Barriers to Trading in the EU-ETS: A Theoretical and Empirical Appraisal*

Marc Baudry[1,2] Anouk Faure[1,2] Simon Quemin[2,3]

[1]EconomiX – *Paris-Nanterre University (UPL)*

[2]Climate Economics Chair – *Institut Louis Bachelier (PSL)*

[3]Grantham Research Institute – *London School of Economics*

Preliminary – Please do not circulate

Abstract

Over Phase II of the EU-ETS (2008-12), nearly a third of liable companies did not record any transactions other than compliance-related. With an original analysis of Phase II transaction and compliance data, we first provide evidence of significant barriers to trading in the EU-ETS. We then develop a model of permit trading in presence of fixed and proportional trading costs in which firms can be initially over-allocated. Both market participation and clearing are endogenous and hinge on firms' characteristics and trading costs. This allows us to characterize the price responses to shifts in the trading costs and allocation levels analytically. We next calibrate our model based on Phase II market data. We find that fixed and proportional trading costs are in the order of k€10-30 and €0.1-1 per permit traded to match with observed prices and firms' market behaviors. We also quantify how supply-curbing policies like the Market Stability Reserve induce a larger price increase than under frictionless market conditions and analyze numerically how the allocation method influences compliance costs.

Keywords: Emissions trading, Transaction costs, Eviction, Market price response, EU-ETS.

JEL classification codes: Q58, Q54.

*Email addresses: Marc.Baudry@ParisNanterre.fr, Anouk.Faure@ChaireEconomieduClimat.org (corresponding author) and S.Quemin@lse.ac.uk.

1 Introduction

Market-based emissions trading programs have become inevitable in industrial environmental regulation and climate change mitigation, a particularly appealing feature of which being the efficiency gains derived from allocating abatement efforts to the lowest abatement cost firms first. This advantage, however, rests on one fundamental hypothesis: perfectly competitive markets.[1] Usually bundled up under the 'transaction costs' notion, market frictions are however acknowledged to be pervasive (Gangadharan, 2000; Jaraitė-Kažukauskė & Kažukauskas, 2015; Naegele, 2018) and impede the mechanism's efficiency and cost effectiveness (Stavins, 1995; Montero, 1998; Cason & Gangadharan, 2003; Singh & Weninger, 2017). In particular, if transaction costs imply substantive distortions in permit markets in practice, this has important implications for policy design, implementation and evaluation.

This paper contributes to both the empirical and theoretical literature on transaction costs in permit markets, with a particular focus on the European Union Emissions Trading Scheme (EU-ETS). First, we provide empirical evidence of the existence of both fixed and proportional trading costs in Phase II (2008-2012) of the EU-ETS based on a 'home-cooked' database aggregating the European Union Transaction Log (EUTL) transaction and compliance data at the consolidated firm level. Specifically, we find that on average a third of covered companies did not complete any transactions other than compliance-related, revealing significant barriers to trading in the market. Moreover, our analysis of transaction data uncovers a volume floor in yearly transactions which supports the presence of recurring costs that could be responsible for the generally low trading frequency we observe – and in turn proportional trading costs – also documented by Jaraitė et al. (2010) and Heindl (2012b).

Second, we develop a model of permit trading in presence of fixed and proportional trading costs. Fixed trading costs influence firms' decision to participate in the market while proportional costs further affect firms' emission choices by driving a wedge between sellers' and buyers' marginal abatement costs. For given trading costs, the sets of selling, buying and autarkic firms are endogenous to our characterization of the market equilibrium. This allows us to determine the impact of an shift in the costs levels on the market-clearing price. This depends on how supply and demand are affected by that change, and we characterize how this in turn relates to the distributions of the firms' characteristics (i.e. marginal abatement cost slopes and initial permit deficits).

Third, we calibrate our model to EU-ETS Phase II. Using estimated firms' characteristics

[1] This assumes zero transaction costs, full information and cost-minimizing behavior inter alia.

based on our analysis of transaction and compliance data, we build a calibration strategy to appraise fixed and proportional trading costs. Specifically, we select the cost pair that maximizes the alignment between our model's predictions and both the observed price dynamics and firms' decisions (net position on the market or autarkic compliance, traded volumes). We find a fixed cost in the order of k€10-30 per annum, and a proportional cost lying in the range of €0.1-1 per permit traded, which is in line with recent empirical studies of transaction costs in the EU-ETS (Naegele, 2018; Joas & Flachsland, 2016; Medina et al., 2014). These costs replicate up to 70% (resp. 90%) of the observed firms' behaviors in number of firms (resp. in volume of emissions). We discuss how our strategy controls for – at least partially – the existence in practice of primary (i.e. auctions), derivatives (i.e. futures, forwards), and intertemporal (i.e. banking) trading that our model does not formally account for.

Last but not least, the existence of trading costs has implications for policy design, especially in a context where the EU recently reformed its ETS by tightening its cap and implementing a supply-curbing mechanism, the Market Stability Reserve (Perino, 2018; Quemin & Trotignon, 2019). Therefore, we first use our stylized model to tease out comparative statics results on the price response to a cap reduction in absence vs. presence of trading costs. While this always implies higher price levels, we characterize when such an increase would tend to be magnified or dampened present trading costs relative to frictionless market conditions. Then, with our calibrated model, we find that the price response to a supply cutback is greater than under frictionless market conditions. Besides, testing different allocation methods reveals that targeting the supply crunch on under-allocated firms induces the largest price response, while 'grandfathering' achieves pollution control at least-cost.

On the theoretical front, the closest paper to ours is Singh & Weninger (2017) who characterize the market equilibrium in presence of fixed or proportional trading costs, but in their model firms are ex ante identical so that market clearing and firms' market positions only hinge on idiosyncratic productivity shocks. Like Singh & Weninger, we take trading costs as exogenously given and analyze the properties of the associated market equilibrium – see Liski (2001) for a formal treatment of transaction costs endogenously arising as a function of the market size and initial allocation profile among firms. By contrast, we allow for both types of costs simultaneously as well as heterogeneity in firms' abatement technologies and initial deficits (baseline - allocation) to permit calibration on market data.[2] Moreover, although we are primarily interested in the distortions induced by the trading costs in the permit mar-

[2]Relatedly and as a modeling novelty, we also define the equilibrium and associated firm-level efficiency gains from market participation when some firms are allowed to be initially over-allocated as the data show.

ket, they will in turn affect firms' production decisions and thus product markets. In this context, permit allocation can be used as a strategic trade tool by national regulators in an internationational permit market as analyzed by Constantatos et al. (2014).

Early contributions by Stavins (1995) and Montero (1998) derived comparative statics results on the effects of trading costs at the firm level (and hence taking the market price as constant). Here we harness our formal modeling of the market equilibrium in presence of trading costs to derive new comparative statics results at the market level. For instance we find that the market price may increase or decrease when trading costs increase (and in turn it can be above or below the frictionless market price). We also show that following a cap tightening, the market price response can be amplified or dampened relative to the frictionless case. This has important implications in evaluating the price impacts of pervasive supply-side controls in existing ETSs if trading costs are present.

On the empirical front, our paper provides a new approach to estimating trading costs. To the best of our knowledge, most of if not all empirical papers use econometric models to estimate transaction costs based on firm level transaction data or interviews with firms. Moreover, for the case of the EU-ETS, they are often concerned with subparts of the system, e.g. offsets (Naegele, 2018), German (Heindl & Lutz, 2012; Heindl, 2012a), Swedish (Sandoff & Schaad, 2009) or Irish (Jaraitė et al., 2010) firms. Here, we build a theoretical model of permit trading in presence of fixed and proportional trading costs, which we then calibrate on Phase II data so as to estimate their value by matching our model's predictions with observed market firms' behaviors. That is, we pick the trading costs that rationalize observed firms' market participation and position. The analysis of compliance data at the consolidated firm level also helps better pin down the identity of market participants and outsiders. Besides, our calibrated model helps better understanding the impact of permit supply shocks, both at the market and micro levels. In particular, we show that firms' trading behavior is dependent on the permit distribution method and specify how it relates to their individual characteristics. To our knowledge, such applications are novel in the context of the EU-ETS.

The remainder is organized as follows. Section 2 provides evidence of search and information acquisition/processing costs in Phase II of the EU-ETS. Section 3 develops a stylized permit trading model in presence of fixed and proportional trading costs. Section 4 provides analytical and numerical illustrations of the model to build intuition. Section 5 describes the model calibration to the EU-ETS and resulting policy implications. Section 6 concludes. An Appendix contains the analytical derivations and collected proofs (A), details on our transaction data consolidation methodology (B) and a sensitivity analysis (C).

2 Background

2.1 Permit trading in the EU-ETS

Under the usual assumptions of static permit trading models (i.e. competitiveness, perfect information and absence of transaction costs), trades are entirely driven by complementarity in marginal abatement costs, resulting in cost-efficiency (Montgomery, 1972). If firms are cost-minimizing, any profitable trading opportunity should be exploited. Besides, any market should exhibit a full participation rate so long as firms have heterogeneous abatement costs and the aggregate cap is binding relative to counterfactual emissions.

In the EU-ETS, regulated entities can trade directly at exchanges (such as ICE and LCH Clearnet), use brokers operating in the allowance market or make use of bilateral agreements (referred to as over-the-counter exchanges). During the second phase (2008-2012) of the scheme, the yearly averaged trading volume amounted to 5.6 million tonnes EUAs.[3] Out of these, 40% were exchanged over the counter and 60% through exchanges.

Trading activity is recorded in an electronic registry, the European Union Transaction Log (EUTL), whose aim is to guarantee an accurate accounting of all allowances issued under the system by keeping track of the ownership of allowances held in accounts over time. The EUTL records the activity of account holders such as the transfer of allowances, annual allocation or reconciliation, whereby regulated companies have to surrender enough allowances to cover their verified emissions.

Although transactions are recorded at the account level (i.e. plant or production site), we can safely assume that strategic and economic decisions are made at the firm level, motivating an aggregation of the EUTL data. Unlike previous works we build our own, 'home-cooked' database aggregating installations registered in Phase II to 5,145 consolidated firms.[4] Our consolidation methodology in described in Appendix B. Figure 1 depicts the consolidated firms' annual market participation over Phase II and shows that around a third of them did not record any trading activity if not for the annual allocation and surrender of allowances.[5]

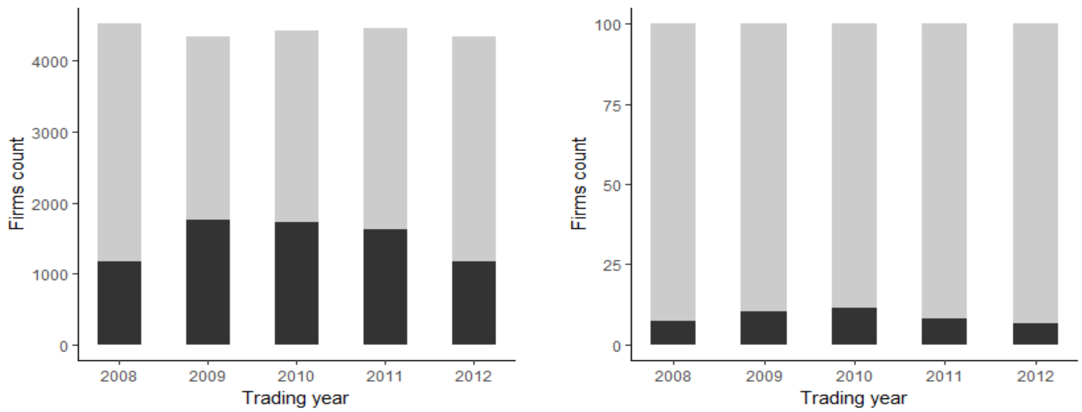
[3]See the EU6ETS handbook, Market Oversight section.

[4]Naegele (2018) and Jaraitė-Kažukauskė & Kažukauskas (2015) use Bureau van Dijk's Orbis company database to connect plants to parent companies. Naegele finds a close 4,578 firms with her method. We also exclude the aviation sector as it was brought under the scheme only from 2012.

[5]Slight year-on-year changes in the number of observations are due to firms continuously exiting or entering the market. We consider that a trading year (or compliance period) begins on April 1st of year t and terminates on March 31st of year $t+1$, in accordance with the compliance calendar of the EU-ETS.

Figure 1: Annual market participation (grey=yes; black=no) over Phase II

(a) Firms count (b) Perc. of annual allocation



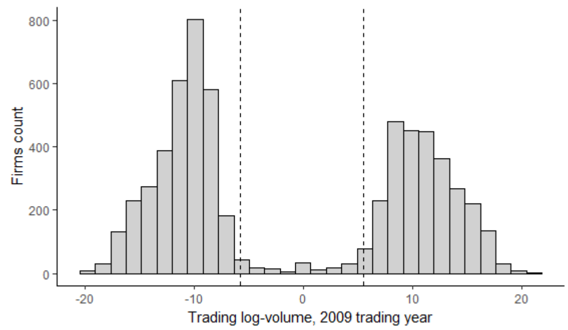
As for firms effectively participating to the market, Figure 2 represents the distribution of their transaction log-volumes, whether it be allowances arriving on their accounts (on the right, positive side) or leaving them (on the left, negative side) in a given trading year, here 2009. The dotted line demarcates the 95$^{\text{th}}$ percentile of the distribution, indicating that firms rarely engage in trades below a certain volume.

Table 5 contains descriptive statistics on consolidated firms, bringing some information about the profiles of market participants and outsiders. In particular, the average number of covered installations is higher for firms participating to the market regardless of their business. Besides, participating firms are quite infrequent traders in general. For instance, we record only 5 to 15 transactions on average in 2009 as in Sandoff & Schaad (2009).

Interestingly, computing firms' permit deficits (that is, the difference between yearly baseline emissions and initial permit endowment) reveals that in four activity sectors out of the seven, autarkic firms hold, on average, more permits than needed to cover their annual emissions.[6] Controlling for banking opportunities, it might be rational for them to sell (some portion of) their surplus allowances on the market though. Martin et al. (2014) notice a similar pattern in their survey of manufacturing firms, specifically firms start to sell permits only if they have an excess supply of 5,000 allowances on average.

[6]See Section 5 for more details about baseline emissions.

Figure 2: Trading log-volumes of firms in 2009



2.2 What shapes covered firms' trading behaviors?

Observed trading behaviors in Phase II of the EU-ETS do not indicate that firms consider allowances as financial assets and as such, sources of profit opportunities. Trading costs have long been help responsible for bringing inefficiencies in pollution permit markets. In their study, Schultz & Swieringa (2014) find that traders may have preferences for particular securities, suggesting frictions exist within the market that impede price discovery, including trading costs. Charles et al. (2013) further shows that the cost-of-carry relationship doesn't apply in the CO_2 market, which can be interpreted as a market inefficiency and may bring arbitrage opportunities.

Two definitions of trading costs emerge in the literature. On the one hand, explicit monetary costs relate to brokerage and clearing fees and come under a bid-ask spread in Medina et al. (2014) or unitary trading cost in Betz (2016), Joas & Flachsland (2016) or Heindl (2012b). On the other hand, a broader definition covers the implicit costs of trading and embodies search and information costs, market surveillance, investment in trading facilities and staff training. Coria & Jaraite (2018), Gangadharan (2000),Heindl (2012b), Jaraitė et al. (2010) and Jaraitė-Kažukauskė & Kažukauskas (2015) use the latter interpretation.

These definitions often overlap in practice. The explicit costs of transacting are difficult to measure indeed, as they depend on market participants' private relationship to their broker or brokers' level of membership at a particular exchange, which may alter their clearing fees (Schultz & Swieringa, 2014). For instance, trading costs differ according to whether regulated entities trade directly at exchanges (such as ICE and LCH Clearnet), use brokers operating in the allowance market or make use of bilateral agreements for over-the-counter exchanges in the EU ETS. In all of these cases they can be broken down into explicit costs including exchange membership, brokers' fees and financial services, and implicit costs including personnel to manage transactions and risk, data/advisory services and financial reporting (Constantatos et al., 2014; Sandoff & Schaad, 2009).

Karpf et al. (2018) provide recent evidence for trading costs in the EU ETS. They find that over Phases I and II the EU-ETS was characterized by a clustering of agents and a marked hierarchical network structure as agents have to resort to local networks or financial intermediaries to exchange permits. As a result, trades were not entirely driven by complementarity in marginal abatement costs (i.e. cost efficiency) and central agents and intermediaries captured informational rents. Unnecessary costs were imposed on industrial actors who often do not have the resources to collect market-related information. This, along with the necessary recourse to intermediaries is suggestive of the existence of proportional trading costs.

Besides, empirical studies have long shown that trading costs influence firms' market behaviors in cap-and-trade schemes. In the context of California's RECLAIM program, Gangadharan (2000) showed that they could have decreased the probability of trading by no less than 32%. This is particularly true for small and medium-size firms who face relatively bigger direct and indirect trading costs, especially when trading is not part of their core business. More recently in the EU-ETS, Jaraitė-Kažukauskė & Kažukauskas (2015) show that firms owning a greater number of plants face less intense search and information costs. In particular, the experience gained through learning-by-doing slightly increases the likelihood of trading over time. Relatedly, Heindl (2012b) highlights economies of scale in firms' management of emissions trading, favoring big emitters.

Trading costs inevitably mitigate gains from trade and in turn undermine a fundamental motive to rely on market-based instruments. In a model of trading decisions with uncertainty and transaction costs, Montero (1998) showed that the volume of trading opportunities came to be depressed and that aggregate control costs were higher relative to the least-cost equilibrium in turn. In the context of the Acid Rain Program, Carlson et al. (2000) also studied the origins of cost-reductions in SO_2 abatement. Though allowance trading did not turn out to be the main factor, it could have achieved hundred million cost savings compared to an emission standard.

Besides, the final allocation of allowances will likely be a function of the initial allocation if some market participants are not cost-minimizing (Hahn & Stavins, 2011). In Stavins (1995), market outcomes depend on the initial distribution of allowances so long as market frictions exist, though Fowlie & Perloff (2013) failed to reject the independence property empirically in the context of California's RECLAIM program. As any real-word complications that imply substantive violations of independence property in practice, this has important implications for policy design and implementation. Montero (1998) showed that an initial allocation close to the least-cost solution is always desirable in theory. More, Burtraw et al. (2001) simulated different allocation scenarios in a national electricity market covered under an allowance trading market. While grandfathering achieved the largest producers' surplus, auctioning won when including consumer surplus and government's revenue raising.

Other reasons could explain the deviation of polluters' trading behavior from that under the perfect markets hypothesis. In particular, the EU ETS compliance cycle allows firms to use their permit endowment to cover previous year's emissions. Borrowing forward has become alarmingly prevalent (Szabo, 2019). Here we rule out inter-temporal borrowing though, as we model a static model of permit trading. Besides, some polluters own multiple plants and in turn can trade within their business's perimeter. Though intra-firm trading may be less costly than inter-firm trading (as to search and information cost in particular), we do not make the distinction between intra and inter-firm trades in our empirical analysis. In turn, we do not consider as autarkic firms who practice intra-firm reshuffling, i.e. 27% of the annual volume of trades on average over Phase II (2008-2012).

3 Model

We consider a unitary-mass continuum $\mathcal{I}$ of firms indexed by $i \in \mathcal{I}$ covered under a competitive market for emissions permits. The model is static and we ignore firms' production decisions, i.e. we assume no incidence or indirect effect of the permit market on the goods markets that the firms serve. We denote by u_i, q_i and α_i firm i's unregulated emission level, permit allocation and slope of linear marginal abatement cost schedule. Specifically, firm i's abatement cost function is quadratic and writes $a \mapsto \alpha_i a^2/2$ where we omit the linear term for convenience and without loss of generality up to an innocuous translation of the results.

We assume that the emissions cap $\mathcal{Q}$ is binding relative to aggregate unregulated emissions $\mathcal{U}$, that is

$$\mathcal{Q} = \int_{\mathcal{I}} q_i \mathrm{d}i \; < \mathcal{U} = \int_{\mathcal{I}} u_i \mathrm{d}i,$$

but we allow for firm-level over-allocation, i.e. there exist some firms such that $u_i < q_i$. We let $\beta_i = u_i - q_i \gtreqless 0$ denote firm i's initial permit deficit. The two firms' characteristics of interest are thus the α_i's and β_i's which we assume have bounded supports $[\underline{\alpha}; \bar{\alpha}]$ and $[\underline{\beta}; \bar{\beta}]$ with $0 < \underline{\alpha} < \bar{\alpha} < \infty$ and $\underline{\beta} < 0 < \bar{\beta} < \infty$. When firms cannot trade permits with one another (i.e. autarky), firm i abates up to $a_i^0 = \max\{0; \beta_i\}$ with $p_i^0 = \alpha_i a_i^0$ the associated autarkic compliance shadow permit price. We next set forth the frictionless benchmark case before introducing fixed and unitary trading costs.

3.1 Frictionless equilibrium

Under frictionless conditions (i.e. unrestricted inter-firm permit trading, no trading costs), all firms equalize their marginal abatement cost to the prevailing market price p, i.e. $\alpha_i(u_i - e_i) = p$ for all $i \in \mathcal{I}$. Note that a feasible market price must be positive as the cap is binding and it can be no larger than $\max_i p_i^0 = \bar{\alpha}\bar{\beta}$ for otherwise no firm would be willing to buy permits on the market. Proposition 1 characterizes firms' net market positions and efficiency gains from permit trading relative to autarky for any given feasible permit price on the market.

Proposition 1. *For any feasible market permit price $p \in (0; \bar{\alpha}\bar{\beta})$, the sets of buying and selling firms are $\mathcal{D}(p) = \{i \,|\, \alpha_i\beta_i \geq p\}$ and $\mathcal{S}(p) = \{i \,|\, \alpha_i\beta_i \leq p\}$, and individual efficiency gains from permit trading on the market (w.r.t. autarky) write, for any firm $i \in \mathcal{I}$*

$$G_i(p) = (p_i^0 - p)^2/(2\alpha_i) + p \max\{0; -\beta_i\} \geq 0, \tag{1}$$

where $p_i^0 = \alpha_i \max\{0; \beta_i\} \geq 0$ is firm i's shadow price of autarkic compliance.

Proof. Relegated to Appendix A.1. □

The first component in (1) is standard and proportional to the squared distance in autarky-market prices. Specifically, selling (resp. buying) firms with $p_i^0 \leq p$ (resp. $p_i^0 \geq p$) find it profitable to abate more (resp. less) than in autarky and sell surplus (resp. purchase missing) permits on the market. This goes on until all trading opportunities are exhausted, i.e. when marginal abatement costs are equalized between all firms (to the market price). The second component in (1) is novel and accrues only to those firms that are initially over-allocated as they sell their entire initial surplus of permits at the market price at no cost.

Adding market closure $\int_{\mathcal{I}} (u_i - e_i^\star)\mathrm{d}i = \mathcal{U} - \mathcal{Q}$ on top of firms' optimality conditions defines the competitive market equilibrium, characterized by the equilibrium price

$$p^\star = (\mathcal{U} - \mathcal{Q}) \Big/ \int_{\mathcal{I}} \mathrm{d}i/\alpha_i > 0. \tag{2}$$

As is well known, $p^\star$ is independent of how the u_i's and q_i's (and thus the β_i's) are distributed among firms. This is the Coasian independence property, i.e. equilibrium outcomes do not hinge on the initial allocation of permits. Note, however, that $p^\star$ depends on the distribution of the α_i's, $\{\alpha_i\}_i$. Specifically, it is proportional to the stringency of the overall constraint on emissions imposed by the cap $\mathcal{U} - \mathcal{Q}$ and the harmonic mean of $\{\alpha_i\}_i$ (the more skewed $\{\alpha_i\}_i$ towards low values, the lower $p^\star$, and vice versa).

The efficiency gains defined in (1) with $p = p^\star$ stem from the cost-efficient allocation of the total abatement effort $\mathcal{U} - \mathcal{Q}$ among all firms. Specifically, all firms abate in equilibrium, even initially over-allocated firms since $e_i^\star = u_i - p^\star/\alpha_i < u_i$ for any firm $i \in \mathcal{I}$ as it abates in inverse proportion to α_i, i.e. $a_i^\star = u_i - e_i^\star = p^\star/\alpha_i > 0$. We say that the frictionless equilibrium is cost-efficient in the sense that (1) all firms are (weakly) better off participating to the market and (2) marginal abatement costs are equalized between them.[7] As described below, this is no longer the case in presence of trading costs.

3.2 Equilibrium with trading costs

Without loss of generality we assume that both permit suppliers and demanders face common market participation and proportional trading costs, denoted by F and T respectively. That is, all firms have to pay the fixed fee F to access the market and trade permits and T is a mark-up on the permit price p (buying firms pay $p + T$ per permit purchased, selling firms receive $p - T$ per permit sold). Thus, when $F > 0$ or $T > 0$ some firms can be better off under autarky. In the presence of trading costs buying (resp. selling) firms in the frictionless equilibrium can either remain buyers (resp. sellers) or prefer not to enter the market. Such autarkic compliance implies that short firms abate just as much as to cover their permit deficits while long firms keep their surplus permits unused.

Specifically, when $F > 0$ and $T = 0$, the market outcome is not cost-efficient at the extensive margin (some firms do not participate in the market and some trades that would otherwise be mutually beneficial go unrealized) but remains cost-efficient at the intensive margin (all mutually beneficial trades materialize between participating firms as their marginal abatement costs are equalized). When $T > 0$ and $F = 0$, cost-efficiency at the intensive margin further drops as participating firms abate in proportion to the actual permit price that they face (that is, inclusive of the per-unit trading fee), which drives a wedge between buyers' and sellers' marginal abatement costs in equilibrium. Specifically, the marginal abatement costs experienced by a buyer exceed those experienced by a seller by $2T$.

[7] In the following we always consider that indifference frontiers are of measure zero. In this case the set $\{i \,|\, p_i^0 = p^\star\}$ of firms indifferent between participating to the market or not has measure zero.

Therefore, given a market permit price p and a per-unit trading cost $T < p$, firm i will find it profitable to buy (resp. sell) permits on the market provided that $p_i^0 > p+T$ (resp. $p_i^0 < p-T$). Additionally, given a market participation cost $F \geq 0$, a buying (resp. selling) firm i will trade permits on the market when its efficiency gains from permit trading net of both trading costs $G_i(p+T) - F$ (resp. $G_i(p-T) - F$) are positive. Lemma 1 thus defines market participation price thresholds for prospective buying and selling firms.[8]

Lemma 1. *Given trading costs F and T and a market permit price $p > T$, it is profitable for firm i to buy permits on the market when $p < \bar{p}_i = \alpha_i\beta_i - T - \sqrt{2\alpha_i F}$. Symmetrically, it is profitable for firm i to sell permits on the market when $p > \underline{p}_i = \alpha_i\beta_i + T + \sqrt{2\alpha_i F}$ if $\beta_i > 0$ or when $p > \underline{p}_i = \alpha_i\beta_i + T + \sqrt{\alpha_i^2\beta_i^2 + 2\alpha_i F}$ if $\beta_i \leq 0$.*

Proof. Relegated to Appendix A.2. □

Intuitively, firm i will purchase permits only if the market price is below its autarkic shadow price $p_i^0 = \alpha_i\beta_i$ corrected for the fixed and unitary trading costs (note that $\bar{p}_i$ is decreasing with F and T). Symmetrically, firm i will sell permits only if the market price is above its cost-adjusted autarkic shadow price $\underline{p}_i$ (which is increasing with F and T). To gain further insight, Proposition 2 relates firms' market participation decisions to their characteristics.

Proposition 2. *Given a market participation cost F, a unitary trading cost T and a market permit price $p > T$, the sets of buying and selling firms are defined by*

$$\begin{aligned}
&\mathcal{D}(p,F,T) = \{i \,|\, \alpha_i \geq \alpha^+(p,F,T;\beta_i) \wedge \beta_i > 0\}, \text{ and} \\
&\mathcal{S}(p,F,T) = \Big\{\{i \,|\, \alpha_i \leq \alpha^-(p,F,T;\beta_i) \wedge \beta_i > 0\} \\
&\qquad \cup \{i \,|\, \alpha_i \leq \alpha^0(p,F,T;\beta_i) \wedge -F/(p-T) < \beta_i \leq 0\} \cup \{i \,|\, \beta_i \leq -F/(p-T)\}\Big\},
\end{aligned}$$

and $\mathcal{A}(p,F,T) = \mathcal{I} \cap \{\mathcal{D}(p,F,T) \cup \mathcal{S}(p,F,T)\}$ denotes the set of autarkic firms, where

$$\begin{aligned}
\alpha^\pm(p,F,T;\beta) &= \Big(F + (p \pm T)\beta \pm \sqrt{F(F + 2(p\pm T)\beta)}\Big)\Big/\beta^2 > 0, \\
\text{and } \alpha^0(p,F,T;\beta) &= (p-T)^2\Big/(2(F + (p-T)\beta)) > 0.
\end{aligned}$$

In particular, $\mathcal{D}$ and $\mathcal{S}$ are of decreasing measure as F or T increases and $\mathcal{D}$ (resp. $\mathcal{S}$) is of decreasing (resp. increasing) measure as p increases, ceteris paribus. Individual efficiency gains are $G_i(p+T) - F \geq 0$ (resp. $G_i(p-T) - F \geq 0$) for firm i in $\mathcal{D}$ (resp. $\mathcal{S}$).

[8] Entry and exit prices in the sense of Dixit (1989) are identical in our static setting, hence both denominated by the term 'participation price threshold'. In a dynamic setting, even absent trading costs, firms' entry or exit decisions are also contingent on the how permits are initially distributed (Fowlie et al., 2016).

Proof. Relegated to Appendix A.3. □

Proposition 2 extends Proposition 1 in presence of fixed and per-unit trading costs F and T. Specifically, for p large (resp. low) enough one has $\mathcal{D} = \varnothing$ (resp. $\mathcal{S} = \varnothing$); for $F = T = 0$ one has $\mathcal{A} = \varnothing$, $\mathcal{D} = \{i|\alpha_i\beta_i \geq p\}$ and $\mathcal{S} = \{i|\alpha_i\beta_i \leq p\}$; for F and T large enough one has $\mathcal{D} = \mathcal{S} = \varnothing$ and $\mathcal{A} = \mathcal{I}$. For intermediate admissible values of F and T, Figure 3 depicts the zones where buying (red), selling (green) and autarkic (grey) firms are located in the (α, β)-space for a given feasible price p. The notions of admissible costs and feasible price will be formalized in Lemma 2 below. Additionally, the blue (resp. yellow) arrows indicate how the participation frontiers move in response to an increase in F and T (resp. p).[9]

Let us describe Figure 3 from left to right. When $\beta_i \leq -F/(p-T)$, firm i more than recovers the entry cost by just selling its initial surplus since $-\beta_i(p-T) \geq F$. Because this comes at no cost for firm i, this is true whatever its marginal abatement cost slope α_i. Additionally, firm i finds it profitable to abate $(p-T)/\alpha_i > 0$ and sell the corresponding amount of freed-up permits on the market. When $-F/(p-T) < \beta_i \leq 0$, selling the initial surplus is not enough to cover the fixed cost. Thus, either firm i can abate at a sufficiently low cost at the margin ($\alpha_i \leq \alpha^0$) and make profits by selling both surplus $-\beta_i$ and freed-up permits $(p-T)/\alpha_i$ on the market, or it prefers leaving its surplus permits sitting in its account, unused ($\alpha_i > \alpha^0$). Finally, when $\beta_i > 0$, three cases can occur. When firm i's deficit is small, only when its abatement cost at the margin is sufficiently high ($\alpha_i > \alpha^+$) is it profitable to enter the market to purchase permits. Otherwise, firm i is better off abating its deficit to meet compliance by itself ($\alpha^- < \alpha_i \leq \alpha^+$) or even so as to make some profits by selling residual freed-up permits $(p-T)/\alpha_i - \beta_i$ ($\alpha_i \leq \alpha^-$). As firm i's deficit increases, it becomes more and more (resp. less and less) profitable to enter the market to purchase (resp. sell) permits.

Importantly, note that with $F = 0$ and $T > 0$, the demand and supply sets in Proposition 2 rewrite $\mathcal{D} = \{i|\alpha_i \geq (p+T)/\beta_i \wedge \beta_i > 0\}$ and $\mathcal{S} = \{\{i|\alpha_i \leq (p-T)/\beta_i \wedge \beta_i > 0\} \cup \{i|\beta_i \leq 0\}\}$. That is, over-allocated firms would always find it profitable to sell their extra permits at price $p > T$. In our model a fixed entry cost is thus necessary to explain the fact that some over-allocated firms may choose autarky over market participation. More generally, when $F > 0$, and both trading costs increase, the last permit supplier has characteristics $(\underline{\alpha}, \underline{\beta})$. It drops out of the market when $\alpha^0(p, F, T; \underline{\beta}) = \underline{\alpha}$. Symmetrically, on the demand side, as trading costs increase, the last buying firm has characteristics $(\bar{\alpha}, \bar{\beta})$. It switches to autarkic compliance when $\alpha^+(p, F, T; \bar{\beta}) = \bar{\alpha}$. Lemma 2 summarizes the above considerations.

[9] The analytical derivations necessary to characterize these movements are contained in Appendix A.3.

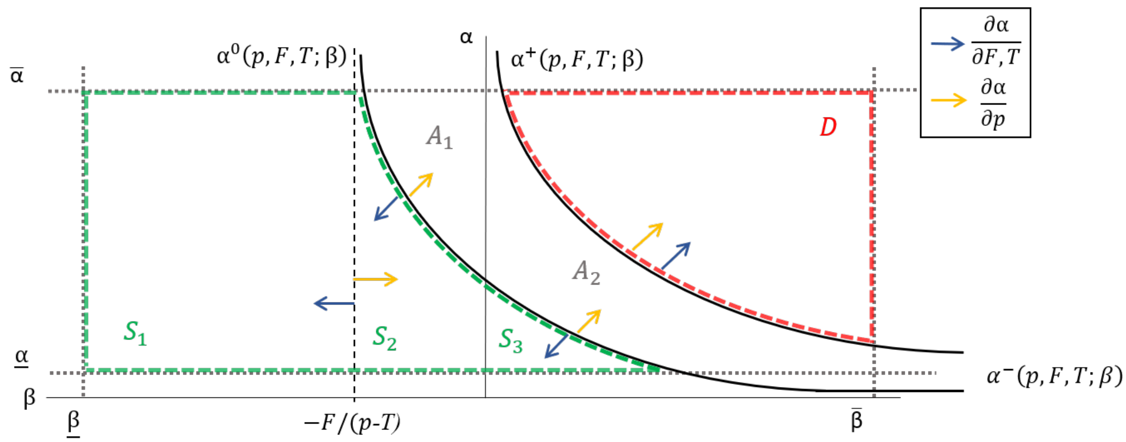


Figure 3: Market participation given a price p, entry cost F and unitary trading cost T

Lemma 2. *The fixed and unitary trading costs F and T are said admissible when*

$$F < \bar{\alpha}\bar{\beta}^2/2 \;\; and \;\; 2T + \sqrt{\underline{\beta}^2\underline{\alpha}^2 + 2\underline{\alpha}F} + \sqrt{2\bar{\alpha}F} < \bar{\alpha}\bar{\beta} - \underline{\alpha}\underline{\beta}. \tag{4}$$

When one the above two conditions does not hold, the market breaks down. Given admissible trading costs F and T, a permit price p is said feasible when

$$\underline{\beta}\underline{\alpha} + T + \sqrt{\underline{\beta}^2\underline{\alpha}^2 + 2\underline{\alpha}F} < p < \bar{\alpha}\bar{\beta} - T - \sqrt{2\bar{\alpha}F}. \tag{5}$$

Proof. Relegated to Appendix A.4. □

Trading costs are said admissible when they are low enough to ensure positive supply and demand on the market. When this is not the case, the market breaks down. A permit price is said feasible when it may clear the market – it is thus necessarily bounded by the participation price thresholds of the last potential permit buyer $(\bar{\alpha}, \bar{\beta})$ and seller $(\underline{\alpha}, \underline{\beta})$ on the market as given in Lemma 1. Note that the range of feasible prices in (5) is not empty provided that trading costs are admissible.

Equipped with Proposition 2, supply and demand functions can then be defined as follows

$$S(p,F,T) = \int_{\mathcal{S}(p,F,T)} (a_i^*(p-T) - \beta_i)\mathrm{d}i \text{ and } D(p,F,T) = \int_{\mathcal{D}(p,F,T)} (\beta_i - a_i^*(p+T))\mathrm{d}i,$$

where $a_i^*(p \pm T) = (p \pm T)/\alpha_i$ is participating firm i's optimal abatement decision. If we let h denote the density function of the β_i's and $g(\cdot|\beta)$ denote the density function of the α_i's conditional on the β_i's, supply and demand then rewrite

$$S(p,F,T) = \int_{\underline{\beta}}^{-F/(p-T)} \int_{\underline{\alpha}}^{\bar{\alpha}} \Big((p-T)/x - y\Big) g(x|y)h(y)\mathrm{d}x\mathrm{d}y \\ + \int_{-F/(p-T)}^{0} \int_{\underline{\alpha}}^{\alpha^0(p,F,T;y)} \Big((p-T)/x - y\Big) g(x|y)h(y)\mathrm{d}x\mathrm{d}y \\ + \int_{0}^{\bar{\beta}} \int_{\underline{\alpha}}^{\alpha^-(p,F,T;y)} \Big((p-T)/x - y\Big) g(x|y)h(y)\mathrm{d}x\mathrm{d}y, \tag{6}$$

$$\text{and } D(p,F,T) = \int_{0}^{\bar{\beta}} \int_{\alpha^+(p,F,T;y)}^{\bar{\alpha}} \Big(y - (p+T)/x\Big) g(x|y)h(y)\mathrm{d}x\mathrm{d}y, \tag{7}$$

We characterize how S and D react to a change in F, T or p in the following Lemma.

Lemma 3. *For any admissible trading costs F and T and feasible market price p, it holds that*

$$\frac{\partial S}{\partial p} > 0,\ \frac{\partial S}{\partial F} < 0,\ \frac{\partial S}{\partial T} < 0,\ \frac{\partial D}{\partial p} < 0,\ \frac{\partial D}{\partial F} < 0,\ \textit{and}\ \frac{\partial D}{\partial T} < 0.$$

Proof. Relegated to Appendix A.5. □

We also let $V = S - D$ denote the net permit supply function. Then, equipped with Lemmas 2 and 3, we can now state the following result.

Proposition 3. *Given admissible trading costs F and T, there exists a unique feasible permit price $\hat{p}$ that clears the market, i.e. $V(\hat{p}, F, T) = 0$.*

Proof. Relegated to Appendix A.6. □

Proposition 3 ensures the existence and uniqueness of the market equilibrium in presence of fixed and proportional trading costs provided they are not too large, i.e. admissible. Save for $F = T = 0$ where the equilibrium collapses to the frictionless one, i.e. $V(p^\star, 0, 0) = 0$, it is otherwise apparent from (6) and (7) that $\hat{p}$ does not admit a simple closed-form solution. We thus seek to derive comparative statics results in the following section.

3.3 Comparative statics

Consider an arbitrary small shift $\mathrm{d}C$ in the trading cost $C = F$ or T. By virtue of the implicit function theorem, the resulting price response $\mathrm{d}\hat{p}$ in the vicinity of the equilibrium reads

$$\frac{\mathrm{d}\hat{p}}{\mathrm{d}C} = -\frac{\partial V(\hat{p},F,T)/\partial C}{\partial V(\hat{p},F,T)/\partial p} \gtrless 0, \tag{8}$$

which cannot unambiguously be signed in general. Indeed, Lemma 3 shows that $\partial V/\partial p > 0$ but the sign of $\partial V/\partial C$ is indefinite and depends on the relative magnitudes of the impacts of the change in C on D and S.[10] For instance, when demand is more responsive than supply, i.e. $|\partial D/\partial C| > |\partial S/\partial C|$, then $\mathrm{d}\hat{p}/\mathrm{d}C < 0$. In words, downward pressure on the equilibrium price is exerted as relatively more buyers than sellers are forced out of the market as a result of higher trading costs – and vice versa. In turn, $\hat{p}$ can be higher or lower than the frictionless market price $p^\star$ depending on the trading cost levels. Note that the magnitudes of the supply and demand responses to higher trading costs also depend on the distribution of the firm parameters α_i's and β_i's, as well as on the initial levels of the trading costs and equilibrium price. In presence of trading costs it thus matters how permits are distributed among firms. This is an illustration of a well-known break-down of the Coasian independence property as equilibrium outcomes hinge on the initial allocation of permits.

Additionally, taking the total differential of supply and demand in response to the cost shift $\mathrm{d}C$ yields

$$\frac{\mathrm{d}S}{\mathrm{d}C} = \underbrace{\frac{\partial S}{\partial C}}_{<0} + \underbrace{\frac{\partial S}{\partial p}}_{>0}\underbrace{\frac{\mathrm{d}\hat{p}}{\mathrm{d}C}}_{\gtrless 0} \quad \text{and} \quad \frac{\mathrm{d}D}{\mathrm{d}C} = \underbrace{\frac{\partial D}{\partial C}}_{<0} + \underbrace{\frac{\partial D}{\partial p}}_{<0}\underbrace{\frac{\mathrm{d}\hat{p}}{\mathrm{d}C}}_{\gtrless 0}, \tag{9}$$

which, on the face of it, cannot unambiguously be signed either. However, it must necessarily be that $\mathrm{d}S/\mathrm{d}C = \mathrm{d}D/\mathrm{d}C$ in equilibrium – this can readily be seen by plugging $V = S - D$ in (8). This fact in conjunction with (9) necessarily implies that $\mathrm{d}S/\mathrm{d}C = \mathrm{d}D/\mathrm{d}C < 0$. In other words, an increase in trading costs always reduces the equilibrium volume of trade.

Relatedly, the total regulatory control costs, which comprise the total abatement costs and trading costs, will always increase as C rises. Indeed, the total control costs are the largest under complete autarky and gradually decrease as the volume of firms' trading increases. In addition, firms choose to trade provided it is profitable to do so relative to autarky. Thus, as trading costs increase – as the trading volume decreases and the number of autarkic firms increases – the total control costs necessarily increase.[11] In other words, larger trading costs negatively affect welfare by consuming more monetary resources and stifling more trades that otherwise would have been mutually beneficial (at both the extensive and intensive margins). Proposition 4 summarizes the above considerations.

Proposition 4. *In response to an increase in the entry or unitary trading cost, the market equilibrium price may increase or decrease. However, the total equilibrium volume of trade*

[10] Note that even taking (8) in the small as F and/or $T \to 0$ or also yields an indefinite sign.

[11] Stavins (1995) and Montero (1998) show and discuss that how total regulatory costs and trade volume are affected by trading costs also hinges upon the way permits are allocated.

(resp. total regulatory control costs) unambiguously decreases (resp. increase).

Now consider an arbitrary small variation in $\{\beta_i\}_i$.[12] Specifically, we let β_i change to $\beta_i + \mathrm{d}\beta_i$ for all $i \in \mathcal{I}$. We then let the price p adapt to these changes so that $V(p, F, T) = 0$ remains satisfied. This amounts to taking the total differential of $V(\hat{p}, F, T) = 0$ with respect to $\hat{p}$ and β_i for all i. With a slight abuse of notation for the derivatives w.r.t. the β_i's, one has

$$\frac{\partial V(\hat{p}, F, T)}{\partial p}\mathrm{d}\hat{p} + \int_{\mathcal{I}} \frac{\partial V(\hat{p}, F, T)}{\partial \beta_i}\mathrm{d}\beta_i \mathrm{d}i = 0. \tag{10}$$

For the sake of the argument, we here only consider induced changes at the intensive margin (i.e. within $\mathcal{S}$ and $\mathcal{D}$) and ignore those at the extensive margin (i.e. entries to and exits from $\mathcal{S}$ and $\mathcal{D}$). The latter are formally treated in Appendix A.7. Using the shorthand notation $\bar{\mathcal{D}} = \mathcal{D}(\hat{p}, F, T)$ and $\bar{\mathcal{S}} = \mathcal{S}(\hat{p}, F, T)$, one has $\dfrac{\partial V}{\partial \beta_i} = -\displaystyle\int_{\bar{\mathcal{S}} \cup \bar{\mathcal{D}}} \delta(j = i)\mathrm{d}j$ where $\delta(\cdot)$ is the Dirac distribution. Therefore, (10) simplifies to

$$\frac{\partial V(\hat{p}, F, T)}{\partial p}\mathrm{d}\bar{p} = \int_{\bar{\mathcal{S}} \cup \bar{\mathcal{D}}} \mathrm{d}\beta_i \mathrm{d}i, \tag{11}$$

meaning that the sign and magnitude of $\mathrm{d}\hat{p}$ hinge upon the overall net change in the β_i's over $\bar{\mathcal{S}} \cup \bar{\mathcal{D}}$. Further consider that the cap is reduced by $-\mathrm{d}\mathcal{Q} < 0$ and assume that the cap tightening is uniformly distributed across all firms.[13] Because $\mathcal{I}$ is of mass one, we have that $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}$ for all $i \in \mathcal{I}$. If we let $|\cdot|$ denote the mass (or measure) of a set, (11) rewrites

$$\frac{\mathrm{d}\hat{p}}{\mathrm{d}\mathcal{Q}} = \frac{|\mathcal{S}(\hat{p}, F, T)| + |\mathcal{D}(\hat{p}, F, T)|}{\partial V(\hat{p}, F, T)/\partial p} > 0. \tag{12}$$

The market price response to the cap tightening is always positive present trading costs, but how does its magnitude compare to the frictionless case? Without trading costs, note that (11) reads $\mathrm{d}p^\star \partial V(p^\star, 0, 0)/\partial p = \int_{\mathcal{I}} \mathrm{d}\beta_i \mathrm{d}i = \mathrm{d}\mathcal{Q}$ so that one gets

$$\frac{\mathrm{d}\hat{p}/\mathrm{d}\mathcal{Q}}{\mathrm{d}p^\star/\mathrm{d}\mathcal{Q}} = \underbrace{\Big(|\mathcal{S}(\hat{p}, F, T)| + |\mathcal{D}(\hat{p}, F, T)|\Big)}_{\text{distribution effect } \leq 1} \underbrace{\frac{\partial V(p^\star, 0, 0)/\partial p}{\partial V(\hat{p}, F, T)/\partial p}}_{\text{price effect } \geq 1} \gtrless 1. \tag{13}$$

The price response to a cap tightening in presence of trading costs can be lower or greater than

[12] With comparative statics at the firm level (i.e. assuming the market price is unchanged), Stavins (1995) finds that with constant marginal trading costs (as is the case here) the firm's initial allocation level does not influence its emission level in equilibrium. Here we consider the case where each firm's allocation is allowed to vary simultaneously and how this impacts the market price.

[13] Alternative distributions of the cap tightening are presented in Appendix A.7.

without frictions. This depends on two effects. First, how much of the tightening effectively affects market participants as allocation changes for autarkic firms have no bearing on the market price. This distribution effect always mitigates the price response with trading costs, all the more so that they are large and the mass of autarkic firms is significant. Second, the relative sensitivity of the net supply to a price change. As Appendix A.7 shows in this case, this price effect magnifies the magnitude of the price response, yielding the overall ambiguous effect in (13).[14] We can state the following result in the general case, inclusive of the induced changes at both the intensive and extensive margins.

Proposition 5. *In response to a cap tightening, the market price increase can be amplified or dampened in presence of trading costs relative to frictionless market conditions. This also depends on how the tightening is distributed among firms.*

Proof. Relegated to Appendix A.7. □

In the general case both the distribution and price effects can be greater or lower than one. Appendix A.7 shows that the extensive margin effects are nil when $F = 0$ so that this case collapses to the above analysis. Otherwise, the price (resp. distribution) effect is likely to be predominantly greater (resp. lower) than one except perhaps for small trading cost levels. Additionally, Appendix A.7 discusses how the magnitude of the distribution effect also hinges on how the overall cap tightening is distributed across firms.

4 Illustrations

The preceding section shows that clear-cut comparative statics results on the effects of trading costs on market equilibrium outcomes are hard to come by. In this section, therefore, we first derive implicit closed-formed solutions for the equilibrium price in presence of trading costs in two special cases for the distributions of the firms' characteristics. In turn, this allows us to analyze how the equilibrium price responds to small shifts in the trading cost or emissions cap levels in the vicinity of the equilibrium. We next provide numerical illustrations in more general cases to gain further insight into the market impacts of trading costs.

[14] Specifically, $\partial V(p^\star, 0, 0)/\partial p = \int_{\mathcal{I}} \mathrm{d}i/\alpha_i$ (also from (2)) and $\partial V(\hat{p}, F, T)/\partial p = \partial V(p^\star, 0, 0)/\partial p - \int_{\mathcal{A}} \mathrm{d}i/\alpha_i$.

4.1 Analytical examples

We are primarily interested in the effects of a fixed entry cost, i.e. we let $F \geq 0$ and $T = 0$. Additionally, we let firms be homogeneous in terms of initial net deficit, i.e. $\beta_i = \beta = \mathcal{U} - \mathcal{Q} > 0$ for all $i \in \mathcal{I}$. We then proceed with two cases regarding the distribution of the α_i's, namely

$$g_1(x) = 2x/(\bar{\alpha}^2 - \underline{\alpha}^2) \text{ and } g_2(x) = \bar{\alpha}\underline{\alpha}/(x^2(\bar{\alpha} - \underline{\alpha})),$$

for $x \in [\bar{\alpha}; \underline{\alpha}]$ and $g_k(x) = 0$ otherwise with $k = 1, 2$. For $k = 2$ (resp. $k = 3$), the distribution of the α_i's is tilted towards high (resp. low) α-values. The following results then obtain. The sketches of all analytical derivations are relegated to Appendix A.8.

Example 1. *Let $\beta_i = \beta > 0$ for all $i \in \mathcal{I}$ and $g = g_1$. Then $\hat{p}_1$ is implicitly defined by*

$$\hat{p}_1 + \frac{2F\sqrt{F(F + 2\beta\hat{p}_1)}}{\beta^3(\bar{\alpha} - \underline{\alpha})} = p_1^\star = \frac{\beta(\bar{\alpha} + \underline{\alpha})}{2}. \tag{14}$$

In this case, $\hat{p}_1 \leq p_1^\star$ with equality in $F = 0$ and $\mathrm{d}\hat{p}_1/\mathrm{d}F < 0$.

Example 2. *Let $\beta_i = \beta > 0$ for all $i \in \mathcal{I}$ and $g = g_2$. Then $\hat{p}_2$ is implicitly defined by*

$$\hat{p}_2 - \frac{4F\bar{\alpha}^2\underline{\alpha}^2}{(\bar{\alpha}^2 - \underline{\alpha}^2)\hat{p}_2^3}\sqrt{F(F + 2\beta\hat{p}_2)} = p_2^\star = \frac{2\beta\bar{\alpha}\underline{\alpha}}{\bar{\alpha} + \underline{\alpha}}. \tag{15}$$

In this case, $\hat{p}_2 \geq p_2^\star$ with equality in $F = 0$ and $\mathrm{d}\hat{p}_2/\mathrm{d}F > 0$.

First note that $p_1^\star > p_2^\star$. Indeed, with a homogeneous deficit β across firms, the frictionless equilibrium price is higher, the more tilted the distribution of the α_i's towards high values. Crucially, the introduction of a fixed entry cost tends to mitigate this. Specifically, when the α_i's are tilted towards high values, introducing a fixed cost tends to evict more firms with high α_i (i.e. demanders) than with low α_i (i.e. suppliers) relative to the average, ceteris paribus. This means that demand is more constricted than supply, hence a downward pressure on the price. The converse holds when the α_i's are tilted towards low values as introducing a fixed cost reduces supply more than demand, ceteris paribus.[15] Mechanically, therefore, the entry cost tends to have a tempering effect on the market price when the frictionless market price is 'high' and conversely it tends to hike the former when the latter is 'low'.

Due to regulatory supply-side control, consider now that the cap $\mathcal{Q}$ is reduced by $-\mathrm{d}\mathcal{Q} < 0$, which translates into an increase in β of $\mathrm{d}\beta = \mathrm{d}\mathcal{Q} > 0$. We have the following result.

[15]When $\{\alpha_i\}_i$ is uniform the ordering between $\hat{p}$ and $p^\star$ depends on the cost level in a complicated way.

Example 3. *For $k = 1, 2$ the price response to a supply squeeze is positive, i.e.* $\mathrm{d}\hat{p}_k/\mathrm{d}\beta > 0$*. However, only when $\hat{p}_k$ is large enough does it hold that* $\mathrm{d}\hat{p}_k/\mathrm{d}\beta > \mathrm{d}p^\star_k/\mathrm{d}\beta > 0$*, specifically* $\hat{p}_1 > (\beta p^\star_1 - 3F)/5\beta$ *and* $\hat{p}_2 > p^\star_2 + \sqrt{p^\star_2(p^\star_2 + 3F/2\beta)}$.

Intuitively, shrinking the cap always implies higher price levels in presence of trading costs, but interestingly the price rise can be magnified or dampened relative to frictionless conditions. Example 3 further suggests that the price response is very much likely (resp. unlikely) to be magnified in presence of an entry cost when $k = 1$ (resp. $k = 2$).

4.2 Numerical examples

We now let $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ follow independent beta distributions $B(\alpha_\alpha, \beta_\alpha)$ and $B(\alpha_\beta, \beta_\beta)$ and consider both entry and proportional trading costs simultaneously. Given trading costs F and T we solve numerically for $\hat{p}$ such that $V(\hat{p}, F, T) = 0$ jointly verifying the admissibility conditions for the costs. Specifically, $\hat{p} = \min p$ such that $D - S > 0$ and $\hat{p}$ is feasible.

We fix $\underline{\alpha} = 1$, $\bar{\alpha} = 10$, $\underline{\beta} = -5$ and $\bar{\beta} = 10$ and consider three different cases: both $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ uniform (i.e. $\alpha_\alpha = \alpha_\beta = \beta_\alpha = \beta_\beta = 1$); $\{\alpha_i\}_i$ tilted towards high values (i.e. $\alpha_\alpha = 3$, $\beta_\alpha = 1$) with $\{\beta_i\}_i$ uniform; and $\{\beta_i\}_i$ tilted towards high values (i.e. $\alpha_\beta = 3$, $\beta_\beta = 1$) with $\{\alpha_i\}_i$ uniform. In Figure 4 the left (resp. right) column depicts the ratio $\bar{p}/p^\star$ (resp. $(\bar{p} - p^\star)/p^\star$ resulting from a uniform cap tightening) as a function of F and T in these three cases.[16] The thick black line delineates the admissible cost range – above it, the market breaks down.

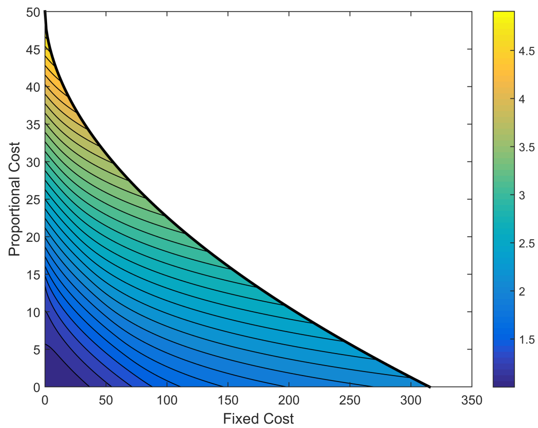
When both $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ are uniform the market price in presence of trading costs is always larger than the frictionless market price (the ratio is always above one in Fig. 4a) and always increases more than the frictionless price in response to the cap tightening (the ratio is always above zero in Fig. 4b). Additionally, because the contour lines are decreasing and convex both the market price level and its response to the cap tightening are larger the higher the entry or unitary trading cost. This is no longer the case in the other two cases. Specifically, for some trading cost levels, the market price can be lower than the frictionless market price and decrease when the trading costs increase given the shape of the contour lines. Indeed, as the frictionless market price is relatively higher in these cases, this illustrates the tempering price effect of the trading costs mentionned in the previous section. In these cost zones also, the market price response to a cap tightening is less than that of the frictionless one.[17]

[16] The uniform cap tightening coincides with a uniform increase in firms' deficits, i.e. $\underline{\beta} = -4$ and $\bar{\beta} = 11$.

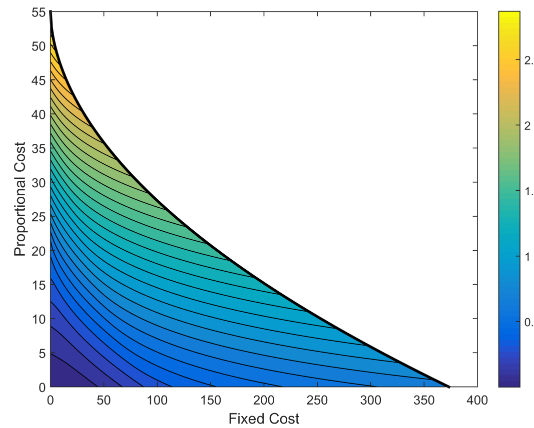
[17] The cases where $\{\alpha_i\}_i$ and/or $\{\beta_i\}_i$ are tilted towards low values yield magnified but qualitatively similar results as for $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ both uniform, hence omitted here.

Figure 4: Ratios $\hat{p}/p^\star$ (left) and $(\hat{p}-p^\star)/p^\star$ (right; following a uniform cap tightening)

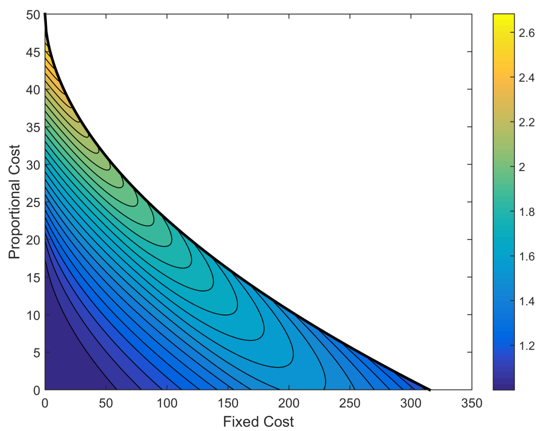
(a) $\alpha_\alpha = \beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p^\star = 9.8$)



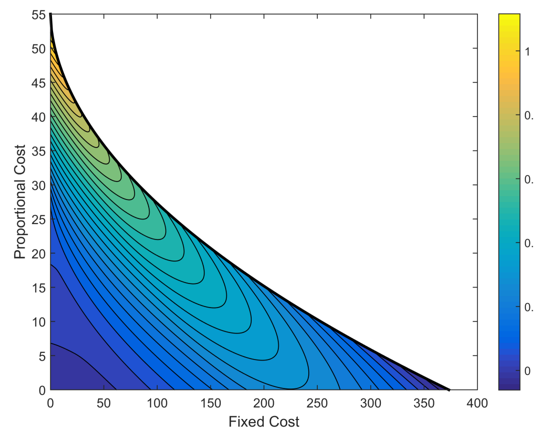
(b) $\alpha_\alpha = \beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p^\star = 13.7$)



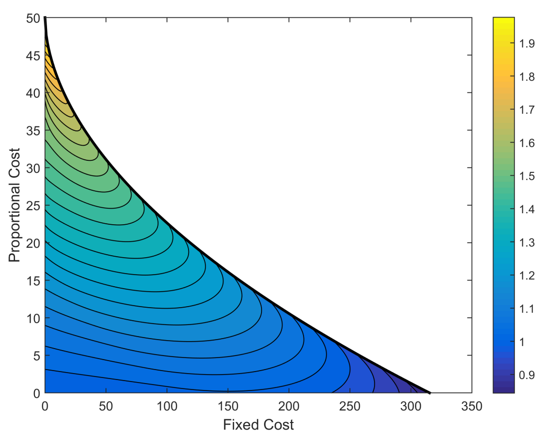
(c) $\alpha_\alpha = 3$, $\beta_\alpha = \alpha_\beta = \beta_\beta = 1 (p^\star = 20.0)$



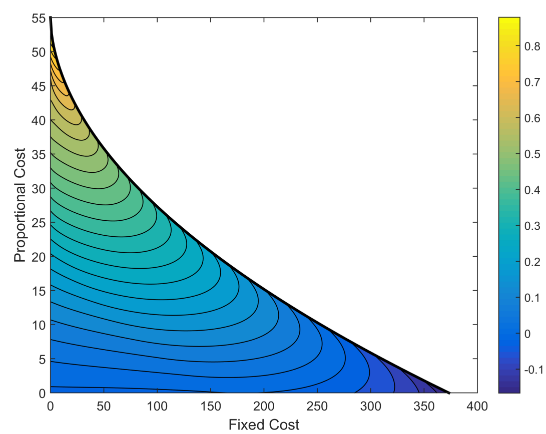
(d) $\alpha_\alpha = 3$, $\beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p^\star = 25.2$)



(e) $\alpha_\beta = 3$, $\alpha_\alpha = \beta_\alpha = \beta_\beta = 1$ ($p^\star = 24.4$)



(f) $\alpha_\beta = 3$, $\alpha_\alpha = \beta_\alpha = \beta_\beta = 1 (p^\star = 28.3)$



5 Applications

The preceding sections highlight that trading costs affect market outcomes, which has important implications for policy design, implementation and evaluation. In this section, we first calibrate fixed and proportional trading costs to the EU-ETS market data over Phase II. Our strategy is to find the pairs (F, T) that yield the best match between our modeling output and 2009-2012 observations in terms of both annual market price and firms' decisions (i.e. market participation and net position). Second, we leverage our model to simulate the impact of a supply-curbing policy intervention on the market price in presence vs. absence of trading costs. We also simulate different permit allocation scenarios to better understand their impact on firms' trading behavior and in turn, on aggregate compliance costs. That is, we compare the situations in which the regulator is aware of the existence of trading costs to that in which they are ignored, and discuss what this implies for policy design and evaluation.

5.1 Calibration to EU-ETS Phase II

Yearly initial allocations and observed verified emissions data are readily available at the regulated source level from the EUTL, which we consolidate at the firm level (see Appendix B) to get rid of non-market, intra-firm permit reshuffling. We respectively denote them $q_{i,t}$ and $e^o_{i,t}$ for consolidated firm i in year t. We compute the yearly-averaged observed permit spot price, denoted p^o_t, and choose observed 2005 emission levels as a reasonable first-pass proxy for annual counterfactual emission levels, denoted $u_{i,t}$, over Phase II. We exclude firms with implied negative abatement (i.e. with $u_{i,t} - e^o_{i,t} < 0$) leading to some year-on-year changes in number of observations in our sample of firms.[18] With this, we compute the firms' marginal abatement cost slopes $\alpha_{i,t}$ and net deficits $\beta_{i,t}$ such that

$$\alpha_{i,t} = (p^o_t \pm T)/(u_{i,t} - e^o_{i,t}) \quad \text{and} \quad \beta_{i,t} = u_{i,t} - q_{i,t}.$$

Note that present the proportional trading cost T, marginal abatement cost slopes need be conditioned on firms' observed market net positions.[19] Table 1 contains some descriptive statistics of the firms' annually calibrated characteristics $(\alpha_{i,t}, \beta_{i,t})$.

[18] In total, after removing firms that did not participate to the market or for which transaction data were not fully provided on the EUTL or computed characteristics did not satisfy $u_{i,t} - e^o_{i,t} \geq 0$, we are left with about a half of all consolidated firms for each year over 2009-2012.

[19] When firms do not record any trading activity on the market, we compute their marginal abatement cost slopes as $p^o_t/(u_{i,t} - e^o_{i,t})$, thereby assuming that they considered the market price as a relevant signal to guide their abatement decisions though not participating to the market effectively.

	#Obs	Min	Max	Mean	Median	% > 0
α_{2008}	1,811	$1.9{\cdot}10^{-6}$	9.8	$4.5{\cdot}10^{-2}$	$5.3{\cdot}10^{-3}$	–
α_{2009}	2,080	$1.8{\cdot}10^{-6}$	13.3	$2.5{\cdot}10^{-2}$	$2.4{\cdot}10^{-3}$	–
α_{2010}	1,783	$2.5{\cdot}10^{-6}$	13.3	$2.9{\cdot}10^{-2}$	$2.3{\cdot}10^{-3}$	–
α_{2011}	1,911	$2.4{\cdot}10^{-6}$	1.02	$1.1{\cdot}10^{-2}$	$2{\cdot}10^{-3}$	–
α_{2012}	1,823	$2.3{\cdot}10^{-6}$	2.7	$1.7{\cdot}10^{-2}$	$1.8{\cdot}10^{-3}$	–
β_{2008}	1,811	-2,138,179	8,534,035	21,286	-212	44.2%
β_{2009}	2,080	-2,150,453	11,270,359	29,022	-356	42.1%
β_{2010}	1,783	-2,150,453	8,534,035	29,543	-354	42.5%
β_{2011}	1,911	-2,150,453	8,534,035	25,808	-387	42.1%
β_{2012}	1,823	-2,261,345	4,720,594	18,491	-425	42.3%

Table 1: Firms' characteristics in Phase II: α in €/$(tCO_2)^2$ when $T = 0$ and β in tCO_2

Next, given a couple (F, T) and a permit price p, we populate the sets $\mathcal{D}_t$, $\mathcal{S}_t$ and $\mathcal{A}_t$ defined in Proposition 2 with the calibrated firms' characteristics in a given year t. This defines yearly supply S_t and demand D_t functions as per (6) and (7). We then solve for the year-t equilibrium price such that $\bar{p}_t = \min p$ such that $D_t - S_t > 0$.

However, it is likely that the theoretical and observed prices $\bar{p}_t$ and p^o_t will differ due to other factors impinging on market price formation in practice. To partly control for this, we introduce yearly fixed effects η_t capturing common shocks to (or common trends in) firm-level counterfactual emissions.[20] Specifically, individual marginal abatement cost schedules are adjusted to $\alpha_{i,t}(u_{i,t} - e_{i,t}) + \eta_t$, which can be interpreted as a shift in individual counterfactual emissions from $u_{i,t}$ to $u_{i,t} + \eta_t/\alpha_{i,t}$. We then choose η_t such that $\bar{p}_t = p^o_t$.[21]

To every pair (F, T) and year t there thus corresponds a η_t ensuring that $\bar{p}_t = p^o_t$. Next, to choose among all candidate trading cost pairs, we proceed with an analysis of the modeling sorting errors, i.e. resulting discrepancies between firms' market participation and net positions as predicted by the model and as observed in practice. Five types of sorting errors are displayed in Table 2. Types 1, 2 and 5 relate to firm's market participation decisions, while types 3 and 4 relate to their net positions on the market.

Appendix C shows the sensitivity of errors to an increase of F or T. First, note that type 1 and 2 errors are mainly driven by the magnitude of F. In particular, higher fixed costs widen the autarky zone, increasing the number of autarkic firms theoretically and, mechanically, type 2 errors. On the opposite, more firms with no observed trading activity are accurately placed in the autarkic zone, diminishing type 1 errors. As a subclass of type 1, type 5

[20] Technically, $\{\eta_t\}$ follows a white noise process, i.e. for all t and $t' \neq t$, $\mathbb{E}\{\eta_t\} = 0$ and $\mathrm{Cov}\{\eta_t, \eta_{t'}\} = 0$.
[21] When $\bar{p}_t$ is initially higher (resp. lower) than p^o_t, η_t need be shifted downwards (resp. upwards).

		Observations		
		Autarkic	*Buyer*	*Seller*
Model	*Autarkic*	–	Type 2	Type 2
	Buyer	Type 1	–	Type 4
	Seller	Type 1 or 5	Type 3	–

Table 2: Typology of sorting errors at the consolidated firm level

changes in the same direction. Recall that firms for which $\beta_i < -F/(p - T - \eta_t)$ should rationally sell their surplus allowances on the carbon market irrespective of their marginal cost of abatement. Type 3 and 4 errors rather depend on T and both decrease with its magnitude. Indeed, firms with extreme β's generally have a clear-cut market position. As T increases, the thresholds of the autarkic zone move away to the corners of the (α, β)-space, thus chances to mistakenly take a buyer for a seller (and vice versa) drop.

In order to choose among the remaining (F, T) pairs, we want to minimize the total number of sorting errors as well as favor balanced distributions of these errors among error types. The latter objective corresponds to maximizing Shannon's entropy applied to the distribution of errors $\{\epsilon_1, \ldots, \epsilon_5\}$. If we let $|\epsilon_i|$ denote the cardinality of type-i errors and $\theta_i = |\epsilon_i| \Big/ \sum_{j=1}^{5} |\epsilon_j|$ the probability of occurrence of a type-i error, the entropy H is defined by

$$0 \leq H = -\sum_{i=1}^{5} \theta_i \log(\theta_i) \leq \log(5),$$

and is maximal when the ϵ_i's are equiprobable. Denoting N the total number of firms in the sample, we thus select the pair (F, T) so as to maximize the index

$$\Big(H/\log(5)\Big) \cdot \Big(\Big(N - \sum_{i=1}^{5} |\epsilon_i|\Big) \Big/ N\Big),$$

which is comprised between 0 and 1. This index reflects the arbitrage between having a higher (resp. lower) entropy associated with lower (resp. higher) values of F and/or T, and the corresponding increase (resp. decrease) in the total number of sorting errors.

Table 3 contains our calibration results over Phase II. First note that $\eta_t > 0$ for all t considered, meaning that an upward shift in counterfactual emissions is necessary to replicate observed permit prices. Alternatively, a positive η can be indicative of a market-wide incentive to bank distributed but yet non-remitted permits for future compliance. Indeed, our static model static leaves aside the intertemporal dimension of the EU-ETS and, in aggregate,

Year	p_t^o (€/tCO$_2$)	η_t (€/tCO$_2$)	F (k€)	T (€/tCO$_2$)	T (% of p_t^o)
2008	19.6	3.9	10	1.7	8.6
2009	13.3	5.1	30	0.25	1.8
2010	14.3	5.6	20	0.80	5.5
2011	13.1	5.7	30	0.05	0.4
2012	7.4	4.9	10	0,01	0.1

Table 3: Observed price p_t^o and calibrated model parameters η_t, F and T over Phase II

firms had an incentive to bank permits over Phase II.[22] Introducing yearly fixed effects η_t's allows us to partly control for this intertemporal incentive and capture the associated price rise in our static setup.[23] We also note that our framework does not account for primary (i.e. auctions) and derivatives trading (e.g. forwards, futures) as the EUTL only records the physical movement of EUAs on the secondary market.[24]

That said, our calibrated fixed costs are in the same order of magnitude as other empirical studies, like Naegele (2018) who finds average fixed trading costs of €100,000 approximately. Our calibrated proportional trading costs are also in line with the existing literature. Joas & Flachsland (2016) find that trading costs ranged from €ct0.1/tCO$_2$ to €ct10/tCO$_2$, and Heindl (2012b) evaluates trading costs to be on average €ct15/tCO$_2$ for small firms. In their study, Medina et al. (2014) find that the bid-ask spread for Phase II futures contracts range from 1 to 10% of the permit price. Although our results need to be refined, our model is able to replicate the observed market behavior (participation and net position) of 70% of all the firms in the sample on average over Phase II.

Although Table 3 shows the pair (F,T) corresponding to the index's global maximum, local maxima can be determined for every value of T.[25] Specifically, the index is maximized by the same F for any given T in the range of values considered. The difference in value between the global and local maxima can be very small, meaning that while F can be pinpointed with a fair degree of confidence, the exact value of T is more uncertain to the nearest tenths. Quantifying and reducing this margin of error is to be done in future work.

[22]This is a well-documented empirical fact, see inter alia Ellerman et al. (2016) and Fuss et al. (2018).

[23]Indeed, $\eta_t > 0$ can alternatively be seen as capturing a reduction in effective initial allocation levels as some allocated permits are carried over to subsequent compliance periods. Thus, η_t effectively captures the total time-fixed effect on the net deficit $\beta_{i,t} = u_{i,t} - q_{i,t}$ from $\beta_{i,t}$ to $\beta_{i,t} + \eta_t/\alpha_{i,t}$. Directly adjusting initial allocation levels to correct for observed banking yields quantitatively similar calibration results.

[24]However, note that more than 90% of all permits were freely allocated in Phase II.

[25]We consider F ranging from 1€/tCO$_2$ to 100,000€/tCO$_2$ and T ranging from 0€/tCO$_2$ to 1€/tCO$_2$ except in 2008 where we had to raise T further.

5.2 Policy design in presence of trading costs

The fact that transaction costs affect market outcomes has important implications for policy design and implementation. For instance, the newly operational Market Stability Reserve (MSR) is bound to induce a noticeable supply squeeze over the short to mid term in the EU-ETS, and thereby to raise the permit price. In this context, we use our model to examine numerically the magnitudes of the market price responses in presence vs. absence of trading costs. In other words, we take the perspective of two regulators in evaluating the price impact of their supply-curbing reform: one who believes that the market operates without frictions; the other who is aware of the existence of trading costs.

Specifically, we analyze the price impacts of a one-sixth slash in the annual overall cap on emissions in 2009 and 2012 for the selected samples of firms. We consider that the cutback impinges on individual allocations uniformly or that it is targeted on over-allocated or under-allocated firms alternatively. Table 4 summarizes our results. As expected, first note that the distribution of free allocation is neutral in terms of market clearing absent transaction costs. Indeed, the frictionless price change induced by the cutback in a given year is identical irrespective of its distribution among firms.

Second, price increases are higher present trading costs regardless of the incidence of the supply squeeze on firms. In the context of our calibration exercise, therefore, the mass of sellers is relatively more constricted than that of buyers, leaving the market shorter than absent trading costs. Additionally, note that targeting the cap slash on firms that are initially under-allocated (resp. over-allocated) generates a greater (resp. smaller) price increase than when the cutback is evenly distributed among firms. Intuitively, this by construction increases (resp. decreases) the mass of buyers (resp. sellers) as some autarkic (resp. selling) firms would now find it profitable to enter (resp. exit) the market.

Interestingly, price increases are relatively larger in 2009 than 2012 regardless of the allocation method. This difference can possibly be attributed to the initial observed price levels: recall that Claim 3 shows that trading costs amplify the price response to a supply shock only if price levels are high enough to start with. In 2012 and 2009, observed prices were on average 7.4 €/tCO_2 and 13.3 €/tCO_2, respectively. Thus 2012 price levels likely approached the threshold below which trading costs would rather buffer supply shocks than amplify them.

More generally, our results suggest that the presence of trading costs might amplify the price impacts of supply shocks relative to frictionless conditions. In the context of the MSR, not only will the induced price increase be under-estimated by assuming away trading costs as

	2009			2012		
	Uniform cutback	*Targeted on* $\beta < 0$	*Targeted on* $\beta > 0$	*Uniform cutback*	*Targeted on* $\beta < 0$	*Targeted on* $\beta > 0$
$p^\star$	22.8	22.8	22.8	11.6	11.6	11.6
$\bar{p}$	31.5	31.9	37.2	15.1	14.6	17.3

Table 4: Price responses (in €/tCO_2) to a one-sixth cap slash with different distributions

is common practice in modeling approaches, but it will also under-estimate the associated increase in price variability. Indeed, Perino & Willner (2016) show that if the MSR is sufficiently stringent and a demand shock occurs before the end of the banking period, it tends to increase price volatility. More, Kollenberg & Taschini (2016) highlight the risk of an immediate depletion of the permit bank under risk aversion. This exposure might be even more relevant present trading costs. Indeed, if the price response to a supply squeeze is greater than expected, the aggregate bank might turn to be too small to cushion the shock.

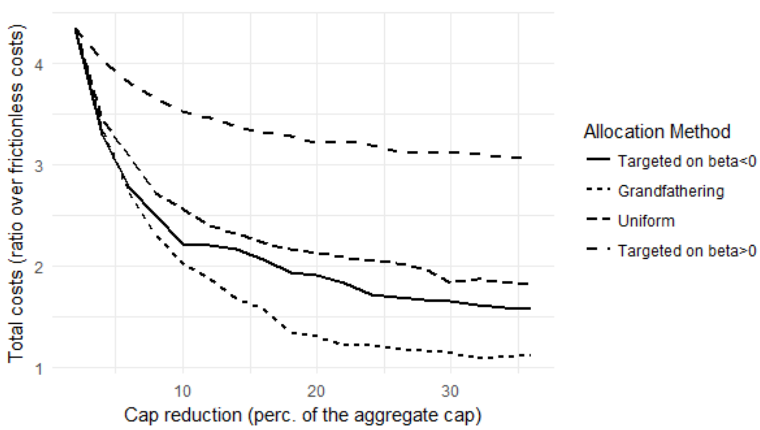


Figure 5: Evolution of pollution control costs with different permit allocation methods

Next, we use the calibrated (F, T, η_t) to simulate pollution control costs under different permit allocation methods. Figure 5 depicts the evolution of aggregate compliance costs when the sample's cap becomes tighter. The ordinate is the ratio of compliance costs present trading costs over costs that would have prevailed in the frictionless case. In turn, a ratio close to 1 reflects similar pollution control costs.

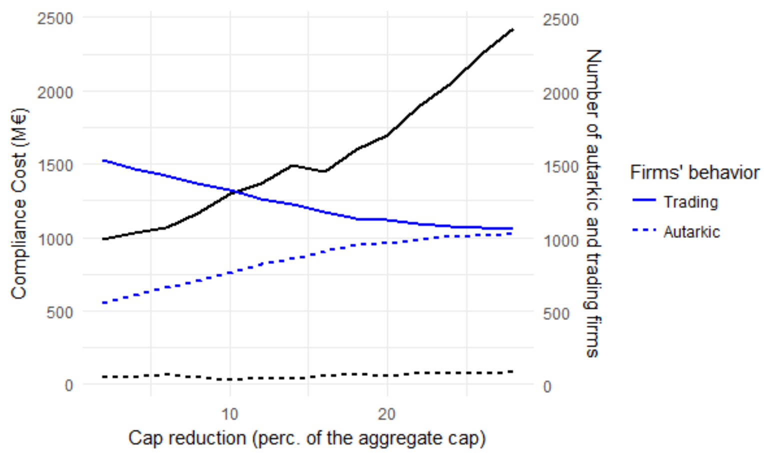


Figure 6: Decomposition of compliance costs by trading behavior under grandfathering

First, it looks that 'grandfathering' (i.e. a reduction in permit supply proportional to individual endowments), is the preferred allocation method from a cost perspective. Intuitively, it is the least distortive as every firm would have to furnish the same additional abatement effort relative to their emission volume. This result is consistent with Burtraw et al. (2001) who finds that grandfathering achieves the biggest producer surplus relative to auctioning or a performance standard. Second and interestingly, the ratio tends to decrease with the stringency of the aggregate cap before stabilizing, regardless of the allocation method. Recall from Section 3 that the cost of compliance depend on a price effect (i.e. the sensitivity of the permit price to a supply cutback) and a distribution effect (i.e. how the supply cutback affects firms' market participation).

Figure 6 decomposes the effect of a supply cutback in the least-cost scenario. First note that the number of autarkic firms increases with the cap reduction, revealing a transfer of firms from the seller zone to the autarkic zone. Second, the aggregate cost of pollution control rises for market participants while that of autarkic firms remains relatively stable despite their increasing number. These trends reveal that on the one hand, the compliance costs of firms who were already in the set of buyers (i.e. at the intensive margin) increased dramatically as they fell even shorter of permits with the supply cutback. On the other hand, sellers who were forced into the autarkic zone (i.e. at the extensive margin) bore a relatively small increase in compliance costs. An interpretation is that these firms held a small permit surplus before the supply crunch, so the forgone gain from selling it on the market was modest anyway. These results need careful interpretation though, as they crucially depend on the specific distribution of firms' permit deficits and marginal abatement costs in the EU-ETS.

6 Conclusion

Over Phase II of the EU-ETS, nearly a third of liable companies did not record any transactions other than compliance-related. With an original analysis of 2008-2012 transaction and compliance data, we first provide evidence of significant barriers to trading in the EU-ETS. We then develop a model of permit trading in presence of fixed and proportional trading costs in which firms can be initially over-allocated. Both market participation and clearing are endogenous and hinge on firms' characteristics and trading costs. This allows us to characterize the price responses to shifts in the trading costs and allocation levels analytically. We next calibrate our model using market data. We find that the permit price is increasing in both fixed and proportional trading costs, which are in the order of k€10-30 and €0.1-1 per permit traded respectively on average to match with observed prices and firms' market behaviors over Phase II. We also quantify how supply-curbing policies like the Market Stability Reserve induce a larger price increase than under frictionless market conditions and specify how this relates to the firm-level distribution of the aggregate supply cutback. Finally, we simulate different permit allocation scenarios and find that grandfathering achieves the smallest increase in compliance costs in our context.

We finally highlight two alleys for future research. First, introducing auctioning as a complementary allocation method would allow us to characterize the price impacts of more realistic supply squeezes on auction like those induced by the MSR instead of on free allocation. Second, it would be interesting to account for intertemporal trading in presence of trading costs, as already hinted at in Section 5. This would allow us to investigate to which extent trading costs may incentivize firms to participate to the market closer to compliance dates.

References

Betz, L., Cludius (2016). *Trading costs in the first phase of the EU ETS: Estimation and explanations - Evidence from CITL data*. Working Paper, SCCER - CREST.

Burtraw, D., Palmer, K. L., Bharvirkar, R. & Paul, A. (2001). *The effect of allowance allocation on the cost of carbon emission trading*. Technical report.

Carlson, C., Burtraw, D., Cropper, M. & Palmer, K. L. (2000). Sulfur Dioxide Control by Electric Utilities: What are the Gains from Trade? *Journal of Political Economy*, **108**(6), 1292–326.

Cason, T. N. & Gangadharan, L. (2003). Transactions Costs in Tradable Permit Markets: An Experimental Study of Pollution Market Designs. *Journal of Regulatory Economics*, **23**(2), 145–65.

Charles, A., Darné, O. & Fouilloux, J. (2013). Market efficiency in the european carbon markets. *Energy policy*, **60**, 785–792.

Constantatos, C., Filippiadis, E. & Sartzetakis, E. S. (2014). Using the Allocation of Emission Permits for Strategic Trade Purposes. *Journal of Regulatory Economics*, **45**(3), 259–80.

Coria, J. & Jaraite, J. (2018). Transaction costs of upstream versus downstream pricing of co2 emissions. *Environmental and Resource Economics*.

Dixit, A. (1989). Entry and Exit Decisions under Uncertainty. *Journal of Political Economy*, **97**(3), 620–38.

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.

Fowlie, M. & Perloff, J. M. (2013). Distributing Pollution Rights in Cap-and-Trade Programs: Are Outcomes Independent of Allocation? *Review of Economics & Statistics*, **95**(5), 1640–52.

Fowlie, M., Reguant, M. & Ryan, S. P. (2016). Market-Based Emissions Regulation and Industry Dynamics. *Journal of Political Economy*, **124**(1), 249–302.

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.

Gangadharan, L. (2000). Transaction Costs in Pollution Markets: An Empirical Study. *Land Economics*, **76**(4), 601–14.

Hahn, R. W. & Stavins, R. N. (2011). The Effect of Allowance Allocations on Cap-and-Trade System Performance. *Journal of Law & Economics*, **54**(S4), S267–94.

Heindl, P. (2012a). Financial intermediaries and emissions trading: Market development and pricing strategies. *ZEW Discussion Papers*, **12**(64).

Heindl, P. (2012b). Transaction Costs and Tradable permits: Empirical Evidence from the EU Emissions Trading Scheme. *ZEW Disucssion Papers*, (21).

Heindl, P. & Lutz, B. (2012). Carbon Management-Evidence from Case Studies of German

Firms under the EU ETS. *ZEW Discussion Papers*, (79).

Jaraitė, J., Convery, F. & Di Maria, C. (2010). Transaction Costs for Firms in the EU ETS: Lessons from Ireland. *Climate Policy*, **10**(2), 190–215.

Jaraitė-Kažukauskė, J. & Kažukauskas, A. (2015). Do Transaction Costs Influence Firm Trading Behaviour in the European Emissions Trading System? *Environmental & Resource Economics*, **62**(3), 583–613.

Joas, F. & Flachsland, C. (2016). The (Ir)Relevance of Transaction Costs in Climate Policy Instrument Choice: An Analysis of the EU and the US. *Climate Policy*, **16**(1), 26–49.

Karpf, A., Mandel, A. & Battiston, S. (2018). Price and Network Dynamics in the European Carbon Market. *Journal of Economic Behavior & Organization*, **153**, 103–22.

Kollenberg, S. & Taschini, L. (2016). *Dynamic Supply Adjustment and Banking under Uncertainty: the Market Stability Reserve*. Working paper, Grantham Research Institute.

Liski, M. (2001). Thin versus Thick CO_2 Market. *Journal of Environmental Economics & Management*, **41**(3), 295–311.

Martin, R., Muûls, M. & Wagner, U. (2014). Trading behavior in the eu emissions trading scheme.

Medina, V., Pardo, Á. & Pascual, R. (2014). The Timeline of Trading Frictions in the European Carbon Market. *Energy Economics*, **42**, 378–94.

Montero, J.-P. (1998). Marketable Pollution Permits with Uncertainty and Transaction Costs. *Resource & Energy Economics*, **20**(1), 27–50.

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

Naegele, H. (2018). Offset Credits in the EU ETS: A Quantile Estimation of Firm-Level Transaction Costs. *Environmental & Resource Economics*, **70**(1), 77–106.

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

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

Quemin, S. & Trotignon, R. (2019). *Intertemporal Emissions Trading and Market Design: An Application to the EU-ETS*. Working Paper, Grantham Research Institute.

Sandoff, A. & Schaad, G. (2009). Does EU ETS Lead to Emission Reductions through Trade? The Case of the Swedish Emissions Trading Sector Participants. *Energy Policy*, **37**(10), 3967–77.

Schultz, E. & Swieringa, J. (2014). Catalysts for price discovery in the european union emissions trading system. *Journal of Banking & Finance*, **42**, 112–122.

Singh, R. & Weninger, Q. (2017). Cap-and-Trade under Transactions Costs and Factor Irreversibility. *Economic Theory*, **64**(2), 357–407.

Stavins, R. N. (1995). Transaction Costs and Tradeable Permits. *Journal of Environmental Economics & Management*, **29**(2), 133–48.

Szabo, M. (2019). Eu emitters that borrow forward for annual ets compliance face rude awakening, experts warn. *Carbon Pulse*.

Appendices

A Analytical derivations and collected proofs

A.1 Proof of Proposition 1

Let $K_i : x \mapsto \alpha_i x^2/2$ denote firm i's abatement cost function. Given a market price $p > 0$, i abates up to $a_i^*(p) = p/\alpha_i$ and its individual efficiency gains from permit trading are defined by

$$G_i(p) = K_i(a_i^0) - \Big(K_i(a_i^*(p)) + p(\beta_i - a_i^*(p))\Big),$$

where $a_i^0 = \alpha_i \max\{0; \beta_i\}$. Recalling that $p_i^0 = \alpha_i a_i^0$, the above rewrites as follows

$$\begin{aligned}
G_i(p) &= K_i(a_i^0) - \Big(K_i(a_i^*(p)) + p(\max\{0; \beta_i\} + \min\{0; \beta_i\} - a_i^*(p))\Big) \\
&= K_i(a_i^0) - \Big(K_i(a_i^*(p)) + p(a_i^0 + \min\{0; \beta_i\} - a_i^*(p))\Big) \\
&= \Big((p_i^0)^2 - p^2 - 2pp_i^0 + 2p^2\Big)/(2\alpha_i) - p\min\{0; \beta_i\} \\
&= (p_i^0 - p)^2/(2\alpha_i) - p\min\{0; \beta_i\} = (p_i^0 - p)^2/(2\alpha_i) + p\max\{0; -\beta_i\}.
\end{aligned}$$

Firm i is better off buying (resp. selling) permits when $p_i^0 > p$ (resp. $p_i^0 < p$) which defines the sets $\mathcal{D}$ and $\mathcal{S}$. Consequently, no firm is willing to buy (resp. sell) permits on the market when $p \geq \bar{\alpha}\bar{\beta}$ (resp. $p = 0$). Thus a market price is feasible when $p \in (0; \bar{\alpha}\bar{\beta})$.

A.2 Proof of Lemma 1

For $\beta_i \leq 0$, i sells permits on the market if $G_i(p - T) - F > 0$, i.e. $X^2 - 2\alpha_i\beta_i X - 2\alpha_i F \geq 0$ with $X = p - T$. Only keeping the positive root, this occurs if $p - T > \alpha_i\beta_i + \sqrt{\alpha_i^2\beta_i^2 + 2\alpha_i F}$, which is always non-negative.

For $\beta_i > 0$, i buys (+) or sells (−) permits on the market if $X^2 - 2\alpha_i(F + \beta_i X) + \alpha_i^2\beta_i^2 > 0$ with $X = p \pm T$. For a seller, we only keep the relevant root $X = p - T > \alpha_i\beta_i$ so i partakes in the market if $p - T > \alpha_i\beta_i + \sqrt{2\alpha_i F}$, which is always non-negative. For a buyer, we only keep the relevant root $X = p + T < \alpha_i\beta_i$ so i partakes in the market if $p + T < \alpha_i\beta_i - \sqrt{2\alpha_i F}$. This is non-negative provided that F is not too large, i.e. $F \leq \alpha_i\beta_i^2/2$. This must at least be true for the last potential buyer so $F \leq \bar{\alpha}\bar{\beta}^2/2$ (see Lemma 2).

A.3 Proof of Proposition 2

Expanding firm i's market participation $G_i(p \pm T) - F \geq 0$ constraint gives

$$\begin{aligned}
&(p_i^0)^2 - 2p_i^0(p \pm T) + (p \pm T)^2 - 2\alpha_i(p \pm T)\min\{0;\beta_i\} - 2\alpha_i F \geq 0 \\
\Leftrightarrow\ &\alpha_i^2(\max\{0;\beta_i\})^2 - 2\alpha_i(p \pm T)(\max\{0;\beta_i\} + \min\{0;\beta_i\}) - 2\alpha_i F + (p \pm T)^2 \geq 0 \\
\Leftrightarrow\ &\alpha_i^2(\max\{0;\beta_i\})^2 - 2\alpha_i(\beta_i(p \pm T) + F) + (p \pm T)^2 \geq 0.
\end{aligned}$$

Substantiating the three different cases depending on the pairs (α_i, β_i), this rewrites

$$\begin{aligned}
\alpha_i^2\beta_i^2 - 2\alpha_i(F + (p+T)\beta_i) + (p+T)^2 \geq 0 &\ \text{ when } \alpha_i\beta_i \geq p + T, \\
\alpha_i^2\beta_i^2 - 2\alpha_i(F + (p-T)\beta_i) + (p-T)^2 \geq 0 &\ \text{ when } 0 < \alpha_i\beta_i \leq p - T, \text{ or} \\
-2\alpha_i(F + (p-T)\beta_i) + (p-T)^2 \geq 0 &\ \text{ when } \beta_i \leq 0.
\end{aligned}$$

Thus, when $\beta_i \leq 0$ (resp. $\beta_i > 0$) the α_i-thresholds obtain by solving a first-order (resp. second-order) polynomial inequality and keeping the relevant roots. When $\beta_i \leq -F/(p-T)$, the third inequality above holds for all $\alpha_i > 0$. This defines the sets $\mathcal{D}(p,F,T)$ and $\mathcal{S}(p,F,T)$.

Next, we show that $\mathcal{S}(p,F,T)$ (resp. $\mathcal{D}(p,F,T)$) effectively contains all selling (resp. buying) firms. To see this, note that i is a seller (resp. buyer) i.f.f. $\beta_i - a_i^*(p-T) \leq 0 \Leftrightarrow \alpha_i \leq (p-T)/\beta_i$ (resp. $\beta_i - a_i^*(p+T) \geq 0 \Leftrightarrow \alpha_i \geq (p+T)/\beta_i$). This suffices to prove our claim since i's thresholds can easily be shown to satisfy $\alpha_i^0, \alpha_i^- \leq (p-T)/\beta_i$ and $(p+T)/\beta_i \leq \alpha_i^+$.

Below, we provide the partial derivatives of the α-thresholds with their signs:

$$\begin{aligned}
\partial\alpha^\pm/\partial p = (1 \pm F/X_1^\pm)\big/\beta > 0 \quad & \partial\alpha^\pm/\partial\beta = -X_2^\pm(1 \pm F/X_1^\pm)/\beta^3 < 0 \\
\partial\alpha^\pm/\partial F = (1 \pm X_3^\pm/X_1^\pm)\big/\beta^2 \gtreqless 0 \quad & \partial\alpha^\pm/\partial T = (\pm 1 + F/X_1^\pm)\big/\beta^2 \gtreqless 0 \\
\partial\alpha^0/\partial p = (p-T)X_2^-\big/2(X_3^-)^2 > 0 \quad & \partial\alpha^0/\partial\beta = -(p-T)^3\big/2(X_3^-)^2 < 0 \\
\partial\alpha^0/\partial F = -(p-T)^2\big/2(X_3^-)^2 < 0 \quad & \partial\alpha^0/\partial T = -(p-T)X_2^-\big/2(X_3^-)^2 < 0
\end{aligned}$$

where $X_1^\pm = \sqrt{F(F + 2(p \pm T)\beta)} > F$, $X_2^\pm = 2F + (p \pm T)\beta > 2F$ and $X_3^\pm = X_2^\pm - F > X_1^\pm$. This proves our claim on the changes in the measures of $\mathcal{D}(p,F,T)$ and $\mathcal{S}(p,F,T)$ as p, F or T increases. Note also that $\lim_{\beta\to 0^+}\alpha^+ = \lim_{\beta\to 0^+} 2F/\beta^2 = +\infty$. To get at $\lim_{\beta\to 0^+}\alpha^-$, we first compute the second-order Taylor expansion of the numerator in α^-, namely

$$F + (p-T)\beta - F\left(1 + \frac{1}{2}\frac{2(p-T)\beta}{F} - \frac{1}{8}\left(\frac{2(p-T)\beta}{F}\right)^2\right) = \frac{(p-T)^2\beta^2}{2F},$$

so that $\alpha^- \sim_{\beta\to 0^+} (p-T)^2/(2F) = \alpha^0(p,F,T;0)$, i.e. there is continuity between α^- and α^0 in $\beta = 0$. By a similar token, $\partial\alpha^-/\partial\beta \sim_{\beta\to 0^+} = -(p-T)^3/2F^2 = \partial\alpha^0/\partial\beta(p,F,T;0)$, i.e. there is also continuity in slope. Finally, $\lim_{\beta\to 0^+} \partial\alpha^+/\partial\beta = \lim_{\beta\to 0^+} -4F/\beta^3 = -\infty$, $\lim_{\beta\to+\infty} \alpha^\pm = 0$, $\lim_{\beta\to -F/(p-T)} \alpha^0 = +\infty$ and $\lim_{\beta\to -F/(p-T)} \partial\alpha^0/\partial\beta = -\infty$, which completes the description of the behaviors of the supply and demand frontiers in Figure 3.

A.4 Proof of Lemma 2

A price is feasible as long as there is at least one buyer and one seller in the market. Hence (5) follows from Lemma 1 applied to the last potential buyer $(\bar\alpha, \bar\beta)$ and seller $(\underline{\alpha}, \underline{\beta})$. Alternatively, the two price bounds obtain by solving $\alpha^0(p,F,T;\underline{\beta}) = \underline{\alpha}$ and $\alpha^+(p,F,T;\underline{\beta}) = \bar\alpha$. Trading costs are admissible if there exist feasible prices. From (5) this requires $\underline{\beta}\underline{\alpha} + T + \sqrt{\underline{\beta}^2\underline{\alpha}^2 + 2\underline{\alpha}F} < \bar\alpha\bar\beta - T - \sqrt{2\bar\alpha F}$ and $\bar\alpha\bar\beta - \sqrt{2\bar\alpha F} > p + T > 0$, which gives (4).

A.5 Proof of Lemma 3

First note that D and S are continuous and differentiable in p, F and T.

Partial derivatives w.r.t. p: D (resp. S) is strictly decreasing (resp. increasing) with p. In the case of D, it suffices to see that $p \mapsto y - (p+T)/x$ is strictly decreasing with p and that α^+ is strictly increasing with p. A similar argument follows for S, although the behavior of the second term in (6) is unclear as the bound $-F/(p-T)$ is increasing with p. To clarify this, we compute the partial derivatives of the two terms of interest in (6) using Leibniz's rule in conjunction with Fubini's theorem yielding

$$F/(p-T)^2 \int_{\underline{\alpha}}^{\bar\alpha} \Big((p-T)/x + F/(p-T)\Big) g(x|y = -F/(p-T)) h(-F/(p-T)) \mathrm{d}x$$
$$-F/(p-T)^2 \int_{\underline{\alpha}}^{\alpha^0} \Big((p-T)/x + F/(p-T)\Big) g(x|y = -F/(p-T)) h(-F/(p-T)) \mathrm{d}x,$$

which concludes since by Chasles' rule the above simplifies to

$$F/(p-T)^2 \int_{\alpha^0}^{\bar\alpha} \Big((p-T)/x + F/(p-T)\Big) g(x|y = -F/(p-T)) h(-F/(p-T)) \mathrm{d}x \; > \; 0.$$

Partial derivatives w.r.t. F: A qualitative argument as in the above could suffice but formal calculus will prove helpful in the following. Differentiating (7) and (6) w.r.t. F gives

$$\frac{\partial D}{\partial F} = -\int_0^{\bar\beta} \frac{\partial \alpha^+(p,F,T;y)}{\partial F}\Big(y-(p+T)/\alpha^+(p,F,T;y)\Big)g(\alpha^+(p,F,T;y)|y)h(y)\mathrm{d}y < 0,$$

$$\begin{aligned}\frac{\partial S}{\partial F} = &-1/(p-T)\int_{\underline\alpha}^{\bar\alpha}\Big((p-T)/x+F/(p-T)\Big)g(x|y=-F/(p-T))h(-F/(p-T))\mathrm{d}x\\ &+1/(p-T)\int_{\underline\alpha}^{\alpha^0}\Big((p-T)/x+F/(p-T)\Big)g(x|y=-F/(p-T))h(-F/(p-T))\mathrm{d}x\\ &+\int_{-F/(p-T)}^{0}\frac{\partial\alpha^0(p,F,T;y)}{\partial F}\Big((p-T)/\alpha^0(p,F,T;y)-y\Big)g(\alpha^0(p,F,T;y)|y)h(y)\mathrm{d}y\\ &+\int_0^{\bar\beta}\frac{\partial\alpha^-(p,F,T;y)}{\partial F}\Big((p-T)/\alpha^-(p,F,T;y)-y\Big)g(\alpha^-(p,F,T;y)|y)h(y)\mathrm{d}y.\end{aligned}$$

By Chasles' rule again, the first two terms in $\partial S/\partial F$ reduce to

$$-1/(p-T)\int_{\alpha^0}^{\bar\alpha}\Big((p-T)/x+F/(p-T)\Big)g(x|y=-F/(p-T))h(-F/(p-T))\mathrm{d}x \;<\; 0.$$

Thus $\partial S/\partial F < 0$ as the last two terms in $\partial S/\partial F$ are also negative.

Partial derivatives w.r.t. T: Similar arguments show that $\partial S/\partial T$ and $\partial D/\partial T$ are negative.

A.6 Proof of Proposition 3

The proof relies on the intermediate value theorem applied to $V = S - D$ which Lemma 3 shows to be continuous and strictly increasing with p.

Denote the upper (resp. lower) feasible price bound in (5) by $\bar p$ (resp. $\underline p$). Assume trading costs are admissible as in (4), thus $\bar p > \underline p$. By definition, $D(\bar p, F, T) = 0$ since $\alpha^+(\bar p, F, T; \bar\beta) = \bar\alpha$ and $\alpha^+ > \bar\alpha$ for all $0 < \beta < \bar\beta$ as α^+ is strictly decreasing with β. Because D is strictly decreasing with p, $D(p, F, T) > 0$ for any $p < \bar p$. Similarly, by definition $S(\underline p, F, T) = 0$. Indeed the first integral in S is nil since $\underline\beta > -F/(\underline p - T)$; the second and third integrals are also nil since $\alpha^0(\underline p, F, T; \underline\beta) = \underline\alpha$ so that $\alpha^0 < \underline\alpha$ and $\alpha^- < \underline\alpha$ for all $\beta > \underline\beta$ since α^0 and α^- are decreasing with β. Because S is strictly increasing with p, $S(p, F, T) > 0$ for any $p > \underline p$.

Therefore, $V(\underline p, F, T) = -D(\underline p, F, T) < 0$ and $V(\bar p, F, T) = S(\bar p, F, T) > 0$. The intermediate value theorem concludes: there exists $\hat p \in (\underline p; \bar p)$ such that $V(\hat p, F, T) = 0$ and it is unique.

A.7 Proof of Proposition 5

We want to compute and determine the magnitudes of both the price and distribution effects in face of a small change in the cap level accounting for induced changes at both the intensive and extensive margins. We study the two effects in turn.

Price effect: We want to rank $\partial V(\hat{p}, F, T)/\partial p$ relative to $\partial V(p^\star, 0, 0)/\partial p$. First note that

$$\frac{\partial D(\hat{p}, F, T)}{\partial p} = \underbrace{-\int_0^{\bar{\beta}} \int_{\alpha^+}^{\bar{\alpha}} (1/x) g(x|y) h(y) \mathrm{d}x\mathrm{d}y}_{\text{intensive margin} \leq 0} \underbrace{- \int_0^{\bar{\beta}} \frac{\partial \alpha^+}{\partial p} \Big(y - (\hat{p} + T)/\alpha^+ \Big) g(\alpha^+|y) h(y) \mathrm{d}x\mathrm{d}y}_{\text{extensive margin} \leq 0},$$

where we omit some functional arguments to avoid clutter. The intensive margin captures the decrease in demand on the part of firms in $\mathcal{D}(\hat{p}, F, T)$. The extensive margin captures what happens at the $\mathcal{A}$-$\mathcal{D}$ frontier, e.g. the additional decrease in demand due to firms exiting $\mathcal{D}$ and entering $\mathcal{A}$ as a result of the price increase. Note that on the frontier the net demand $y - (\hat{p} + T)/\alpha^+$ is nil when $F = 0$ for any $T \geq 0$ since $\alpha^+ = (\hat{p} + T)/y$; and positive whenever $F > 0$. Indeed, with $F > 0$, there exist jumps in aggregate demand as firms enter or exit $\mathcal{D}$ with positive individual demand as the market price varies – such discontinuities do not exist when $F = 0$. This means that the extensive margin drops to zero with $F = 0$.

We proceed similarly for S, see also Appendix A.5 for details. In turn, we obtain

$$\begin{aligned}\frac{\partial V(\hat{p}, F, T)}{\partial p} &= \overbrace{\frac{\partial V(p^\star, 0, 0)}{\partial p} - \int_0^{\bar{\beta}} \int_{\alpha^-}^{\alpha^+} (1/x) g(x|y) h(y) \mathrm{d}x\mathrm{d}y - \int_0^{\bar{\beta}} \int_{\alpha^0}^{\bar{\alpha}} (1/x) g(x|y) h(y) \mathrm{d}x\mathrm{d}y}^{\text{intensive margin}} \\ &\quad + \underbrace{\text{sum of positive terms}}_{\text{extensive margin}} \gtrless \frac{\partial V(p^\star, 0, 0)}{\partial p} = \int_{\underline{\beta}}^{\bar{\beta}} \int_{\underline{\alpha}}^{\bar{\alpha}} (1/x) g(x|y) h(y) \mathrm{d}x\mathrm{d}y.\end{aligned}$$

When $F = 0$ and the extensive margin effects are nil, $\partial V(\hat{p}, F, T)/\partial p < \partial V(p^\star, 0, 0)/\partial p$, i.e. the price effect is above one. The price effect can be below one for some pairs (F, T) for which the extensive margin effects are large enough. This is more likely to occur for small trading costs as the intensive margin term is decreasing with the trading cost levels.

Distribution effect: Consider the collection of individual deficit shifts $\{\beta_i + \gamma(\beta_i)\}_i$ for some bounded function γ such that $|\gamma| \ll 1$. Subsequent demand D_1 evaluated at $(\hat{p}, F, T)$ reads

$$D_1(\hat{p}, F, T) = \int_0^{\bar{\beta}} \int_{\alpha^+(y_0 + \gamma(y_0))}^{\bar{\alpha}} \Big(y_0 + \gamma(y_0) - (\hat{p} + T)/x \Big) g(x|y_0) h(y_0) \mathrm{d}x\mathrm{d}y_0,$$

where we omit all arguments in α^+ that are irrelevant for the proof to reduce clutter. For all

y_0 we can expand α^+ in powers of γ as follows

$$\alpha^+(y_0 + \gamma(y_0)) = \alpha^+(y_0) + \gamma(y_0)\frac{\partial \alpha^+}{\partial y}\bigg|_{y=y_0} + \mathcal{O}(|\gamma(y_0)|^2).$$

Denoting equilibrium demand prior to small cap change by D_0, one gets

$$\begin{aligned} D_1(\hat{p}, F, T) = D_0(\hat{p}, F, T) + \int_0^{\bar{\beta}} \int_{\alpha^+(y_0)}^{\bar{\alpha}} \gamma(y_0) g(x|y_0) h(y_0) \mathrm{d}x \mathrm{d}y_0 \\ - \int_0^{\bar{\beta}} \int_{\alpha^+(y_0)}^{\alpha^+(y_0)+\gamma(y_0)\frac{\partial \alpha^+}{\partial y}|_{y=y_0}} \Big(y_0 + \gamma(y_0) - (\hat{p}+T)/x\Big) g(x|y_0) h(y_0) \mathrm{d}x \mathrm{d}y_0 + \mathcal{O}(|\gamma(y_0)|^2). \end{aligned}$$

The last line in the above expression can be approximated by

$$-\int_0^{\bar{\beta}} \gamma(y_0) \frac{\partial \alpha^+}{\partial y}\bigg|_{y=y_0} \Big(y_0 - (\hat{p}+T)/\alpha^+(y_0)\Big) g(\alpha^+(y_0)|y_0) h(y_0) \mathrm{d}x \mathrm{d}y_0 + \mathcal{O}(|\gamma(y_0)|^2),$$

where the approximation becomes exact in the limit as $|\gamma| \to 0$. Further assuming a uniformly distributed cap tightening, i.e. γ is constant and positive, $\lim_{\gamma\to 0}(D_1 - D_0)/\gamma$ writes

$$\underbrace{\int_0^{\bar{\beta}} \int_{\alpha^+(y_0)}^{\bar{\alpha}} g(x|y_0) h(y_0) \mathrm{d}x \mathrm{d}y_0}_{\text{intensive margin} \geq 0} \underbrace{- \int_0^{\bar{\beta}} \frac{\partial \alpha^+}{\partial y}\bigg|_{y=y_0} \Big(y_0 - (\hat{p}+T)/\alpha^+(y_0)\Big) g(\alpha^+(y_0)|y_0) h(y_0) \mathrm{d}x \mathrm{d}y_0}_{\text{extensive margin} \geq 0}$$

as $\lim_{\gamma\to 0} \mathcal{O}(\gamma) = 0$. The intensive margin captures the increase in demand on the part of firms in $\mathcal{D}$ prior to the tightening. The extensive margin captures what happens at the $\mathcal{A}$-$\mathcal{D}$ frontier, i.e. the novel demand on the part of firms exiting $\mathcal{A}$ and entering $\mathcal{D}$. Note again that the extensive margin component drops for any $T \geq 0$ with $F = 0$.

We proceed similarly for S – computations are longer but follow the same lines. Then, all the terms in $\lim_{\gamma\to 0}(V_1 - V_0)/\gamma$ can be regrouped into two categories

$$\lim_{\gamma\to 0}(V_1(\hat{p}, F, T) - V_0(\hat{p}, F, T))/\gamma = \underbrace{|\mathcal{I}| - |\mathcal{A}(\hat{p}, F, T)|}_{\text{intensive margin}} + \underbrace{\text{positive terms}}_{\text{extensive margin}} \gtrless |\mathcal{I}|,$$

where $|\mathcal{I}| = \lim_{\gamma\to 0}(V_1(p^\star, 0, 0) - V_0(p^\star, 0, 0))/\gamma$. Roughly put, the larger the set of autarkic firms, i.e. the larger the trading costs, the more likely the distribution effect is below one, i.e. $\lim_{\gamma\to 0}(V_1(\hat{p}, F, T) - V_0(\hat{p}, F, T))/\gamma < |\mathcal{I}|$ holds. With $F = 0$, this holds for all $T \geq 0$.

Finally, we consider alternative distributions of the cap tightening in the intensive margin only case treated in the body of the paper. When it is uniformly distributed among all firms, $\mathrm{d}\beta_i =$

$\mathrm{d}\mathcal{Q}$ holds for all $i \in \mathcal{I}$. When it is uniformly targeted on all firms with positive (resp. negative) deficits, $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}/(|\bar{\mathcal{S}}_3| + |\bar{\mathcal{D}}| + |\bar{\mathcal{A}}_2|) > \Delta\mathcal{Q}$ (resp. $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}/(|\bar{\mathcal{S}}_1| + |\bar{\mathcal{S}}_2| + \bar{\mathcal{A}}_1|) > \Delta\mathcal{Q}$) holds for all i with $\beta_i > 0$ (resp. $\beta_i < 0$) where the sets $\mathcal{S}_k$ and $\mathcal{A}_k$ are defined in Figure 3 and the upper bar means evaluated at $(\hat{p}, F, T)$. In these three cases the distribution effect is worth

$$
\begin{aligned}
&(|\bar{\mathcal{S}}_1| + |\bar{\mathcal{S}}_2| + |\bar{\mathcal{S}}_3| + |\bar{\mathcal{D}}|)/|\mathcal{I}| < 1, \\
\text{or } &(|\bar{\mathcal{S}}_3| + |\bar{\mathcal{D}}|)/(|\bar{\mathcal{S}}_3| + |\bar{\mathcal{D}}| + |\bar{\mathcal{A}}_2|) < 1, \\
\text{or } &(|\bar{\mathcal{S}}_1| + |\bar{\mathcal{S}}_2|)/(|\bar{\mathcal{S}}_1| + |\bar{\mathcal{S}}_2| + |\bar{\mathcal{A}}_1|) < 1.
\end{aligned}
$$

The magnitude of the price increase in the face of a given reduction in the cap hinges on the way of allocating it among firms. The ranking between the three types of distribution presented above is unclear prima facie – it depends on the levels of the trading costs F and T, and the distribution of the firms' characteristics $\{\alpha_i\}_i$ and $\{\beta_i\}_i$.

A.8 Proofs of Examples 1, 2 and 3 (sketch)

After standard calculus (14) and (15) obtain by solving $D(\hat{p}, F, 0) = S(\hat{p}, F, 0)$ for $\hat{p}$ while $p^\star$ obtain by solving $D(p^\star, 0, 0) = S(p^\star, 0, 0)$. Below we sketch the steps in the computations for the case $k = 1$ and omit those for $k = 2$ as they follow the same lines. Define function K_1 by

$$
K_1 = \hat{p}_1 + \frac{2F\sqrt{F(F + 2\beta\hat{p}_1)}}{\beta^3(\bar{\alpha} - \underline{\alpha})} - p_1^\star \ (= 0),
$$

so that the implicit function theorem then yields

$$
\frac{\mathrm{d}\hat{p}_1}{\mathrm{d}F} = -\frac{\partial K_1/\partial F}{\partial K_1/\partial \hat{p}_1} \ \text{ and } \ \frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta} = -\frac{\partial K_1/\partial \beta}{\partial K_1/\partial \hat{p}_1}.
$$

One has $\mathrm{d}\hat{p}_1/\mathrm{d}F < 0$ and $\mathrm{d}\hat{p}_1/\mathrm{d}\beta > 0$ since it is easy to check that $\partial K_1/\partial F > 0$, $\partial K_1/\partial \hat{p}_1 > 0$ and $\partial K_1/\partial \beta < 0$. Then, the second equality above can rewrite as follows

$$
\frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta}(1 + \beta^2 X) = X(3F + 5\beta\hat{p}_1) + \frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta} \ \text{ with } \ X = 2F^2\Big/\Big(\beta^4(\bar{\alpha} - \underline{\alpha})\sqrt{F(F + 2\beta\hat{p}_1)}\Big),
$$

so that it comes

$$
\frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta} - \frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta} > 0 \ \Leftrightarrow \ \hat{p}_1 > (\beta p_1^\star - 3F)/5\beta.
$$

B Consolidation methodology

The data recorded through the European Union Transaction Log tells us about compliance-related information on the one hand and trading activity on the other hand. Two databases can be consolidated in turn: the 'compliance' database keeps track of initial permit allocation to polluting plants while the 'transaction' database records every transaction completed on the market, including account holder names of sellers and buyers, the date and volume of allowances exchanged. Account holders, who can be managers of production sites or traders, also inform plants' characteristics such as industry, account type, registry, etc. Two issues arise when trying to match plants by company. First, no particular field indicates the name of the parent company. Usually, account holders fill an 'Account Holder Name' field, which is very uneven as to the precision of company information (one company usually has at least as many accounts as production sites). Second, there is no key to match the two databases, so that we need to create our own 'company' key beforehand.

We proceed by first working on the 'compliance' database to extract a list company names, that we use thereafter as a key to associate the trading information from the 'transaction' database. To extract the list of parent firms, we first proceed by cleaning the 'Account Holder Name' (around 17,000 accounts, all years included) field of the 'compliance' database, like removing punctuation, business names, prefixes and suffixes etc. We then separate words in the character string. We run a first matching round of duplicates on the first words of the cleaned 'Account Holder Names' to get a draft parent-company list and associate the corresponding plants name to it (in practice, the company name is often specified in the first or second word in the search field). Next, we scan the second words of the cleaned 'Account Holder Names' to try and associate plants to their parent company. We repeat this operation on the list of third and fourth words. Finalized with a final manual check, this simple method allows us to get a reasonably good idea of which company own which accounts. Around 10% of accounts remains 'single- as we couldn't associate them with a parent-company. We manually assign them to the right one. Finally, we obtain a list of companies associated with all corresponding 'Account Holder Names' in the compliance database (roughly 8,500 firms all years accumulated). Since the 'Account Holder Name' is also present in the transaction database, the matching of the two databases at the company level is straightforward.

Over the second phase of the EU-ETS, 7215 firms were recorded in the 'Compliance' database. Besides, 7210 businesses recorded some activity in the 'Transaction' database, some of them with no compliance obligations under the EU-ETS. Cutting those back (in practice, this amounts to keep 'Operator Holding Accounts' or 'Former Operator Holding Accounts' only) boils the 'Transaction' database down to 5145 polluting firms in total over Phase II. Figures 1 is based on these observations. Separating years leads to slight changes in the number of observations (± 150) due to firms entering or exiting the market.

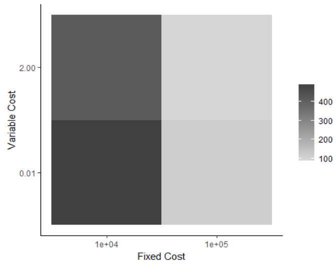
Furthermore, the quantitative application in Section 5 requires to match the 'Compliance' and 'Transaction' databases since information about permit allocation, verified emissions and trading activity are needed at the firm level. This new aggregated database can theoretically include 5145 observations at most. It amounts to 3615 observations in practice, due to matching errors. This new consolidated database is our basis in Figure 2, but still needs some cleaning to be used as an input in Section 5. Out of the 3615 observations, we exclude firms whose 2005 emissions are null, as this is a crucial information for computing baseline emissions. Next, conditional on the year considered, we exclude firms whose emissions are null, whose initial permit allocations and market positions are not available. Without this information, we cannot calibrate trading costs the way it is done in Section 5. This leads to slight changes in the number of observations year-to-year. Table 5 is compiled upon this clean database. Finally, we remove firms with a negative abatement and our yearly dataset is ready for use. This leads, again, to slight changes in the number of yearly observations (see Table 1 for details). Table 1 summarizes descriptive statistics of the 'turnkey' databases used as inputs in Section 5.

Activity	**Trading Firms**	*Average # of accounts*	*% of total emission*	*Median deficit*	*Average # of transactions*	*Average size of transactions (tCO_2)*
# Obs	(1451)	-	-	-	-	-
Combustion	872	2.6	71.2%	70,654	15.8	39,517
Refining	23	4	8.5%	72,063	10.9	211,449
Metallurgy	37	3.4	3.2%	-114,510	6.8	218,805
Cement & Lime	344	3.1	15.1%	-25,474	5.5	34,095
Chemicals	8	3.1	0.2%	-11,056	6.1	52,201
Paper & Glass	164	2.6	1.6%	-8,272	8.3	23,242
Other	3	6	≈ 0	-5,600	4	1,494

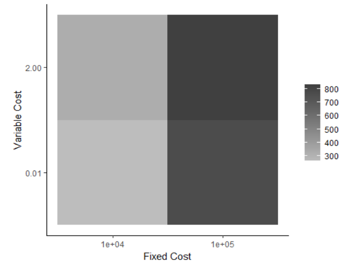
Activity	**Non-Trading Firms**	*Average # of accounts*	*% of total emission*	*Median deficit*	*Average # of transactions*	*Average size of transactions (tCO_2)*
# Obs	(1290)	-	-	-	-	-
Combustion	848	2.3	62.6%	3,665	-	-
Refining	14	2.1	11%	267,178	-	-
Metallurgy	41	2	3.7%	-52,835	-	-
Cement & Lime	165	2.3	13.6%	-7,749	-	-
Chemicals	4	2	1.2%	-13,266	-	-
Paper & Glass	217	2.4	8.9%	-3,114	-	-
Other	1	1	-	-11,139	-	-

Table 5: Descriptive statistics on consolidated firms in 2009

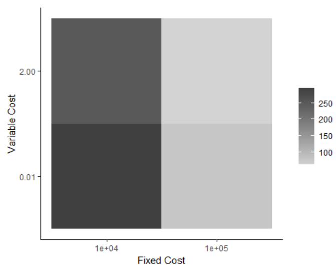
C Sensitivity analysis (preliminary)



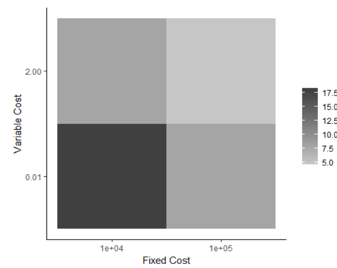
(a) Type 1 errors



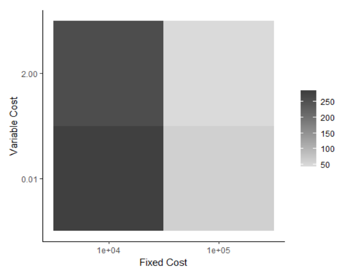
(b) Type 2 errors



(d) Type 3 errors



(e) Type 4 errors



(c) Type 5 errors

Figure 7: Preliminary sensitivity analysis, 2009 sample