Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Using supply-side policies to raise ambition: the case of the EU ETS and the 2021 review

Simon Quemin

March 2020

Centre for Climate Change Economics and Policy Working Paper No. 366
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment
Working Paper No. 335
ISSN 2515-5717 (Online)

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The Grantham Foundation for the Protection of the Environment

University of Leeds

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its third phase started in October 2018 with seven projects:

1. Low-carbon, climate-resilient cities
2. Sustainable infrastructure finance
3. Low-carbon industrial strategies in challenging contexts
4. Integrating climate and development policies for 'climate compatible development'
5. Competitiveness in the low-carbon economy
6. Incentives for behaviour change
7. Climate information for adaptation

More information about CCCEP is available at www.cccep.ac.uk

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and a number of other sources. It has six research themes:

1. Sustainable development
2. Finance, investment and insurance
3. Changing behaviours
4. Growth and innovation
5. Policy design and evaluation
6. Governance and legislation

More information about the Grantham Research Institute is available at www.lse.ac.uk/GranthamInstitute

Suggested citation:
Quemin S (2020) *Using supply-side policies to raise ambition: the case of the EU ETS and the 2021 review*. Centre for Climate Change Economics and Policy Working Paper 366/Grantham Research Institute on Climate Change and the Environment Working Paper 335. London: London School of Economics and Political Science

Using Supply-Side Policies to Raise Ambition: The Case of the EU ETS and the 2021 Review

Simon Quemin*

Grantham Research Institute – *London School of Economics and Political Science*

Climate Economics Chair – *Paris-Dauphine University, PSL Research University*

This version: March 2020

[latest version]

Abstract

Unlike standard supply-side policies, the market stability reserve (MSR) can be used as a potent means of raising climate ambition. We calibrate an emissions trading model to the EU ETS and show that allowing firms to use rolling finite planning horizons can replicate past annual price and banking developments well compared to a standard infinite horizon, including the 2018 price rally. When firms have bounded foresight, indirectly raising ambition through the MSR is not equivalent to directly raising ambition through the emissions cap trajectory. Leveraging the MSR to raise ambition can be efficiency improving as it partially compensates for bounded foresight by frontloading abatement efforts so we analyze its interdependence with the cap trajectory to exploit synergies and minimize regulatory costs. Additionally, we quantitatively assess and compare changes in the MSR parameters for the 2021 review. In any case, MSR-induced resilience to demand shocks remains limited and one-sided by design.

Keywords Emissions trading, Rolling horizon, Raising ambition, Market stability reserve, EU ETS.

JEL classification D47, D81, G41, H32, Q58.

*S.Quemin@lse.ac.uk | London School of Economics, Houghton Street, London WC2A 2AE, UK.

1 Introduction

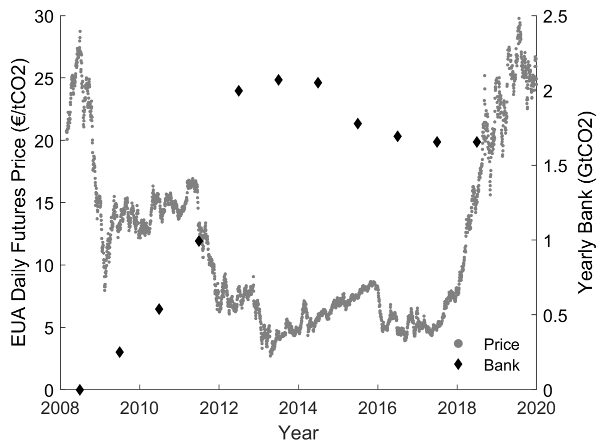
In emissions trading systems, supply-side policies typically introduce price steps in otherwise inelastic supply curves (e.g. Roberts & Spence, 1976; Weitzman, 1978) in order to constrain price variability and regulatory costs in the face of ex-ante uncertainty in demand for emissions permits and abatement costs (e.g. Fell et al., 2012; Borenstein et al., 2019). In 2018, the EU ETS regulators opted for a different, quantity-based control policy – the market stability reserve (MSR) and a companion cancellation mechanism (CM). Even though the purported stabilizing properties of this approach have already been called into question (e.g. Perino & Willner, 2016; Kollenberg & Taschini, 2019), the MSR and CM have the appealing potential to raise both short- and long-term ambition levels (e.g. Perino, 2018; Quemin & Trotignon, 2019). In fact, the recent fourfold price rally (Figure 1) is mainly attributable to this aspect of the 2018 market reform, which increased and brought permit scarcity back to the market earlier than what was originally anticipated, if at all, by market participants.

This potential is particularly important in a context where ambition needs to be ramped up to align with the EU commitments under the Paris Agreement and the objectives of the EU Green Deal (e.g. carbon neutrality by 2050). Imputing a higher ambition target to the ETS perimeter appears as an appealing policy option to that end, and the upcoming 2021 market review constitutes an adequate policy window to carry this out. Indeed, Parry (2020) finds that relying on more robust pricing in the ETS is the policy option yielding the largest welfare gains in raising ambition, and it is a Pareto improvement for Member States. Yet, his model does not capture the intricacies and specificities of the EU ETS, notably its intertemporal dimension and its interaction with the MSR and CM.

In this paper, we evaluate and compare the impacts of realistic regulatory changes to inform the 2021 review and raise ambition. The two main policy levers are an increase in the rate at which the emissions cap declines annually (the linear reduction factor, LRF) and enhancing the MSR through changes in its parameters.[1] Crucially, we will see that these levers are not equivalent in terms of increased stringency over time (as perceived by firms) and that they interact (as complements or substitutes). Both these aspects have not yet been addressed in the literature and we will explore synergies to reduce the costs of an ambition ramp-up.

[1] In the EU ETS, firms can bank (i.e. store) past and present vintage permits for future compliance usage. The MSR is a supply-side trigger mechanism adjusting current auctions a_t as a function of the bank 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. On top of that, the CM yearly cancels permits stored in the MSR in excess of auctions in the previous year. As a result, cumulative emissions are endogenous, i.e. no longer solely determined by the cap trajectory set by the LRF, and can only be reduced.

Figure 1: EUA daily futures prices and yearly bank (01 Jan 08 – 31 Dec 19)



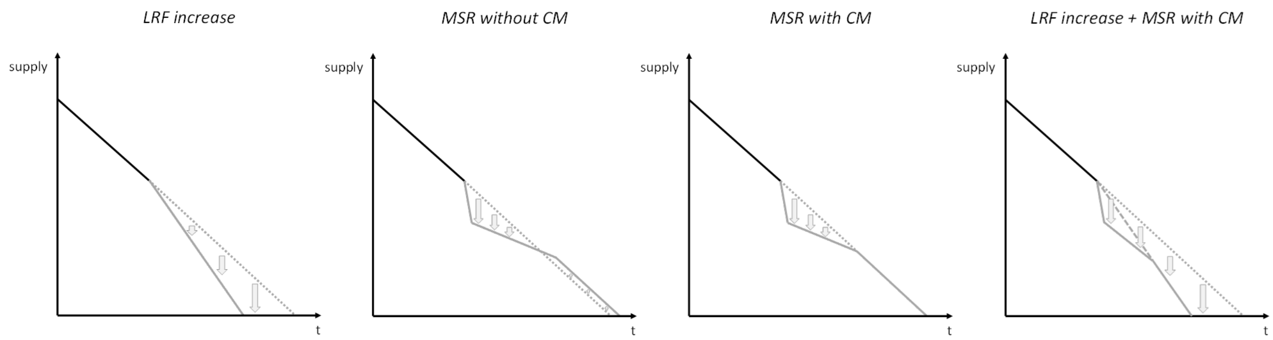
Note: EUA = EU Allowance. Data compiled from the IntercontinentalExchange and EU Transaction Log.

Specifically, we first develop a model of competitive intertemporal emissions permits trading in the presence of the MSR and CM under uncertainty about the future demand for permits. Building on Quemin & Trotignon (2019, henceforth QT19), a key novelty is that firms can use a rolling finite horizon procedure in their decision making as a way of dealing with uncertainty (e.g. Goldman, 1968; Spiro, 2014; Grüne et al., 2015; van Veldhuizen & Sonnemans, 2018).[2] Rolling horizons are able to (1) reconcile annual banking developments over 2008-2017 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 2008-2017 twice better than an infinite horizon, and (3) pick up the 2018 price rally in the wake of the market reforms where an infinite horizon falls short of it. In this respect, we augment the ex-post analysis of banking behavior in the US Acid Rain Program by Ellerman & Montero (2007) and apply it in the context of the EU ETS.

We interpret these results in the spirit of Friedman's (1953) black box approach. That is not to say that firms actually use rolling finite horizons, but this assumption has the comparative advantage of reproducing past market outcomes more satisfactorily than an infinite horizon. If market actors 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. Ideally, policy should be designed to account for these aspects in order to mitigate inefficient outcomes and associated undesirable outcomes, e.g. low prices prevailing early on may not spur investments in line with the long-term ambition target with a risk of lock-in (e.g. Fuss et al., 2018).

[2]See QT19 (Section 2) 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.

Figure 2: Stylized representation of LRF vs. MSR impacts on supply over time



In this context, the MSR has potential to make the long-term ambition target embedded in the cap trajectory more tangible in the short to mid term by frontloading abatement efforts. Coupled with the CM, it further increases long-term ambition. By contrast, higher scarcity induced by an LRF increase is more prevalent in the long term than in the short term. These properties are depicted in Figure 2. In fact, our simulations suggest that the 2018 price rally is largely attributable to the sizable MSR-induced supply cutback in the coming years rather than the companion LRF increase from 1.74 to 2.20%. Therefore, transitional stringency is as important as cumulative stringency for policy design and impacts.

Next, we utilize our calibrated model to evaluate various realistic options in revising the MSR parameters (see footnote 1) and the cap trajectory (LRF) to provide some guidance for the 2021 review. Because the MSR is a trigger mechanism, when the intake-release thresholds are constant over time as in the status quo, increasing the intake rate generates more volatility due to stronger oscillations around the thresholds without leading to larger cancellations and ambition. As a potent avenue for the review, combining thresholds that are declining over time (e.g. based on the LRF) with a higher intake rate can keep induced oscillations in check, thereby leading to higher prices and ambition without destabilizing the market.

Two policy levers can ramp up ambition: enhancing the MSR and increasing the LRF. They can be utilized hand in hand but because they have different impacts in terms of perceived scarcity for firms, the policy mix should be chosen to ensure complementarity and minimize regulatory costs. Our analysis of the LRF-MSR 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. This corresponds to the LRF-MSR combination that involves the smallest LRF increase. Indeed, as the MSR partially compensates for bounded foresight by frontloading abatement efforts, leaving more traction to an enhanced MSR than to an LRF increase can be efficiency-improving.

In spite of this and even after changes in its parameters, the ability of the MSR to improve the market resilience to demand shocks remains limited and one-sided by design. The MSR is essentially geared toward supply contraction and can limitedly respond to shocks in the short and longer terms alike. Therefore, the MSR and CM should be thought of as a potent, efficiency-improving though indirect means of raising ambition rather than as a more standard supply-side instrument à la Roberts & Spence (1976) and Weitzman (1978). This also implies that the complementary introduction of such an instrument could be contemplated, e.g. in the form of a price floor (e.g. Newbery et al., 2019; Flachsland et al., 2020).

Since the early contributions by Fell (2016) and Perino & Willner (2016), the literature has recently witnessed a flurry of papers on the MSR and related aspects – inter alia Perino & Willner (2017), Chaton et al. (2018), Beck & Kruse-Andersen (2019), Bocklet et al. (2019), Bruninx et al. (2019), Carlén et al. (2019), Gerlagh et al. (2019), Gerlagh & Heijmans (2019), Kollenberg & Taschini (2019), Mauer et al. (2019), Quemin & Trotignon (2019) and Tietjen et al. (2019). 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.

Our paper contributes to this literature on three fronts. First, our model has the comparative advantage of a finer-grained calibration. Specifically, if and where others calibrate their models by fitting parameters so as to replicate the market price the year before their simulations, our calibration is based on both annual price and banking observations over 2008-2017. On top of that, a rolling horizon offers a better calibration (both in and out of sample) compared to the infinite (or similar) horizon typically assumed in other models. Second, our paper is the first to propose a thorough quantitative assessment of realistic regulatory changes for the 2021 review, where others evaluate the consequences of the 2018 reform (i.e. keeping MSR design unchanged). Third, our paper provides a detailed analysis of the hitherto unexplored LRF-MSR interaction, which is pivotal in exploiting synergies when raising ambition.

The remainder is structured as follows. Section 2 sets forth our model of emissions trading with infinite vs. rolling finite horizons and its calibration. 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 further details on the calibration.

2 Model

In this section, we first describe our emissions trading model in the presence of the MSR and CM, where firms can use rolling finite horizons to deal with uncertain future supply-demand conditions. We next calibrate the model to annual banking and price outcomes over 2008-2017 and compare the relative merits of infinite and rolling horizons in their ability to reproduce these observed dynamics in sample. Building on QT19, this section is as parsimonious as can be and only provides key building blocks and insights.

2.1 Description

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 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 & 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).

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. 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 & 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 & 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 unregulated (i.e. baseline) emissions in year t, respectively. End-of-pipe abatement $u_t - e_t \geq 0$ is costly and we let C_t denote its abatement cost function in year t with C'_t and $C''_t > 0$. In year t, u_t, f_t, a_t and o_t[3] as well as the state variable b_{t-1} are given to the firm. It selects 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), where the constraint on the bank ensures that cumulative supply equals cumulative emissions (i.e. overall market clearing).

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

Rolling horizons. To address uncertainty the firm can choose to use a rolling finite horizon procedure (e.g. Goldman, 1968; Spiro, 2014; Grüne et al., 2015; van Veldhuizen & Sonnemans, 2018). With horizon $h \geq 0$, the firm selects year-t emission e_t and bank b_t by solving

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

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

where $\hat{x}_\tau^t$ denotes the year-t forecast for $x = \{u, f, a, o\}$ in year $\tau \geq t$. Under 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 and the 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 QT19, 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 with a given discount rate or by the infinite horizon with a higher discount rate are observationally equivalent. This equivalence, however, breaks down in the presence of supply-side control.

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[4]

- if $B_t > IT_t$ then $a_t \leftarrow \max\{a_t - IR_t \times B_t; 0\}$ and $s_{t+1} = s_t + \min\{IR_t \times B_t; a_t\}$,
- else if $B_t < RT_t$ then $a_t \leftarrow a_t + \min\{RQ_t; s_t\}$ and $s_{t+1} = \max\{s_t - RQ_t; 0\}$,
- else the MSR is inactive.

[4]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.

The MSR endogenizes the auction schedule $\{a_t\}_t$, i.e. it rearranges annual auctions over time as a function of (past and future) market outcomes but in principle leaves cumulative supply as defined by the cap trajectory unchanged (Perino & Willner, 2016).

In addition, the cancellation mechanism (CM) may further adjust the MSR stock as follows

- $s_t \leftarrow s_t - \max\{s_t - L_t; 0\}$,

where L_t is the maximum number of permits allowed in the MSR in year t. In words, the CM shaves off all reserve permits in excess of a predefined upper bound from the MSR stock 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 & Prescott, 1971). At each step in this recursive procedure, the firm holds beliefs about future control impacts and supply schedules, optimizes w.r.t. its beliefs, and updates them based on the resulting control impact stream. The equilibrium obtains 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.

Current parametrization. In the status quo, i.e. pursuant to the 2018 reform (European Parliament & 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, specifically 0.9 billion ad-hoc backloaded permits (European Commission, 2014) and an estimated 0.65 billion unallocated Phase-III permits.[5] As per the CM, the reserve is capped by realized auctions in the previous year and 'excess permits' are cancelled from 2023 onward, that is

[5] 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).

- $L_t = \infty$ 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 itself determined by current MSR actions and anticipated future MSR and CM actions. Given the MSR and CM parameters and the initial 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 Section 3.1, these amounts hinge on firm's behavior, e.g. its planning horizon. Finally, since the reform was finalized in late 2017 and enacted in early 2018, we consider that the impacts of the MSR and CM are anticipated and factored in by the firm from 2018 onward.

2.2 Calibration

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 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.[6] 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. Figure 3a depicts the reconstructed and projected paths for these three indexes between 1990 and 2050, and Figure 3b 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 so that the year-t forecast for baseline emissions in year $t+1$ is defined by

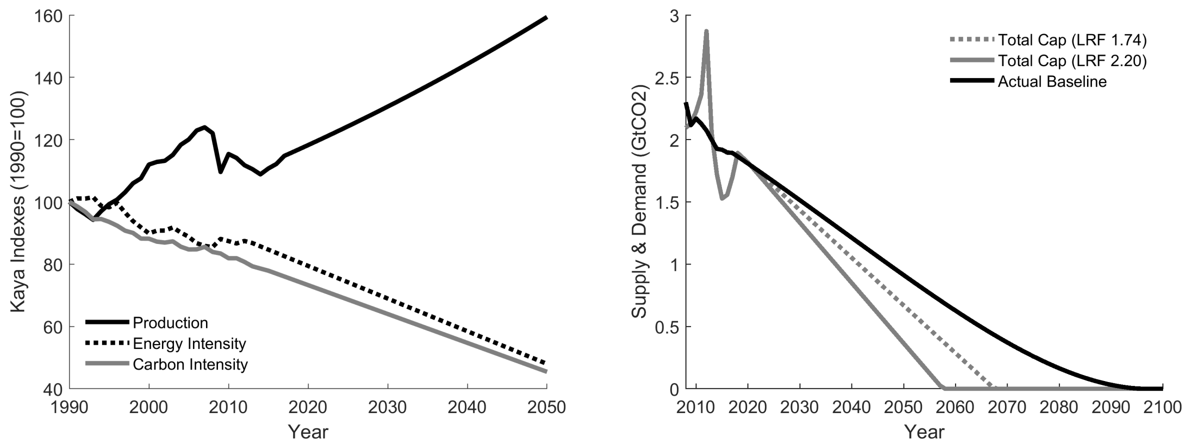
$$\hat{u}^t_{t+1} = \varphi(1+\gamma_t)u_t + (1-\varphi)\bar{u}^t_{t+1}, \tag{4}$$

where u_t is the realized baseline in year t, $\varphi \in [0; 1]$ captures some persistence, γ_t is the annual

[6] For production, our assumption is tenable as there is no evidence of carbon leakage (e.g. Naegele & Zaklan, 2019). For renewables, it is supported by evidence that permit-price equivalents of renewable subsidies have been significantly higher than market prices (e.g. Marcantonini & Ellerman, 2015; Abrell et al., 2019).

Figure 3: Kaya indexes, baseline emissions and total cap on emissions

(a) Kaya Indexes (b) Supply & Demand



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.

growth rate as expected in year t, and the trend $\bar{u}$ can vary over time. Specifically, the trend in year t for some future year $t' > t$ ($\bar{u}^t_{t'}$) is in line with the achievements of the prevailing Climate Energy Package in year t. We set $\varphi = 0.9$ as in Fell (2016) and γ_t in line with GDP growth forecasts by the European Commission over 2008-2020, with a 1% p.a. growth rate afterward. Since actual baselines u_t in Figure 3b differ from their forecasts $\hat{u}^{\tau<t}_t$ in (4), the firm adjusts its demand forecasts and thus 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), i.e. 1.74% before and 2.2% after the 2018 reform (from 2021 on, 57% of the cap is auctioned off). The grey line in Figure 3b depicts total annual supply $\{f_t + a_t + o_t\}_t$.[7] The peak in 2011-12 is due to a massive offsets use, totalling about 1 $GtCO_2$. The following 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.

[7]Kyoto offsets are authorized only over 2008-2020 up to an overall limit $O \approx 1.6$ $GtCO_2$. We assume that offset usage is exogenous to the firm, i.e. o_t is given in year t and equal to observed offset usage. In year t, the firm forecasts that the remainder (i.e. $O - \sum_{\tau=2008}^{\tau=t} o_\tau$) is equally split across the remaining years.

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 to exploit the regulatory change offered by the 2018 reform to compare how in-sample calibrated infinite and rolling horizons fare out of sample, i.e. in capturing the observed 2018 regime shift and price rally (Section 3.1).

Our calibration methodology is described in Appendix A. We use a two-step procedure in the spirit of standard least squares maximum likelihood estimations with one free parameter. In the fist step, we select the couple (r, h) to minimize annual deviations between the simulated and observed bank levels over 2008-2017. In the 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. With an infinite horizon $h = \infty$, a discount rate $r \approx 8\%$ best replicates past banking. This aligns with standard rates of return on risky assets (Jordà et al., 2019) but is in the higher range of the rates inferred from futures' yield curves (Appendix A). With a central such rate $r = 3\%$, a rolling horizon of $h = 11$ years yields a similar banking fit. That is, the rolling horizon can reconcile past bank dynamics with implicit discount rates where the infinite horizon cannot. Moreover, the price fit obtained with the rolling horizon is more than twice as good (in size and sign) than with the infinite horizon.

In the absence of conclusive evidence on how forward-looking firms plan, we use an approach in the spirit of Friedman's (1953) black box model and compare the merits of two alternative assumptions (infinite vs. rolling horizons) in how well replicate 2008-2017 market outcomes. Our results lend more support to the latter but do not imply that market actors actually use a rolling 11-year horizon with a 3% discount rate. For instance, our representative firm model is blind to prevailing heterogeneity in behavior and risk preferences. Rather, our calibration results should be taken as first-pass assessments at the market level, which will prove crucial for policy design and ex-ante evaluation as the next sections show.

3 Simulations

In this section, we first compare simulated market outcomes with the calibrated infinite and rolling finite horizons out of sample in the status quo (i.e. post 2018 reform). We next 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 (Section 2.2) and out of (Section 3.1) sample, we will focus on this case to limit the number of scenarios.[8]

3.1 Assessing the market impacts of the 2018 reform

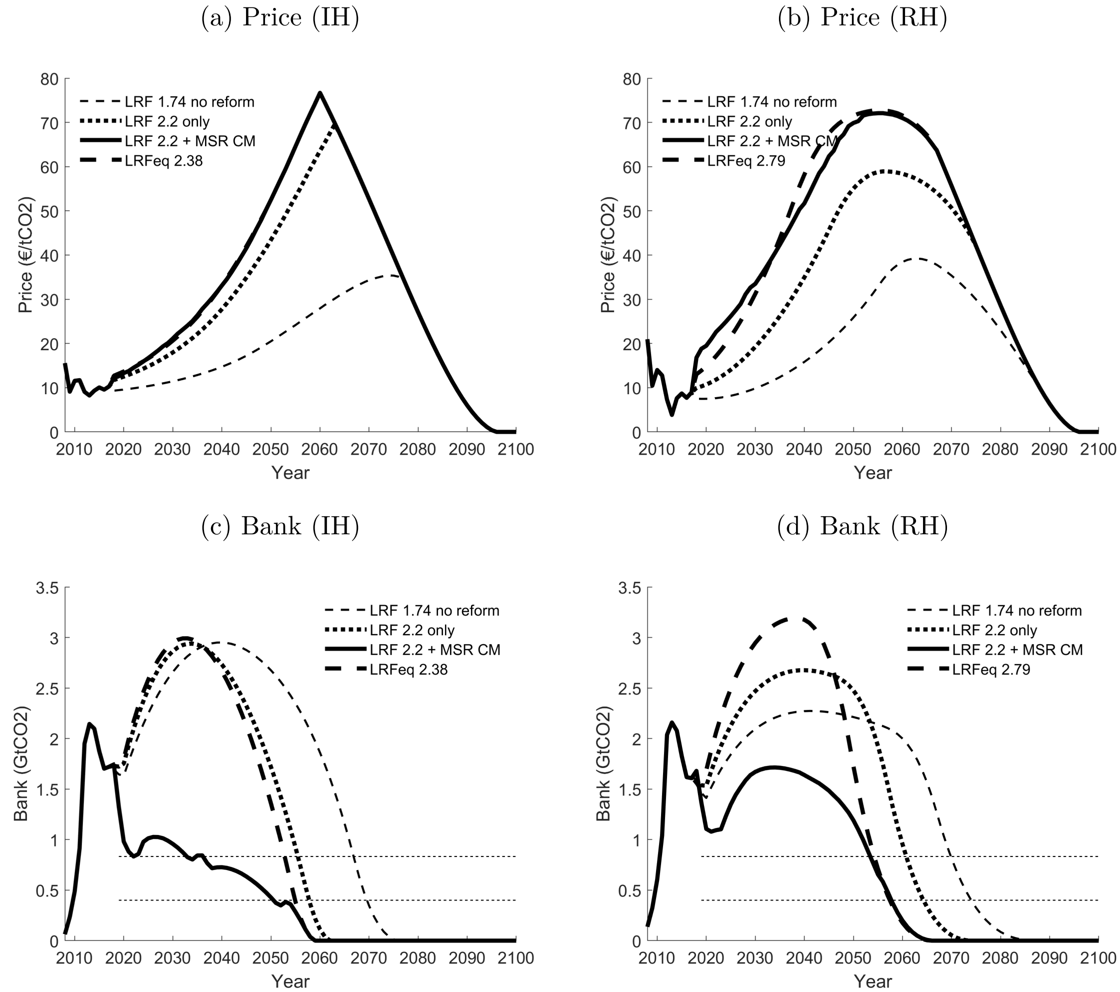
We compare market outcomes generated by the calibrated infinite and rolling horizons out of sample by exploiting the regime shift induced by the 2018 reform (an LRF increase from 1.74 to 2.2% in 2021, a reinforcement of the MSR and the introduction of the CM). The left (right) hand side of Figure 4 depicts the equilibrium price and bank paths with the infinite (rolling) horizon.[9] Relative to no reform, the sole LRF increase leads to higher prices and a shorter banking period, albeit with higher banked volumes early on. With the infinite horizon, prices reach a peak when the bank becomes empty. This is also when emissions become nil since the cap has already shrunk to zero (in 2058 with a 2.2% LRF) due to intertemporal optimization. After the peak, there are no permits left in circulation so the firm can no longer emit and has to abate its baseline emissions. As baseline emissions gradually decline to zero, so do the yearly abatement efforts and associated costs at the margin, hence the declining price paths. We observe similar though less clear-cut patterns with the rolling horizon.

A rolling finite horizon fares better than an infinite horizon out of sample

Introducing the MSR and CM on top of the 2.2% LRF further hikes the price and reduces the bank. While this is true for both the infinite and rolling horizons, specific market outcomes differ noticeably: price and bank levels are both higher with the latter than with the former. Importantly, the rolling horizon captures the observed 2018 price rally (regime shift), where the infinite horizon falls short of it – the price rise is more than four times bigger with the

[8] See QT19 for a quantitative assessment of market outcomes under infinite vs. rolling finite horizons.
[9] Prices are in current Euros, using observed inflation rates over 2008-2018 and 1.5% p.a. afterward.

Figure 4: Price and bank paths with calibrated infinite vs. rolling horizons



Note: Release-intake thresholds set at 400-833 $MtCO_2$ constant over time. Intake rate of 24% over 2019-2023 and 12% afterward. LRF_{eq} is the LRF that generates the same 2008-2100 cumulative emissions on its own as those generated by a 2.2% LRF with the MSR+CM.

rolling horizon than with the infinite horizon. Hence, in addition to better in-sample price and discount rate fits, the rolling horizon also yields a better price fit out of sample.

The underlying mechanism is the following (see also Section 3.3). 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 bank levels than with the infinite horizon. In addition, as the firm foresees a sizable supply tightening over its horizon, it drives up abatement and banking, which in turn inflates future MSR intakes. This raises the firm's overall abatement forecast, which leads to higher banking and future MSR intakes, and so forth.[10] As a result, MSR intakes are larger (due to higher banking) and last longer (the intake cut-off year occurs two decades later) with the rolling horizon. This translates into twice as large cancellations with the rolling horizon, i.e. 8.7 vs. 4.2 $GtCO_2$ with the infinite horizon.

Transitional stringency is as important as cumulative stringency

In a context where market actors tend to focus more on the short to mid term than on the long term, the system's transitional stringency is as important as its cumulative stringency, if not more. Importantly, the MSR has potential to increase both transitional and cumulative stringency. The former aspect occurs as the MSR postpones some auctions, i.e. it de facto frontloads abatement efforts and makes longer-term scarcity more tangible for agents with bounded foresight. The latter aspect occurs as the CM cancels the bulk of MSR-withdrawn permits. In comparison, 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. These properties are depicted on Figure 2.

We illustrate these properties with two examples. First, we simulate what the price response would have been had the 2018 reform only comprised an LRF increase from 1.74 to 2.2% (no MSR). Although the induced reduction in cumulative supply is substantial (9 GtCO2), the price response is less than commensurate with the rolling horizon (the impact on transitional scarcity is less marked). Specifically, our calibrated model suggests that the price would have increased to 10€/tCO_2 only (as compared to about 20€/tCO_2). In our model, this represents 13% of the price rise witnessed in conjunction with the MSR and CM. This is because the MSR brings into present times – as well as augments through the CM – the scheme's long-term scarcity. By contrast, with the infinite horizon, the LRF increase represents 80% of the simulated 2018 price rise obtained in conjunction with the MSR and CM.

Second, to further get a sense of why transitional stringency is key when firms are boundedly farsighted, we compare price paths with those obtained under a sole LRF increase yielding the same 2008-2100 cumulative emissions as under a 2.2% LRF with MSR and CM. These equivalent LRF, or LRF_{eq}, are of 2.38 and 2.79% under the infinite and rolling finite horizons, respectively. With the rolling horizon, as Figure 4b 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

[10] This self-reinforcing effect gradually subsides and stops because $IR < 1$.

and MSR+CM. By contrast, price paths are almost identical with the infinite horizon because the MSR-driven auction backloading is essentially irrelevant in this case (Figure 4a).[11]

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 parameters, we do not quantify their interactions to limit the number of scenarios. However, how MSR 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.[12] 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.[13] 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[14] 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 (Figure 5b), in turn leading to more erratic streams of yearly auctions and MSR intakes (Appendix B).

As oscillations in the bank and annual supply are transmitted to prices, a higher intake rate

[11] Negligible differences occur due to changes in the stringency of the limited borrowing constraint. Indeed, a higher LRF implies reduced free allocations, thus less borrowing opportunities. This affects intertemporal arbitrage and hence market equilibrium paths at the margin. See Appendix F for a related discussion.

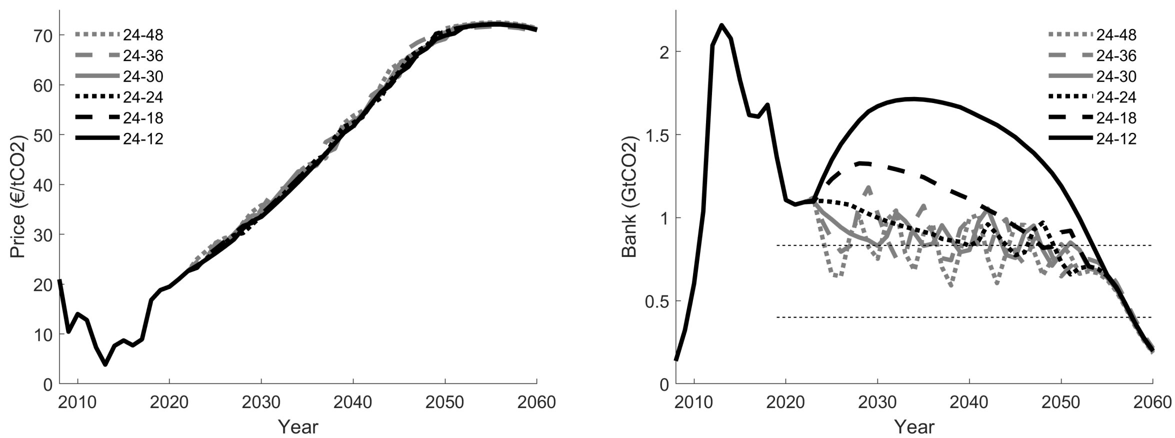
[12] See QT19 for a quantitative assessment of the MSR impacts with and without the CM.

[13] 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 $MtCO_2$.

[14] 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.

Figure 5: Different intake rates from 2024 on with thresholds constant over time

(a) Price paths (b) Bank paths



Note: Release-intake thresholds set at 400-833 $MtCO_2$ constant over time. Intake rate of 24% over 2019-2023 and 12, 18, 24, 30, 36 or 48% afterward.

is conducive to larger price variability with a negligible price increase on average (Figure 5a). 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 (Appendix B).

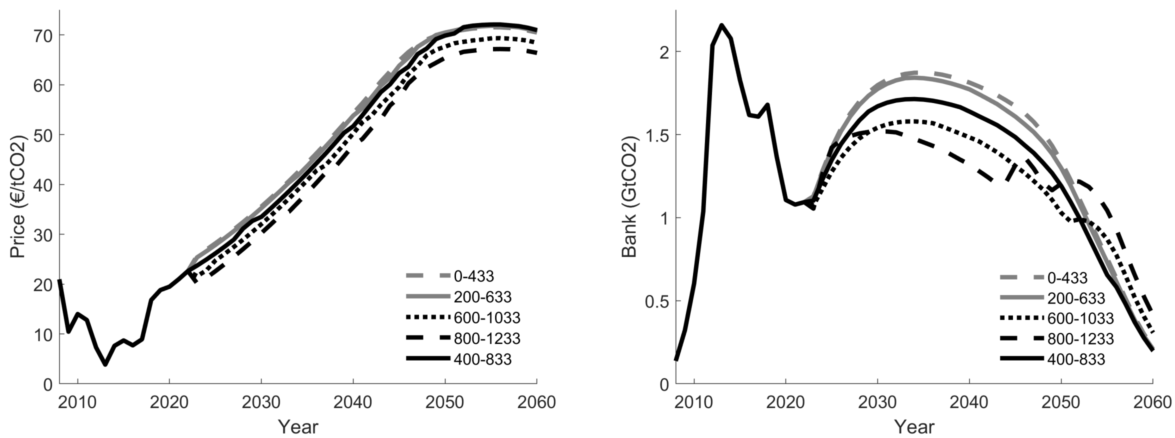
With the current thresholds (400-833 $MtCO_2$ constant over time), a higher intake rate than in the status quo (12% from 2024 on) does not increase ambition and tends to destabilize the market by making annual supply bumpier and prices more volatile. In practice, as a result of more pronounced threshold effects, a higher intake rate makes future supply conditions harder to gauge for market participants. One may also argue that this could make the MSR more prone to manipulation or strategic arbitrage not related to fundamentals.

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

Figure 6: Different threshold positions from 2024 on with given intake rates

(a) Price paths (b) Bank paths



Note: Release-intake thresholds set at 400-833 $MtCO_2$ over 2019-2023; their position varies afterward keeping a breadth of 433 $MtCO_2$ constant. Intake rate of 24% over 2019-2023 and 12% afterward.

433 and 1233 $MtCO_2$ (Figure 6a), 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.[15] 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.[16]

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 to 433) $MtCO_2$. Decreasing returns obtain 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.

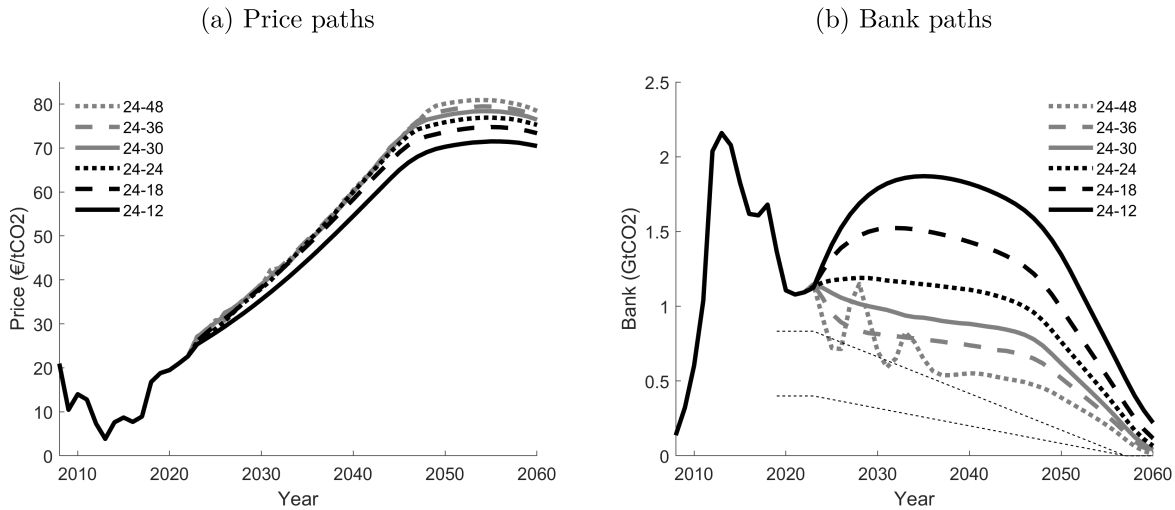
Together with the intake rate, the position of the intake threshold is a pivotal policy handle. In comparison, the position of the release threshold has a negligible bearing on market and ambition outcomes.[17] This is because reinjections may only occur when the release threshold

[15] 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.

[16] This self-reinforcing anticipation effect has been discussed in Section 3.1. Forward-looking anticipation of MSR impacts always occurs but is less visible graphically through banking when intake rates differ.

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

Figure 7: Linearly declining thresholds with different intake rates from 2024 on



Note: Release-intake thresholds set at 400-833 $MtCO_2$ over 2019-2023, which then linearly decline to 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.

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

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).[18]

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 Figures 5 and 7: 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

[18] A similar argument is that thresholds should be adjusted to the volume of the market. Because the LRF dictates how the annual supply volume evolves over time, it could readily be used for the thresholds.

and auction streams are more regular, see Appendix B). Oscillations may still 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 Figure 7a: 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 Figure 8).

3.3 Raising ambition through the LRF and the MSR

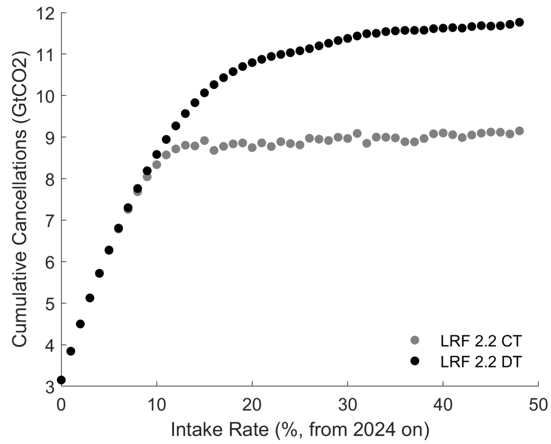
There are two ways of increasing climate ambition within the ETS perimeter: increasing the LRF and leveraging the MSR and CM. Because they are not equivalent in terms of perceived supply impacts and associated abatement decisions when firms have bounded foresight, synergies can be exploited by utilizing 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 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 (Figure 8). 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-

Figure 8: 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 fixed at 2.2%. CT (DT) indicates that release-intake thresholds are set at 400-833 $MtCO_2$ and constant over time (over 2019-2023 and then linearly declining to reach 0 in 2058, the year the cap becomes nil given the 2.2% LRF).

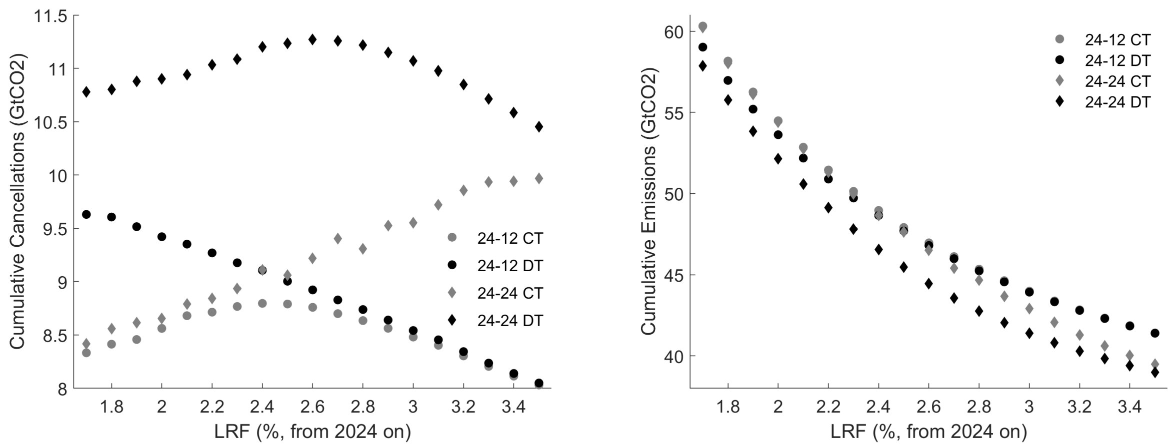
2023 intake rate from 12 to 24% leads to +1.8 $GtCO_2$ cancellations with declining thresholds compared to +0.1 $GtCO_2$ with constant thresholds. Additionally, oscillatory behavior and volatility are mitigated (see Section 3.2 and Appendix B).

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 Figure 9 displays both total cancellations and emissions. Figure 9a 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,

Figure 9: LRF-MSR interaction as a function of the LRF

(a) Cumulative cancellations (b) Cumulative emissions



Note: Intake rate of 24% over 2019-2023 and of 12 or 24% afterward. The post-2023 LRF is varied along the x-axis. CT (DT) indicates that release-intake thresholds are set at 400-833 $MtCO_2$ and constant over time (over 2019-2023 and then linearly declining to reach 0 in the same year as the cap given the prevailing LRF).

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 Figure 9a, 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.

Finally, Figure 9b 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 Figure 3b, 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 QT19). 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.[19] 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 $GtCO_2$ (Figure 9a).

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

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)	⋆ (2.466)	⋆ (2.366)	50.9
	24-24	⋆ (2.647)	⋆ (2.539)	⋆ (2.450)	⋆ (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-intake thresholds fixed at 400-833 $MtCO_2$ constant over time, 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 ⋆ indicates the lowest cost combination to attain a given target, and percentages measure the additional NPV of compliance costs for the other possible combinations over the course of the program. 'Ref' indicates cumulative emissions in the status quo, namely with a 2.2% LRF and the given MSR parametrizations.

3.4 Assessing the MSR-induced resilience to demand shocks

The first objective of the MSR (and CM) is to eliminate past oversupply w.r.t. dampened and 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 hence to improve market resilience should similar 'supply demand imbalances' materialize in the future.

We evaluate the ability of the MSR to respond to demand shocks in terms of induced changes in the supply schedule and resulting immediate and longer-term price responses with an illustrative symmetric example. Specifically, we consider an unanticipated permanent positive or negative shock (±150 $MtCO_2$) on demand occurring in 2025.[20] Our approach to introducing shocks is hence similar to that in Perino & Willner (2016) but simpler than the full-blown analysis in Fell (2016). Figure 10 compares MSR-sustained price paths with those without the shocks and those obtained with a sole equivalent 2.8% LRF_{eq}.

[20] 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.

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

We begin with the negative shock (Figure 10a) 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.[21]

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 (Appendix C).[22]

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 QT19). 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 & Laroque, 1992)

Whatever the direction of the shock, the MSR offers a one-sided response

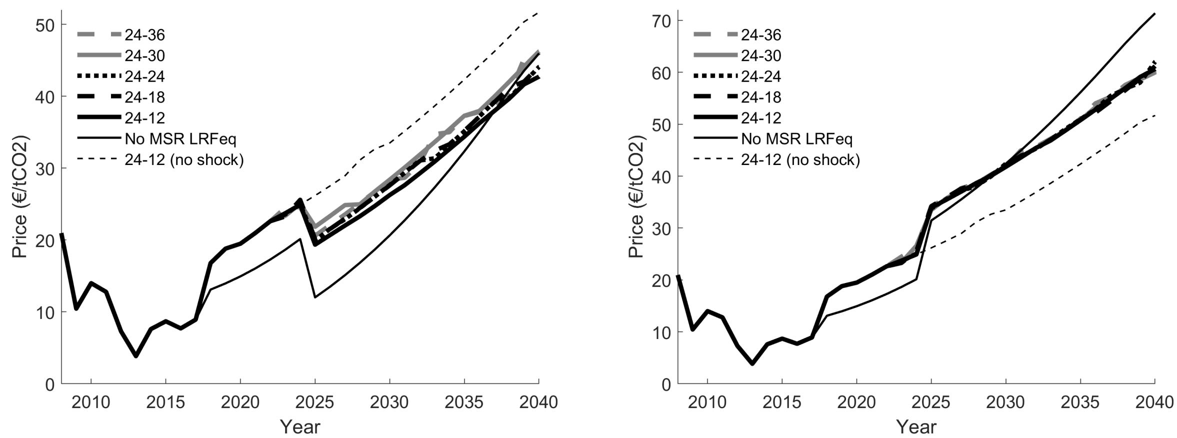
We now turn to the positive shock and observe similar patterns. In the year of the shock, the price jump is contained by 20 up to 39% w.r.t. no MSR (which is less than for the negative shock) depending on the intake rate, again non monotonically. Similarly, there is no sign of

[21] 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 (Figure 5b), which reduces (anticipated future) intakes and thus the buffer.

[22] 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.

Figure 10: Price impacts of permanent demand shocks with and without MSR

(a) Negative shock (b) Positive shock



Note: Unanticipated permanent shocks on yearly permit demand from 2025 onward: negative (-150 $MtCO_2$, left); positive (+150 $MtCO_2$, right). Release-intake thresholds of 400-833 $MtCO_2$ constant over time. Intake rate of 24% over 2019-2023, and of 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).

reversion over time as MSR-driven price paths remain parallel to that with no shock. This is because the MSR continues to cut back on supply despite the positive shock, albeit at a lesser extent as yearly intakes are slightly reduced (Appendix C), and never releases permits into the market.[23] 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 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 oversupply, the MSR has been engineered for supply contraction, not expansion. Irrespective of the shock structure, the MSR always cuts back

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

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-intake thresholds of 400-833 $MtCO_2$ constant over time. Intake rate of 24% over 2019-2023, and of 12, 18, 24, 30 or 36% afterward. 'Δ Cancellations' and 'Δ 2030 price' are measured relative to no shock in $GtCO_2$ and € per tCO_2, respectively.

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 purposes of improving market resilience.

4 Conclusion

In a first step, we build a model of emissions permits trading tailored to the EU ETS which features the market stability reserve (MSR) and its companion cancellation mechanism (CM). A pivotal difference with the literature is that firms can employ rolling finite horizons in their decision making as a procedure to deal with uncertainty about the future demand for permits. Rolling horizons are found to (1) reconcile annual banking developments over 2008-2017 with discount rates derived from futures' yield curves, (2) replicate average annual prices over 2008-2017 twice better than a standard 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 the first 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 regulatory 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 potential to permanently curb supply via the CM, it could be utilized hand in hand with an LRF increase to ramp up ambition. But because these two policy levers have different repercussions in terms of perceived scarcity for boundedly farsighted firms, the policy mix should be chosen to guarantee complementarity 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. This is the LRF-MSR mix which leverages an LRF increase the least. Indeed, as the MSR partially compensates for bounded foresight by frontloading abatements, leaving more traction to an enhanced MSR than to the LRF can be efficiency-improving.

In spite of this and even after changes in its parameters, the ability of the MSR to improve the market resilience to demand shocks remains limited and one-sided by design. The MSR is geared toward supply contraction and weakly responds to shocks in the short and longer terms. Whatever the shock, the MSR always cuts back on supply and the CM cancels withdrawn permits later on – what changes is the size and duration (albeit to a lesser extent) of the MSR intakes. Implementing declining thresholds cannot overcome that inherent limitation – rather, this would amplify the MSR asymmetry in responding to positive vs. negative demand shocks. To get around this, one could contemplate flanking the MSR with another supply-side policy like a price corridor (e.g. Newbery et al., 2019; Flachsland et al., 2020).

Acknowledgments

The author thanks seminar attendees at the College of Europe (Bruges), Potsdam Institute for Climate Impact Research (PIK) and European University Institute (FSR Climate 2019) for their comments. Special thanks go to Raphaël Trotignon. The usual disclaimers apply.

The author acknowledges funding from 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, the

Grantham Foundation for the Protection of the Environment, the UK Economic and Social Research Council (ESRC) through the Centre for Climate Change Economics and Policy, and the Climate Economics Chair (Paris-Dauphine University, PSL Research University).

References

Abrell, J., Kosch, M. & Rausch, S. (2019). Carbon Abatement with Renewables: Evaluating Wind and Solar Subsidies in Germany and Spain. *Journal of Public Economics*, **169**, 172–202.

Baudry, M., Faure, A. & Quemin, S. (2020). *Emissions Trading with Transaction Costs*. Working Paper, Grantham Research Institute, London School of Economics.

Beck, U. & Kruse-Andersen, P. (2019). *Endogenizing the Cap in a Cap-and-Trade System: Assessing the Agreement on EU-ETS Phase 4*. Arbejdspapir Revised, De Økonomiske Råd.

Bocklet, J., Hintermayer, M., Schmidt, L. & Wildgrube, T. (2019). The Reformed EU ETS – Intertemporal Emission Trading with Restricted Banking. *Energy Economics*, **84**.

Böhringer, C., Löschel, A., Moslener, U. & Rutherford, T. F. (2009). EU Climate Policy up to 2020: An Economic Impact Assessment. *Energy Economics*, **31**, S295–305.

Böhringer, C., Rosendahl, K. E. & Storrøsten, H. B. (2017). Robust Policies to Mitigate Carbon Leakage. *Journal of Public Economics*, **149**, 35–46.

Borenstein, S., Bushnell, J., Wolak, F. A. & Zaragoza-Watkins, M. (2019). Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design. *American Economic Review*, **109**(11), 3953–77.

Bruninx, K., Ovaere, M. & Delarue, E. (2019). *The Long-Term Impact of the Market Stability Reserve in the EU Emission Trading System*. TME Working Paper, KU Leuven.

Bushnell, J. B., Chong, H. & Mansur, E. T. (2013). Profiting from Regulation: Evidence from the European Carbon Market. *American Economic Journal: Economic 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**, 816–25.

Chaton, C., Cretì, A. & Sanin, M.-E. (2018). Assessing the Implementation of the Market Stability Reserve. *Energy Policy*, **118**, 642–54.

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 Journal*, **41**(4).

Deaton, A. & Laroque, G. (1992). On the Behaviour of Commodity Prices. *Review of Economic Studies*, **59**(1), 1–23.

Ellerman, A. D., Marcantonini, C. & Zaklan, A. (2016). The European Union Emissions Trading System: Ten Years and Counting. *Review of Environmental Economics & 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 Journal*, **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 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). Comparing Policies to Confront Permit Over-Allocation. *Journal of Environmental Economics & Management*, **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. *Journal of Environmental Economics & Management*, **64**(2), 183–98.

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. *Climate Policy*, **20**(1), 133–42.

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

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. *Review of Environmental Economics & Policy*, **12**(2), 220–41.

Gerlagh, R., Heijmans, R. J. R. K., & Rosendahl, K. E. (2019). *Endogenous Emission Caps Always Induce a Green Paradox*. Working Paper 7862, CESifo.

Gerlagh, R. & Heijmans, R. J. R. K. (2019). Climate-Conscious Consumers and the Buy, Bank, Burn Program. *Nature Climate Change*, **9**, 431–3.

Goldman, S. M. (1968). Optimal Growth and Continual Planning Revision. *Review of Economic Studies*, **35**(2), 145–54.

Grüne, L., Semmler, W. & Stieler, M. (2015). Using Nonlinear Model Predictive Control for Dynamic Decision Problems in Economics. *Journal of Economic Dynamics & Control*, **60**, 112–33.

Hintermann, B., Peterson, S. & Rickels, W. (2016). Price and Market Behavior in Phase II of the EUETS: A Review of the Literature. *Review of Environmental Economics & Policy*, **10**(1), 108–28.

Hotelling, H. (1931). The Economics of Exhaustible Resources. *Journal of Political Economy*, **39**(2), 137–75.

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M. & Taylor, A. M. (2019). The Rate of Return on Everything, 1870–2015. *Quarterly Journal of Economics*, **134**(3), 1225–98.

Kollenberg, S. & Taschini, L. (2019). Dynamic Supply Adjustment and Banking under Uncertainty in an Emissions Trading Scheme: The Market Stability Reserve. *European Economic Review*, **118**, 213–26.

Landis, F. (2015). *Final Report on Marginal Abatement Cost Curves for the Evaluation of the Market Stability Reserve*. Dokumentation 15-01, ZEW Mannheim.

Lucas, R. E. & Prescott, E. C. (1971). Investment Under Uncertainty. *Econometrica*, **39**(5), 659–81.

Marcantonini, C. & Ellerman, A. D. (2015). The Implicit Carbon Price of Renewable Energy Incentives in Germany. *Energy Journal*, **36**(4), 205–39.

Mauer, E.-M., Okullo, S. J. & Pahle, M. (2019). *Evaluating the Performance of the EU ETS*

MSR. Working Paper, Postdam Insitute for Climate (PIK).

Mehling, M. A., van Asselt, H., Das, K., Droege, S. & Verkuijl, C. (2019). Designing Border Carbon Adjustments for Enhanced Climate Action. *American Journal of International Law*, **113**(3), 433–81.

Montgomery, W. D. (1972). Markets in Licenses and Efficient Pollution Control Programs. *Journal of Economic Theory*, **5**(3), 395–418.

Naegele, H. & Zaklan, A. (2019). Does the EU ETS Cause Carbon Leakage in European Manufacturing? *Journal of Environmental Economics & Management*, **93**, 125–47.

Newbery, D. M., Reiner, D. M. & Ritz, R. A. (2019). The Political Economy of a Carbon Price Floor for Power Generation. *Energy Journal*, **40**(1).

Parry, I. (2020). Increasing Carbon Pricing in the EU: Evaluating the Options. *European Economic Review*, **121**.

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

Perino, G. & Willner, M. (2016). Procrastinating Reform: The Impact of the Market Stability Reserve on the EU ETS. *Journal of Environmental Economics & Management*, **80**, 37–52.

Perino, G. & Willner, M. (2017). EU ETS Phase IV: Allowance Prices, Design Choices and the Market Stability Reserve. *Climate Policy*, **17**(7), 936–46.

Quemin, S. & Trotignon, R. (2019). *Emissions Trading with Rolling Horizons*. GRI & CCCEP Working Paper 316, London School of Economics.

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

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

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

Spiro, D. (2014). Resource Prices and Planning Horizons. *Journal of Economic Dynamics & Control*, **48**, 159–75.

Tietjen, O., Lessmann, K. & Pahle, M. (2019). *Hedging and the Temporal Permit Issuance in Cap-and-Trade Programs*. Working Paper, Postdam Institute for Climate (PIK).

van Veldhuizen, R. & Sonnemans, J. (2018). Nonrenewable Resources, Strategic Behavior and the Hotelling Rule: An Experiment. *Journal of Industrial Economics*, **66**, 481–516.

Venmans, F. (2016). The Effect of Allocation Above Emissions and Price Uncertainty on Abatement Investments under the EU ETS. *Journal of Cleaner Production*, **126**, 595–606.

Weitzman, M. L. (1978). Optimal Rewards for Economic Regulation. *American Economic Review*, **68**(4), 683–91.

Appendices

A Calibration to 2008-2017 price and banking data

We consider that permit demand is linear in the permit price, which is a standard assumption (e.g. Ellerman & Montero, 2007; Kollenberg & Taschini, 2019). We thus assume that $C_t''' \approx 0$, which can be viewed as a local Taylor approximation of more general functional forms. We also assume that the slope of the linear marginal abatement cost functions is time invariant, i.e. $C_t'' = c$ 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. Nevertheless, the linear intercept is gradually lowered over time as the actual baseline path is downward sloping (Figure 3b).

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 2008-2017. In the second step, given r and h, we select c so that the simulated price path deviates the least from the yearly-averaged spot price path over 2008-2017. In each step, the free parameter is calibrated by minimizing the distance between simulated and observed paths. Table A.1 reports the best-fit results and Figure A.1 depicts the observed and best-fit simulated paths over 2008-2017 for visual comparison.

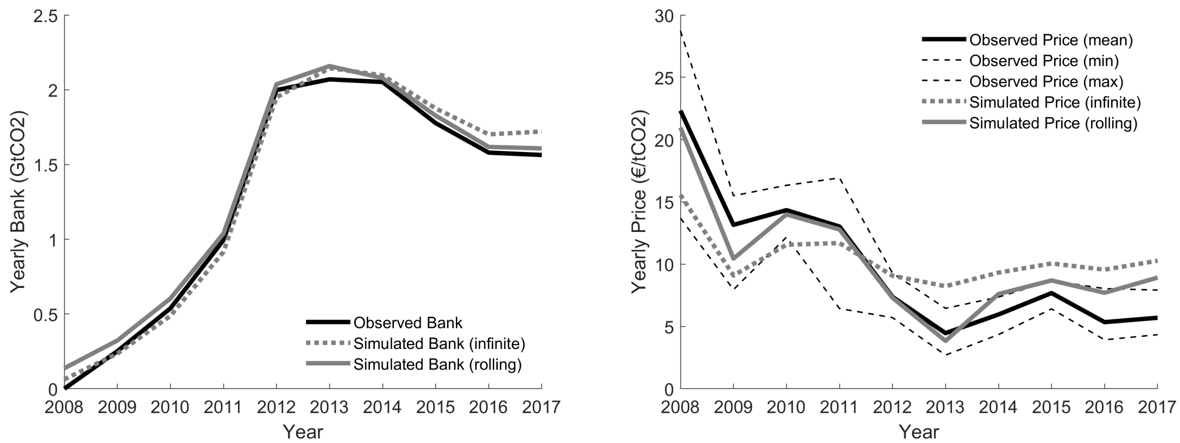
Table A.1: Best-fit results based on 2008-2017 bank and price data

Horizon type	Horizon & discount rate	Marginal abatement cost
Infinite	$h = \infty \quad r = 7.83\%$ (std.dev = 59.4 $MtCO_2$)	$c = 5.77 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 3.80 €/tCO_2)
Rolling	$h = 11 \quad r = 3\%$ (std.dev = 76.4 $MtCO_2$)	$c = 6.10 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 1.71 €/tCO_2)

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.2). Additionally, one might argue that

Figure A.1: Best-fit results based on 2008-2017 bank and price data

(a) Bank (b) Price



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 with the calibrated infinite horizon.

Table A.2: Discount rates inferred from daily futures' yield curves over 2008-2017

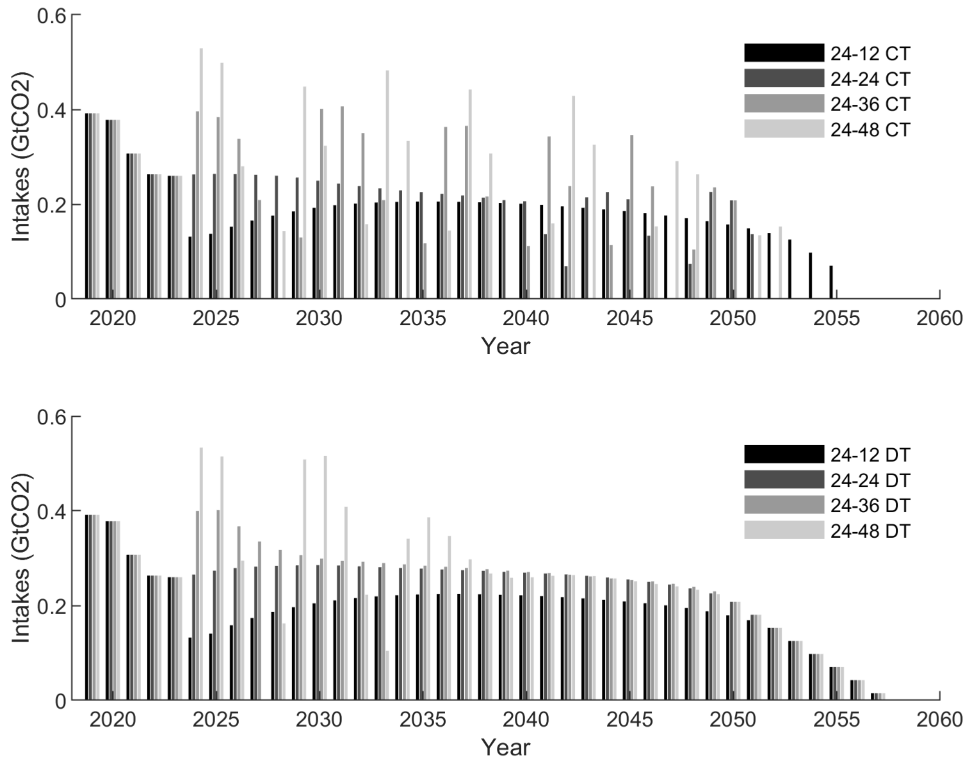
Daily yield curve	Mean	Median	Std.Dev	Min	Max
Fut. Dec Y+1 / Spot	2.4%	2.5%	1.5%	0.2%	7.0%
Fut. Dec Y+1 / Fut. Dec Y	2.9%	2.6%	1.8%	0.3%	8.7%
Fut. Dec Y+2 / Fut. Dec Y+1	3.6%	3.7%	2.0%	0.2%	8.7%

Note: Daily price data compiled from the IntercontinentalExchange. 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 $\ln(p_b^{t_2}/p_a^{t_1})/(t_2 - t_1)$ since storage costs are nil.

With a similar approach in the US Acid Rain Program, Ellerman & 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.

B Streams of annual MSR intakes

Figure B.1: Annual MSR intakes with different intake rates and thresholds



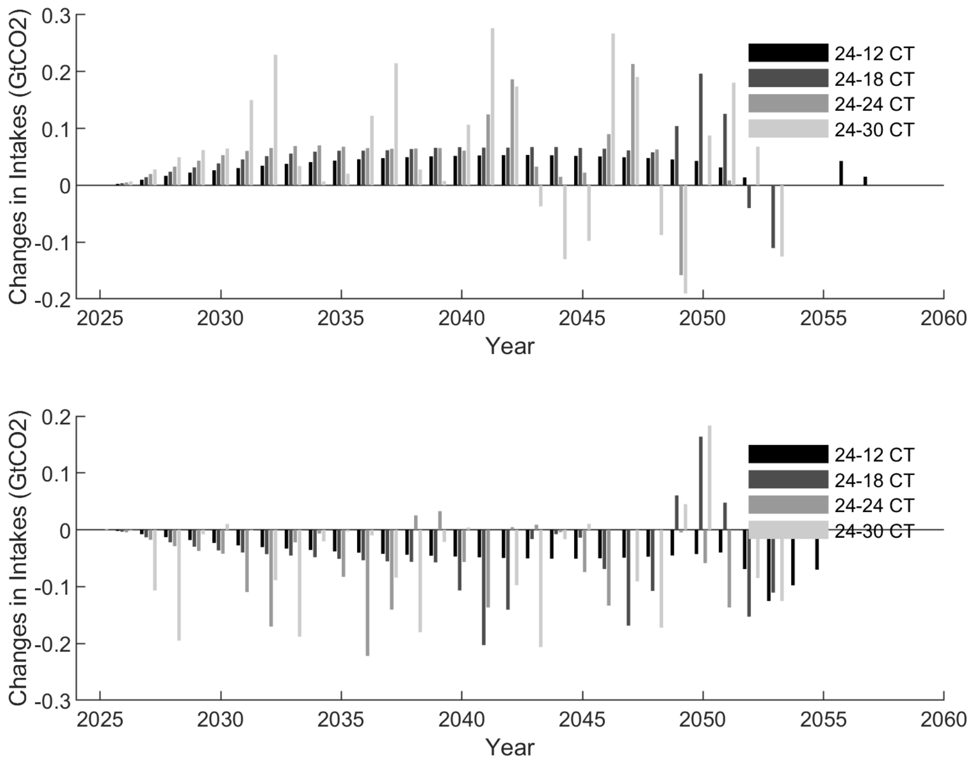
Note: Intake rate of 24% over 2019-2023, and 12, 24, 36 or 48% afterward. (upper) Constant 400-833 $MtCO_2$ thresholds; (lower) linearly declining thresholds from 400-833 in 2023 to 0-0 $MtCO_2$ in 2058. Reinjections (i.e. negative intakes) never occur in our simulations since the MSR is empty when the release threshold is passed. The plots do not include the 1.55 $GtCO_2$ exogenously seeded into the MSR.

With constant thresholds, cumulative intakes (and thus 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. As the intake rate increases, they become more erratic as thresholds are repeatedly being hit and the intake period is shorter – but overall, cumulative impacts are similar across intake rates.

In 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 (Figure 7b) mitigate the absolute impacts of higher intake rates. This notwithstanding, cumulative impacts vary more across intake rates.

C MSR-induced resilience to demand shocks

Figure C.1: 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-intake thresholds set at 400-833 $MtCO_2$ constant over time. Unanticipated permanent negative –0.15$GtCO_2$ (positive +0.15$GtCO_2$) demand shock in 2025 in the upper (lower) plot.

Figure C.1 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.

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. Venmans, 2016). In any case, the implementation of border adjustments should imply the removal of these measures, at least to avoid double protection mechanisms.

Table D.1: 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%').

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 is beyond the scope of our framework, and focus on the sole allocation method impacts. Table D.1 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., 2020), have potential to dwarf the price impacts reported in Table D.1.

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.1). As an illustration, we consider an ambition ramp-up to -62%. This is more ambitious than what is currently on the negotiation table (50-55%). However, this assumption does not change the qualitative nature of our results and suggests that a more ambitious target may 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 since some banked permits may still be used to cover emissions after the cap has shrunk to zero. These aspects are even more convoluted with the MSR in place.

Table E.1 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

Table E.1: LRF-MSR settings to reach a -62% target in 2030 w.r.t. 2005

Thresholds	IR	LRF	Emissions ($GtCO_2$) 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-intake thresholds set at 400-833 $MtCO_2$ constant over time, 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 ⋆ denotes the hypothetical 2030 target of -62% relative to 2005 levels. 'Cancel' reports cumulative cancellations in $GtCO_2$.

2050 would require a much higher LRF, above 4% in all cases.

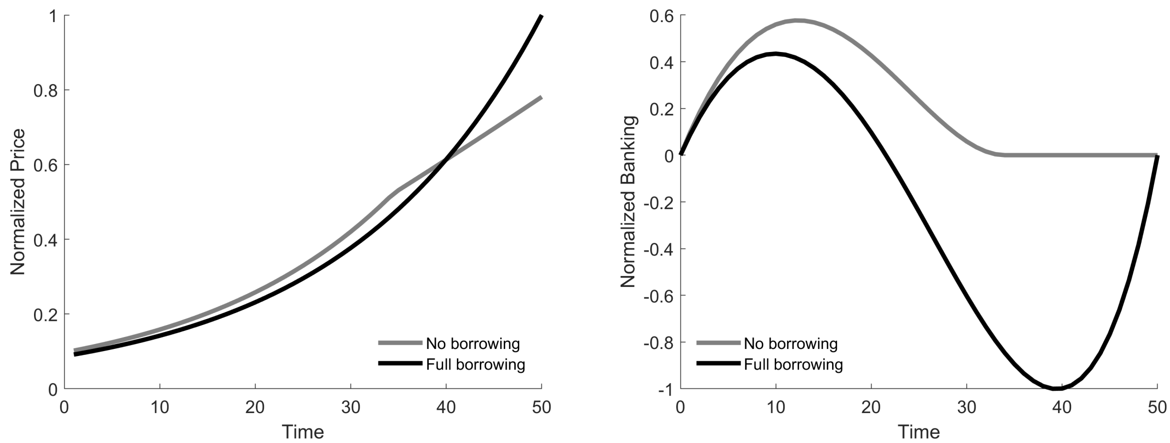
Table E.1 also frames the LRF-MSR interaction analyzed in Section 3.3 in a specific context. With both constant or declining thresholds, we see that an increase in the intake rate or in the LRF always shortens the intake period. As Section 3.3 suggests, only for the smallest intake rate (12%) does the 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 in Section 3.3.

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

Figure F.1: Stylized price and bank paths with and without borrowing (no MSR)

(a) Price paths (b) Bank paths



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 (Figure F.1a) and the optimal intertemporal reallocation of permits (w.r.t the cap trajectory) involves a banking phase followed by a borrowing phase (Figure F.1b). 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.