR can partially compensate for inefficiently low abatements early on that otherwise result from bounded foresight. In turn, leaving more traction to the MSR than to the LRF can lessen the costs of an ambition ramp-up.

Table 1 considers four possible LRF-MSR parametrizations to attain given ambition targets. The latter are specified in cumulative 2008–2100 emissions so as to meaningfully compare the net present value of compliance costs across combinations for a given target. Three results emerge. First, declining thresholds are always less costly than constant thresholds, whatever the intake rate. Second, the more ambitious the target, the larger the cost differences relative to the least expensive combination.[27] Third, except for low ambition targets (48–49 $GtCO_2$), the least expensive combination leverages the LRF increase the least and involves declining thresholds with a 24% intake rate. This shows that with boundedly farsighted firms, utilizing the MSR to raise ambition can be efficiency-improving relative to a higher LRF increase with a less enhanced MSR. At the same time, some LRF increase can also be desirable: for the lower 48–49 $GtCO_2$ targets, the least expensive combination leverages more the LRF and the MSR's ambition-raising potential is not at its maximum.

A key policy takeaway is that declining thresholds are always less costly than constant thresholds for a given target, especially when coupled with an increase in the intake rate from 12% to 24% for the more ambitious targets. In looser terms, we note that in the region of interest for the review – say for an LRF between 2.2% and 2.8% – cancellations under a 24% intake rate with declining thresholds are at their maximum (11 $GtCO_2$) while those in the other three studied cases are in a similar range of 9 $GtCO_2$ (Fig. 8a).

3.4. Assessing the MSR-induced resilience to demand shocks

The first objective of the MSR (and CM) is to eliminate historical oversupply with respect to lower than expected demand conditions. As we have seen (e.g. in Section 3.1), this objective is met. However, future baseline emissions strongly depend on exogenous factors such as the economic conjuncture (e.g. Chèze et al., 2020 and complementary policies like renewables and energy efficiency targets or nuclear and coal phase-outs (e.g. Borenstein et al., 2019. To avoid history repeating itself, the second objective of the MSR is to improve market resilience and price buoyancy, should similar supply-demand imbalances materialize in the future.

---

[27] This pattern is less clear for constant thresholds with a 24% intake rate due to induced oscillations.

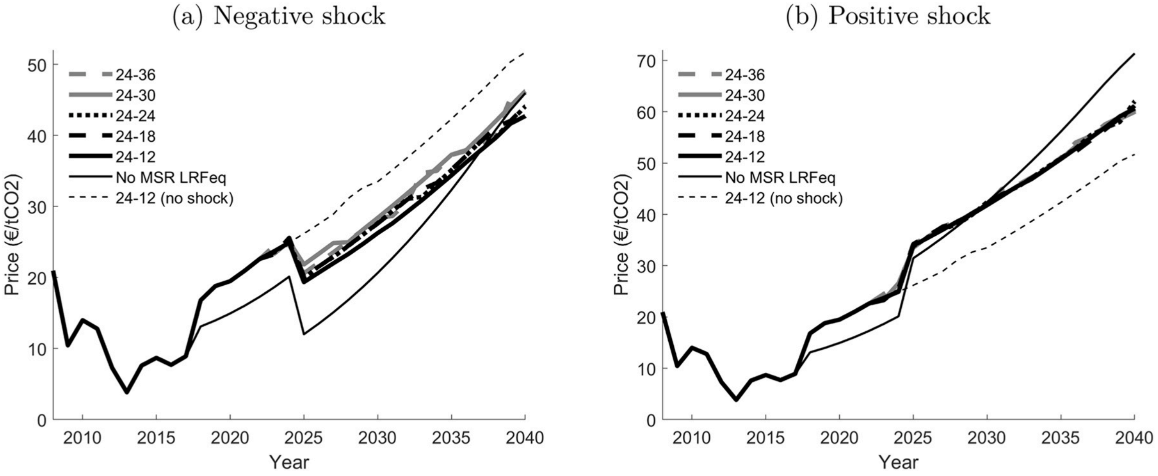


Fig. 9. Price impacts of permanent demand shocks with and without MSR. *Note:* Unanticipated permanent shocks on yearly permit demand from 2025 onward: negative (−150 $MtCO_2$, left); positive (+150 $MtCO_2$, right). Release and intake thresholds set at 0.400 and 0.833 $GtCO_2$, constant over time. Intake rate of 24% over 2019–2023 and 12, 18, 24, 30 or 36% afterward. LRF_{eq} (2.8%) is the LRF that generates the same cumulative emissions without MSR as those obtained under an LRF of 2.2% with MSR + CM (on average across the various intake rates and without the shock).

We assess the ability of the MSR to respond to demand shocks in terms of induced changes in the supply schedule as well as resulting immediate and longer-term price responses. We use an illustrative symmetric example whereby an unanticipated permanent positive or negative shock (± 150 $MtCO_2$) on demand occurs in 2025.[28] Our approach to introducing shocks is hence similar to that in Perino and Willner (2016) but simpler than the full-blown analysis in Fell (2016). Fig. 9 compares MSR-sustained price paths with those without the shocks and those obtained with a sole equivalent 2.8% LRF_{eq}.[29]

The MSR induces some, but limited, resilience to future demand shocks.

We begin with the negative shock (Fig. 9a) and see that the MSR has a limited cushioning capacity. In the year of the shock, the MSR can contain the price fall by 32 up to 60% w.r.t. no MSR (LRF_{eq}) depending on the intake rate. Importantly, because of a minimum one-year lag in MSR operations (they are a function of the bank in the two previous years), the buffer results from the anticipation of MSR-driven supply adjustments in future years, but not from contemporaneous adjustments. Its extent hinges on, but is not monotonic in, the intake rate. Non-monotonicity results from increased oscillations around the intake threshold, and thus more erratic and shock-unrelated intakes, as the intake rate increases.[30]

The MSR does not foster price recovery over time, irrespective of the intake rate which does not make much of a difference in terms of price levels. Indeed, MSR-driven price paths remain parallel to that without shock, with no sign of recovery. This is simply explained in terms of supply-demand imbalance as the MSR only absorbs between 10% and 17% of the cumulative shock depending on the intake rate. Therefore, prices cannot return to the levels that would have prevailed absent the shock. This is despite the fact that yearly MSR intakes are higher in response to the shock, albeit by a less than commensurate amount (see Appendix C).[31]

MSR-induced resilience hinges on: (1) the extent to which shocks are transmitted to the bank and (2) the duration between the shock and the end of the intake period (once intakes have stopped, the MSR can no longer cut back on supply). Point (1) essentially depends on firms' behavior in the face of the shock and the transmission to the bank is always less than one-to-one (see QT21). Point (2) depends on changes in the intake cut-off year, a key element of the MSR responsiveness, but Table 2 shows that changes are marginal. Points (1−2) also depend on the bank level when the shock hits because the MSR is based on arbitrary threshold levels and because there is a inherent asymmetry between adjusting the bank upward or downward due to the limited borrowing condition (e.g. Deaton and Laroque, 1992).

The MSR acts more as a price support provider than as a price stabilizer

We now turn to the positive shock and observe similar patterns. In the year of the shock, the price jump is contained by 20 up to 39% w.r.t. no MSR (which is less than for the negative shock) depending on the intake rate, again non monotonically. Similarly, there is no sign of reversion over time as MSR-driven price paths remain parallel to that with no shock. This is because

---

[28] Roughly speaking, the sooner the shock hits, the more time the MSR has to potentially respond. When the shock occurs, the firm updates its demand forecast as per (4) and adjusts its decisions accordingly.

[29] See Bruninx and Ovaere (2020) and Gerlagh et al. (2020) for specific analyses of MSR responses to the Covid-19 recession. Here, we stick to a generic case to provide a general appraisal of MSR-induced resilience.

[30] Here the buffer is maximal for a 30% intake rate. With a 36% rate the bank is below the intake threshold in 2026–2027 without the shock (Fig. 4b), which reduces (anticipated future) intakes and thus the buffer.

[31] The fact the price path obtained under the 2.8% LRF_{eq} catches up with MSR-driven paths despite the absence of supply-side control may be surprising. As explained in Section 3.1, however, this catch-up effect is driven by bounded foresight and materializes independently of the shock.

Table 2
Shock-driven changes in MSR intake cut-off years, cancellations and 2030 prices.

	Intakes stop in			Δ Cancellations		Δ 2030 price	
IR	No shock	– shock	+ shock	– shock	+ shock	– shock	+ shock
24–12	2053	2056	2051	+1.11	− 1.33	− 7.24	+8.30
24–18	2052	2051	2050	+1.50	− 1.38	− 6.39	+8.06
24–24	2050	2050	2049	+1.42	− 1.41	− 6.03	+8.72
24–30	2052	2051	2049	+1.71	− 1.45	− 5.51	+7.99
24–36	2049	2051	2050	+1.64	− 1.28	− 6.69	+6.76

Note: Release and intake thresholds set at 0.400 and 0.833 $GtCO_2$, constant over time. Intake rate of 24% over 2019–2023 and 12, 18, 24, 30 or 36% afterward. Shock-induced changes in 2030 price levels ('Δ 2030 price') and cancelled volumes ('Δ Cancellations') are measured in € per tCO_2 and $GtCO_2$, respectively.

the MSR continues to cut back on supply despite the positive shock (albeit to a lesser extent since yearly intakes are slightly reduced, see Appendix C), and never releases permits into the market.[32] This reveals an asymmetry to negative vs. positive shocks inherent to the MSR design, which is further reflected by the facts that: (1) price paths are more distinct across intake rates for the negative shock than for the positive shock and (2) the absolute changes in 2030 price levels are always higher for the positive shock (Table 2).

Intuition suggests that relative to no shock, the MSR should withdraw more (less) permits on an annual basis, and do so over a longer (shorter) period under a negative (positive) shock. While our intuition for annual intakes holds most of time (except at times for large intake rates and induced oscillations, see Appendix C), Table 2 shows that for the duration of the intake period only holds for a 12% intake rate, although changes are always small. Table 2 also indicates that changes in cumulative withdrawals and cancellations is small relative to the size of the cumulative shock (± 0.15 $GtCO_2$ yearly).

In line with its first objective of tackling historical oversupply, the MSR has been engineered for supply contraction, not expansion. That is, irrespective of the shock structure and direction, the MSR offers a one-sided response – in this sense, it acts more as an (unconditional) price support provider than as a price stabilizer. Indeed, the MSR always cuts back on supply and the CM cancels withdrawn permits later on, although some responsiveness is reflected in changes in the magnitude of the MSR intakes. Implementing declining thresholds cannot overcome that inherent design asymmetry. On the contrary, it would be amplified by the associated increase in the size and duration of intakes irrespective of the shock structure. More generally, our analysis calls into question the adequateness of basing supply-side control on past banking for the purpose of improving market resilience.

4. Conclusion

In this paper, we first tailor and calibrate an emission permit trading model to the EU ETS featuring the market stability reserve (MSR) and its add-on cancellation mechanism (CM). A pivotal difference with the literature is that firms can use rolling finite horizons in their decision making, e.g. as a procedure to deal with uncertainty about future permit demand. Rolling horizons are found to: (1) reconcile annual banking developments over 2008–17 with discount rates derived from actual futures' yield curves, (2) replicate average annual prices over 2008–17 with a twice better fit than an infinite horizon, and (3) reproduce most of the 2018 price rally where an infinite horizon falls short of it. If firms are de facto or behave as if boundedly farsighted, this has important ramifications for policy design and implementation, which we explore and quantify in the context of the upcoming 2021 market review.

In a second step, we leverage our calibrated model to provide a detailed quantitative assessment of policy-relevant options in revising the MSR parameters and cap trajectory (LRF) to inform the 2021 review. We find that: (1) with intake-release thresholds constant over time, a higher intake rate generates higher volatility due to more pronounced oscillatory behavior around the intake threshold without leading to higher ambition, (2) the position of the intake threshold matters more than that of the release threshold in terms of market outcomes, a lower intake threshold sustaining higher prices and ambition, and (3) as a potent amendment, combining thresholds that are declining over time (e.g. based on the LRF) with higher intake rates leads to higher price and ambition levels without destabilizing the market.

Because the MSR has the potential to permanently curb supply through the CM, it could be utilized hand in hand with an LRF increase to raise ambition. But because these two policy levers have different consequences in terms of perceived scarcity for boundedly farsighted firms, the policy mix should be chosen to reap synergies and minimize costs. This requires us to investigate the LRF-MSR interaction, the nature of which depends on the LRF value and MSR parameters. We find that declining thresholds are always less costly than constant thresholds for a given ambition target, especially when coupled with a higher intake rate for more ambitious targets. In general, the cost-optimal mix of LRF-MSR parameters leverages more the MSR's indirect ambition-raising potential than the LRF's direct one. Indeed, as the MSR partially compensates for firms' bounded foresight by frontloading abatement efforts, leaving more traction to an enhanced MSR than to the LRF can be efficiency-improving.

---

[32] Were it able to release permits, the MSR would only do so in predetermined chunks of 100 $MtCO_2$.

In spite of this and even after changes in its parameters, the ability of the MSR to improve market resilience to permit demand shocks is limited and one-sided by design. The MSR is inherently geared towards supply contraction (it is a price support provider) but only weakly responds to shocks, both in the short and long term (it is not a price stabilizer). Implementing declining thresholds cannot overcome this built-in limitation. Rather, this would even amplify the MSR asymmetry in responding to positive vs. negative demand shocks. Essentially, this is because the MSR adjusts supply based on a poor indicator – market-wide banking – for expected permit scarcity and market stability (e.g. Gerlagh et al., 2021; Perino et al., 2020). As a result, the MSR can even be counterproductive (e.g. erode abatement in case of an anticipated increase in scarcity) or undermine market self-stabilizing forces (e.g. self-fulfilling prophecies). To obviate this problem, one could flank or replace the MSR with a price-based supply-side control instrument as in the UK and North America (e.g. Newbery et al., 2019; Burtraw et al., 2020; Flachsland et al., 2020; Perino et al., 2021). The permit price is indeed the most reliable measure of scarcity that is readily available: it aggregates information from all market participants and reflects how much effort firms put into curbing emissions. Thus, conditioning permit supply on price levels certainly has the potential to create more robust and stable signals for deep decarbonization than a banking-based MSR does.

Funding information

The author acknowledges funding from the UK Economic and Social Research Council (ESRC, grant ES/R009708/1) through the Centre for Climate Change Economics and Policy, the Grantham Foundation for the Protection of the Environment, the LABEX Louis Bachelier Finance and Sustainable Growth (ANR 11-LABX-0019 Project) under the 'Investments for the Future' programme in accordance with Article 8 of the Assignment Agreements for Communication Assistance, and the Climate Economics Chair (Paris-Dauphine University, PSL Research University).

Declarations of interest

None. This work is not commissioned by, nor does it reflect the views of, Électricité de France. All errors and views expressed in this article are the author's.

Appendix A. Calibration to 2008–17 price and banking data

We consider that permit demand is linear in the permit price, which is a standard assumption (e.g. Ellerman and Montero, 2007; Kollenberg and Taschini, 2019). We thus assume that $C''' \approx 0$, which can be viewed as a local second-order approximation of general functional forms. We also assume that the slope of the linear marginal abatement cost functions is time invariant, i.e.

Table A.3
Best-fit results based on 2008–17 bank and price data.

Horizon type	Horizon & discount rate	Marginal abatement cost
Infinite	$h = \infty$ 7D1$r = 7.83\%$ (std.dev = 59.4 $MtCO_2$)	$c = 5.77 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 3.80 €/tCO_2)
Rolling	$h = 117$D1$r = 3\%$ (std.dev = 76.4 $MtCO_2$)	$c = 6.10 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 1.71 €/tCO_2)

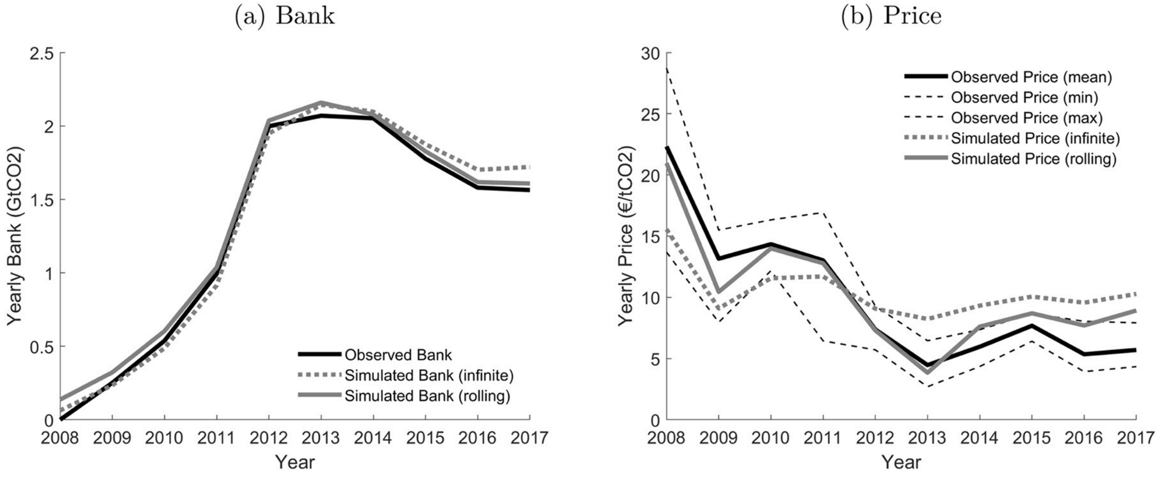

(a) Bank
(b) Price

Fig. A.10. Best-fit results based on 2008–17 bank and price data.

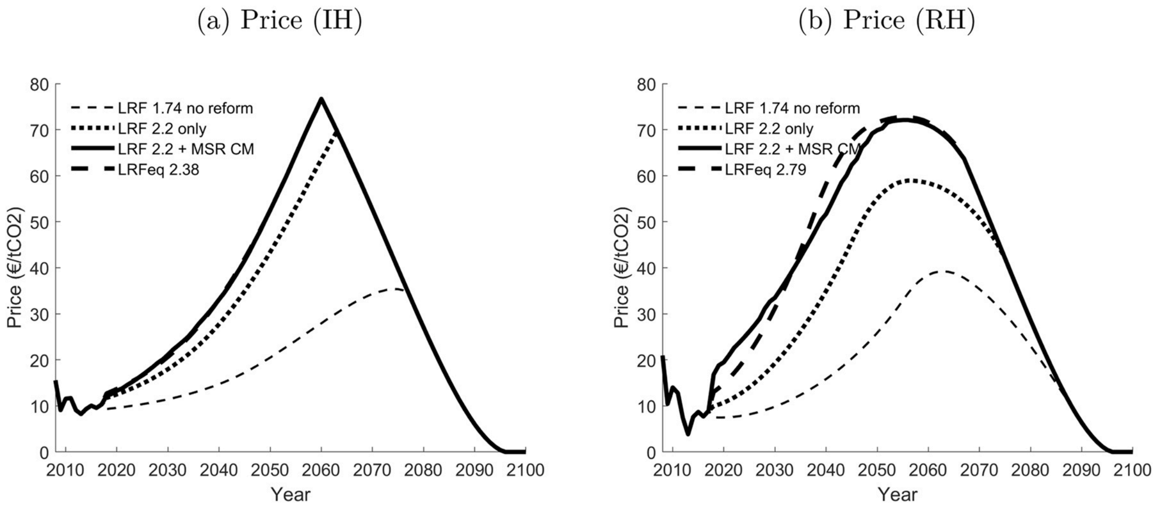
Table A.4
Discount rates inferred from daily futures' yield curves over 2008–17.

Daily yield curve	Mean	Median	Std.Dev	Min	Max
Fut. Dec Y+1 / Spot	2.4%	2.5%	1.5%	0.2%	7.0%
Fut. Dec Y+1 / Fut. Dec Y	2.9%	2.6%	1.8%	0.3%	8.7%
Fut. Dec Y+2 / Fut. Dec Y+1	3.6%	3.7%	2.0%	0.2%	8.7%

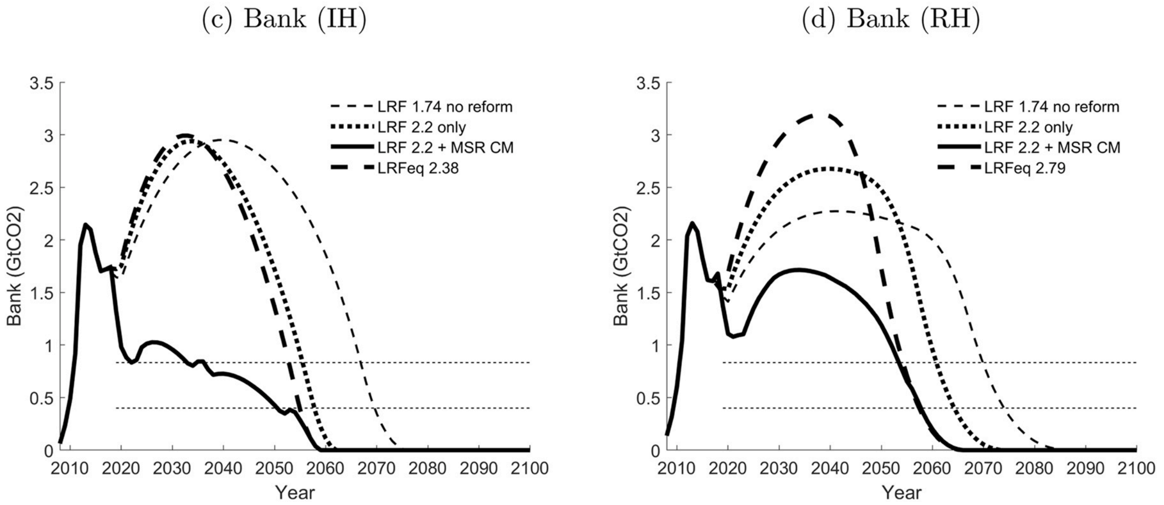
Note: Daily EUA futures price data compiled from the IntercontinentalExchange (ICE). With t_1 the day's date (for spot) or maturity (for futures) of asset a with price $p_a^{t_1}$ and $t_2 > t_1$ the maturity of futures b with price $p_b^{t_2}$ the inferred discount rate is given by $\ln\left(p_b^{t_2}/p_a^{t_1}\right)/(t_2 - t_1)$ since storage costs are nil.

$C'' = c$ for all t. We do so for three reasons. First, it ensures that our two-step calibration approach is legitimate as a constant c does not influence the firm's banking strategies, which only depend on its discount rate and horizon. Second, as a fixed scaling parameter, c only affects the levels, but not the shapes, of the simulated price paths. Third, it is a conservative assumption given that we have limited empirical and theoretical guidance on the evolution of the marginal abatement cost slope over time. Yet, note that the linear intercept is gradually lowered over time as the actual baseline path is downward sloping (see Fig. 2b).

We calibrate the model parameters following a two-step procedure in the spirit of a standard least squares maximum likelihood estimation with one free parameter. In the first step, we select r given h or h given r so that the simulated banking path deviates the least from the observed banking path over 2008–17. In the second step, given r and h, we select c so that the


(a) Price (IH)

(b) Price (RH)


(c) Bank (IH)

(d) Bank (RH)

Fig. A.11. Price and bank paths with calibrated infinite vs. rolling horizons. *Note:* Release and intake thresholds are respectively set at 0.400 and 0.833 $GtCO_2$ and remain constant over time. Intake rate of 24% over 2019–2023 and 12% afterward. LRF_{eq} is the LRF which, on its own, generates the same 2008–2100 cumulative emissions as those generated by a 2.2% LRF with the MSR+CM.

simulated price path deviates the least from the yearly-averaged spot price path over 2008–17. In each step, the free parameter is calibrated by minimizing the distance between simulated and observed paths. Table A.3 reports the best-fit results and Fig. A.10 depicts the observed and best-fit simulated paths over 2008–17 for visual comparison.

With the infinite horizon $h = \infty$, we find that the discount rate $r = 7.83\%$ best replicates past banking with a fit of 59 MtCO$_2$/year. This aligns with general rates of return on risky assets (e.g. Jordà et al., 2019) but is in the higher range of the rates that can be inferred from futures' yield curves since Phase II (see Table A.4). Additionally, one might argue that since permits can be banked for hedging purposes, required returns should be less than those for standard risky assets. With a rolling horizon, we set $r = 3\%$, which is a central value for inferred discount rates, and find $h = 11$ years with a similar fit of 76 MtCO$_2$/year. The values we get for c are similar with the calibrated infinite and rolling horizons, in the order of $6 \cdot 10^{-8}$€/(tCO$_2$)^2 and in line with dedicated studies (Böhringer et al., 2009; Landis, 2015). However, the price fit is more than twice as good with the calibrated rolling finite horizon, i.e. 1.7 vs. 3.8 €/tCO$_2$/year with the calibrated infinite horizon.

With a similar approach in the US Acid Rain Program, Ellerman and Montero (2007) compare observed and simulated banking paths for various given pairs of discount rates and expected demand growth rates to guess at which pair might have governed the dynamics. While they analyze the permit-specific CAPM beta to select appropriate values for the discount rate, we use information provided by futures trading to elicit how market participants value present vs. future permits. Moreover, we augment their approach by endogenizing changes in firms' expectation about future demand, which they note is key in driving banking decisions.

Finally, Fig. A.11 offers a visual aid to compare how the in-sample infinite vs. rolling finite horizons fare out of sample. A short description of the results obtained with the calibrated infinite horizon are provided in footnotes ,[18] [19] and .[20] The reader is also referred to QT21 for a detailed comparative analysis.

B. Streams of annual MSR intakes

With constant thresholds, cumulative intakes (and hence cancellations) are similar and only marginally increasing with the intake rate. Flows, however, differ substantially across intake rates: with a 12 or 24% rate, annual intakes are relatively stable over time but as the intake rate increases, they become more erratic as thresholds are repeatedly being hit and the intake period is shorter. Overall, however, cumulative impacts are similar across intake rates.

By contrast, with declining thresholds, annual intakes are more evenly distributed over time for all intake rates – except for a 48% rate early on as it is too high to avert threshold effects. While intake rates vary by a factor of 4, yearly intakes vary in volume only by a factor of 2 at most. This is because lower bank levels (Fig. 6b) mitigate the absolute impacts of higher intake rates. This notwithstanding, cumulative impacts vary more across intake rates.Fig. B.12

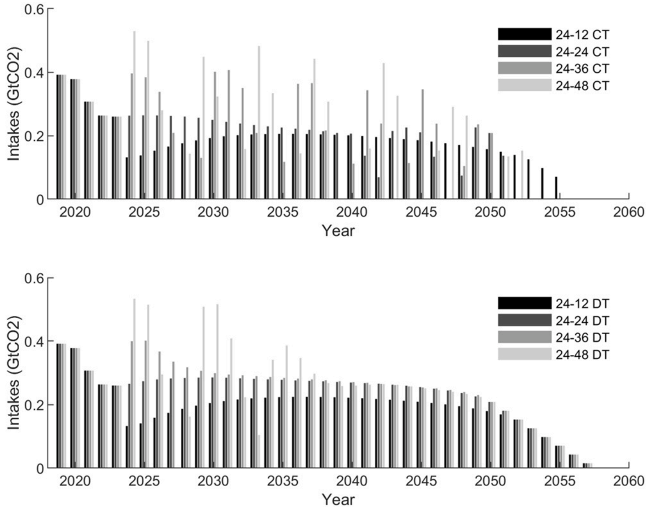


Fig. B.12. Annual MSR intakes with different intake rates and thresholds. *Note:* Intake rate of 24% over 2019–2023, and 12, 24, 36 or 48% afterward. (upper) Constant 400–833 MtCO$_2$ thresholds; (lower) linearly declining thresholds from 400 to 833 in 2023–0–0 MtCO$_2$ in 2058. Reinjections (i.e. negative intakes) never occur in our simulations since the MSR is already empty when the release threshold is passed. The plots do not include the 1.55 GtCO$_2$ that are exogenously seeded into the MSR.

C. MSR-induced resilience to demand shocks

Fig. C.13 shows how annual MSR intakes change in the presence of the positive (upper plot) and negative (lower plot) shocks, relative to no shock. In general, we observe that annual changes in intakes are less than the size of the annual demand changes (± 0.15 GtCO$_2$). This is because the shocks are not entirely transmitted to and reflected in the bank. Additionally, we note that most of the time the sizes of the variations are ranked by increasing intake rate. Oscillations, which are more likely to occur with a higher intake rate, can sometimes lead to: (1) a more than proportional response, (2) a reversed response, and (3) perturbations in the ordering of the responses by increasing intake rate.

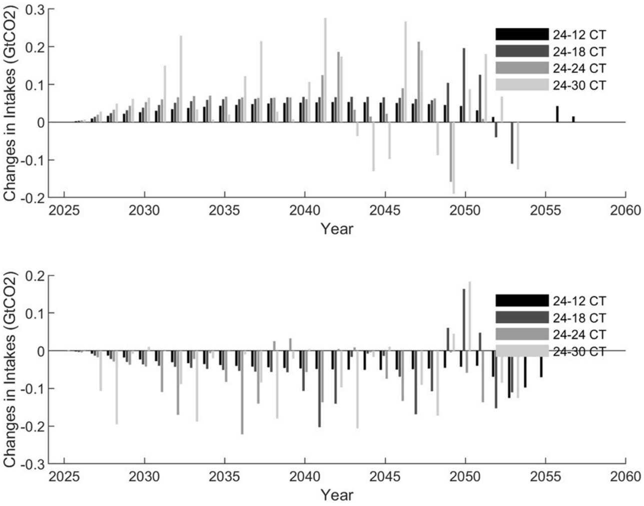


Fig. C.13. Changes in annual MSR intakes with shocks relative to no shock. *Note:* Intake rate of 24% over 2019–2023, and 12, 18, 24 or 30% afterward. Release and intake thresholds are set at 0.400 and 0.833 GtCO$_2$ and stay constant over time. Unanticipated permanent negative –0.15 GtCO$_2$ (positive +0.15 GtCO$_2$) demand shock in 2025 in the upper (lower) plot.

D. Impacts of phasing out free allocations

Due to near-term differences in the stringency of domestic climate policies, the EU Green Deal mentions the possible introduction of border carbon adjustments to safeguard a level playing field for vulnerable and trade-exposed industries and counteract induced carbon leakage, i.e. the displacements of production, investment and GHG emissions (e.g. Mehling et al., 2019). Domestic measures – chiefly free allocations – have so far mostly been used to address these risks, albeit with mixed and disputable results. Indeed, a growing body of evidence suggests that free allocations do not perform as intended, e.g. with windfall and overallocation profits (e.g. Bushnell et al., 2013; Hintermann et al., 2016). Another behavioral limitation of free allocations is that they partly conceal the price signal, thereby eroding the opportunity cost of free permits and associated uptake of abatement options (e.g. Martin et al., 2015; Venmans, 2016; Baudry et al., 2021). In any case, the implementation of border adjustments should imply the removal of these measures, at least to avoid double protection mechanisms.

We evaluate the cancellation and price repercussions of a free allocation phase-out, all else constant. We abstract from border adjustment design considerations (see e.g. Böhringer et al., 2017) which are beyond the scope of our framework, and focus on the sole allocation method impacts. Table D.5 reports our results for a complete phase-out in 2024, namely the constant share of the total cap that was set to be auctioned off from 2024 onward (57%) becomes 100%. By considering this extreme case, we quantify an upper bound on MSR-driven impacts. We see that the free allocation phase-out leads to an increase in MSR intakes, cancellations and thus price levels. Specifically, the average increase in cancellations is larger with declining thresholds than constant thresholds (+0.53 vs. +0.38 GtCO$_2$) while the average 2030 price increase is of similar magnitude (about +3€/tCO$_2$.)

These changes are driven by an increase in the stringency of the limited borrowing constraint. Specifically, the transition to no borrowing affects firms' intertemporal decisions by making them bank more and longer (see Appendix F), which translates into larger MSR intakes over a longer period. Arguably, the changes in MSR impacts we capture are marginal. However, we note that behavioral aspects associated with a free allocation phase-out, such as the end of endowment effects Venmans (2016) and autarkic compliance behavior relying on borrowing Baudry et al. (2021), have potential to dwarf the price impacts reported in Table D.5.

Table D.5
Cancellation and price impacts of a transition to full auctioning in 2024.

		Cancellations (GtCO$_2$)			2030 price (€/tCO$_2$)		
Thresholds	IR	57%*	100%*	Δ	57%*	100%*	Δ
Constant	24–12	8.71	9.10	0.39	33.5	36.3	2.8
	24–18	8.84	9.13	0.29	34.0	37.1	3.1
	24–24	8.84	9.18	0.34	33.7	36.9	3.2
	24–30	8.96	9.33	0.37	33.9	37.1	3.2
	24–36	8.97	9.48	0.51	35.1	37.4	2.3
Declining	24–12	9.27	9.75	0.48	35.5	37.9	2.4
	24–18	10.57	11.27	0.70	38.3	41.3	3.0
	24–24	11.04	11.83	0.79	38.1	41.5	3.4
	24–30	11.39	12.02	0.63	39.0	41.1	2.1
	24–36	11.57	12.34	0.77	39.2	42.1	2.9

Note: Intake rate of 24% over 2019–2023 and 12, 18, 24, 30 or 36% afterward. Constant thresholds are fixed at 400–833 MtCO$_2$ over time, declining thresholds are set at 400–833 MtCO$_2$ over 2019–2023 and then linearly decline to 0 in 2058 (2.2% LRF). The $\star$ indicates the constant proportion of auctions (out of the total cap) from 2024 onward and Δ measures the difference between the two cases analyzed ('100%–57%').

E. Pathways to a higher 2030 ambition target

The current ambition target is to reduce covered emissions by 43% in 2030 w.r.t. 2005 levels (2.32 GtCO$_2$). In the status quo, our simulations show that the ETS is bound to overachieve this target with a 48% cut (Table E.6). As an illustration, we consider an ambition ramp-up to − 62%. This is more ambitious than what has been recently agreed upon (−55%). However, this assumption does not change the qualitative nature of our results and suggests that a more ambitious target could be within reach. In passing, we underline that setting an emission target for a given year is tricky due to the market's intertemporal dimension. For instance, a zero target for 2050 requires that the cap be zero before 2050 as some banked permits may still be used to cover emissions after the cap has shrunk to zero. Importantly, these aspects are even more convoluted with the MSR in place. Table E.6 lists possible LRF-MSR parametrizations to attain the − 62% target. The required LRF is always lower with the MSR than without (4.16%) and varies with the MSR parameters. Specifically, with constant thresholds, the required LRF is around 2.9% and slightly decreases with the intake rate. With declining thresholds, it is even lower, especially with a 24 or 36% intake rate where it lies around 2.6%. This was to be expected because declining thresholds allow for higher ambition (Section 3.2) but observe the decreasing returns in raising the intake rate, e.g. the required LRF is lowered by.01% only when the rate goes from 24% to 36%. Note also that in all cases the − 62% target does not lead to carbon neutrality by 2050, with more than 100 MtCO$_2$ of residual emissions. Reaching exactly zero emissions by 2050 would require a much higher LRF, above 4% in all cases.

Table E.6 also frames the LRF-MSR interaction analyzed in Section 3.3 in a specific context. With both constant or declining thresholds, we see that an increase in the intake rate or in the LRF always shortens the intake period. As Section 3.3 suggests, only for the smallest intake rate (12%) does the long-term indirect effect of an LRF increase dominates its short-term indirect effect, with resulting smaller cancellations. Although all LRF-MSR settings achieve the same 2030 emissions levels, cumulative emissions differ due to the LRF-MSR interaction. Hence, we cannot meaningfully compare their relative costs as we do in Section 3.3.

Table E.6
LRF-MSR settings to reach a − 62% target in 2030 w.r.t. 2005.

			Emissions (GtCO$_2$)					
Thresholds	IR	LRF	2030	2040	2050	Cumul	Intakes end	Cancel
No MSR	–	2.20	1.28	0.85	0.42	58.6	–	–
		4.16	0.88*	0.41	0.15	45.1	–	–
Constant	24–12	2.20	1.11	0.67	0.29	51.5	2053	8.71
		2.96	0.88*	0.40	0.15	44.2	2047	8.51
	24–24	2.20	1.11	0.67	0.28	51.3	2050	8.89
		2.89	0.88*	0.39	0.12	43.8	2042	9.51
	24–36	2.20	1.10	0.68	0.28	51.2	2049	8.97
		2.83	0.88*	0.42	0.13	44.0	2043	9.77
Declining	24–12	2.20	1.08	0.64	0.28	50.9	2066^b	9.27
		2.94	0.88*	0.41	0.15	44.3	2058^b	8.60
	24–24	2.20	1.05	0.59	0.23	49.1	2063^b	11.0
		2.63	0.88*	0.40	0.14	44.2	2058^b	11.3
	24–36	2.20	1.04	0.59	0.21	48.6	2061^b	11.6
		2.62	0.88*	0.38	0.12	43.7	2056^b	11.8

Note: Intake rate of 24% over 2019–2023 and 12, 24 or 36% afterward. Release and intake thresholds are set at 0.400 and 0.833 GtCO$_2$: permanently constant or constant over 2019–2023 and then linearly declining to reach 0 in the same year as the cap given the LRF. The superscript b indicates that intakes stop only when the bank becomes zero, i.e. banking never passes below the intake threshold. The superscript $\star$ denotes the hypothetical 2030 target of − 62% relative to 2005 levels. 'Cancel' reports cumulative cancellations in GtCO$_2$.

F. The limited borrowing constraint

We consider a stylized example to understand the implications of changing the restrictions on borrowing. We separate out this aspect by focusing on the simplest environment possible (i.e. certainty, perfect foresight, cost-minimizing behavior, and no MSR) and the two polar cases where unlimited borrowing is authorized vs. borrowing is completely prohibited. All that is required for our qualitative results to hold is that the distance between baseline emissions and the emissions cap be increasing over time. Note, however, that the shape of the banking path hinges on those of the baseline and cap trajectories.

With full banking and borrowing, Hotelling's rule holds (see Section 2.1). The price always rises at the discount rate (Fig. F.14a) and the optimal intertemporal reallocation of permits (w.r.t the cap trajectory) involves a banking phase followed by a borrowing phase (Fig. F.14b). With full banking but no borrowing, the price starts from a higher level and rises at the discount rate as long as the no borrowing constraint is not binding, and the banking path follows an inverted U curve. Exactly when it becomes binding, the bank becomes empty and the price can but rise at a rate lower than the discount rate from there on.

When borrowing is unrestricted, firms find it optimal to shift some abatement from the short run to the long run relative to no borrowing. This implies that without borrowing firms bank more permits and stop banking at a later date. All else constant, this would lead to larger yearly MSR intakes over a longer period – as would be the case for a complete free allocation phase out in Appendix D, which is tantamount to a transition from limited to no borrowing.

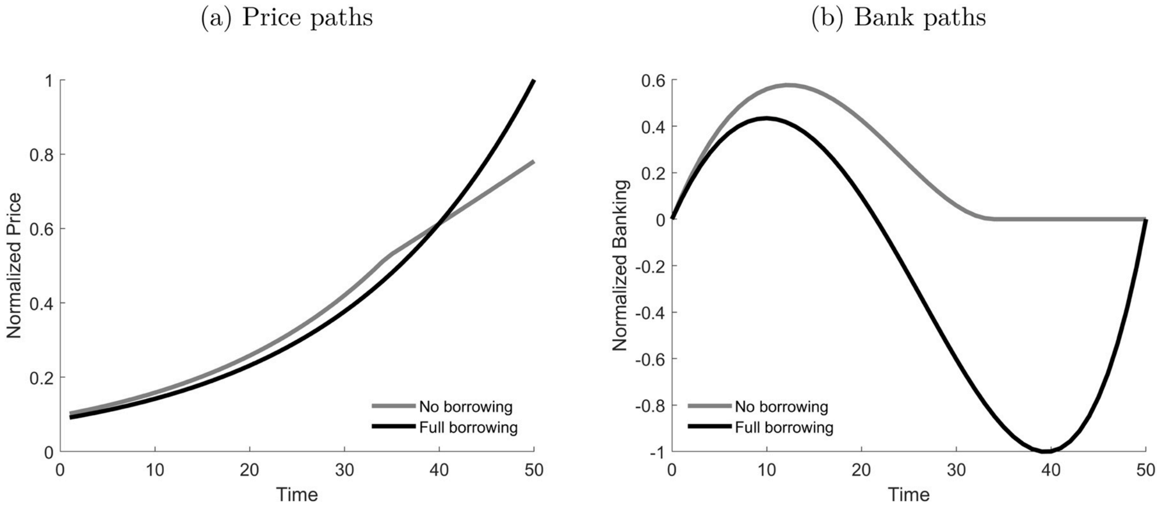


Fig. F.14. Stylized price and bank paths with and without borrowing (no MSR).

References

Abrell, J., Kosch, M., Rausch, S., 2019. Carbon abatement with renewables: evaluating wind and solar subsidies in Germany and Spain. J. Public Econ. 169, 172–202.
Baudry, M., Faure, A., Quemin, S., 2021. Emissions trading with transaction costs. J. Environ. Econ. Manag. 108, 102468.
Beck, U., Kruse-Andersen, P.K., 2020. Endogenizing the cap in a cap-and-trade system: assessing the agreement on EU ETS Phase 4. Environ. Resour. Econ. 77, 781–811.
Bocklet, J., Hintermayer, M., Schmidt, L., Wildgrube, T., 2019. The reformed EU ETS - intertemporal emission trading with restricted banking. Energy Econ. 84, 104486.
Böhringer, C., Löschel, A., Moslener, U., Rutherford, T.F., 2009. EU climate policy up to 2020: an economic impact assessment. Energy Econ. 31, S295–305.
Böhringer, C., Rosendahl, K.E., Storrøsten, H.B., 2017. Robust policies to mitigate carbon leakage. J. Public Econ. 149, 35–46.
Borenstein, S., Bushnell, J., Wolak, F.A., Zaragoza-Watkins, M., 2019. Expecting the unexpected: emissions uncertainty and environmental market design. Am. Econ. Rev. 109 (11), 3953–3977.
Bruninx, K., Ovaere, M., 2020. Estimating the impact of COVID-19 on emissions and emission allowance prices under EU ETS. IAEE Energy Forum 40–42 Covid-19 Issue.
Bruninx, K., Ovaere, M., 2022. Waterbed leakage drives EU ETS emissions: COVID-19, the green deal & the recovery plan. Nat. Commun (in press).
Bruninx, K., Ovaere, M., Delarue, E., 2020. The long-term impact of the market stability reserve on the EU emission trading system. Energy Econ. 89, 104746.
Burtraw, D., Holt, C.A., Palmer, K.L., Paul, A., Shobe, W.M., 2020. Quantities with prices: price-responsive allowance supply in environmental markets. Work. Pap. 20-17, Resour. Future.
Bushnell, J.B., Chong, H., Mansur, E.T., 2013. Profiting from regulation: evidence from the european carbon market. Am. Econ. J.: Econ. Policy 5 (4), 78–106.
Carlén, B., Dahlqvist, A., Mandell, S., Marklund, P., 2019. EU ETS emissions under the cancellation mechanism - effects of national measures. Energy Policy 129, 816–825.
Chaton, C., Cretì, A., Sanin, M.-E., 2018. Assessing the implementation of the market stability reserve. Energy Policy 118, 642–654.
Chèze, B., Chevallier, J., Berghmans, N., Alberola, E., 2020. On the CO_2 emissions determinants during the EU ETS phases I and II: a plant-level analysis merging the EUTL and platts power data. Energy J. 41 (4), 153–184.

Deaton, A., Laroque, G., 1992. On the behaviour of commodity prices. Rev. Econ. Stud. 59 (1), 1–23.

Ellerman, A.D., Marcantonini, C., Zaklan, A., 2016. The European union emissions trading system: ten years and counting. Rev. Environ. Econ. Policy 10 (1), 89–107.

Ellerman, A.D., Montero, J.-P., 2007. The efficiency and robustness of allowance banking in the U.S. acid rain program. Energy J. 28 (4), 47–71.

European Commission (2014). Commission Regulation (EU) 176/2014. Brussels, 16.2.2014.

European Commission (2015). Impact Assessment SWD(2015) 135. Brussels, 15.7.2015.

European Commission (2021). COM(2021) 551 final. Brussels, 14.07.2021.

European Parliament & Council (2003). Directive 2003/87/ec. OJ L 275, 13.10.2003.

European Parliament & Council (2018). Directive (eu) 2018/410. OJ L 76, 19.03.2018.

Fell, H., 2016. Comp. Policies Confront. Permit -Alloc. J. Environ. Econ. Manag. 80, 53–68.

Fell, H., Burtraw, D., Morgenstern, R.D., Palmer, K.L., 2012. Soft and hard price collars in a cap-and-trade system: a comparative analysis. J. Environ. Econ. Manag. 64 (2), 183–198.

Flachsland, C., Pahle, M., Burtraw, D., Edenhofer, O., Elkerbout, M., Fischer, C., Tietjen, O., Zetterberg, L., 2020. How to avoid history repeating itself: the case for an EU emissions trading system (EU ETS) price floor revisited. Clim. Policy 20 (1), 133–142.

Friedman, M., 1953. Essays in Positive Economics. University of Chicago Press.

Friedrich, M., Fries, S., Pahle, M., Edenhofer, O., 2020. Rules vs. discretion in cap-and-trade programs: evidence from the EU emission trading system. Work. Pap. No. 8637, CESifo.

Fuss, S., Flachsland, C., Koch, N., Kornek, U., Knopf, B., Edenhofer, O., 2018. A framework for assessing the performance of cap-and-trade systems: insights from the european union emissions trading system. Rev. Environ. Econ. Policy 12 (2), 220–241.

Gerlagh, R., Heijmans, R.J.R.K., Rosendahl, K.E., 2021. An endogenous emissions cap produces a green paradox. Econ. Policy 36 (107), 485–522.

Gerlagh, R., Heijmans, R.J.R.K., Rosendahl, K.E., 2020. COVID-19 tests the market stability reserve. Environ. Resour. Econ. 76, 855–865.

Goldman, S.M., 1968. Optimal growth and continual planning revision. Rev. Econ. Stud. 35 (2), 145–154.

Grüne, L., Semmler, W., Stieler, M., 2015. Using nonlinear model predictive control for dynamic decision problems in economics. J. Econ. Dyn. Control 60, 112–133.

Hintermann, B., Peterson, S., Rickels, W., 2016. Price and market behavior in phase II of the EUETS: a review of the literature. Rev. Environ. Econ. Policy 10 (1), 108–128.

Hotelling, H., 1931. The economics of exhaustible resources. J. Political Econ. 39 (2), 137–175.

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M., Taylor, A.M., 2019. The rate of return on everything, 1870-2015. Q. J. Econ. 134 (3), 1225–1298.

Kollenberg, S., Taschini, L., 2019. Dynamic supply adjustment and banking under uncertainty in an emissions trading scheme: the market stability reserve. Eur. Econ. Rev. 118, 213–226.

Landis, F., 2015. Final report on marginal abatement cost curves for the evaluation of the market stability reserve. Dok. 15-01, ZEW Mannh.

Leard, B., 2013. The welfare effects of allowance banking in emissions trading programs. Environ. Resour. Econ. 55, 175–197.

Lucas, R.E., Prescott, E.C., 1971. Investment under uncertainty. Econometrica 39 (5), 659–681.

Marcantonini, C., Ellerman, A.D., 2015. The implicit carbon price of renewable energy incentives in Germany. Energy J. 36 (4), 205–239.

Marcu, A., Cabras, S., 2021. Rev. EU ETS- Overv. EU Comm. Propos. 14. 07. 2021, ERCST.

Martin, R., Muûls, M., Wagner, U.J., 2015. Trading behavior in the EU emissions trading scheme. In: Gronwald, M., Hintermann, B. (Eds.), Emissions Trading as a Policy Instrument: Evaluation and Prospects, chapter 9. MIT Press, pp. 213–238.

Mauer, E.-M., Okullo, S.J., and Pahle, M. (2019). Evaluating the Performance of the EU ETS MSR. Working Paper, PIK.

Mehling, M.A., van Asselt, H., Das, K., Droege, S., Verkuijl, C., 2019. Designing border carbon adjustments for enhanced climate action. Am. J. Int. Law 113 (3), 433–481.

Montgomery, W.D., 1972. Markets in licenses and efficient pollution control programs. J. Econ. Theory 5 (3), 395–418.

Naegele, H., Zaklan, A., 2019. Does the EU ETS cause carbon leakage in european manufacturing? J. Environ. Econ. Manag. 93, 125–147.

Newbery, D.M., Reiner, D.M., Ritz, R.A., 2019. The political economy of a carbon price floor for power generation. Energy J. 40 (1), 1–24.

Osorio, S., Tietjen, O., Pahle, M., Pietzcker, R., Edenhofer, O., 2020. Reviewing the market stability reserve in light of more ambitious EU ETS emission targETS. Work. Pap., ZBW - Leibniz Inf. Cent. Econ.

Pahle, M., Quemin, S., 2020. EU ETS: the market stability reserve should focus on carbon prices, not allowance volumes. Energypost.eu (16 June 2020).

Parry, I., 2020. Increasing carbon pricing in the EU: evaluating the options. Eur. Econ. Rev. 121, 103341.

Perino, G., 2018. New EU ETS phase IV rules temporarily puncture waterbed. Nat. Clim. Change 8, 262–264.

Perino, G., Pahle, M., Pause, F., Quemin, S., Scheuing, H., Willner, M., 2021. EU ETS stability mechanism needs new design. CEN Policy Brief., Univ. Hambg.

Perino, G., Ritz, R.A., van Bentem, A., 2020. Overlapping Clim. Policies. Discuss Paper 13569, CEPR.

Perino, G., Willner, M., 2016. Procrastinating reform: the impact of the market stability reserve on the EU ETS. J. Environ. Econ. Manag. 80, 37–52.

Perino, G., Willner, M., 2017. EU ETS Phase IV: allowance prices, design choices and the market stability reserve. Clim. Policy 17 (7), 936–946.

Quemin, S. and Pahle, M. (2021). Financials threaten to undermine the functioning of emissions markets. SSRN Working Paper 3985079.

Quemin, S., Trotignon, R., 2021. Emissions trading with rolling horizons. J. Econ. Dyn. Control 125, 104099.

Roberts, M.J., Spence, A.M., 1976. Effluent charges and licenses under uncertainty. J. Public Econ. 5 (3), 193–208.

Rubin, J.D., 1996. A model of intertemporal emission trading. Bank. Borrow. J. Environ. Econ. Manag. 31 (3), 269–286.

Sahin, F., Narayanan, A., Robinson, E.P., 2013. Rolling horizon planning in supply chains: review, implications and directions for future research. Int. J. Prod. Res. 51 (18), 5413–5436.

Schennach, S.M., 2000. The economics of pollution permit banking in the context of title iv of the 1990 clean air act amendments. J. Environ. Econ. Manag. 40 (3), 189–210.

Spiro, D., 2014. Resource prices and planning horizons. J. Econ. Dyn. Control 48, 159–175.

Tietjen, O., Lessmann, K., Pahle, M., 2021. Hedging and temporal permit issuances in cap-and-trade programs: the market stability reserve under risk aversion. Resour. Energy Econ. 63, 101214.

van Veldhuizen, R., Sonnemans, J., 2018. Nonrenewable resources, strategic behavior and the hotelling rule: an experiment. J. Ind. Econ. 66, 481–516.

Venmans, F., 2016. The effect of allocation above emissions and price uncertainty on abatement investments under the EU ETS. J. Clean. Prod. 126, 595–606.

Weitzman, M.L., 1978. Optimal rewards for economic regulation. Am. Econ. Rev. 68 (4), 683–691.