Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Emissions trading with rolling horizons

Simon Quemin and Raphaël Trotignon

August 2019

Centre for Climate Change Economics and Policy Working Paper No. 348
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment Working Paper No. 316
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 the Global Green Growth Institute. 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 and Trotignon R (2019) *Emissions trading with rolling horizons*. Centre for Climate Change Economics and Policy Working Paper 348/Grantham Research Institute on Climate Change and the Environment Working Paper 316. London: London School of Economics and Political Science

Emissions Trading with Rolling Horizons*

Simon Quemin[1,2] Raphaël Trotignon[2]

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

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

This version: August 2019

[latest version]

Abstract

We build a model of competitive emissions trading under uncertainty with supply-side control. Firms can use rolling planning horizons to deal with uncertainty and can also exhibit bounded responsiveness to the control. We tailor the model to the EU ETS, calibrate it to 2008-2017 market developments and find that a rolling horizon is able to reconcile the banking dynamics with discount rates implied from futures' yield curves. We evaluate the 2018 market reform, decompose the impacts of its main features and quantify how they hinge on the firms' horizon and responsiveness. We highlight important implications for policy design and evaluation.

Keywords Emissions trading, Rolling horizons, Bounded rationality, Decision-making under uncertainty, Supply-side control, EU ETS.

JEL classification D25, D81, E63, H32, Q58.

*Quemin: S.Quemin@lse.ac.uk and Trotignon: Raphael.Trotignon@ChaireEconomieduClimat.org.

«Agents may be easily assumed to be forward-looking, but do their horizons extend thirty years and even more into the future? And, if the reality is truncated horizons that are updated and moved forward as time progresses, what is the relevant time span? More importantly [...] how would such behavior change the equilibrium paths [...]?»

— Ellerman et al. (2015), discussing on intertemporal trading in the EU ETS.

1 Introduction

Emissions trading systems have become a widely used climate policy instrument. Most allow for some form of flexibility in the regulated firms' emission streams via banking and, albeit to a lesser extent, borrowing of emission permits over time (ICAP, 2019; World Bank, 2019). That is, firms must comply with the cumulative cap on emissions over the length of the program, but how they manage their emissions' calendars is left to their discretion, at least to a certain extent. In many respects, determining the optimal timing of emissions and permit usage is thus isomorphic to Hotelling's problem of efficiently extracting an exhaustible resource sold in a competitive market under conditions of uncertainty.[1] In principle, as firms cost minimize over time, intertemporal trading and absence of arbitrage entail that the permit price reflects the expected discounted long-term scarcity of permits implied by the cap trajectory and that the cumulative cap is attained at least discounted cost in expectation (Rubin, 1996; Kling & Rubin, 1997; Schennach, 2000; Ellerman & Montero, 2007).

As the opening quote recognizes, however, the market equilibrium outcomes prevailing under this cost-efficient approach to timing emissions crucially hinge on the planning horizon firms effectively employ. In this paper, we formalize and answer the three questions it raises with a specific focus on the EU ETS. That is, we first provide ample anecdotal evidence suggesting that firms use 'truncated horizons that are updated and moved forward as time progresses' in their decision-making process, what we call *rolling horizons*. Second, we develop a model of competitive emissions trading where firms can use rolling horizons, which we calibrate to market developments over 2008-2017 to obtain a measure of the 'relevant time span'. Third, we analyze how rolling horizons 'change the equilibrium paths' relative to infinite horizons. This has important implications for policy design and evaluation, some of which we highlight by assessing the impacts of the 2018 EU ETS reform with our calibrated model.

[1] One difference from Hotelling's problem arises when borrowing is restricted, if not prohibited as is often the case in practice, since the feasibility of the cost-efficient path then depends on the temporal availability of permits and thus on the allocation stream, see Hasegawa & Salant (2014) for a discussion. Other differences include the existence of supply-side controls or the absence of storage costs and depreciation.

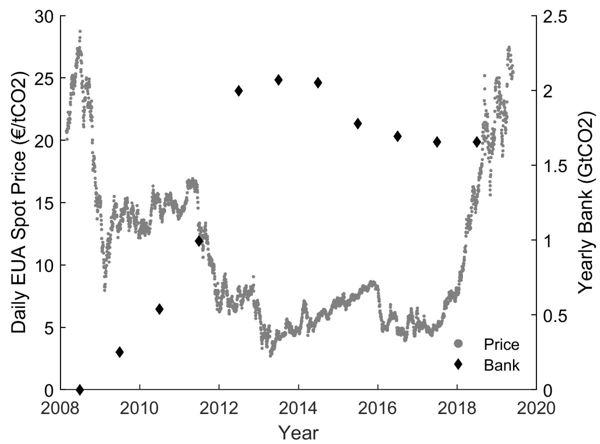
Specifically, the equilibrium outcomes when firms use a rolling horizon with a given discount rate may prima facie be seen as qualitatively equivalent to those when firms have an infinite horizon with a larger discount rate. Crucially, however, this may only hold assuming market design considerations away while in practice permit markets are often equipped with supply-side controls, typically in the form of a price collar (e.g. Roberts & Spence, 1976), or a banking corridor in the post-reform EU ETS (e.g. Perino & Willner, 2016).[2] In the presence of a supply control, short- and long-term equilibrium outcomes will depend on the interaction between the control's design features and the firms' horizon. Additionally, they will also depend on the firms' sophistication in understanding the implications of the control on their intertemporal decision making, what we call firms' *responsiveness*. In this paper, we quantify the interplay between the post-reform EU ETS design elements and the firms' horizon and responsiveness degree. The above delineates the three main contributions of this paper, which we present below in further detail in relation with the existing literature.

As a first contribution, we develop a model of competitive emissions trading under uncertainty with supply-side control which departs from the existing literature along two key dimensions. First, firms can use rolling horizons in the face of uncertainty and limited information about future permit supply and demand. As a case in point, future demand may vary substantially depending on the economic activity and the achievements of complementary policies (Borenstein et al., 2016; Burtraw & Keyes, 2018). This procedure essentially involves an iterative optimization over a sliding truncated horizon given realistic forecasts of relevant exogenous factors. Only the first period of the plan is implemented, then a new plan is formulated over an equally long horizon with updated forecasts. Second, firms can exhibit two polar degrees of responsiveness. With zero (resp. full) responsiveness, firms have no (resp. a perfect) understanding of how their intertemporal decisions in the competitive equilibrium interact with the control impacts on supply over time. The full responsiveness case obtains as a fixed point between the firms' beliefs about the control impact stream and the equilibrium stream.

By contrast, the archetypal permit trading model considers fully rational firms with an infinite horizon (or at least running up to the end of the program), see Hasegawa & Salant (2015) for a recent review. The concept of rolling horizon was first introduced by Goldman (1968) and can be seen as a way of addressing increasing uncertainty and costly informational requirements (e.g. Easley & Spulber, 1981; Grüne et al., 2015). Additionally, rolling planning is extensively used by firms in their routine production and supply management procedures (see Sahin et al.

[2] Section 3.3 shows that this equivalence also breaks down when yearly caps on emissions are not binding (w.r.t. unregulated emission levels) over some years. In this case, a rolling horizon may remove the perceived cumulative scarcity constraint, resulting in a zero permit price, which cannot occur with an infinite horizon.

Figure 1: Observed daily EUA spot price and yearly bank (Jan 2008 – May 2019)



Note: Compiled by the authors from the IntercontinentalExchange and the EU Transaction Log.

(2013) for a review) and has been applied in a similar context – exhaustible resource extraction – to help rationalize the lack of empirical support for the Hotelling rule (Spiro, 2014; van Veldhuizen & Sonnemans, 2018).[3] As another source of bounded rationality, we consider that firms may face cognitive or computational limitations in optimally adjusting their decisions in response to supply control. We believe that introducing limited responsiveness is an alley worth exploring, especially as it fits well into the context of the EU ETS. Indeed, there is evidence that, even in the absence of supply control, firms do not fully appreciate the trading and profit opportunities created by the market (e.g. Martin et al., 2015; Baudry et al., 2019), and the implications of a banking corridor such as the one introduced by the 2018 reform may arguably not be as transparent and easily translatable into concrete decisions as, say, those of a price signal conveyed by a price collar (Perino, 2018).[4]

As a second contribution, we tailor the model to the EU ETS design and parametrize permit supply and demand using historical emissions and allocation data and to align with prevailing regulation (e.g. ETS Directives, EU-wide renewable or energy efficiency objectives, etc.). For instance, this implies that, in line with the observed trend, baseline emissions and resulting demand are declining over time and become nil by the end of the century. We calibrate the market-wide discount rate and planning horizon by fitting the model outputs to the observed annual banking dynamics with discount rates implied from futures' yield curves over 2008-

[3]See Section 2.1 for a wider literature overview on rolling planning and its applications, and Section 2.2 for anecdotal evidence of shortsightedness in the EU ETS.

[4]In the words of Perino (2018), «the [control] rules should be simple and stable and their impacts predictable so that both market participants and regulators can understand them readily and respond accordingly. Such mechanisms do exist – but the new rules for [EU ETS] Phase IV are not among them».

Table 1: Discount rates implied 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%
Fut. Dec Y+3 / Fut. Dec Y+2	4.1%	2.5%	2.0%	0.6%	9.2%

Note: Summary statistics for four year-on-year futures' yield curves with daily price data. With t_1 the day's date (spot) or maturity (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 implied discount rate is given by $\ln(p_b^{t_2}/p_a^{t_1})/(t_2 - t_1)$ since storage costs are nil.

2017 (see Figure 1 and Table 1) and compare the merits of two alternative assumptions – infinite vs. rolling horizon – in how well they can rationalize observations.[5] With an infinite horizon, a discount rate of about 7% can replicate the bank dynamics with an average fit of ± 50MtCO_2/year. This is line with general rates of return on risky assets (Jordà et al., 2019) but higher than the implicit rates observed for EU allowances. If, instead, we fix the discount rate to a central value for the implicit rates (3%), we find that a rolling horizon of about a dozen years replicates past banking with a similar fit.[6] Besides, the calibrated infinite horizon does not capture the size and sign of yearly-averaged observed price variations as well as the calibrated rolling horizon, by a factor of about two.

These estimates are a novel and valuable contribution to the empirical literature on the EU ETS (e.g. Ellerman et al., 2016; Hintermann et al., 2016; Fuss et al., 2018), which indicates that although firms' behaviors are to some extent consistent with intertemporal optimization, their optimization degrees, foresight levels and horizons remain open empirical questions. Our calibration exercise also substantiates how rolling horizons can be key in understanding price and banking dynamics and evaluating ETSs' performances. Ellerman & Montero (2007) and Ellerman et al. (2015) use a similar approach to assessing the efficiency of banking behaviors ex post in the US Acid Rain Program and the EU ETS, respectively. Namely, they compare the fit between observed banking and simulated banking for various given pairs of discount and expected demand growth rates to guess at which pair might have governed the dynamics. In this paper, we augment this approach by endogenizing changes in firms' expectation about future demand, which is key in driving banking decisions and the resulting market price, and

[5]Firms are assumed to apply the same discount rate and horizon so our results should be taken as first-pass market-wide average estimates. We consider linear marginal abatement cost functions whose slopes are assumed to be time invariant so that firms' banking decisions only hinge on their discount rate and horizon, but whose linear intercepts are declining over time as the baseline emissions path is downward sloping.

[6]The calibrated horizon is quite insensitive to changes in the assumed discount rate around the central value, specifically ±2 years for ±2% with similar fits, see Section 4.4.

by introducing rolling horizons and leveraging observed futures' yield curves.

As a third contribution, we use our calibrated model to assess the impacts of the 2018 EU ETS reform (European Parliament & Council, 2018). We decompose the impacts of its three main elements, that is (1) an increase in the yearly Linear Reduction Factor of the cap from 1.74% to 2.20% from 2021 on, (2) the implementation of the Market Stability Reserve from 2019 on, a banking corridor that adjusts current auctions downwards (resp. upwards) when the bank in the previous year was above (resp. below) a predetermined threshold, and (3) its reinforcement with an add-on cancellation mechanism from 2023 on, which cancels permits stored in the MSR in excess of auctioned volumes in the previous year. This paper hence relates and adds to the burgeoning literature on the MSR initiated by Richstein et al. (2015), Fell (2016) and Perino & Willner (2016).[7] In this respect, our two main contributions are (1) the use of a calibrated model able to replicate past banking, which is important in a context where the firms' discount rate and planning horizon play a key role in determining the size of the bank and thus the quantitative impacts of the MSR (see Fell (2016) for a discussion) and (2) an exploration of the comparative implications of infinite vs. rolling horizon and zero vs. full responsiveness on equilibrium outcomes. We highlight a few notable results:

1. A price increase of a magnitude similar to the one observed over 2018 (see Figure 1) is congruent with firms using the rolling horizon and being responsive to the introduction of the MSR but crucially, irrespective of the cancellation mechanism.

2. The MSR withdraws permits annually during two up to three decades with the infinite and rolling horizons, respectively. Contrary to conventional results (Perino & Willner, 2016) the MSR without cancellations does not preserve the cumulative cap on emissions as it does not have time to empty before the program ends, which is significantly more pronounced under the rolling horizon.[8] With cancellations, cumulative emissions can be reduced by 5 up to 10 $GtCO_2$ with the infinite and rolling horizons, respectively.

3. Taken separately, both the LRF increase and the introduction of the MSR raise prices and reduce cumulative emissions. The LRF increase induces a fixed reduction in cumulative emissions of 9 $GtCO_2$. The impact of the introduction of the MSR on cumulative

[7] For other dynamic supply adjustment mechanisms based on firms' banking similar in spirit to the MSR, see Kollenberg & Taschini (2016, 2019), Gerlagh & Heijmans (2018) and Lintunen & Kuusela (2018).

[8] Because baseline emissions – calibrated to align with the current Climate Energy Package and a linear extrapolation of its objectives afterwards – decrease over time and eventually become nil at the end of the century, the ETS comes to an end naturally. With the infinite horizon, the MSR without cancellations does not preserve the cumulative cap because firms can make but imprecise forecasts about future baselines.

emissions is largely independent of the LRF increase but crucially depends on the cancellation mechanism and the firm's horizon and responsiveness.

4. The cancellation mechanism has negligible impacts on equilibrium outcomes until the 2050's when fixed annual reinjections begin. Essentially, its implementation reduces cumulative emissions and increases the MSR cost-efficiency, i.e. lowers the extra overall costs imputable to the MSR of achieving the resulting cumulative emissions relative to an adjusted linear cap path. Interestingly, in one instance, with responsive firms using the rolling horizon, the MSR augmented by cancellations is more cost-efficient than a sole linear cap path adjusted to yield the same cumulative emissions.

5. Relative to a linear cap path adjusted to yield the same cumulative emissions, the MSR sustains higher price levels early on that are counterbalanced by lower price levels later on. This holds with and without cancellations under both infinite and rolling horizons.

6. The MSR has potential to adjust cumulative emissions in the face of exogenous reductions in demand.[9] With the infinite horizon, this adjustment is partial and temporary, i.e. decreasing with the date when reductions occur (e.g. Perino, 2018; Beck & Kruse-Andersen, 2019; Carlén et al., 2019). With the rolling horizon, it is twice as small in size for early reductions but rather invariant with the date of the reduction.

Other papers introduce risk aversion on the part of firms (Kollenberg & Taschini, 2016, 2019; Tietjen et al., 2019) or an output market (Chaton et al., 2018), typically electricity, in which case some further analyze firms' investment decisions (Bruninx et al., 2019; Mauer et al., 2019; Tietjen et al., 2019). As with a rolling horizon, the mere postponement of auctions by the MSR (i.e. without cancellations) has noticeable impacts on market outcomes, either because it affects the temporal availability of permits and its variability and thus the permit-specific risk premium and the equilibrium price path, or because it affects equilibrium investment decisions and thus the speed and path of low-carbon capacity deployment.

The remainder is structured as follows. Section 2 sets forth the concept of rolling horizons in relation with the literature and contains the anecdotal evidence alluded to above. Section 3 introduces rolling horizons and limited responsiveness into the archetypal permit market modeling framework. Section 4 describes the model parametrization and calibration to the EU ETS. Section 5 analyzes the impacts of the 2018 reform, focusing on the interplay between market design elements and the firms' horizon and responsiveness. Section 6 concludes.

[9] In our setting, this holds even without cancellations, albeit to a lesser extent, as the MSR never empties.

2 Planning with rolling horizons

In this section, we begin by describing the concept of rolling planning procedures with finite horizons in relation with the literature. Then, we provide various pieces of anecdotal evidence suggesting that firms regulated under the EU ETS might use rolling finite horizons.

2.1 Concept and literature

The farther away one looks into the future the more that future is uncertain in terms of possible outcomes, their probabilities and how to incorporate them into one's planning today. As a rational way of dealing with increasing uncertainty, informational constraints or cognitive limitations, agents may conceivably use heuristics or rules of thumb in their decision making to contain the associated computational complexity and informational requirements, all the while performing comparably to more complex procedures (Simon, 1955; Gigerenzer & Todd, 1999; Gigerenzer & Selten, 2003). As a case in point, rolling horizons are widely used by firms in their planning decisions. Essentially, this involves making the most immediate decisions, i.e. in the first period, by optimizing over a finite number of periods ahead for which agents can make reasonably good forecasts about the relevant exogenous factors entering into their decision making. Only the first-period optimal decisions are implemented and the procedure is sliding, i.e. it repeats every period thereafter, each time over the rolling horizon. Agents can also revise or update their forecasts as they initiate each new planning cycle.

A rolling horizon can thus be seen as a particular form of decision making under uncertainty with limited information, either because information is scarce if not unavailable, or because agents can but imprecisely respond to it. Specifically, forecasts can either be too costly or unreliable or both, and the more distant the future, the costlier or less reliable the forecasts. It may even be that forecasts beyond a certain point are simply unavailable. Because of these constraints, agents may rationally choose to postpone decisions regarding the distant future now and only address these later on when things become clearer.

Admittedly, sliding truncated horizons may be seen as a crude way of modeling the behavior of agents facing uncertainty. Micro-foundations for similar types of behavior may involve, inter alia, ambiguity aversion especially through the maxmin decision rule (Gilboa & Schmeidler, 1989), sparsity-based bounded rationality (Gabaix, 2014) or rational inattention in the face of information acquiring and processing costs (Reis, 2006). Although these modeling approaches also restrict informational requirements, a rolling horizon has the comparative advantage of being the most simple expedient for capturing the main driving mechanism that plans are finite without inducing other biases.

Goldman (1968) first delineated and formalized the concept of continual finite planning revision. It was later extended by Easley & Spulber (1981) and Sethi & Sorger (1991) to account for stochasticity and stationarity, by Jehiel (1995) to account for strategic interactions, and by Kaganovich (1985) in the context of capital accumulation, where rolling finite plans are approximately optimal in the infinite horizon sense. More recently, Grüne et al. (2015) have proposed a procedure called nonlinear model predictive control which leverages an iterative solution of optimal control problems with a receding horizon – a widely used procedure in control engineering. Grüne et al. provide a convergence result for discounted optimal control problems and characterize when a rolling horizon approximately yields the infinite horizon optimal paths depending on the length of the horizon and the discount rate.

Rolling horizons have also been extensively developed in the production planning and supply chain management literature, see e.g. Sahin et al. (2013) for a review. Importantly, they have helped rationalize quantitative puzzles in terms of saving behaviors (Caliendo & Aadland, 2007), social security choices (Findley & Caliendo, 2009), and more in line with our paper, the long-run price dynamics of exhaustible resources such as oil (Spiro, 2014; van Veldhuizen & Sonnemans, 2018). In the latter context, a rolling horizon suppresses long-term scarcity considerations when the resource stock is sufficiently large so that the dynamic nature of the problem vanishes and resource extraction does not conform to Hotelling's rule.

Additionally, existing literature on expectation formation indicates that agents use heuristics to forecast future factors relevant to their decision problems. For instance, as an alternative to rational expectations à la Muth (1961), Brock & Hommes (1997) consider heterogeneity in expectations and adaptive switching between heuristics. Such models of behavioral expectations perform well in representing expectation dynamics based on experimental or survey data (Branch, 2004; Hommes, 2011; Hommes et al., 2019). Relatedly, there is a large experimental body of literature documenting violations of rational behaviors in dynamic decision problems (e.g. Carbone & Hey, 2001; Johnson et al., 2002) or limitations on how far ahead people can plan (e.g. Hey & Knoll, 2007). In all these cases, agents typically attach insignificant if not no (salvage) value to outcomes beyond their horizon. Finally, experiments in asset markets show that traders can be myopic (Smith et al., 1988) and base their expectations on the extrapolation of past trends (Haruvy et al., 2007).

2.2 Anecdotal evidence in the EU ETS

The empirical literature on the EU ETS indicates that regulated firms behave consistently with intertemporal cost minimization, at least to a certain extent (e.g. Ellerman et al., 2016; Hintermann et al., 2016; Fuss et al., 2018). However, their degrees of optimizing behavior, levels of foresight and time horizons largely remain open empirical questions. In particular, we list below many reasons to think that EU ETS participants resort to rolling horizons.

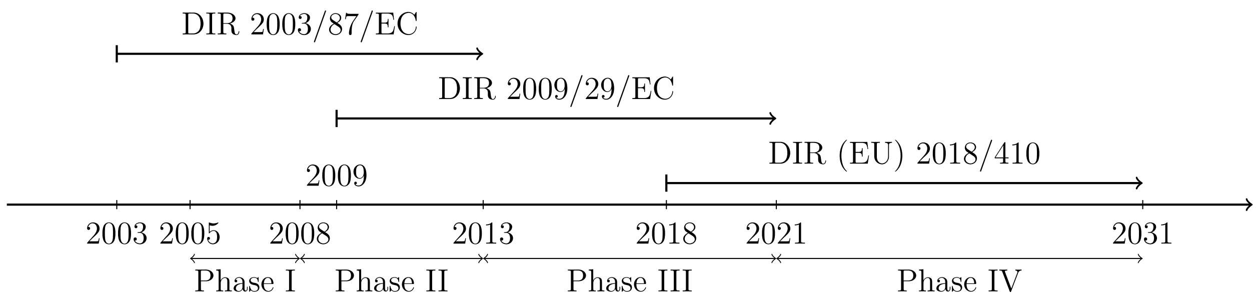
As a rule, business plans in general and production plans in particular are formulated over a finite horizon and updated on a regular basis. This is relevant for emission permits which are one factor of production (e.g. Zhang & Xu, 2013) and thereby part of the supply chain management process, which is known to rely on rolling horizon procedures (e.g. Sahin et al., 2013). For instance, large industry and power companies are known to hedge future production. More specifically for EU ETS regulated power companies, the hedging target (achieved e.g. via banking of permits or via the purchase of derivative contracts from financial actors) is stated as a percentage of future emissions generated by forward power sales: on average 80%, 40% and 20% of expected emissions in the following year, two and three year's time, respectively (Eurelectric, 2009). Beyond the hedging target, banking or acquiring futures can only happen at much higher discount rates, if at all (Schopp et al., 2015). This is due to the small leeway in deviating from standard in-house risk management procedures.

Indeed, intra-firm managerial or accounting constraints can restrict the reach of intertemporal considerations. Essentially, stockpiling is de facto limited by the willingness to tie up capital in banked permits over some years.[10] This also echoes with informal discussions we have had with carbon trading representatives of regulated entities: they are ultimately held accountable vis-à-vis their shareholders, to whom it may indeed be arduous to justify trading and banking strategies in a domain which does not belong to their firms' core activities. Accumulating a sizable bank can be deemed suspicious for national supervisory authorities (e.g. securities and markets authorities) for the same reason. Moreover, such hoarding behavior might also trigger concerns about market cornering and price manipulation.[11] Intuitively, even shorter horizons and more limited foresight can be expected from smaller entities which lack in-house trading desks and for which there is ample evidence of autarkic compliance, see e.g. Trotignon

[10] Symmetrically, companies may liquidate their stock of permits should they need to raise money quickly. This incentive is particularly salient during an economic downturn and de facto shortens firms' horizons. See Dardati & Riutort (2016) for evidence on the impacts of financial constraints on covered firms' behavior.

[11] In the California-Québec cap-and-trade scheme, such concerns are behind administrative limits on banking in the form of holding limits, i.e. a maximum number of permits each participant can hold at one time.

Figure 2: EU ETS regulatory timeline and trading phases



Note: The two ends of an arrow ↦ respectively indicate when the Directive is passed and the time horizon associated with the quantitative objectives the Directive should help meet (viz. the 2012 Kyoto targets for DIR 2003/87/EC, the 2020 objectives for DIR 2009/29/EC and the 2030 objectives for DIR (EU) 2018/410. At that time, if no new Directive is implemented, the current Directive continues to apply by default. An overlap between arrows means that the more recent Directive supersedes the older.

& Martino (2013), Martin et al. (2015) and Baudry et al. (2019).[12]

Additionally, another suggestive piece of evidence is the time span of futures markets, which can be construed as a proxy for participants' foresight. Maturities typically run from the end of the current year up to ten years in the future, but liquidity quickly decreases with time-to-maturity and trades in long-term contracts remain thin. Even more suggestive perhaps is the prevalence of regulatory uncertainty, which by the very discretionary nature of policy choices and government interventions, can be reduced but never eliminated (Kydland & Prescott, 1977; Salant, 2016). As a result, market participants may excessively if not exclusively focus on the short term, because of longer-term credibility issues or simply owing to the fact that regulation is never set in stone. Figure 2 illustrates how regulation has evolved over time by depicting the EU ETS regulatory timeline. Note that after endorsement, it took 6 to 9 years for regulation to be amended. Moreover, credibility issues may arise when participants doubt the regulator's ability or willingness to intervene to increase ambition or 'fix the market',[13] or when regulatory language remains indefinite.[14]

[12] In Phase II of the EU ETS, about a third of liable entities did not engage in trading, accounting for just less than 10% of the total cap, implying they passively banked their excess endowment or borrowed permits from their allocation of the following year in order to meet present compliance.

[13] A clear example is a 2014 sentiment survey of market participants which found that 30% of respondents thought the ETS would no longer exist after 2020 (Thomson Reuters, 2014). See also The Economist (2013).

[14] A clear example of vagueness in language in our case is that the add-on cancellation mechanism might in fact not cancel permits or even be enforced. The first source of vagueness relates to the practical interpretation of the notion of permit validity vs. cancellation as the official terminology is 'no longer be *valid*'. The second source of vagueness relates to the implementation of the measure itself, as it might be '*decided otherwise* in the first review' in 2021, see (23) in Directive 2018/410 (European Parliament & Council, 2018). Relatedly, there are discrepancies within the Directive itself ('*should* no longer be valid' in (23) vs. *shall* in Art. 2(5a))

Finally, some have argued that a high discount rate is 'rough way to reflect concerns over the degree of farsightedness of market participants and the market specific economic and political risks' (Perino & Willner, 2017), while others have suggested that firms apply a low discount rate in the short term and then a very high rate for later years due to an 'insufficient regard for long-term strategy' (Climate Strategies, 2015). First note that sliding truncated horizons can be seen as an extreme version of the second approach in the limit as the longer-term rate becomes large. Second, our analysis will show that these two approaches do not imply the same qualitative effects on equilibrium outcomes, especially in presence of supply control.[15] Moreover, Table 1, which contains year-on-year discount rates implied from futures' yield curves over 2008-17 for short-term maturities, indicates an average rate of about 3%, thereby lending some support to the second approach. We investigate this further in Section 4.

3 Modeling framework

We consider a competitive emissions trading system with full banking and limited borrowing of emission permits over time. Time is discrete and indexed by $t \geq 1$. The system starts at date 1 and compliance is required at each date t. The regulator sets a cap on system-wide emissions for each date t, which consists of freely allocated and auctioned permits f_t and a_t. Some fixed quantity of offset credits O may also be surrendered for compliance over a given time period and o_t denotes the volume of offsets used at date t with $\sum_t o_t \leq O$. At each date t, we assume that regulated firms fully acquit their compliance obligations, i.e. they remit as many permits and offsets as needed to cover their current emissions.[16]

The decentralized market equilibrium can be characterized indirectly as the solution to joint compliance cost minimization among all firms (Montgomery, 1972; Rubin, 1996). This well-known result has been employed pervasively in the literature (e.g. Schennach, 2000; Fell, 2016; Perino & Willner, 2016). Here as well, we take the perspective of the regulated perimeter as a whole – hereafter the firm.[17] We let e_t and u_t denote its levels of realized and unregulated (i.e. baseline) emissions at date t, respectively. Abatement $u_t - e_t \geq 0$ is costly and we let

and with a more recent Communication which resorts to *will* (European Commission, 2018a).

[15] See Section 3.3 for a broader discussion on the relative impacts of these two approaches.

[16] Penalties incurred for various permit and self-reporting violations are implicitly assumed to be adequately designed and enforced. See Stranlund et al. (2005) for a discussion.

[17] With competitive trading, this is warranted only when firms' baseline emissions are private information, which can be aggregated via the market by a sufficient statistic (the equilibrium price), while informational efficiency breaks down when firms' abatement costs are private information (Cantillon & Slechten, 2018).

C_t denote its minimum total abatement cost function at date t with $C_t', C_t'' > 0$.[18]

The firm's baseline emissions u_t depend on business cycle fluctuations (Koch et al., 2014; Bel & Joseph, 2015) and the variable performances of companion policies, that is complementary climate and energy policies flanking the permit market that can affect counterfactual emission levels independently of the permit price (Borenstein et al., 2016; Burtraw & Keyes, 2018). On the supply side, there exist small annual discrepancies between announced and realized cap levels, the future cap trajectories $\{f_t\}_t$ and $\{a_t\}_t$ can be affected by regulatory changes, and $\{o_t\}_t$ depends on external offset market conditions (de Perthuis & Trotignon, 2014; Ellerman et al., 2016). In the following sections, we use the tilde notation to signify that future permit demand and supply conditions are uncertain in nature for the firm.

We next describe the properties of the competitive intertemporal market equilibrium before introducing in turn rolling horizon planning, supply-side control and bounded responsiveness to the control, along with the associated resolution procedures. We also discuss the relative implications and induced intertemporal behaviors of infinite and rolling finite horizons.

3.1 Competitive intertemporal equilibrium

The firm's annual demand for permits has two components, one for annual compliance and the other for intertemporal arbitrage.[19] That is, at any date t, given the prevailing permit price p_t and realized baseline u_t, the firm's emission level, or demand for permits for compliance $e_t^\star(p_t, u_t)$, satisfies the usual first-order necessary condition

$$C_t'(u_t - e_t^\star(p_t, u_t)) - p_t = 0. \tag{1}$$

In addition, the firm can over-comply relative to total contemporaneous available supply and carry over (i.e. bank) unremitted permits into future dates. Likewise, it can under-comply and front-load (i.e. borrow) yet-unallocated permits from its future self and use them to achieve full compliance today. While banking is unlimited, we consider that borrowing at date t is authorized up to a limit $l_t \geq 0$. As the firm cost minimizes over time, limited intertemporal trading opportunities imply a no-arbitrage condition closely following the rationale of com-

[18] Formally, the market-wide abatement cost function C_t is the envelope of all firm-level abatement cost functions, i.e. its derivative C_t' obtains by horizontal summation of the latters' (Montgomery, 1972).

[19] We do not explicitly account for forwards and futures as the aggregate demand for such bilateral contracts is nil in equilibrium, so p_t is de facto the date-t spot price (Laffont & Tirole, 1996; Seifert et al., 2008). Note, however, that (2) can legitimately hold provided an active market for futures exists (Pindyck, 1993). Besides, prices on the primary and secondary markets have no reason to differ given the uncertainty structure assumed here, specifically u_t, f_t, a_t and o_t are revealed at the beginning of date t (Kling & Rubin, 1997).

petitive commodity storage with negligible storage costs and no stock depreciation over time (Wright & Williams, 1982; Deaton & Laroque, 1992).[20] Permit banking, whose level at date t is denoted b_t, thus constitutes the second determinant of permit demand and satisfies the following two conditions with complementary slackness

$$b_t + l_t \geq 0 \quad \perp \quad p_t - \beta \mathbb{E}_t\{p_{t+1}\} \geq 0, \tag{2}$$

where $\mathbb{E}_t\{\cdot\}$ denotes expectation conditional on all information available to the firm at date t and $\beta = (1+r)^{-1}$ is the firm's discount factor with r the discount rate, possibly inclusive of a permit-specific risk premium.[21] As long as $\beta \mathbb{E}_t\{p_{t+1}\} > p_t$, banking is profitable and increases date-t permit demand, which raises p_t and lowers $\mathbb{E}_t\{p_{t+1}\}$ until all arbitrage opportunities are exhausted and the firm breaks even so the cost-of-carry price coincides with the spot price grown at the discount rate, i.e. $\beta \mathbb{E}_t\{p_{t+1}\} = p_t$. Conversely, as long as $\beta \mathbb{E}_t\{p_{t+1}\} < p_t$, borrowing is profitable but only authorized up to l_t. As soon as this constraint is binding, the connection between current and expected future prices ceases to hold and the price rises at a rate less than the discount rate. In sum, the price can rise at a rate at most as high as the discount rate in a rational expectations equilibrium.

Note that banked (resp. borrowed) permits add to (resp. subtract from) future permit supply, i.e. total available supply at date t amounts to $f_t + a_t + o_t + b_{t-1}$. Market clearing at date t, which implies that total supply equalizes total demand, thus reads

$$f_t + a_t + o_t + b_{t-1} = e_t + b_t. \tag{3}$$

Combining the compliance, no-arbitrage, and market-clearing conditions in (1), (2) and (3) then leads to two regimes in the equilibrium price and emission dynamics

$$p_t = \max\{\beta \mathbb{E}_t\{p_{t+1}\};\ C_t'(u_t - (l_t + f_t + a_t + o_t + b_{t-1}))\}, \tag{4a}$$

$$e_t = \min\{e_t^\star(\beta \mathbb{E}_t\{p_{t+1}\}, u_t);\ l_t + f_t + a_t + o_t + b_{t-1}\}, \tag{4b}$$

$$\text{with}\ \ b_t = f_t + a_t + o_t + b_{t-1} - e_t \geq -l_t. \tag{4c}$$

The first regime features intertemporal flexibility in the firm's emission stream in line with

[20] Schennach (2000) first pointed out the tight connection between commodity storage and permit banking. One key difference is that stockouts, which here coincide with positive borrowing ($b_t < 0$), are not feasible.

[21] Schennach (2000), Perino & Willner (2016) and Kollenberg & Taschini (2019) introduce an exogenous risk premium on top of the risk-free rate to capture the impacts of future emissions uncertainty on the firm's decisions – in the spirit of the Capital Asset Pricing Model. See Tietjen et al. (2019) for a derivation of an endogenous premium formally accounting for firms' hedging demand and investment decisions.

Hotelling's rule in expectation with $b_t > -l_t$.[22] In the second regime the borrowing constraint is binding ($b_t = -l_t$) so that permit demand becomes solely determined by annual compliance requirements in (1) and emissions are pegged to the contemporaneous amount of permits on hand. Limited borrowing implies an asymmetric demand shock dampening potential for the market,[23] and induces a non-linearity so that no closed-form solution to (4) can be derived.[24] Our resolution procedures are described in the following sections.

3.2 Introducing rolling finite horizons

At the beginning of any date t, u_t, f_t, a_t and o_t are given and known to the firm, which also keeps track of the state variable, i.e. the bank b_{t-1} (with $b_0 = 0$).[25] The firm selects its date-t emission e_t and implied bank b_t by minimizing its expected present value of compliance costs. That is, with an infinite horizon, the firm solves the following program

$$\min_{\{e_\tau\}_{\tau \geq t}} \mathbb{E}_t \Big\{ \sum_{\tau \geq t} \beta^{\tau - t} C_\tau(\tilde{u}_\tau - e_\tau) \Big\} \tag{5a}$$

$$\text{subject to } 0 \leq e_\tau \leq \tilde{u}_\tau, \tag{5b}$$

$$\text{and } b_\tau = b_{\tau-1} + \tilde{f}_\tau + \tilde{a}_\tau + \tilde{o}_\tau - e_\tau \geq -l_\tau, \tag{5c}$$

where (5b) contains feasibility constraints for the emission path and (5c) describes the law of motion for the state variable (i.e. annual market clearing), where the constraint on the bank ensures the cumulative emissions cap is fulfilled (i.e. overall market clearing).

To deal with uncertainty the firm may choose to optimize over a finite horizon $h \geq 0$ within which it is confident about its forecasts of future exogenous variables (see Section 2).[26] Hence

[22] This regime is always finite in time but may not be unique under uncertainty, see e.g. Schennach (2000).

[23] In principle, the firm can entirely smooth out the price impact of a downward demand shock (temporary glut) by stockpiling more permits while it can only be so for a symmetric upward shock (temporary shortage) to the extent that the bank is not too small. See also Kollenberg & Taschini (2019) for a discussion.

[24] Deaton & Laroque (1992) use a fixed-point approach in a similar commodity storage problem, for which they prove that with time-independent linear consumption demand (which translates to time-independent linear marginal abatement cost functions here) there is a unique stationary rational expectations equilibrium.

[25] Offset usage is assumed exogenous to the firm's problem. In Section 4.3 we explain how we tackle offset usage for the ex-post model calibration and why this assumption is innocuous for our analysis in Section 5. See e.g. Koch et al. (2017) for a treatment of joint permit and offset usage decisions.

[26] We treat the horizon length as given in the model, although it could endogenously emerge as a result of informational barriers and computational cost containment. See Section 4.4 for a calibration of h.

it selects its date-t emission e_t and implied bank b_t by solving the following program

$$\min_{\{e_\tau\}_{\tau=t}^{t+h}} \sum_{\tau=t}^{t+h} \beta^{\tau-t} C_\tau(\hat{u}_\tau^t - e_\tau) \tag{6a}$$

$$\text{subject to } \; 0 \leq e_\tau \leq \hat{u}_\tau^t, \; b_\tau = b_{\tau-1} + \hat{f}_\tau^t + \hat{a}_\tau^t + \hat{o}_\tau^t - e_\tau \geq -l_\tau, \tag{6b}$$

$$\text{and } \; \sum_{\tau=t}^{t+h} \left[\hat{u}_\tau^t - e_\tau\right] \geq \sum_{\tau=t}^{t+h} \left[\hat{u}_\tau^t - (\hat{f}_\tau^t + \hat{a}_\tau^t + \hat{o}_\tau^t)\right] - b_{t-1}, \tag{6c}$$

where $\hat{x}_\tau^t$ denotes the date-t forecast for $x = \{u, f, a, o\}$ at date $\tau \geq t$. Although redundant, (6c) is added for clarity: it specifies the firm's date-t assessment of the system stringency over the horizon h, that is the sum of forecasted yearly raw abatement efforts $\hat{u}_\tau^t - (\hat{f}_\tau^t + \hat{a}_\tau^t + \hat{o}_\tau^t)$ corrected for the initial bank b_{t-1}. Then, (6a) dictates that this forecasted overall abatement effort be spread over the horizon in accordance with the equimarginal value principle.

Under a rolling horizon, the firm solves for the equilibrium path from date t up to $t+h$ given its current forecasts $\{\hat{x}_\tau^t\}_\tau$, but only implements the first date of the plan, which pins down the state variable for the next date. At date $t+1$ the firm revises its forecasts based on new information about x and initiates a new planning cycle from date $t+1$ to $t+h+1$ taking the state variable b_t as given.[27] This date-on-date solving and updating of finite plans and the sequential execution of the first date of these plans then unfolds over time. Technically, a rolling horizon shrinks the firm's problem dimensionality from uncertainty to certainty given appropriate forecasts. Note also that the current plan is not contingent on future plans, as this would otherwise call for high informational requirements and computational complexity, perhaps even more so than for an infinite horizon.

Last but not least, note that terminal conditions matter with a finite horizon. In our setting, they can be construed as the belief about the continuation value of banked permits, in terms of market value, at the end of the horizon. Because of informational constraints, this value can reasonably be based on an educated guess at best, which is unlikely to correspond to the market equilibrium value of permits at this date that would obtain under an infinite horizon. Program (6) hence assumes a zero salvage value for simplicity and without loss of generality for our results.[28] Indeed, a non-zero salvage value would affect our quantitative results in a predictable way: a positive value implies higher price and bank paths, and vice versa.[29]

[27] We consider that the firm can revise its forecasts and initiate a new planning cycle at each date, although in practice the re-planning periodicity can be a choice variable for the firm. See Sections 4.2 and 4.3 for a description of future demand and supply forecast heuristics and how they are updated.

[28] This assumption also reflects empirical and experimental evidence presented in Section 2.1.

[29] When assessing the reform impacts in Section 5 a positive value would induce larger permit withdrawals.

3.3 Infinite versus rolling finite horizons

Intuitively, as h grows, the equilibrium paths obtained with a rolling horizon should eventually converge to those of an infinite horizon. In our setting, we make two additional assumptions to arrive at this convergence result and ensure some comparability between rolling and infinite horizons. First, we derive the expected equilibrium paths under an infinite horizon invoking a first-order approximation along certainty-equivalent paths for the exogenous variables x.[30] This approach was suggested but not operationalized by Schennach (2000). In our setting, it reduces the dimensionality of the infinite horizon problem and makes it similar to that of a rolling horizon. Second, we impose that any date-t certainty-equivalent paths coincide with the corresponding forecasts over the relevant time span, i.e. $\hat{x}_\tau^t = \mathbb{E}_t\{\tilde{x}_\tau\}$.

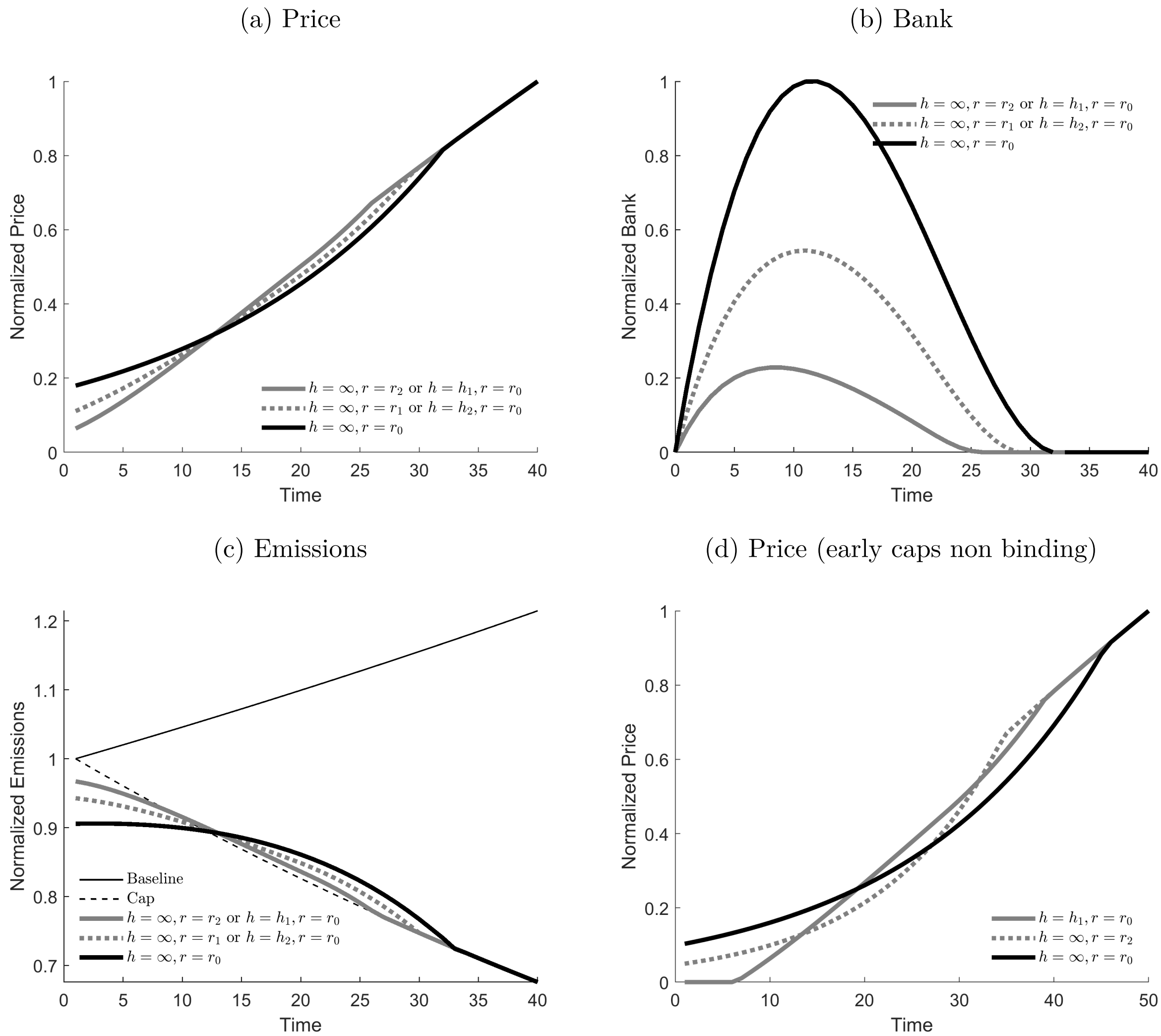
More specifically, by chaining (4a) over time and the tower rule, the date-t expected price path under an infinite horizon satisfies $\mathbb{E}_t\{p_{t'} - \lambda_{t'}\} = \beta^{t-t'} p_t$ for any $t' > t$, where $\lambda_{t'} \geq 0$ is the Lagrange multiplier associated with the limited borrowing constraint at t', i.e. $b_{t'} \geq -l_{t'}$. When there is a zero probability of a binding constraint in the future, i.e. $\mathbb{E}_t\{\lambda_{t'}\} = 0$ for all $t' \in [\![t_1; t_2]\!]$, the date-t expected price grows at rate r over $[\![t_1; t_2]\!]$. When this probability is positive, i.e. $\mathbb{E}_t\{\lambda_{t'}\} > 0$ for some $t' \in [\![t_1; t_2]\!]$, the expected price rises at a rate less than r (which is not uniquely pinned down) with the downward offset rising over $[\![t_1; t_2]\!]$. When it is unity, the expected price is uniquely determined by $\mathbb{E}_t\{p_{t'}\} = \mathbb{E}_t\{C_{t'}(\tilde{u}_{t'} - (\tilde{f}_{t'} + l_{t'} + \tilde{a}_{t'} + \tilde{o}_{t'}))\}$. An *expected* equilibrium path thus exhibits the above three-regime dynamics while the *actual* path only features the two regimes in (4).

Schennach (2000) proposed to invoke the certainty equivalence property and solve (5) assuming that the expected equilibrium paths follow the same two-regime dynamics as the actual paths in (4). We follow the suggested approach. That is, we by construction only consider the size of the bank in expectation. Therefore, a second-order bias between our approximate and the exact expected paths arises as we do not capture the possibility that $\mathbb{E}_t\{\lambda_{t'}\}$ may become positive for some $t' > t$ although the bank is still expected to satisfy the limited borrowing condition by that time. Schennach (2000) also showed that the approximate expected price path is slightly biased downward in the first regime and early on in the second one, biased upward for the rest of it, and unbiased in the third regime.

To develop insight into the relative implications of infinite versus rolling finite horizons, we describe and compare the associated market outcomes in a simple setup shutting down other

[30] The certainty equivalence property entails that up to a first-order approximation optimal decisions at date t coincide with those under full information provided that random variables equal their date-t expected values. This naturally comes about with linear marginal abatement cost functions (Schennach, 2000).

Figure 3: Qualitative implications of infinite vs. rolling horizons



Note: $r_2 > r_1 > r_0 > 0$ and $0 < h_1 < h_2 < \infty$. In Fig. 3d the early yearly caps are not binding w.r.t. baselines.

components that might come into play. Specifically, we assume perfect foresight on the part of the firm and time invariant abatement cost functions to single out the impacts of different time horizons and discount rates. When yearly caps on emissions are declining over time and increasingly binding relative to baseline emissions (as shown in Figure 3c), it is rational for the cost-minimizing firm to cut emissions below yearly caps early on and accumulate a bank of permits which is drawn down later on when the cap becomes more stringent, see Figure 3b. As shown in Figure 3a, the associated price path rises at the discount rate as long as the bank is positive, and then it equals the marginal cost of meeting annual caps.

The optimal price, bank and emission paths depend on the combination of the firm's horizon and discount rate. Specifically, for h (resp. r) given, the larger r (resp. smaller h) the shorter the banking period and the lower the banked volumes at all points in time. Accordingly, the permit price is lower early on but rises at a higher rate over time during the banking period as it is the vehicle that equalizes supply and demand over the horizon. This suggests that the intertemporal behavior of a firm using a rolling finite horizon is observationally equivalent to that of a firm using an infinite horizon with a higher discount rate.

Crucially, however, this qualitative equivalence holds as long as (1) yearly caps are binding and (2) there is no supply-side control. Regarding (1), Figure 3d shows how the equivalence breaks down when early annual caps are not binding: when the horizon is sufficiently short, the firm perceives no overall constraint on emissions implying that the price is constant and nil for a few years, while with an infinite horizon and a larger discount rate, the price exhibits the same behavior as in Figure 3a.[31] Regarding (2), we will explore the interaction between the firm's horizon and discount rate and a specific supply control in Section 5.

3.4 Introducing supply control via a banking corridor

The regulator can induce some resilience in its system through supply-side control (Roberts & Spence, 1976; Pizer, 2002). These mechanisms adjust contemporaneous supply based on the value of a given market indicator, typically the permit price, relative to some predefined thresholds.[32] This usually requires the creation of a reserve of set-aside permits whose stock at date t we denote $s_t \geq 0$. We here consider a banking corridor which automatically adjusts current auctions a_t based on banking history $\{b_\tau\}_{\tau<t}$ according to

$$a_t \longleftarrow a_t - \min\Big\{a_t; R \cdot \sum_{\tau<t} \mathbb{1}\{b_\tau > \bar{b}\} x_\tau b_\tau\Big\} + \min\{I; s_{t-1}\} \cdot \sum_{\tau<t} \mathbb{1}\{b_\tau < \underline{b}\} x_\tau, \tag{7}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, $\underline{b} > 0$ and $\bar{b} > \underline{b}$ lower and upper bank thresholds, $I > 0$ an injection quantity, $R \in [0; 1]$ an absorption rate and historical weights $\{x_\tau\}_{\tau<t}$ are such that $x_\tau \in [0; 1]$ for all $\tau < t$ and $\sum_{\tau<t} x_\tau = 1$. In parallel, the stock of permits stored in the reserve follows the complementary dynamics

$$s_t = s_{t-1} + \min\Big\{a_t; R \cdot \sum_{\tau<t} \mathbb{1}\{b_\tau > \bar{b}\} x_\tau b_\tau\Big\} - \min\{I; s_{t-1}\} \cdot \sum_{\tau<t} \mathbb{1}\{b_\tau < \underline{b}\} x_\tau. \tag{8}$$

[31] See Spiro (2014) for an extensive discussion in the context of exhaustible resource extraction.

[32] A standard approach to controlling supply consists in introducing steps in otherwise vertical (i.e. inflexible) supply schedules and typically takes the form of a price corridor, i.e. a combination of a price floor and ceiling, see Grüll & Taschini (2011), Fell et al. (2012) and Abrell & Rausch (2017) inter alia.

In words, when $b_{\tau<t}$ is above $\bar{b}$, a predefined share $x_\tau R$ thereof is withheld from auctions at date t and placed in the reserve. Symmetrically, when $b_{\tau<t}$ is below $\underline{b}$ and the current stock of the reserve allows, a fixed quantity of reserve permits $x_\tau I$ is added to auctions at date t. Otherwise, the banking corridor is inactive. Because the shift in auctions is determined by the banking history it is fixed once and for all at the beginning of each date.[33]

The banking corridor defined in (7-8) can in principle be expected to preserve the cumulative emissions cap as it essentially rearranges the auction schedule. Indeed, after a certain period the bank will pass below $\underline{b}$ (see Figure 3b) implying that the reserve should eventually empty. This ceases to be the case, however, if the corridor is equipped with an add-on cancellation mechanism that retires and invalidates permits stored in the reserve in excess of a certain threshold. In this case, the reserve stock is further adjusted such that

$$s_t \longleftarrow s_t - \max\{0; s_t - k_t\}, \tag{9}$$

where $k_t \geq 0$ is the maximum number of permits that can be stored in the reserve at date t. The cumulative emissions cap is endogenized and becomes a market outcome. Whether and by how much it is reduced depends on the thresholds $\{k_t\}_t$ and the equilibrium dynamics of the reserve stock, which itself hinges on the initial bank and reserve stock conditions and is ultimately governed by the firm's horizon, discount rate and responsiveness to the control.

3.5 Introducing bounded responsiveness to the control

The key decision quantity that the firm has to appraise at each date is its required cumulative abatement effort over its horizon in (6c) which hinges on its future permit supply and demand forecasts. With supply control, crucially, this further depends on the perceived future control-driven reshuffling of yearly auction volumes, if not net shift in the overall cap. We consider two polar degrees of responsiveness on the part of the firm in anticipating these adjustments, namely zero and full responsiveness. In the former case, the firm does not account for control impacts on future supply when assessing its abatement effort. At each date, the unresponsive firm is simply discovering and factoring in annual control impacts on auction volumes while remaining completely oblivious as to what the future impacts might be. In the latter case, the firm fully comprehends the interplay between its decisions in the competitive equilibrium and the induced control impacts over time. Thus the expected total abatement effort depends

[33] This is typically not the case for a price corridor whereby auctioned volumes are continuously adjusted until the contemporaneous auction price falls within the predefined price band.

on the control design and impacts, which are correctly perceived by the fully responsive firm and can prompt it to adjust its intertemporal decision making accordingly.

Our indirect approach to solving for the recursive competitive equilibrium as the outcome of a planning problem by a representative firm is viable under laissez-faire (e.g. Samuelson, 1971). However, Salant (1983) showed in the wider commodity storage context that such an approach may mischaracterize the rational expectations equilibrium in presence of supply-side controls aimed at stabilizing prices.[34] Indeed, forward-looking rational agents can take advantage of the control rules which the regulator adheres to, which may engender speculative attacks on the scheme and result in policy failure. Under a banking corridor for instance, firms would collectively like the policy handle (the bank) to fall below the intake threshold for the control to induce a minimal, if any, contraction in overall supply.[35] Individually, however, they have a negligible impact on the bank and cannot coordinate their banking decisions to 'game the system'. In a competitive equilibrium, therefore, supply controls cannot alter intertemporal efficiency (i.e. the equalization of discounted expected marginal abatement costs across firms and periods) although they do have an impact on annual market clearing.

Given our indirect planning approach, the fully responsive firm must be able to understand the interplay between its decisions in the recursive competitive equilibrium and the associated control actions over time. To this end, we develop a procedure to solve for the equilibrium as the fixed point of a mapping between the firm's beliefs about the control impact profile and optimal beliefs.[36] That is, the equilibrium obtains when a given such belief coincides with the actual law of motion for the control actions generated by intertemporally efficient choices induced by this belief. At each step of this procedure, the firm has a forecast for both the control action and annual supply profiles to evaluate its required cumulative abatement effort over its horizon, and then cost minimizes over time. The associated first-order conditions are thus congruent with those obtained in a competitive equilibrium.

Specifically, starting from the zero responsiveness case at any date t with horizon h, the firm

[34] See Hasegawa & Salant (2015) for a transposition of Salant's original point to emissions trading.

[35] Specifically, banking levels below intertemporally efficient ones could yield a smaller supply tightening and hence lower compliance costs, but would not conform to a competitive equilibrium. See Stocking (2012) for an example of manipulation under a hard price ceiling where purchasing permits at the ceiling price when the market price is lower is not cost efficient but may reduce the overall compliance cost by relaxing the cap.

[36] Such a fixed-point approach is hardly new. Lucas & Prescott (1971) were the first to use one to determine a rational expectations equilibrium in a Bellmanized indirect planning optimization program.

first derives the date-t expected equilibrium path by recursively solving the program

$$\min_{\{e_\tau\}_{\tau=t'}^{t'+h}} \sum_{\tau=t'}^{t'+h} \beta^{\tau-t'} C_\tau(\hat{u}_\tau^t - e_\tau) \tag{10a}$$

$$\text{subject to } \ 0 \leq e_\tau \leq \hat{u}_\tau^t \text{ and } b_\tau = b_{\tau-1} + \hat{f}_\tau^t + \hat{a}_\tau^t + \hat{o}_\tau^t - e_\tau \geq -l_\tau. \tag{10b}$$

Program (10) is solved for each $t' \in [\![t; t+h]\!]$ iteratively: only the first-date optimal outputs for each t' are implemented, assuming that (1) the firm's date-t forecasts are kept unchanged throughout and materialize as forecasted; and (2) the initial bank condition at t' (i.e. $b_{t'-1}$) is set by the previous optimization round. This iterative process captures that the responsive firm is aware of its rolling horizon and accounts for it when computing the date-t expected equilibrium path. Throughout this process, the control further affects annual auctions as per (7), although future control impacts remain unforeseen to the firm.

This process yields a path for annual permit flows into and out of the reserve as per (8) and (9). The firm then adjusts its beliefs about the future auctions stream based on the obtained flow path, and repeats the above process. If this is not neutral vis-à-vis the previously optimal emission and bank paths, the firm will revise them. This changes the annual control impact profile, which in turn again affects the firm's intertemporal decisions, and so forth. At each step in this gradual adjustment procedure the firm holds beliefs about future control impacts, behaves rationally with respect to these beliefs, and updates them in between each step. The procedure gradually attains a fixed point yielding the date-t competitive equilibrium with a fully responsive firm.

4 Calibration to the EU ETS

In this section, we first set the model parameters to align with the EU ETS design. We next describe how we use historical data and regulatory texts to construct the permit demand and supply schedules, as well as the corresponding forecasts the firm can form and revise over time. We then calibrate the firm's discount rate, planning horizon and marginal abatement cost parameter based on the observed banking and price dynamics over 2008-2017.

4.1 Market design parametrization

In the EU ETS, compliance is required on a calendar-year basis, by 30 April of the following year. Thus each date t in the model coincides with a calendar year. Banking within and across trading phases is authorized and unlimited as there is no vintage restriction on the temporal validity of issued permits as compliance instruments since the beginning of Phase II in 2008 (European Parliament & Council, 2003). Year-on-year borrowing is tacitly authorized since freely-allocated year-t vintage permits are issued two to four months prior to the year-$(t-1)$ compliance deadline. Hence we set $l_t = f_{t+1}$ as firms can effectively tap into their allotment for two consecutive years.

We parametrize the banking corridor to conform with the features of the Market Stability Reserve as adopted in 2018 (European Parliament & Council, 2018). The MSR begins operations in 2019 and is initially seeded with backloaded permits over 2014-2016, i.e. $s_{2018} = 900$ million (European Commission, 2014). Non-issued Phase III permits up to 2017 are also placed in the reserve in 2021, the number of which we estimate at 581 million.[37] The MSR thresholds are set at $\bar{b} = 833$ million and $\underline{b} = 400$ million. Moreover, we set $I = 100$ million and $R = 0.24$ until 2023 with $R = 0.12$ afterwards (European Commission, 2018a). Finally, we set the historical weights such that $x_{t-2} = 2/3$ and $x_{t-1} = 1/3$ to account for the mismatch between the MSR and compliance calendars.[38]

The cancellation mechanism defined in (9) is active from 2023 on and operates such that any permits stored in the reserve in a given year in excess of the number of auctioned permits in the previous year are invalidated (European Parliament & Council, 2018). Hence we set $k_t = a_{t-1}$. Importantly, note that a_{t-1} is endogenously determined via the MSR as per (7). Finally, since the reform was finalized in late 2017 and enacted in early 2018, we consider that the impacts of the MSR and the cancellation mechanism on annual auction volumes can be anticipated and factored in by the responsive firm from 2018 on.

[37] According to the European Commission (2015) the expected number of unallocated Phase III permits is between 550 and 700 million. The cumulated difference between announced annual caps and effective annual caps mostly arises due to non-distributed permits from the New Entrants Reserve and Article 10(c), or from plant closures and production capacity changes. We are thus in the lower range of the estimate as we assume 2018-20 allocation to coincide with the announced caps, but note that the exact number of non-issued permits or date at which they enter the reserve is not crucial for our quantitative results. Over the entire Phase III, note that as much as 1.2 billion permits may not be distributed in total.

[38] The official bank value, or 'total number of allowances in circulation', for year $t-1$ is published in May of year t, and is used for MSR operations over a twelve-month period from 1 September of year t onwards.

4.2 Permit demand

Our aim is to construct a counterfactual scenario for CO_2 emissions of the EU ETS perimeter, i.e. emissions as they would be absent the scheme but accounting for industrial production growth and with all complementary energy and climate policies in place. To that end we use a simple decomposition of baseline CO_2 emissions into three Kaya indexes

$$CO_2\,emissions = \underbrace{Production}_{economic\ activity} \times \underbrace{\frac{Energy}{Production}}_{energy\ intensity} \times \underbrace{\frac{CO_2\,emissions}{Energy}}_{carbon\ intensity}. \tag{11}$$

Figure 4a depicts the reconstructed and projected trajectories of these three indexes between 1990 and 2050. We assume that the permit price has negligible impacts on both production and energy intensity.[39] Specifically, we compute the production index ex post from Eurostat sector-level production data and consider a 1% p.a. production growth from 2018 on.[40] Next, we compute the energy intensity index ex post from total final energy consumption time series for a proxy of EU ETS sectors (i.e. electricity, heat, industry, and energy industry own use and losses) obtained from the 1990-2015 balance sheets of the International Energy Agency.[41] From 2016 on we use a linear interpolation so that the EU energy efficiency targets for 2020 and 2030 are met and we assume this linear trend to be valid afterwards.

Next, we compute the carbon intensity index ex post by reconstructing EU ETS CO_2 emissions over 1990-2004 based on IEA primary energy consumption data and standard EU-level emission factors.[42] From 2005 on, we need to calculate emissions as they would have been absent the ETS. We assume that the permit price may have driven some fuel switching (thus impacting the carbon content of energy) but not the development of renewable energy.[43] To isolate and account for renewable deployment while neutralizing fuel switching in the baseline emissions, we fit a linear relationship between renewable deployment and the carbon content of energy prior to Phase II. We extrapolate this first-pass relationship for later years using observed renewable deployment for 2008-2015 and then assuming that EU renewable targets set for 2020 and 2030 (and their continuation afterwards) are attained linearly. Computing

[39] Regarding production, this seems like a reasonable assumption ex post as there is no evidence of carbon leakage in the EU ETS (Joltreau & Sommerfeld, 2019; Naegele & Zaklan, 2019). Regarding energy intensity, Figure 4a shows that it declines less over 2005-2015 with the ETS in place than over 1990-2005 without the ETS, which provides some a posteriori support to our assumption.

[40] Link to Eurostat data (`http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sts_inpr_a`).

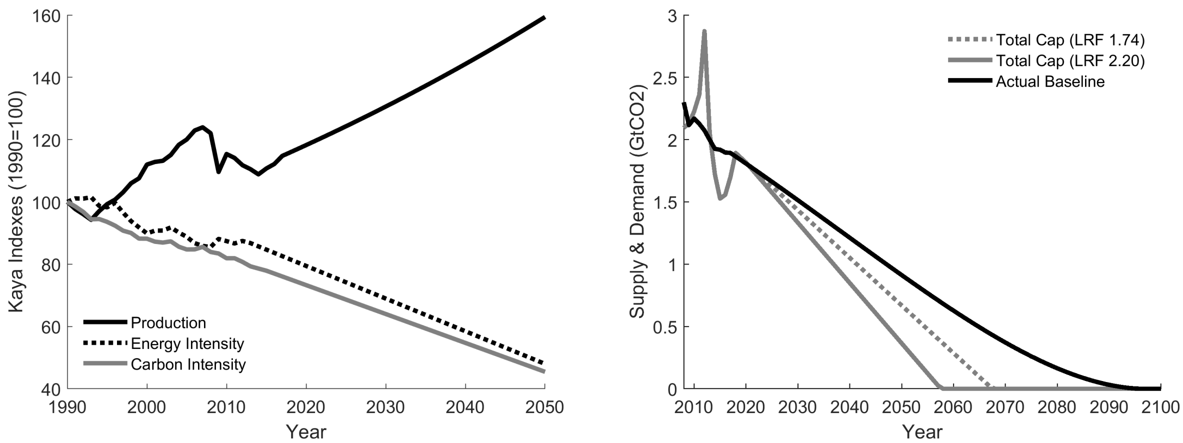
[41] Link to IEA data (`https://www.iea.org/statistics/?country=EU28`).

[42] Link to IEA data. Emission factors are 4.2, 3.1 and 2.4 tCO_2/toe for coal, oil and gas, respectively.

[43] This is supported by evidence that the EUA-price equivalent of renewable subsidies has been significantly higher than EUA prices (Marcantonini & Ellerman, 2015; Marcantonini & Valero, 2017; Abrell et al., 2019).

Figure 4: 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 2.20%, respectively.

the ratio of CO_2 emissions to energy consumed finally yields the carbon intensity index.

Thus, baseline emissions at any point in time are by construction independent of the history of permit prices.[44] Graphically, the black line in Figure 4b depicts the resulting baseline path, which we obtain by plugging the evolution of the three Kaya indexes in (11). It is downward sloping, which is in line with the steady decline in ETS perimeter emissions observed prior to the launch of the ETS. Regarding demand forecasts, we assume that the firm uses a simple heuristic congruent with the deterministic part of an AR(1) process, which is slightly tweaked to accommodate for growth and varying trend.[45] That is, the date-t forecast for demand at date $t+1$ is defined by

$$\hat{u}_{t+1}^{t} = \varphi(1+\gamma_t)u_t + (1-\varphi)\bar{u}_{t+1}^{t}, \tag{12}$$

where u_t is the realized baseline at date t, $\varphi \in [0;1]$ captures some persistence in baseline emissions and γ_t is the expected annual growth rate at date t for future years. We allow the trend $\bar{u}$ to vary over time: typically, it can be thought of as declining over time due to the

[44] This independence assumption becomes less tenable when prices reach higher levels than those observed up to now. Note, however, that endogenous baselines are not considered in similar modeling approaches. Fell (2016) and Perino & Willner (2016, 2017) assume given baseline paths, respectively increasing and constant over time. Similarly, baselines in Beck & Kruse-Andersen (2019) are decreasing over time due to increasing renewable deployment, which slows down over time but nonetheless remains independent of the permit price.

[45] We consider one representative firm using one practically relevant forecast rule, and sticking to it over time. More generally, models of behavioral expectations consider a set of rules with agents rationally switching from one to another based on their relative performances in the recent past, see e.g. Hommes et al. (2019). Additionally, choosing a rule close to an AR(1) process is a sensible choice if one wants to ensure $\hat{x}_{\tau}^{t} = \mathbb{E}_t\{\tilde{x}_{\tau}\}$.

Table 2: Forecasted trends of baseline emissions

Forecast period	Climate Energy Package	$\bar{u}_{2050}/e_{2008}$	$\bar{u}_t = 0$ in
2008-2013	CEP#1	57.5%	2115
2013-2017	CEP#2	50.7%	2105
2018-2100	Reinforced CEP#2	39.7%	2096

Note: 2050 baseline trend is given as a percentage of 2008 verified emissions (2.12 $GtCO_2$).

achievements of complementary policies outside the system's perimeter. Importantly, note that realized baseline emissions u_t differ from their past forecasts $\hat{u}_t^{\tau<t}$. This induces the firm to revise its demand forecasts yearly, and adjust its intertemporal decisions accordingly.

We assume that the trend at date t for some future year $t' > t$ ($\bar{u}_{t'}^t$) is set to align with the attainment of the prevailing Climate Energy Package at date t. These computed trends are reported in Table 2. Moreover, we set γ_t in line with GDP growth forecasts by the European Commission over 2008-2017 and we consider a 2% p.a. growth rate afterwards.[46] Finally for the persistence parameter, we follow Fell (2016) and take $\varphi = 0.9$.[47]

4.3 Permit supply

In year t the firm observes annual permit supply $f_t + a_t$ and it forecasts future permit supply to coincide with the cap trajectory as given in currently prevailing regulatory texts (e.g. EU Directives, Decisions or Communications).[48] As soon as regulation is revised (see e.g. Figure 2) or upon release of actual supply data (e.g. EC Carbon Market Reports), the firm corrects its forecast. Hence yearly forecast updates and year-on-year discrepancies between forecasted and actual supply naturally come about. For instance, the firm considers a total cap path from Phase IV on based on the currently effective linear reduction factor, i.e. 1.74% before and 2.20% after the 2018 reform. From 2021 on, 57% of the total cap is auctioned off.

Usage of Kyoto offsets (viz. CERs and ERUs) is authorized in Phases II and III, up to the cumulative limit $O \approx 1.6$ $GtCO_2$.[49] Within that period, we assume that the firm does not

[46] Link to EC forecasts published in spring of year t for year $t + 1$.

[47] In related contexts, Heutel (2012) and Lintunen & Kuusela (2018) use $\varphi = 0.95$ and $\varphi = 0.8$ respectively. Roughly speaking, the lower φ the more the firm expects future baselines to coincide with the trend. As the trends happen to be relatively close to actual baselines, a lower φ thus implies 'better foresight'.

[48] We only consider stationary sources and exclude the aviation sector (intra-EEA flights) as it was brought under the ETS in 2012, represents a small fraction of the 'regular cap' (∼2%) and regulatory uncertainty lingers pending ICAO's adoption of the relevant CORSIA instruments (European Commission, 2018b).

[49] EU legislation specifies qualitative and maximum limits on offset usage (European Parliament & Council, 2003; European Commission, 2013). Our estimate for O is based on aggregated individual entitlements.

decide how many offsets to surrender, but note this simplification is innocuous for our ex-ante analysis (i.e. from Phase IV on) in Section 5. Specifically, o_t is given at the beginning of year t and equal to observed offset usage. Moreover, the firm forecasts in year t that the remaining allowed quantity of offsets that can be surrendered from year $t+1$ on (i.e. $O - \sum_{\tau=2008}^{\tau=t} o_\tau$) is equally split across the remaining years of the period. Again, discrepancies between realized and forecasted offset usage naturally arise, which causes forecasts to shift over time.

Graphically, the grey line in Figure 4b depicts total annual supply $\{f_t + a_t + o_t\}_t$. The peak in 2011-12 is due to a massive use of offsets, totalling about 1 $GtCO_2$ over Phase II (Trotignon, 2012; de Perthuis & Trotignon, 2014). The following dip is due to the 900 $MtCO_2$ backloading over 2014-16 and to non-issued Phase-III permits, totalling about 600 $MtCO_2$ over 2013-17 (e.g. European Commission, 2015). From Phase IV on, supply coincides with the announced cap which is set to decline at a yearly linear reduction factor of 2.20%, implying that supply is nil from 2058 on. Figure 4b also shows the pre-reform cap path with an LRF of 1.74%.

4.4 Ex-post calibration

We choose to calibrate the firm's discount rate and horizon on 2008-2017 data for two practical reasons. First, we leave aside the trial Phase I (2005-2007) since banking and borrowing across Phases I and II was not authorized, de facto restricting the firm's horizon, at least with regard to Phase-I permits usage.[50] Second, we deliberately exclude 2018 so that we need not take a stance on the firm's responsiveness degree for the calibration (as the reform passed in early 2018, its effects can be anticipated and factored in from this point on).

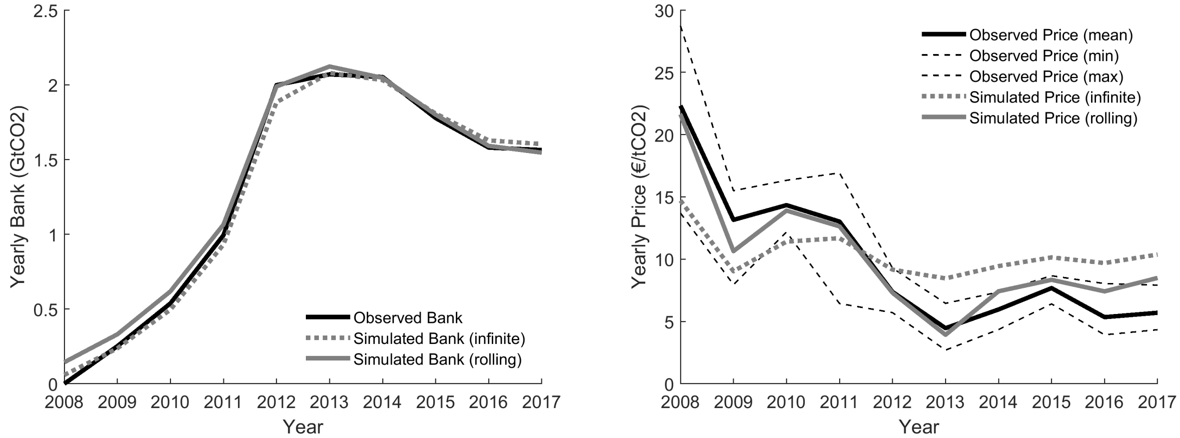
We consider linear marginal abatement cost functions with time invariant slopes, i.e. $C_t'' = c$ for all t.[51] We take this is as a conservative assumption given that we have limited empirical and theoretical guidance regarding the evolution over time of the marginal abatement cost slope due to interplay between the exhaustion of low-hanging abatement opportunities and low-carbon innovation, see e.g. Bréchet & Jouvet (2008) for a discussion. Note however that the linear intercept of the marginal abatement cost curve is gradually lowered over time as the actual baseline path is downward sloping.

[50] As market participants gradually understood that the Phase-I cap was in fact not binding and permits could not be banked into Phase II and beyond, the EUA spot price dropped to zero.

[51] This can be viewed as a local Taylor approximation of more general functional forms and implies that compliance demand in (1) is linear in the permit price. This is a standard assumption in our case (Schennach, 2000; Ellerman & Montero, 2007; Perino & Willner, 2016; Lintunen & Kuusela, 2018; Kollenberg & Taschini, 2019). As a constant scaling parameter, c is neutral vis-à-vis both the firm's intertemporal decision making and our analysis in Section 5 (i.e. it only changes the levels, but not the shapes, of the price paths).

Figure 5: Calibration results based on 2008-2017 bank and price data

(a) Bank (b) Price



We calibrate the firm's parameters following a two-step procedure in the spirit of a standard least squares maximum likelihood estimation with one free parameter. First, we select r given h or h given r so that the simulated bank path deviates the least from the observed bank path over 2008-2017. Second, given these select 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.[52] In these two steps the free parameter is calibrated by minimizing the distance between simulated and observed paths. Our best-fit results are reported in Table 3 and Figure 5 depicts the observed and calibrated simulated paths over 2008-2017.

With an infinite horizon $h = \infty$, we find that a discount rate $r \approx 7\%$ would best replicate past banking with a fit of 53 $MtCO_2$/year. This aligns with general rates of return on risky assets (Jordà et al., 2019) although in the higher range of the rates that can be implied from futures' yield curves since early Phase II, see Table 1.[53] Additionally, one might argue that since permits can be banked for hedging purposes, required returns should be below those for standard risky assets. With a rolling horizon, we thus set $r = 3\%$, which is a central value for inferred discount rates, and find $h \approx 13$ years with a similar fit of 65 $MtCO_2$/year.[54] The

[52]Note that this two-step calibration approach is legitimate since a constant c over time does not influence the firm's banking strategies, which thus only depend on its discount rate and horizon.

[53]Ellerman & Montero (2007) analyze the permit-specific CAPM beta to select appropriate values for the discount rate. We choose a different approach and impute implicit rates from futures' yield curves. Indeed, futures markets can provide information about the discount rates applied by market participants in valuing present vs. future permits. Bredin & Parsons (2016) and Trück & Weron (2016) find similar values.

[54]Fell (2016) and Kollenberg & Taschini (2016) pick the same value while Beck & Kruse-Andersen (2019) and Perino & Willner (2016, 2017) use 5% and 10%, respectively. In terms of sensitivity, our calibrated h is increasing with the select value of r: 12, 14, 15 and 16 years for an r of 1 (or 2), 4, 5 and 6%, respectively.

Table 3: Calibration results based on 2008-2017 bank and price data

Horizon type	Horizon & discount rate	Marginal abatement cost
Infinite	$h = \infty^\star \quad r = 7.06\%$ (std.dev = 52.9 $MtCO_2$)	$c = 5.53 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 4.04 €/tCO_2)
Rolling	$h = 13 \quad r = 3\%^\star$ (std.dev = 64.9 $MtCO_2$)	$c = 5.72 \cdot 10^{-8}$ €/$(tCO_2)^2$ (std.dev = 2.12 €/tCO_2)

Note: Parameters are calibrated by minimizing the distance between annual simulated and observed banking (center column) and price (right column) levels over 2008-2017. A $\star$ indicates parameters fixed exogenously.

values we get for c are similar with the calibrated infinite and rolling horizons, in the order of $5.5 \cdot 10^{-8}$€/$(tCO_2)^2$ and in line with dedicated studies, e.g. $4.3 \cdot 10^{-8}$€/$(tCO_2)^2$ (Böhringer et al., 2009) or $5.7 \cdot 10^{-8}$€/$(tCO_2)^2$ (Landis, 2015).[55] That said, the price fit is twice as good with the calibrated rolling horizon, i.e. 2.1 vs. 4.0 €/tCO_2/year with the calibrated infinite horizon. Our calibrated estimates for r and h have no counterparts in the empirical literature dedicated to the EU ETS and thus constitute first-pass assessments at the market level.[56]

From an ex-post evaluation perspective, the regulator has two options to capture the alleged shortsightedness on the part of market participants: an infinite horizon with a 'high' discount rate vs. a rolling horizon with a lower discount rate (see Section 2.2). Our calibration results give more support to the second option since it can reconcile the past bank dynamics with implicit discount rates. Additionally, the rolling horizon is better able to pick up the observed yearly-averaged price changes, in sign and amplitude (see Figure 5b). Coincidentally or not, the calibrated horizon extends more or less as far as the closest policy target regulation aims at achieving (see Figure 2). In any case, distinguishing between these two options is not only important for ex-post evaluation, but it will also prove crucial for policy design and ex-ante assessment, as the next section shows.

5 Simulations: The 2018 EU ETS reform

In this section, we use our calibrated model to provide a quantitative assessment of different aspects of the 2018 EU ETS reform. We offer a twofold perspective by comparing outcomes

For r=7.06%, any horizon $h \geq 27$ years yields the same calibration output as the infinite horizon.

[55]From Böhringer et al. (2009), we obtain 1 $GtCO_2$ of abatement in 2020 for 43 €/tCO_2, which pins down the marginal abatement cost slope. See Table 4 for 2020 in Landis (2015).

[56]Our representative firm approach implicitly assumes that all firms apply the same horizon and discount rate. In practice, they might use different horizons and rates so our results should be seen average estimates.

under the calibrated infinite and rolling finite horizons. In other words, we take the perspective of two regulators in assessing the reform impacts ex ante: one who believes that market participants use an infinite horizon in conjunction with a discount rate of about 7%, the other who believes that they use a 13-year rolling horizon with a discount rate of 3%. These two regulators can also make two polar assumptions about the ability of market participants to understand the interplay between their decisions in the competitive equilibrium and the control impacts on supply over time: full or zero responsiveness (see Section 3.5).

Specifically, we first analyze the general impacts of the reform on equilibrium price and bank paths as well as on cumulative emissions. We then focus on the MSR efficiency in achieving the induced cumulative volume of emissions as well as on its potential to adjust cumulative emissions in the face of exogenous reductions in permit demand. We consider several scenarios separating out the impacts of the three main reform elements (the increase in the Linear Reduction Factor from 1.74% to 2.20% from 2021 on, the introduction of the Market Stability Reserve from 2019 on, and its reinforcement with the cancellation mechanism from 2023 on) and the firm's horizon and responsiveness: the status quo, i.e. the continuation of Phase III rules without reform (NO REF), the sole LRF increase (NO MSR) and the same scenario with the MSR (MSR). The MSR scenario is divided into four sub-scenarios, depending on whether the cancellation mechanism is on (C) or off (N) and whether the firm is fully responsive to the control (F) or not at all (Z). For instance, the scenario MSR F+C features a fully responsive firm and the MSR augmented by cancellations. We present our results until 2100 since all permits are used and emissions are zero in all scenarios by that time.

5.1 General impacts of the reform

The left (resp. right) hand side of Figure 6 depicts the equilibrium price, bank and cumulative withdrawals (MSR stock + cancelled volumes) paths with the calibrated infinite (resp. rolling) horizon.[57] Comparing NO REF with NO MSR, we see that the sole LRF increase induces higher price levels throughout and a shorter banking period, albeit with higher bank volumes. We also note that all post-reform price paths with the infinite horizon reach a peak the year the bank becomes empty (and emissions become nil). This always occurs a few years after the cap effectively shrank to zero as a result of intertemporal cost minimization. After the peaks, as there are no permits left in circulation (both the cap and bank are nil), the firm does no longer emit and abates its baseline emissions. As baseline emissions drop to zero, so

[57] All price paths are given in current € values where we use the observed annual inflation rates between 2008 and 2018 and take 1.5% per annum afterwards.

do the yearly abatement efforts and associated costs at the margin, hence the declining price path. We observe similar though less clear-cut patterns with the rolling horizon.

Introducing the MSR on top of the larger LRF further hikes the price and reduces the bank. Crucially, the firm's horizon matters so let us begin with the infinite horizon. A noticeable result is that price and banking paths with the MSR are very much alike irrespective of both cancellations and the firm's responsiveness, at least before set-aside permits are released into circulation starting in the 2050's. From this point onwards only, cancellations sustain higher price levels. Specifically, with zero responsiveness, price and bank paths are identical with and without cancellations before reinjections effectively materialize. With full responsiveness, price and bank paths are slightly higher with than without cancellations. Indeed, as the firm foresees larger future reinjections without cancellations, it forecasts lower overall abatement efforts which translate into lower bank and price levels before then. This in turn implies that the MSR withdraws less permits (in volume and duration) and starts reinjecting sooner.

Relatedly, another important result is that, despite the fact that the MSR eats away at the bank, it takes about two decades for the latter to fall below the intake threshold $\bar{b}$. As Figure 6e shows, the MSR withdraws permits each year until then. As the bank then remains above the outtake threshold $\underline{b}$ for another 15 years or so, which translates into a plateau in Figure 6e, reinjections do not occur before the 2050's.[58] Because most of the set-aside permits have been cancelled by then, only about 100 million permits are reinjected with cancellations. In turn, cumulative emissions are reduced by about 5 $GtCO_2$. Importantly, they are also reduced without cancellations, albeit by a lower amount, since the MSR does not have time to empty before the market terminates.[59] Finally, we note that cumulative MSR intakes (and ensuing cancellations when applicable) are smaller when the firm is responsive.

Let us now turn to the rolling finite horizon. Introducing the MSR induces a more pronounced price increase and reduces the bank less sharply than with the infinite horizon.[60] Importantly,

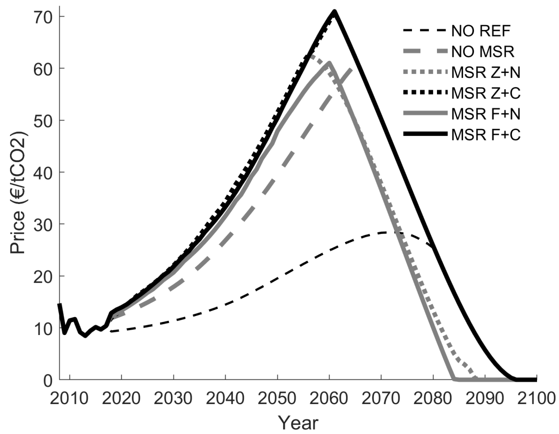
[58]The slight bank upticks occurring after the intake and outtake thresholds are passed arise as the MSR suddenly stops withdrawing or starts reinjecting permits, respectively. Without cancellations, annual reinjections of 100 million permits induce a second banking period when the firm is responsive (see Perino & Willner (2016) for a similar effect) while the bank never drops to zero in the first place and fluctuates around the outtake threshold when the firm is not responsive.

[59]With higher baseline emissions the MSR would have time to empty, thereby recovering the conventional result that the MSR without cancellations preserves the cumulative cap (Perino & Willner, 2016). That said, note that when the firm is responsive, the MSR reinjects more permits than effectively withdrawn as the terminal MSR stock is lower than the amount of permits the MSR is initially seeded with.

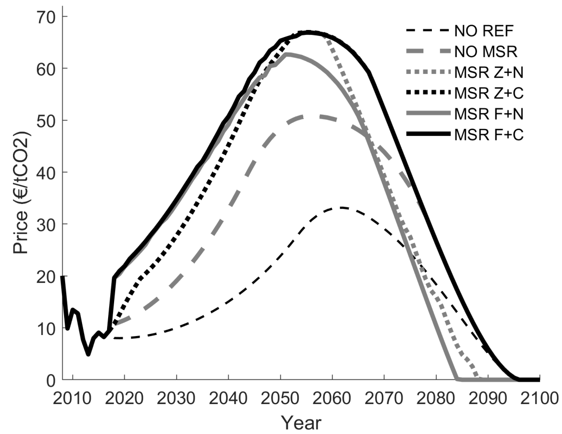
[60]With the rolling horizon, yearly MSR-driven supply cutbacks have a larger relative impact on the firm's perceived overall abatement effort and more of it is being abated earlier on given the lower discount date. These two effects concur to yield a higher price and bank than with the infinite horizon.

Figure 6: MSR impacts as a function of the horizon, responsiveness and cancellations

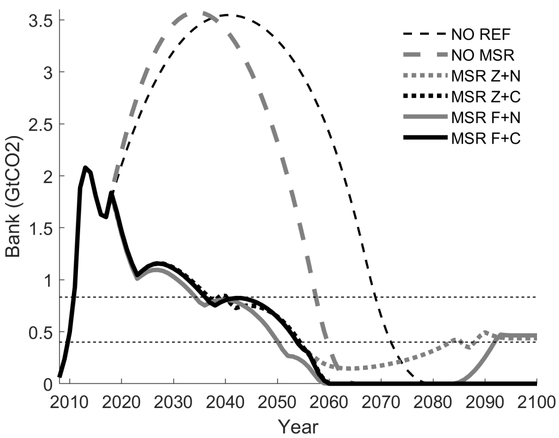
(a) Price ($h = \infty$, $r = 7.06\%$)



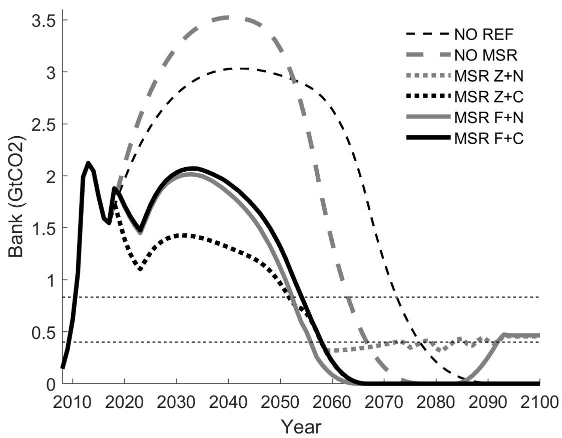
(b) Price ($h = 13$, $r = 3\%$)



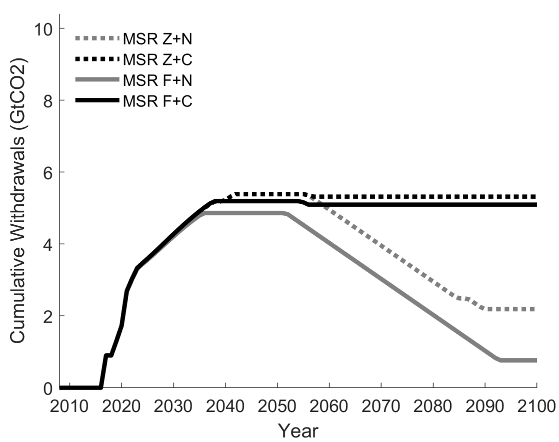
(c) Banking ($h = \infty$, $r = 7.06\%$)



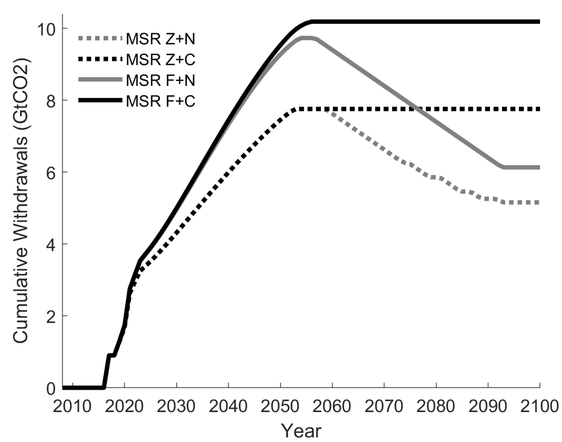
(d) Banking ($h = 13$, $r = 3\%$)



(e) Cumul. withdrawals ($h = \infty$, $r = 7.06\%$)



(f) Cumul. withdrawals ($h = 13$, $r = 3\%$)



Note: All scenarios have an LRF of 2.20% (except NO REF for which the LRF is 1.74%) and feature full (F) or zero (Z) responsiveness to the control with (C) or without (N) cancellations alternatively.

the price and bank paths are significantly higher when the firm is responsive but, crucially, irrespective of cancellations. Indeed, as the MSR initially cuts back on supply and reinjections are far off into the future and beyond the firm's horizon, cancellations are initially irrelevant for the responsive firm.[61] As it 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.[62] This implies higher price and bank paths with responsiveness than without until the 2050's. In particular, only a responsive firm with the rolling horizon leads to a price surge in 2018 of a magnitude similar to the one observed in practice (see Figure 1).

Moreover, as the bank remains above the intake threshold for more than three decades with higher levels, the rolling horizon implies larger withdrawals (in volume and duration) than the infinite horizon.[63] Even if the bank decreases from the intake to the outtake thresholds in just a few years, there are no permits left in the MSR with cancellations by that time and no reinjections occur. In total, this translates into a larger contraction in cumulative suppy, in the order of 6 to 10 $GtCO_2$ when the firm is responsive, respectively with and without cancellations. Finally, we stress again that the rolling horizon coupled with responsiveness leads to larger MSR intakes (and ensuing cancellations when applicable) than without responsiveness.

5.2 Focus on cumulative emissions and cost efficiency

The introduction of the MSR endogenizes the cumulative emissions cap: it is now a function of the implementation of the cancellation mechanism as well as of the firm's horizon, discount rate and responsiveness to the control. Table 4 contains the 2008-2100 cumulative emissions that obtain with the MSR and an LRF of 2.20% in each of the eight possible combinations, and the equivalent LRF (LRF_{eq}) used from 2021 on that would, without the MSR, yield the same cumulative emissions. With the rolling horizon, full responsiveness, and cancellations, the LRF_{eq} can be as high as 2.95%, meaning that the cap shrinks by 64.9 $MtCO_2$ per annum and becomes nil in 2048 (w.r.t. 48.4 $MtCO_2$ and 2058 with an LRF of 2.20%).

Relative to the MSR, the LRF_{eq} has the advantage of specifying a clear supply path without

[61]As with the infinite horizon, price and bank paths are identical with and without cancellations when the firm is not responsive to the MSR for as long as reinjections do not effectively kick in.

[62]This adjustment procedure eventually converges to a fixed point, see Section 3.5. As time goes by and reinjections enter the horizon, the responsive firm forecasts less of an abatement effort without cancellations, implying lower bank and price levels than with cancellations.

[63]Given the respective bank paths, annual intakes are larger when the firm is responsive, though note that the duration of the intake period hardly depends on the firm's responsiveness.

Table 4: MSR impacts on cumulative emissions and total compliance costs

Horizon	Respons.	Cancel.	Cumul. emissions	LRF_{eq}	Efficiency loss	Interaction
Infinite	Zero	Off	57.86 $GtCO_2$	2.28%	9.0%	16.4%
		On	55.16 $GtCO_2$	2.48%	0.2%	4.6%
	Full	Off	59.26 $GtCO_2$	2.18%	11.5%	11.1%
		On	55.39 $GtCO_2$	2.46%	0.2%	3.1%
Rolling	Zero	Off	54.87 $GtCO_2$	2.50%	9.0%	11.7%
		On	52.72 $GtCO_2$	2.70%	0.6%	0.6%
	Full	Off	53.88 $GtCO_2$	2.59%	7.9%	1.7%
		On	50.29 $GtCO_2$	2.95%	-2.2%	-5.2%

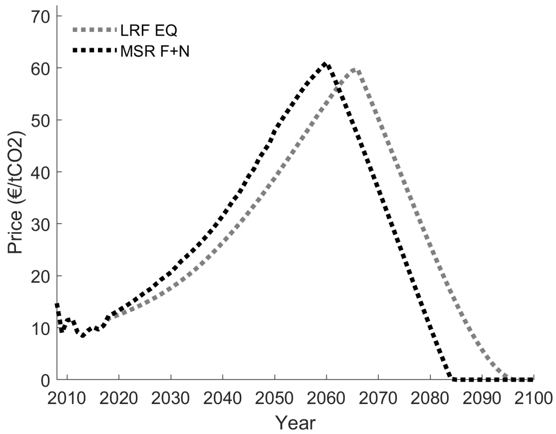
Note: LRF_{eq} is the Linear Reduction Factor from 2021 on yielding the same cumulative emissions over 2008-2100 without MSR as with the MSR and an LRF of 2.20%. Efficiency loss measures the additional cumulative compliance costs implied by the MSR with an LRF of 2.20% relative to the sole LRF_{eq}. Interaction measures the reductions in cumulative emissions due to the MSR with an LRF of 1.74% that are additional to those obtained with the MSR and an LRF of 2.20%, as a percentage of the foregone cumulative emission reductions from keeping the LRF constant at 1.74% instead of raising it 2.20%, without the MSR (9.0 $GtCO_2$).

inducing a priori distortions in the firm's intertemporal allocation of abatement. We compute the additional total costs of achieving a given cumulative cap with the MSR relative to with the LRF_{eq}, reported in the second-to-last column of Table 4. With the infinite horizon, we see that the MSR-induced efficiency loss is negligible with cancellations and in the order of 10% without. We observe similar results with the rolling horizon, with the notable exception that implementing the MSR can even be less costly than the LRF_{eq}. Indeed, given the pre-existing distortions due to the rolling horizon, there is room for improving efficiency and the MSR augmented by cancellations is found to achieve a 2% cost reduction when the firm is responsive. The above suggests that from a regulatory cost point of view, the cancellation mechanism is a key feature of the 2018 reform to flank the MSR with.

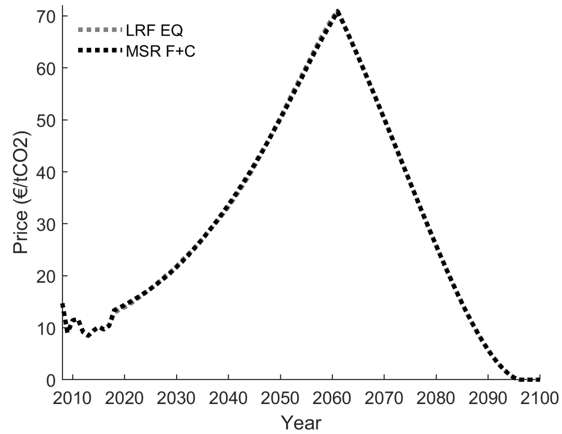
To get a better understanding of these results, Figure 7 depicts the equilibrium price paths that obtain with the MSR vs. the LRF_{eq} in the four cases where the firm is responsive. We see that without cancellations, for a given cumulative cap, the MSR induces too much abatement early on and too few later on (because of fixed, inflexible yearly reinjections) with both the infinite and rolling horizons, hence the higher costs. This 'imbalance' is partly leveled out by the cancellation mechanism. Quite remarkably in this case, with the infinite horizon, the price paths with the MSR and the LRF_{eq} almost exactly overlap, hence the negligible cost difference. With the rolling horizon, in addition of lower costs, we see that the MSR with cancellations can bring about another comparative advantage over the LRF_{eq}. Specifically, with the LRF_{eq} the firm does not properly factor in the long-term constraint on emissions,

Figure 7: Comparative price paths with the MSR and an equivalent LRF without MSR

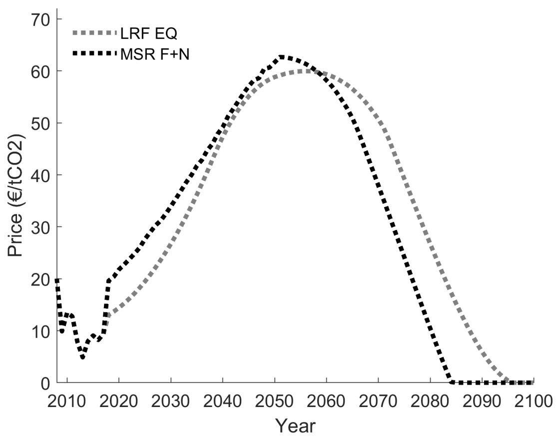
(a) MSR F+N ($h = \infty$, $r = 7.06\%$)



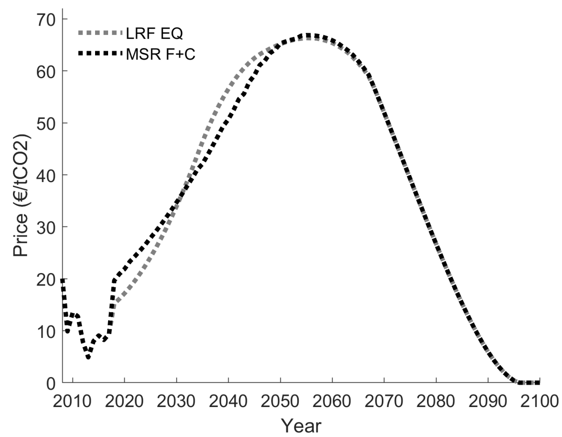
(b) MSR F+C ($h = \infty$, $r = 7.06\%$)



(c) MSR F+N ($h = 13$, $r = 3\%$)



(d) MSR F+C ($h = 13$, $r = 3\%$)



Note: MSR scenarios use an LRF of 2.20%. LRF EQ scenarios use the corresponding LRF_{eq} in Table 4.

resulting in 'too low' price levels early on. The MSR more than counteracts this aspect by forcing the firm to anticipate a more stringent constraint, resulting in higher price levels early on, which can be a policy objective in itself (e.g. to spur low-carbon investments).

It is also worth analyzing the interaction between the LRF increase from 1.74% to 2.20% and the implementation of the MSR. That is, are these two elements complements, substitutes or independent in terms of resulting cumulative emissions? First note that independently of the introduction of the MSR, the LRF increase reduces cumulative emissions by 9 $GtCO_2$ (from 68 down to 59 $GtCO_2$ over 2008-2100). To evaluate the LRF-MSR interaction, we measure how much of these missing 9 $GtCO_2$ would be permanently withdrawn from circualtion by the

MSR with an LRF of 1.74%, i.e. on top of the cumulative withdrawals the MSR achieves with an LRF of 2.20%. As the last column of Table 4 shows, the small magnitudes of the additional cumulative withdrawals suggest that in large part the LRF increase and the introduction of the MSR are independent measures, especially with cancellations. In other words, the MSR on its own would not make up for much of the foregone cumulative emission reductions due to an unchanged LRF. Noticeably, the MSR even has less of an impact on cumulative emissions without the LRF increase in one instance, viz. in presence of cancellations with a responsive firm that uses the rolling horizon.

5.3 Focus on cumulative emissions and exogenous abatement

Given a fixed cumulative emissions cap, if one additional tCO_2 under the cap is not emitted today because of some event, factor, measure or policy external to the scheme, i.e. independently of the permit price, it will necessarily be emitted at some point in the future, and vice versa. This is generally referred to as a 'waterbed effect' over time.[64] As a consequence of the reform, however, the cumulative emissions cap is no longer a key fixed policy element but has rather become a market outcome, hence uncertain ex ante. As such, it can be affected by any exogenous factors that curtail permit demand, e.g. complementary policies. Specifically, some share of that additional abated tCO_2 can be made permanent, i.e. not emitted in the future, and this share typically decreases with the date when the tCO_2 is not emitted initially. In other words, the reform partially and temporarily 'punctures the waterbed' (Perino, 2018). This aspect of the reform has received attention in the policy debate because it allows some adjustments reflecting the interaction between the market and exogenous factors.[65]

We contribute to this debate by quantifying and comparing the partial and temporary nature of the puncture with our calibrated infinite and rolling finite horizons. To this end, we analyze the impacts on cumulative emissions of one-shot marginal shifts in baseline emission levels, which can be attributed to marginal changes in economic activity, renewable deployment, coal phase-out, etc. Specifically, we consider unanticipated one-shot 10 $ktCO_2$ drops in baseline emissions occurring in different years. The arbitrary size and sign of the shift do not affect

[64]A waterbed effect also exists over space within the system in the form of carbon leakage or fuel switch but is beyond the scope of this paper. That is, if one tCO_2 is abated independently of the permit price in some location, the induced permit price decrease will trigger one tCO_2 emission elsewhere, e.g. in a more CO_2-intensive location or via fuel switching as CO_2-intensive technologies become relatively cheaper.

[65]The prolonged price downturn that led to the reform has in large part been attributed to the economic recession and the achievements of overlapping renewable and energy efficiency policies (e.g. de Perthuis & Trotignon, 2014; Bel & Joseph, 2015; Ellerman et al., 2016). Some scholars also see such cap adjustments as a way of partially preserving the environmental integrity of these complementary policies.

Table 5: Long-term effects of one-shot marginal shifts in baseline emissions

			Year of shift				
Horizon	Respons.	Cancel.	2020	2025	2030	2035	2040
Infinite	Zero	Off/On	53%	42%	33%	19%	6%
	Full	Off	49%	38%	24%	0%	0%
		On	54%	43%	32%	12%	0%
Rolling	Zero	Off/On	14%	14%	15%	17%	20%
	Full	Off	22%	24%	25%	27%	28%
		On	23%	24%	26%	27%	28%

Note: Permanent impacts on cumulative emissions as a percentage of the shift's magnitude.

our results provided that it is small enough in absolute terms so that the MSR intake cut-off date is preserved. This allows us to measure and compare marginal impacts in a meaningful way by ruling out threshold effects.[66] Table 5 contains the permanent shares of the shift as a function of cancellations and the firm's horizon and responsiveness. The higher this share, the more significant the puncture of the waterbed.

Our results with the infinite horizon, full responsiveness and cancellations are consistent with other studies (Beck & Kruse-Andersen, 2019; Carlén et al., 2019; Perino et al., 2019). That is, the puncture is partial (notice we find smaller punctures in size) and smaller the closer the shift from the intake cut-off date (a zero puncture obtains for shifts occurring after this date). By contrast, because the MSR never empties completely without cancellations, we find that the puncture is not zero in this case, although it remains smaller than with cancellations. As a novelty, we also provide results in other cases. First, with zero responsiveness, the puncture is independent of cancellations as they have no bearing on the bank paths before the 2050's. Second, a smaller puncture (twice as small for early shifts) obtains with the rolling horizon but it is remarkably stable and even slightly increasing with the year of the shift.

How can we interpret these results? Note that the decrease in baseline emissions in year t does not translate into a one-to-one increase in banking in year t, which itself does not lead to a one-to-one increase in cumulative MSR withdrawals (terminal MSR stock or cancellations).[67] That is

$$\overbrace{-X_t}^{\text{baseline shift}} \longrightarrow \overbrace{Y_t \cdot X_t}^{\text{bank increment}} \longrightarrow \overbrace{W_{cumul} \cdot Y_t \cdot X_t}^{\text{cumulative withdrawals}}, \tag{13}$$

[66] We note that the intake cut-off date is a key indicator of the MSR-induced resilience to external demand shocks. Analyzing how it depends on the demand shock characteristics is left for future work.

[67] With (resp. without) the cancellation mechanism, these extra withdrawn permits are cancelled (resp. still sitting in the MSR when the program ends) and never return to the market.

Table 6: Decomposition of the effects of one-shot marginal shifts in baseline emissions

		Year of shift				
Horizon	Share	2020	2025	2030	2035	2040
Infinite	Y_t	60%	57%	54%	53%	52%
	W_{cumul}	90%	76%	59%	23%	0%
Rolling	Y_t	24%	28%	31%	34%	40%
	W_{cumul}	92%	87%	85%	80%	70%

Note: Y_t and W_{cumul} are defined in (13). Case with the cancellation mechanism and full responsiveness.

with X_t the absolute value of the size of the negative baseline shift, $Y_t \in [0,1]$ the share of the shift that is passed into the bank as a result of intertemporal smoothing, and $W_{cumul} \in [0,1]$ the share of the bank increment converted into cumulative withdrawals via the MSR.[68] Table 6 gives Y_t and W_{cumul} in the case of cancellations and full responsiveness, the product of which yields the overall impact on cumulative emissions listed in Table 5. We see that for a given baseline shift year t, Y_t is always larger with the infinite horizon. This is because with the rolling horizon the firm has less room to spread the shift over time and because its bank is larger to start with (due to the MSR). Both elements concur to a lower incentive to bank the freed-up permits, and note that the wedge in Y_t between the infinite and rolling horizons decreases with the shift year. Conversely, for a given shift year, W_{cumul} is always larger with the rolling horizon. This is because the intake cut-off always occurs later in this case so that the MSR has more time to absorb the initial bank increment.[69] Naturally, W_{cumul} decreases with the shift year as there is less time before the intake cut-off occurs.

6 Concluding remarks

In this paper, we have made three main contributions to the literature on emissions trading as a climate policy tool. As a first contribution, we have built a model of competitive emissions trading under uncertainty with supply control departing from the existing literature along two dimensions: firms can (1) use rolling horizons as a way of addressing uncertainty, i.e. optimize over a truncated horizon given realistic forecasts of all relevant exogenous factors, implement only the first-period optimal decisions, and repeat the process at all subsequent periods with

[68] In the first order, $W_{cumul} \approx 1-(1-0.24)^x(1-0.12)^y$ where x and y are the number of years between the year of the shift and the intake cut-off year with an R of 24% and 12% respectively (Perino, 2018).

[69] Given its exponential property (see footnote 68), W_{cumul} does not (1) decrease by much over 2020-40 with the rolling horizon and (2) differ by much between the infinite and rolling horizons for early shifts.

updated forecasts and an equally long horizon and (2) exhibit bounded responsiveness to the control, i.e. various degrees of sophistication in understanding the interplay between their decisions in a competitive equilibrium and the control impacts on supply over time. We have motivated the introduction of rolling horizons based on anecdotal evidence in the context of the EU ETS and by drawing from the existing literature in other related fields.

As a second contribution, we have tailored the model to the EU ETS design and calibrated the market-wide discount rate and planning horizon to replicate 2007-2018 market developments. A rolling horizon of about a dozen years is able to reconcile the past bank dynamics with a central value for discount rates implied from futures' yield curves (∼3%). By contrast, with an infinite horizon, only a rate of about 7% can replicate past banking but is higher than the implicit rates. Moreover, the calibrated infinite horizon does not capture the size and sign of yearly-averaged past price variations as well as the calibrated rolling horizon. Our first-pass estimates are a valuable addition to the empirical literature on the EU ETS. More generally, our calibration exercise allows us to infer useful information (1) substantiating claims that shortsightedness could be crucial in understanding price and banking dynamics and assessing systems' performances ex post (e.g. Ellerman et al., 2015; Fuss et al., 2018) and (2) increasing the quantitative relevance of model simulations for ex-ante policy assessment.

As a third contribution, we have used our calibrated model to assess the impacts of the 2018 EU ETS reform depending on what is assumed about the firms' horizon and responsiveness. We have decomposed the impacts of its three main elements (viz. the increase in the Linear Reduction Factor from 1.74% to 2.20%, the introduction of the Market Stability Reserve and its reinforcement with the cancellation mechanism) and quantified the interactions between them and the firms' horizon and responsiveness. This has important implications for policy design and evaluation, some of which are listed in the Introduction.

More broadly, our framework as well as our calibration and simulation results in the context of the EU ETS contribute to improving one's understanding of the intertemporal performances of ETSs in general – a topic which is high on both policy and research agendas (Hasegawa & Salant, 2015; Ellerman et al., 2016) – and the interactions between intertemporal trading and supply-side controls. Specifically, our framework can serve as a good basis for an assessment of the MSR for the upcoming review in 2021 by extending our ex-post analysis when additional data is available and investigating the impacts of changing the MSR parameters $\bar{b}$, $\underline{b}$, R and I. This could also be an opportunity to analyze more profound design changes in the MSR or the introduction of a price floor in place or on top of the MSR, e.g. via an auction reserve price as is currently discussed (Newbery et al., 2019). Relatedly, our framework is amenable to amendments and calibration to other systems, e.g. the Regional Greenhouse Gas Initiative or the linked California-Québec ETS where specific forms of price corridors, intertemporal trading provisions and compliance cycles are in place.

Finally, our paper constitutes a first step towards analyzing hybrid emissions trading systems when firms employ rolling finite horizons or exhibit bounded responsiveness to supply-control instruments. Several alleys for future research seem fruitful. A first line of important research would be optimal policy selection when the regulator is aware that firms are shortsighted or limitedly responsive. This would imply revisiting the normative analysis of price vs. quantity vs. hybrid policies and establishing welfare preferences over hybrid policy designs, e.g. price vs. banking corridors. Second, follow-up work could endogenize the firms' horizon as a result of a trade-off between compliance cost minimization and forecasting or trading costs. Because our results suggest that shortsighted firms would be better off using longer horizons with the MSR in place – this would lead to a smaller contraction in cumulative supply and thus smaller overall compliance costs – a third line of research could analyze the interdependence between design elements and the firms' horizon. Interestingly, this aspect could also be approached empirically with an ex-post analysis of the first years of functioning of the MSR.

Acknowledgments

We would like to thank Anna Cretì, Christian Flachsland, Peter Kruse-Andersen, Michael Pahle, John Parsons, Grischa Perino, Christian de Perthuis, Katheline Schubert, Luca Taschini and Oliver Tietjen for valuable conversations at different stages of this project. We also thank an anonymous referee from FAERE and received helpful comments from participants at various conferences (FSR Climate 2018, AFSE-DGT 2018, IAEE 2019, EAERE 2019) and seminars (LSE, MCC-PIK, Universität Hamburg). All remaining errors are ours.

Simon Quemin acknowledges funding from 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 Assigment Agreements for Communication Assistance, and from 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. Both authors further acknowledge funding from 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.

Abrell, J. & Rausch, S. (2017). Combining Price and Quantity Controls under Partitioned Environmental Regulation. *Journal of Public Economics*, **145**, 226–42.

Baudry, M., Faure, A. & Quemin, S. (2019). *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.

Bel, G. & Joseph, S. (2015). Emission Abatement: Untangling the Impacts of the EU ETS and the Economic Crisis. *Energy Economics*, **49**, 531–9.

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.

Borenstein, S., Bushnell, J., Wolak, F. A. & Zaragoza-Watkins, M. (2016). *Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design*. Working Paper 20999, NBER.

Branch, W. A. (2004). The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations. *Economic Journal*, **114**, 592–621.

Bréchet, T. & Jouvet, P.-A. (2008). Environmental Innovation and the Cost of Pollution Abatement Revisited. *Ecological Economics*, **65**(2), 262–5.

Bredin, D. & Parsons, J. (2016). Why is Spot Carbon so Cheap and Future Carbon so Dear? The Term Structure of Carbon Prices. *Energy Journal*, **37**(3), 83–107.

Brock, W. A. & Hommes, C. H. (1997). A Rational Route to Randomness. *Econometrica*, **65**(5), 1059–95.

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 Revised, KU Leuven.

Burtraw, D. & Keyes, A. (2018). Recognizing Gravity as a Strong Force in Atmosphere Emissions Markets. *Agricultural & Resource Economics Review*, **47**(2), 201–19.

Caliendo, F. & Aadland, D. (2007). Short-Term Planning and the Life-Cycle Consumption Puzzle. *Journal of Economic Dynamics & Control*, **31**(4), 1392–415.

Cantillon, E. & Slechten, A. (2018). Information Aggregation in Emissions Markets with Abatement. *Annals of Economics & Statistics*, **132**, 53–79.

Carbone, E. & Hey, J. D. (2001). A Test of the Principle of Optimality. *Theory & Decision*, **50**(3), 263–81.

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.

Climate Strategies (2015). *Is a Market Stability Reserve Likely to Improve the Functioning of the EU ETS?* Report, DIW Berlin.

Dardati, E. & Riutort, J. (2016). Cap-and-Trade and Financial Constraints: Is Investment Independent of Permit Holdings? *Environmental & Resource Economics*, **65**(4), 841–64.

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

Easley, D. & Spulber, D. F. (1981). Stochastic Equilibrium and Optimality with Rolling Plans. *International Economic Review*, **22**(1), 79–103.

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.

Ellerman, A. D., Valero, V. & Zaklan, A. (2015). *An Analysis of Allowance Banking in the EU ETS*. Working Paper RSCAS 2015/29, European University Institute.

Eurelectric (2009). EU ETS Phase 3 Auctioning – Timing and Futures versus Spot. Brussels.

European Commission (2013). *Commission Regulation (EU) 1123/2013*. Brussels, 9.11.2013.

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

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

European Commission (2018a). *Communication C(2018) 2801*. Brussels, 15.5.2018.

European Commission (2018b). *Report COM(2018) 842 final*. Brussels, 17.12.2018.

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**.

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.

Findley, T. S. & Caliendo, F. N. (2009). Short Horizons, Time Inconsistency, and Optimal Social Security. *International Tax & Public Finance*, **16**(4), 487–513.

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.

Gabaix, X. (2014). A Sparisty-Based Model of Bounded Rationality. *Quarterly Journal of Economics*, **4**, 1661–710.

Gerlagh, R. & Heijmans, R. J. R. K. (2018). *Regulating Stock Pollutants*. Working Paper 7383/2018, CESifo.

Gigerenzer, G. & Selten, R. (2003). *Bounded Rationality: The Adaptative Tooolboox*. MIT

Press.

Gigerenzer, G. & Todd, P. M. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gilboa, I. & Schmeidler, D. (1989). Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics*, **18**(2), 141–53.

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

Grüll, G. & Taschini, L. (2011). Cap-and-Trade Properties under Different Hybrid Scheme Designs. *Journal of Environmental Economics & Management*, **61**(1), 107–18.

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.

Haruvy, E., Lahav, Y. & Noussair, C. N. (2007). Traders' Expectations in Asset Markets: Experimental Evidence. *American Economic Review*, **97**(5), 1901–20.

Hasegawa, M. & Salant, S. W. (2014). Cap-and-Trade Programs under Delayed Compliance: Consequences of Interim Injections of Permits. *Journal of Public Economics*, **119**, 24–34.

Hasegawa, M. & Salant, S. W. (2015). The Dynamics of Pollution Permits. *Annual Review of Resource Economics*, **7**, 61–79.

Heutel, G. (2012). How Should Environmental Policy Respond to Business Cycles? Optimal Policy under Persistent Productivity Shocks. *Review of Economic Dynamics*, **15**(2), 244–64.

Hey, J. D. & Knoll, J. A. (2007). How Far Ahead Do People Plan? *Economics Letters*, **96**(1), 8–13.

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.

Hommes, C. (2011). The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab. *Journal of Economic Dynamics & Control*, **35**(1), 1–24.

Hommes, C., Massaro, D. & Weber, M. (2019). Monetary Policy under Behavioral Expectations: Theory and Experiment. *European Economic Review*, **118**, 193–212.

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

ICAP (2019). *Status Report 2019*. Berlin: International Carbon Action Partnership.

Jehiel, P. (1995). Limited Horizon Forecast in Repeated Alternate Games. *Journal of Economic Theory*, **67**(2), 497–519.

Johnson, E. J., Camerer, C., Sen, S. & Rymon, T. (2002). Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory*, **104**(1), 16–47.

Joltreau, E. & Sommerfeld, K. (2019). Why does Emissions Trading under the EU Emissions Trading System (ETS) not Affect Firms' Competitiveness? Empirical Findings from the

Literature. *Climate Policy*, **19**(4), 453–71.

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.

Kaganovich, M. (1985). Efficiency of Sliding Plans in a Linear Model with Time-Dependent Technology. *Review of Economic Studies*, **52**(4), 691–702.

Kling, C. & Rubin, J. (1997). Bankable Permits for the Control of Environmental Pollution. *Journal of Public Economics*, **64**(1), 101–15.

Koch, N., Fuss, S., Grosjean, G. & Edenhofer, O. (2014). Causes of the EUETS Price Drop: Recession, CDM, Renewable Policies or a Bit of Everything?−New Evidence. *Energy Policy*, **73**, 676–85.

Koch, N., Reuter, W. H., Fuss, S. & Grosjean, G. (2017). Permits vs. Offsets under Investment Uncertainty. *Resource & Energy Economics*, **49**, 33–47.

Kollenberg, S. & Taschini, L. (2016). Emissions Trading Systems with Cap Adjustments. *Journal of Environmental Economics & Management*, **80**, 20–36.

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.

Kydland, F. E. & Prescott, E. C. (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, **85**(3), 473–91.

Laffont, J.-J. & Tirole, J. (1996). Pollution Permits and Compliance Strategies. *Journal of Public Economics*, **62**(1), 85–125.

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

Lintunen, J. & Kuusela, O.-P. (2018). Business Cycles and Emission Trading with Banking. *European Economic Review*, **101**, 397–417.

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.

Marcantonini, C. & Valero, V. (2017). Renewable Energy and CO_2 Abatement in Italy. *Energy Policy*, **106**, 600–13.

Martin, R., Muûls, M. & Wagner, U. J. (2015). Trading Behavior in the EU Emissions Trading Scheme. In M. Gronwald & B. Hintermann (eds.) *Emissions Trading as a Policy Instrument: Evaluation and Prospects*, chapter 9, (pp. 213–38). MIT Press.

Mauer, E.-M., Okullo, S. J. & Pahle, M. (2019). *Evaluating the Performance of the EU ETS MSR*. Working Paper, Postdam Insitute for Climate (PIK).

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

Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, **29**(3), 315–35.

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**.

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

Perino, G., Ritz, R. A. & van Bentem, A. (2019). *Understanding Overlapping Policies: Internal Carbon Leakage and the Punctured Waterbed*. Working Paper 25643, NBER.

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.

de Perthuis, C. & Trotignon, R. (2014). Governance of CO_2 Markets: Lessons from the EU ETS. *Energy Policy*, **75**, 100–6.

Pindyck, R. S. (1993). The Present Value Model of Rational Commodity Pricing. *Economic Journal*, **103**(418), 511–30.

Pizer, W. A. (2002). Combining Price and Quantity Controls to Mitigate Global Climate Change. *Journal of Public Economics*, **85**(3), 409–34.

Reis, R. (2006). Inattentive Producers. *Review of Economic Studies*, **73**(3), 793–821.

Richstein, J. C., Chappin, E. J. & de Vries, L. J. (2015). The Market (In-)Stability Reserve for EU Carbon Emission Trading: Why It Might Fail and How to Improve It. *Utilities Policy*, **35**, 1–18.

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.

Sahin, F., Narayanan, A. & Robinson, E. P. (2013). Rolling Horizon Planning in Supply Chains: Review, Implications and Directions for Future Research. *International Journal of Production Research*, **51**(18), 5413–36.

Salant, S. W. (1983). The Vulnerability of Price Stabilization Schemes to Speculative Attack. *Journal of Political Economy*, **91**(1), 1–38.

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

Samuelson, P. A. (1971). Stochastic Speculative Price. *Proceedings of the National Academy of Sciences*, **68**(2), 335–7.

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.

Schopp, A., Acworth, W., Hauptmann, D. & Neuhoff, K. (2015). *Modelling a Market Stability Reserve in Carbon Markets*. Discussion Paper 1483, DIW Berlin.

Seifert, J., Uhrig-Homburg, M. & Wagner, M. (2008). Dynamic Behavior of CO_2 Spot Prices. *Journal of Environmental Economics & Management*, **56**(2), 180–94.

Sethi, S. & Sorger, G. (1991). A Theory of Rolling Horizon Decision Making. *Annals of Operations Research*, **29**(1), 387–415.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, **69**(1), 99–118.

Smith, V. L., Suchanek, G. L. & Williams, A. W. (1988). Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets. *Econometrica*, **56**(5), 1119–51.

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

Stocking, A. (2012). Unintended Consequences of Price Controls: An Application to Allowance Markets. *Journal of Environmental Economics & Management*, **63**(1), 120–36.

Stranlund, J. K., Costello, C. & Chávez, C. A. (2005). Enforcing Emissions Trading when Emissions Permits are Bankable. *Journal of Regulatory Economics*, **28**(2), 181–204.

The Economist (2013). ETS, RIP? 20 April 2013.

Thomson Reuters (2014). Carbon Market Survey – Into Smoother Waters? Point Carbon.

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).

Trotignon, R. (2012). Combining Cap-and-Trade with Offsets: Lessons from the EU ETS. *Climate Policy*, **12**(3), 273–87.

Trotignon, R. & Martino, V. (2013). *Back to the Future: A Comprehensive Analysis of Carbon Transactions in Phase I of the EU ETS*. ID 27, Climate Economics Chair.

Trück, S. & Weron, R. (2016). Convenience Yields and Risk Premiums in the EU-ETS – Evidence from the Kyoto Commitment Period. *Journal of Futures Markets*, **36**(6), 587–611.

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

World Bank (2019). *State and Trends of Carbon Pricing 2019*. Washington, DC: World Bank.

Wright, B. D. & Williams, J. C. (1982). The Economic Role of Commodity Storage. *Economic Journal*, **92**(367), 596–614.

Zhang, B. & Xu, L. (2013). Multi-Item Production Planning with Carbon Cap and Trade Mechanism. *International Journal of Production Economics*, **144**(1), 118–27.