Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Emissions trading with transaction costs

Marc Baudry, Anouk Faure and Simon Quemin

June 2020

Centre for Climate Change Economics and Policy Working Paper No. 371
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment Working Paper No. 341
ISSN 2515-5717 (Online)

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

The Grantham Foundation for the Protection of the Environment

University of Leeds

The London School of Economics and Political Science

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

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

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

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and a number of other sources. It has 11 broad research areas:

1. Climate change adaptation and resilience
2. Climate change governance, legislation and litigation
3. Environmental behaviour
4. Environmental economic theory
5. Environmental policy evaluation
6. International climate politics
7. Science and impacts of climate change
8. Sustainable finance
9. Sustainable natural resources
10. Transition to zero emissions growth
11. UK national and local climate policies

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

Suggested citation:
Baudry M, Faure A and Quemin S (2020) *Emissions trading with transaction costs*. Centre for Climate Change Economics and Policy Working Paper 371/Grantham Research Institute on Climate Change and the Environment Working Paper 341. London: London School of Economics and Political Science

Emissions Trading with Transaction Costs*

Marc Baudry[1,2] Anouk Faure[1,2] Simon Quemin[2,3,4]

[1]EconomiX (UMR 7235) – *Université Paris Nanterre*

[2]Climate Economics Chair – *Institut Louis Bachelier, Paris*

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

[4]Potsdam Institute for Climate Impact Research – *Member of the Leibniz Association*

June 2020

[latest version]

Abstract

We develop an equilibrium model of emissions permit trading in the presence of fixed and proportional trading costs in which the permit price and firms' participation in and extent of trading are endogenously determined. We analyze the sensitivity of the equilibrium to changes in the trading costs and firms' allocations, and characterize situations where the trading costs alternatively depress or raise permit prices relative to frictionless market conditions. We calibrate our model to annual transaction and compliance data in Phase II of the EU ETS (2008-2012) which we consolidate at the firm level. We find that trading costs in the order of 10 k€ per annum plus 1 € per permit traded substantially reduce discrepancies between observations and theoretical predictions for firms' behavior (e.g. autarkic compliance). Our simulations suggest that ignoring trading costs leads to an underestimation of the price impacts of supply-curbing policies, this difference varying with the incidence on firms.

Keywords Emissions trading, Transaction costs, Policy design and evaluation, EU ETS.

JEL classification D22, D23, H32, L22, Q52, Q58.

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

«Failure to account explicitly for bounded rationality, uncertainty, and informational, contractual and policing costs inherent in all air pollution problems weakens the applicability of [the] conclusions about the allocative efficiency properties of alternative control instruments.»

— Thomas D. Crocker (1972)

«Since standard economic theory assumes transaction costs to be zero, the Coase Theorem demonstrates that the Pigovian solutions are unnecessary in these circumstances. Of course, it does not imply, when transaction costs are positive, that [...] regulation or taxation [...] could not produce a better result than relying on negotiations between individuals in the market. [...] My conclusion: let us study the world of positive transaction costs.»

— Nobel Memorial Prize Lecture, Ronald H. Coase (1992)

1 Introduction

Stemming from the seminal works of Coase (1960), Crocker (1966) and Dales (1968) and later formalized by Montgomery (1972), emissions trading has become pivotal in the environmental and climate change mitigation regulatory toolbox.[1] Purportedly, comparative advantages of this instrument include cost effectiveness, modest information requirements for the regulator and a political-economy lever by means of the initial distribution of emissions permits. The collective optimum can in principle be decentralized via the market price: given a total supply of permits, the same price level obtains in equilibrium and abatement efforts are rerouted to firms with lowest marginal abatement costs irrespective of their initial allocation as a result of market participants' endeavor to ferret out least-expensive abatement sources.

As two of the concept's founding fathers recognize in the opening quotes, however, a variety of barriers to trading – usually grouped under the term of transaction costs – can drive a wedge between theoretical and practical market outcomes. In practice, indeed, frictions of various types are acknowledged to be pervasive as the empirical literature on permit markets attests (e.g. Carlson et al., 2000; Gangadharan, 2000; Hahn & Stavins, 2011; Jaraitė-Kažukauskė & Kažukauskas, 2015; Venmans, 2016; Karpf et al., 2018; Naegele, 2018; Cludius & Betz, 2020) and, more generally, there are costs associated with trading in financial markets (e.g. Gârleanu & Pedersen, 2013; Dávila & Parlatore, 2020, and references therein). We corroborate these findings with a descriptive analysis of trading and compliance patterns in the second phase of

[1] Medema (2020) offers an excellent overview of the historical context and impacts of the Coase theorem. See also Deryugina et al. (2020) for a recent review of its applications to environmental problems.

the EU Emissions Trading System (2008-12) which also suggests the existence of transaction costs. This echoes one conclusion of Hintermann et al. (2016) who review the literature on price and market behaviors in Phase II of the EU ETS: they highlight that transaction costs are a key factor that impinge on price formation, notably its level, and can explain persisting differentials in marginal abatement costs across firms.

Yet, surprisingly, the prevalence of transaction costs and their implications for market outcomes as well as for policy design, evaluation and implementation are largely ignored in the theoretical literature on permit markets, with only a few exceptions discussed below.[2] In this paper, we seek to remedy this gap. Specifically, we incorporate trading costs in an otherwise archetypal emissions trading model to formally analyze how they impact the market equilibrium. We further calibrate the model to observed transactions in Phase II of the EU ETS to offer an illustration based on a relevant real-world example. As we shall see, not only is such a framework better equipped to conduct finer-grained ex-post analyses of firms' trading and compliance behaviors, but it also constitutes a more realistic basis for ex-ante assessments of supply-side management policies, a regular feature in the hybrid ETSs of today.

We articulate three contributions to the literature. To motivate our analysis further, we begin by exploring the universe of annual transactions in EU ETS Phase II, our policy environment in this paper. Data is available at the account (installation) level but we consider the firm as the relevant decision-making unit for our analysis and we concentrate on inter-firm trading.[3] To aggregate the data at the firm level and remove intra-firm permit reallocations, we develop a consolidation methodology matching each installation to a parent company building on an iterative search procedure for duplicates in the accounts' information fields. We then utilize the consolidated dataset to scrutinize firms' annual trading and compliance behaviors. The consolidation methodology and the description of observed firms' market behavior constitute our first contribution to the empirical literature on the EU ETS.

We find evidence of autarkic compliance and signs of impaired trading, i.e. some gains from trade go unrealized at both the extensive and intensive margins, see Ellerman et al. (2010), Martin et al. (2015) and Schleich et al. (2020) for similar descriptive results.[4] At the extensive

[2] In a related context, see Dixit & Olson (2000) and Anderlini & Felli (2006) for formal analyses of Coasean bargaining in the presence of transaction costs.

[3] A firm typically owns several regulated polluting sites and can redistribute allocated permits across sites based on their realized emissions, i.e. it can potentially achieve compliance without effectively trading on the market. Here, we implicitly assume that the costs of intra-firm transfers are negligible compared to those of inter-firm trading. In our dataset, intra-firm transfers represent 27% of all cross-account flows in Phase II.

[4] Other studies have provided indirect evidence of transaction costs relying on surveys and interviews with firms' managers (e.g. Venmans, 2016; Heindl, 2017) or analyzing transactions with network theory tools (e.g. Borghesi & Flori, 2018; Hintermann & Ludwig, 2018; Karpf et al., 2018). See Section 2.1 for more details.

margin, about a third of firms did not trade at all on a yearly basis. Autarkic firms are mostly small (in terms of size of emissions or number of installations) representing 9% of regulated emissions and often hold excess permits w.r.t. realized emissions. At the intensive margin, active firms engaged in trading infrequently (typically a few times per annum) and only for sufficiently high volumes, suggesting that marginal abatement costs are not equalized across firms in equilibrium. Stifled trade at both margins points us to the prevalence of both fixed and variable transaction costs (see also Stavins, 1995; Singh & Weninger, 2017).[5]

Our second contribution to the literature is theoretical in nature. We enrich a standard static and deterministic permit trading model by introducing both fixed and proportional trading costs.[6] The fixed cost impacts firms' decisions to take part in the market (extensive margin) while the proportional cost further affects firms' trading choices by driving a wedge between their marginal abatement costs (intensive margin). In our equilibrium framework, the permit price and firms' participation in and extent of trading are determined endogenously, and they depend on the given trading costs and firms' characteristics (i.e. abatement costs and permit allocations, where we let some firms be initially overallocated).[7] Importantly, this framework enables us to study when a market equilibrium exists and if so, how it is achieved.

Tracking trading cost impacts through buyer-seller interactions and resulting market prices, we can analyze the sensitivity of the market equilibrium to changes in the trading costs and firms' initial allocations. While an increase in trading costs always reduces cost effectiveness and the volume of trade, its price effects are ambiguous and non-monotonic in general as they depend on its relative impacts on the supply and demand sides of the market (i.e. ultimately the distributions of firms' characteristics). As a rule, we find that trading costs are generally conducive to higher price levels when the (theoretical) frictionless market price is 'low', and vice versa. Similarly, the price increase following a reduction in the total number of permits can be amplified or dampened by the presence of trading costs. This hinges on a distribution effect (the overall impact on net permit demand, holding the price constant) and a price effect (the relative price elasticity of net permit demand with vs. without trading costs) which are

[5] In the words of Stavins (1995) «transaction costs can take one of two forms, inputs of resources–including time–by a buyer and/or seller or a margin between the buying and selling price», i.e. fixed and variable costs. More concretely, fixed entry costs can compound exchange membership fees with other resources invested in operating a trading desk, monitoring the market and defining a trading strategy. Variable costs can comprise search, information, brokerage, intermediation and consultancy costs inter alia.

[6] We solely focus on those transaction costs entailed by (or conditional on) permit trading, hence termed trading costs for short. See Section 2.1 for a brief overview of other types of transaction costs.

[7] Following the related literature (e.g. Stavins, 1995; Singh & Weninger, 2017) we take the trading costs as exogenously given, but we discuss in Section 2 how they may emerge in practice. As a notable exception, see Liski (2001) for microfoundations, i.e. a formal treatment of trading costs endogenously arising and evolving over time as a function of the market size and the initial distribution of permits among firms.

generally countervailing. To gain additional insight into the market impacts of trading costs, we then illustrate our theoretical results with analytical and numerical examples for different distributions of firms' characteristics. These show that both the price level and increase (due to a lowered amount of permits) are more often higher with trading costs than without.

The benchmark framework to analyze the impacts of transaction costs in markets permit has been developed by Stavins (1995) and extended by Montero (1998).[8] Crucially, however, this is not an equilibrium framework and the market price is taken as exogenously fixed. That is, Stavins and Montero study the impacts of trading costs on an individual firm's emission and trading choices at the margin but do not formally characterize the market price impacts nor how firms self-select into costly trading in the first place as we do in this paper. As a result, our framework sometimes leads to different results, e.g. market outcomes are sensitive to the initial allocation of permits even with constant marginal trading costs. More recently, Singh & Weninger (2017) have developed a similar equilibrium framework in the presence of fixed or proportional trading costs, alternatively.[9] But in their model, firms are ex-ante identical and differ only in idiosyncratic productivity shocks, the main motive for permit trade.[10] Our analysis is hence different in nature as we choose to focus on what we believe to be the more practically relevant case of ex-ante heterogeneous firms (see Bernard et al. (2012) and Melitz & Redding (2014) in the more general context of international trade in goods).

This brings us to our third contribution to the literature, which exploits firms' heterogeneity in abatement costs and allocations allowed by the model. Specifically, we utilize the universe of yearly allocations, emissions, transactions and prices in EU ETS Phase II to discipline the calibration of model parameters and the selection of practically relevant trading costs values. We propose a selection criterion minimizing the total number of sorting errors (i.e. discrepancies between firms' market participation and net market positions in the model vs. the data) and their dispersion across error types (measured by Shannon's entropy).[11] Respectively, we

[8]Specifically, Montero (1998) offered an extension of Stavins' analysis in the form of uncertainty on trade approval and provided further insights with numerical simulations. Moreover, Cason & Gangadharan (2003) used a laboratory experiment to test (and confirm) the main results implied by Stavins' theory.

[9]Similarly, in a permit trading model with transaction costs, Constantatos et al. (2014) show how permit allocation can be used as a strategic trade instrument on the product market even without market power.

[10]Singh & Weninger invoke an argument in the spirit of Samuelson's Factor Price Equalization theorem whereby in mature ETSs productivity shocks should be the main drivers for trade. While this simplifies their analysis, which accounts for the interaction with the product market, it is our contention that existing ETSs are still far from mature in this respect, and therefore that heterogeneity in abatement costs and allocations remains the main motive for trade (a fortiori in EU ETS Phase II, our policy environment in this paper).

[11]Our calibration methodology also replicates observed annual prices but this cannot be the key selection criterion as it is not robust enough in itself for our purposes (e.g. Carlson et al., 2000) and price depends on a variety of other factors our model does not explicitly account for (but which we control for). Additionally, we focus more on the extensive margin than on the intensive margin impacts of trading costs, as the latter

find fixed and proportional trading costs in the order of 5-20 k€ per annum and 0.5-1.5 € per permit traded (or 3-11% of the permit price) across years. Relative to zero trading costs, the selected trading costs reduce the total number of sorting errors by 40%, their dispersion by 160%, and can rationalize 70% of autarkic compliance cases. Our model calibration exercise thus shows how accounting for trading costs can be important for ex-post policy evaluation. It also provides first-pass estimates of trading costs in the EU ETS where related empirical studies have gathered anecdotal or indirect evidence (e.g. Venmans, 2016; Karpf et al., 2018) or used econometric estimation techniques (e.g. Medina et al., 2014; Jaraitė-Kažukauskė & Kažukauskas, 2015; Naegele, 2018). Interestingly, or perhaps reassuringly for our approach, we obtain similar orders of magnitude as the latter studies.

Finally, we leverage our calibrated model to compare the quantitative results that a modeler or regulator would obtain in assessing the total costs the ETS imposes on firms or the market price impacts of additional supply-curbing policies, depending on whether or not transaction costs are accounted for. In our setting, extra compliance costs resulting from incurred trading costs and foregone efficiency gains are in the order of 7% of the compliance costs in a scenario where transaction costs are ignored. In a similar vein, we find that the price increase following a reduction in the total number of permits would be underestimated if one does not account for transaction costs. This is because in our setting some firms holding excess permits do not offer them for sale due to the transaction costs, implying that the price increase is inefficiently large. Specifically, we find an underestimation factor of two for a one-sixth reduction in the total cap, with variations in size of up to 30-40% depending on its incidence on firms.

The remainder of the paper is organized as follows. We close this section with a brief literature review to contextualize our results and some limitations of our analysis. Section 2 provides further background on transaction costs and trading patterns in EU ETS Phase II. Section 3 develops the emissions trading model in the presence of fixed and proportional trading costs, and provides analytical and numerical illustrations. Section 4.1 describes the calibration of the model to EU ETS Phase II data and the selection of trading costs. Section 4.2 utilizes the calibrated model to evaluate and compare supply-tightening policy impacts in the presence vs. absence of trading costs. Section 5 concludes. An Appendix collects analytical derivations (A) as well as details on the consolidation methodology (B) and calibration results (C).

are hard to quantify meaningfully without precise abatement cost data.

1.1 Contextualization

Before proceeding further, we wish to acknowledge some limitations of our approach in order to put into perspective our modeling choices and results. These relate to temporal, behavioral and institutional factors which may impinge on firms' participation in and extent of inter-firm trading but are not formally treated in the present framework.

Because these interrelated factors can distort and impair inter-firm trading incentives relative to an idealized market environment without frictions, our approach, by exclusively focusing on (pecuniary) transaction costs, may overstate the latter's impacts. That said, the literature on permit markets generally ignores these factors or, more exceptionally, treats them individually as we discuss below. Against this background, we believe that our formal equilibrium analysis of the impacts of transaction costs and underlying mechanisms at play constitutes a welcome addition to the literature, as a fist step towards a more comprehensive framework. Relatedly, our model calibration in Section 4.1 can be thought of as capturing the aggregate impact of these various factors, rather than the exclusive impacts of transaction costs.

Banking and borrowing. Firms generally have some leeway in banking issued permits for future use or borrowing future permits for present use. While this flexibility margin reduces trading incentives, it is not sufficient to rationalize autarkic compliance and one should still expect some potential profits from inter-firm trading. For instance, some firms may still find it too costly to solely abate internally (e.g. when borrowing is restricted) while others may not want to bear the opportunity cost of not selling at least some of their excess permits. Again, this points us towards the prevalence of transaction costs, or other biases discussed below. In this context, intertemporal trading can even serve as a substitute to costly spatial trading. As a case in point, in the US Acid Rain Program (ARP), Toyama (2019) numerically estimates significant trading costs, in a range of 15-35% of the market price per permit traded, which in turn imply excess banking and less dispersed emissions as a result of lower inter-firm trades relative to an idealized counterfactual scenario without trading costs.[12]

Except in Toyama (2019), however, the interplay between trading and banking decisions with transaction costs is seldom accounted for.[13] And more generally, to investigate specific policy

[12]Specifically, buyers (resp. sellers) face a lower (resp. higher) banking incentive as transaction costs drive a wedge between firms' marginal abatement costs – and ultimately a shadow value higher (resp. lower) than the permit price. Thus, aggregate banking could in principle be higher or lower than in the first best.

[13]Rubin (1996) already noted that in the absence of quadratic transaction costs or bounds on firms' trading, a dynamic model suffers from indeterminacy as firms' objective functions are linear in traded volumes – thus, optimal banking and trading decisions cannot be identified. For instance, Kollenberg & Taschini (2016, 2019) introduce quadratic transaction costs to have a well-defined optimization problem and simplify the derivation

impacts, the temporal dimension of permit markets is often set aside in both numerical and analytical approaches. For instance, Fowlie et al. (2016) develop a dynamic oligopoly model to numerically compare firms' entry/exit decisions and capacity investments on the product market under different permit allocation regimes, but do not consider banking provisions to isolate the allocation impacts. Similarly, to assess cost savings and health impacts under the ARP, Chan et al. (2018) use a static model of compliance choices (fuel switch, permit purchase, scrubber installation) to sidestep the complexities associated with modeling banking and permit purchase as an option to install a scrubber or fuel switch at a later date.[14] Both approaches are therefore likely to overemphasize the spatial trading dimension.

As in Singh & Weninger (2017), we develop a static permit trading model in order to be able to derive and exploit novel analytical results on the equilibrium impacts of transaction costs. Since the temporal dimension is not formally treated, our calibration exercise should in turn be seen as providing upper bounds for the transaction costs and their impacts.[15] However, we take firms' observed banking dynamics as a given to adjust their allocations and mitigate this limitation. Additionally, we wish to underline that the temporal dimension is also likely to be subject to other specific costs and limitations, which relate to the other biases discussed below. It is indeed difficult to elicit firms' degree of intertemporal optimization (e.g. Ellerman et al., 2016; Hintermann et al., 2016) and there is evidence of limited farsightedness or biased beliefs (e.g. Chen, 2018; Fuss et al., 2018; Quemin & Trotignon, 2019).[16]

Behavioral biases. Autarkic compliance and stifled trading may also result from behavioral biases. For instance, autarkic firms can be thought of as trading off profits from entering the market with higher associated organizational and decision-making complexity. As heuristics or rules of thumb, autarkic banking and borrowing may thus constitute viable, if not rational, strategies when deemed to perform satisfactorily well relative to more complex and thus costly procedures (e.g. Baumol & Quandt, 1964; Simon, 1979; Radner, 1996; Gigerenzer & Selten, 2003).[17] That is, firms may have to make compromises as they juggle with other objectives,

of the equilibrium in closed form (relative to fixed costs, as they do not affect firms' participation in trading).

[14] Similarly, Carlson et al. (2000) compute long-run gains from trade in a steady state of the ARP, bypassing the issue of modeling intertemporal decisions by taking firms' banking behavior as given.

[15] Similarly, Jaraitė-Kažukauskė & Kažukauskas (2015) and Naegele (2018) employ static frameworks for their econometric estimations of transaction costs and their impacts on firms in the EU ETS.

[16] In the ARP, Chen (2018) shows that it took time for firms to form and adapt their beliefs about future price levels, leading to biased banking strategies especially in the early stages of the program. In the EU ETS, Quemin & Trotignon (2019) show how rolling finite planning horizons help reconcile the observed aggregate banking dynamics with discount rates implied from futures' yield curves.

[17] In the EU ETS for instance, Cludius & Betz (2020) observe that small firms are more likely to pursue a single, simple trading strategy compared to larger, more professionalized firms that are more likely to interact with different partners on different exchanges. Relatedly, Hortaçsu & Puller (2008) explain that large power

perhaps perceived as more essential. For instance, based on interviews with plant managers, Venmans (2016) notes that some perceive the ETS as a command-and-control type of policy, especially when commodity trading is not part of their firms' core business – in this case, the carbon liability is typically dealt with by the accounting department.[18] In other words, firms may rather seek to attain compliance with the least additional complexity and disruption to their routine operations than maximize profits from permit trading.

More generally, an insight from behavioral economics is that firms are likely to be subject to endowment effects w.r.t. their permit holdings, which can also frustrate trading. Indeed, «endowment effects are predicted for property rights [...] such as transferable pollution permits» (Kahneman et al., 1990) because «losses are weighted substantially more than objectively commensurate gains in the evaluation of prospects and trades» (Tversky & Kahneman, 1991). This is confirmed by Murphy & Stranlund (2007) and Venmans (2016) with laboratory experiments and interviews, respectively. As a result, firms' willingness to pay for extra permits is larger than their willingness to sell excess permits, implying that differences in firms' marginal abatement costs can persist post trading independently of transaction costs.[19] Relatedly, we note that other institutional factors may also affect firms' compliance decisions, e.g. the level of trust in institutions as uncovered by Jo (2019).

Intermediaries and trading venues. We further wish to acknowledge that our framework restrains the compliance choice space for firms to a blunt 'autarky vs. trading' and that it does not account for the involvement of non-regulated entities, typically banks and intermediaries. That is, it does not formally distinguish between the different available trading platforms (e.g. auctions, exchanges, over the counter), products, partners or combination thereof that firms may possibly select.[20] Relatedly, it only represents trading as a way of minimizing compliance costs, but it ignores other motives to trade such as hedging or generating additional revenues (e.g. Schleich et al., 2020).[21] Yet, we note that these aspects structure the trading network,

producing firms in Texas perform closer to profit maximization than smaller ones as a result of fixed costs of establishing and maintaining sophisticated auction bidding strategies, with clear economies of scale.

[18]Similarly, Martin et al. (2015), Liu et al. (2017) and Schleich et al. (2020) note that some firms perceive the EU ETS as a pure compliance instrument rather than as a compliance market. Relatedly, Jaraitė et al. (2010) highlight that small firms' reluctance to sell excess permits may be explained by a cautious inclination to keep and use permits for future compliance only rather than by transaction costs per se.

[19]Importantly, Kahneman et al. (1990) show and underline that, in a trade setting, transaction costs alone are not sufficient to explain undertrading – endowment effects also play a role.

[20]As a possible way of accounting for these aspects in an ETS, see Dugast et al. (2019) for a model where banks optimally choose to participate in over-the-counter or centralized markets, or both.

[21]We note that compliance-only trading leads to lower effectiveness and market liquidity relative to more active trading strategies that are more likely to enable learning and thus to reduce information and search costs, an aspect our framework does not capture.

leading to hubs and concentration in trade (e.g. Borghesi & Flori, 2018; Karpf et al., 2018).[22] Cludius & Betz (2020) also document the increasing engagement of banks in the EU ETS and their liquidity-enhancing and trade-facilitating roles (e.g. account manager, hedging partner). This highlights the need to understand how market microstructure shapes transaction costs, which may decrease over time as markets become more mature (e.g. Joskow et al., 1998).

Finally, there exists evidence from bid-ask spreads and anomalies in cost-of-carry relationships that price informativeness might be hampered in the EU ETS, see e.g. Friedrich et al. (2019) for a review. However, in our deterministic framework, we cannot analyze how informational efficiency is affected by the costs of trading and refer the reader to Dávila & Parlatore (2020) for a formal analysis in financial markets. In fact, we underline that even without transaction costs, informational efficiency is likely to break down as dispersed information cannot be fully aggregated by the market in equilibrium (e.g. Grossman & Stiglitz, 1980). For instance, in a frictionless ETS, Cantillon & Slechten (2018) demonstrate that the permit price is already not a sufficient statistic when individual abatement costs are private information.[23]

2 Background

In this section, we first review the related empirical literature on transaction costs in permit markets. We then analyze transaction and compliance data in Phase II of the EU ETS and uncover stylized patterns in firms' market behavior that point us towards the prevalence of costs associated with inter-firm trading.

2.1 Transaction costs and empirical literature

In practice, firms regulated under an ETS face all sorts of non abatement-related costs.[24] On the one hand, implementation, regulatory and other administrative costs associated with the monitoring, reporting and verification process account for a large share of collateral regulatory costs (e.g. Jaraitė et al., 2010; Heindl, 2017). Since they are one-shot, sunk and faced by all

[22] Relatedly, in EU ETS Phase I, Balietti (2016) describes how trading activity by different types of traders is influenced by permit price volatility and vice versa.

[23] For instance, Montagnoli & de Vries (2010) and Crossland et al. (2013) find that the EU ETS's informational efficiency was limited in Phases I and II as a result thin trading and the existence of both momentum and overreaction in prices, leading to profitable trading strategies even in the presence of transaction costs.

[24] A more comprehensive taxonomy of transaction costs and an analysis of their determinants in the context of environmental policy can be found in Coggan et al. (2010). See also Krutilla & Krause (2011) and McCann (2013) for a discussion on how to shape policy to lower transaction costs and alleviate their impacts.

firms, they do not affect compliance costs and choices at the margin and have no bearing on market outcomes. On the other hand, firms incur transaction costs that are associated with their trading activity, the focus of this paper. These include explicit monetary costs such as brokerage and exchange membership fees as well as implicit costs such as search, information, bargaining and internal decision-making costs (e.g. Hahn & Stavins, 2011).[25]

The vast majority of existing empirical analyses of transaction costs is based on the pioneering US cap-and-trade programs or the EU ETS. For instance, they are found to have decreased trading activity in the Wisconsin's Fox River program (Hahn & Hester, 1989) and lowered cost effectiveness by 10-20% in the US lead phasedown program (Kerr & Maré, 1998). Similarly, in the Los Angeles basin (RECLAIM), Foster & Hahn (1995) analyze trading activity and find that large transaction costs altered market behavior. Gangadharan (2000) econometrically tests the existence and magnitude of transaction costs, finding that they were most influential in the early years of the program, with a decrease in the probability of trading of 32%. This notwithstanding, Fowlie & Perloff (2013) cannot reject the Coasean hypothesis that market outcomes were independent of the initial endowments. These observations are corroborated by similar evidence in the ARP where transaction costs were sizable (e.g. Toyama, 2019) but diminished over time as the market developed and firms learned and gained experience (e.g. Joskow et al., 1998; Carlson et al., 2000; Schmalensee & Stavins, 2013; Chen, 2018).

In the EU ETS, we separate how the literature has approached the issue of transaction costs into three strands. The first one describes observed trading and compliance behavior through surveys of managers' practices or network-based analyses of transactions. Targeting subparts of the ETS, viz. Belgian (Venmans, 2016), German (Heindl, 2012a, 2017), Irish (Jaraitė et al., 2010), Swedish (Sandoff & Schaad, 2009) and manufacturing (Martin et al., 2015) firms, these surveys reveal that permit trading is sparse, used mostly for compliance rather than revenue purposes, and often a subsidiary objective in firms' business operations. Analyses of patterns in realized transactions concur to underline the influential role of non-compliance participants in shaping the trading network. For instance, Karpf et al. (2018) bring to light a hierarchical and assortative network structure in which most firms have to resort to local connections or costly intermediaries, the implications of which are then discussed for price discovery, market inefficiency and informational asymmetries.[26] Similarly, Borghesi & Flori (2018) show that some national registries are more central than others in the network, which is corroborated

[25] For instance, EEX, a major organized exchange platform for the EU ETS, charges 2,500 € for an annual trading license plus ∼3 € per bundle of 1,000 permits (permits are only traded in such bundles on exchanges).

[26] In this context, Hintermann et al. (2016) note that the strong presence of intermediaries suggests that transaction costs may be important.

by a home-country bias in permit trades in Hintermann & Ludwig (2018).[27]

Further showing that permit trades are not exclusively driven by complementarity in marginal abatement costs, the second strand consists of financial analyses of permit pricing properties. For instance, Palao & Pardo (2012) find evidence of price clustering in permit futures which they attribute to trading costs. Frino et al. (2010) and Medina et al. (2014) similarly interpret the existence of positive bid-ask spreads. Relatedly, Charles et al. (2013), Schultz & Swieringa (2014) and Friedrich et al. (2019) stress that such frictions could partly explain the observed deviations from cost-of-carry arbitrage between spot and futures prices as predicted by theory. This also suggests that some trading cost pass-through in permit prices may exist.

The third strand gathers three econometric analyses specifically focused on transaction costs. Using transactions data for Phase I (2005-2007) and a set of constructed firm-level proxies for (search and information) transaction costs, Jaraitė-Kažukauskė & Kažukauskas (2015) show that transaction costs have significant impacts on firms' decisions to participate in the market and trade directly vs. indirectly via third parties.[28] They make two important observations. First, there are economies of scale as transaction costs constitute more of an impediment for smaller firms. Second, their proxies also negatively affect firms' extent of trading, suggesting that transaction costs have both a fixed and a variable component. Similarly, Schleich et al. (2020) carry out multivariate analyses of firms' trading behavior (e.g. volume and frequency of transactions, use of intermediaries and derivatives) based on their characteristics (e.g. sector, number of employees and installations, net position) over 2005-2015 (see e.g. their Table 13). For instance, Schleich et al. find that firms having a higher net position prior to trading ($|\beta_i|$ in this paper), facing a higher competitive pressure or belonging to the energy sector make a more active and efficient use of the EU ETS.

Finally, to our knowledge, Naegele (2018) is the only analysis to estimate the magnitude of trading costs, specifically fixed entry costs for both the permit and offset certificate markets. Using transactions data for Phase II (2008-2012), she measures these costs at the firm level as the foregone profits (or opportunity costs) from choosing not to trade.[29] She employs binary quantile regressions, showing that cost distributions are skewed: the median and mean entry

[27] With a cluster-based analysis of transactions in Phase I, Betz & Schmidt (2016) also show that the bulk of market participants are rather passive traders, with some of these hardly trading at all, and that the most active accounts are to a large extent non-compliance ones (see also Cludius & Betz, 2020).

[28] Transaction costs have a greater impact on firms with a smaller number of installations, less experience with trading (e.g. no in-house trading desk), or no specialized units dealing with emissions abatement. These firms are more likely to trade less frequently or lower volumes, or to trade indirectly, or not to trade at all.

[29] As in our model calibration exercise in Section 4.1, Naegele's estimates of fixed trading costs capture all the frictions, i.e. monetary costs as well as the other factors discussed in Section 1.1.

costs on the permit market are 7 and 21 k€ respectively (many firms face rather small costs but a few have very high costs). Interestingly, she also finds that firms holding excess permits are relatively more reluctant to trade, highlighting a key asymmetry between short and long firms: the former are under no compulsion to sell while the latter need to be proactive in one way or another (e.g. purchasing permits) in order to achieve compliance.[30]

2.2 Anecdotal evidence in EU ETS Phase II (2008-2012)

EU ETS. Every year the EU issues emissions allowances (EUA) through free allocations and auctions, whose total number makes up the cap on emissions. On 30 April of year t, regulated entities are required to remit the equivalent number of EUAs to cover their verified emissions in calendar year $t-1$, one EUA accounting for one metric ton of carbon dioxide equivalent. Options to demonstrate compliance include abatement of emissions (e.g. production curtailment, input substitution, technological upgrade, end-of-pipe measure), purchasing EUAs on the market, and tapping into one's bank of EUAs or next-year's free allocation.[31]

Firms can purchase EUAs on primary markets (i.e. auctions) or trade on secondary markets (i.e. organized exchanges such as ICE and EEX, or over the counter). They may have recourse to registered brokers (i.e. intermediaries) to trade on their behalf. Over Phase II (2008-2012), our period of interest, the yearly averaged total trading volume amounted to 5.6 billion EUAs, about three times the size of the annual emissions caps. Out of these, about 40% were traded over the counter and 60% on exchanges (European Commission, 2015).[32]

Data. Trading activity is recorded in an electronic registry, the EU Transaction Log (EUTL), whose aim is to track the EUA ownership structure across accounts and over time to guarantee an accurate accounting of all issued EUAs. That is, the EUTL records the activity of account holders by keeping track of any EUA transfer, allocation and reconciliation. Unfortunately, it only gathers physical movements on secondary markets so we lack direct information about derivative (i.e. forwards, futures or options) and primary (i.e. auctions) trading.[33] In Phase

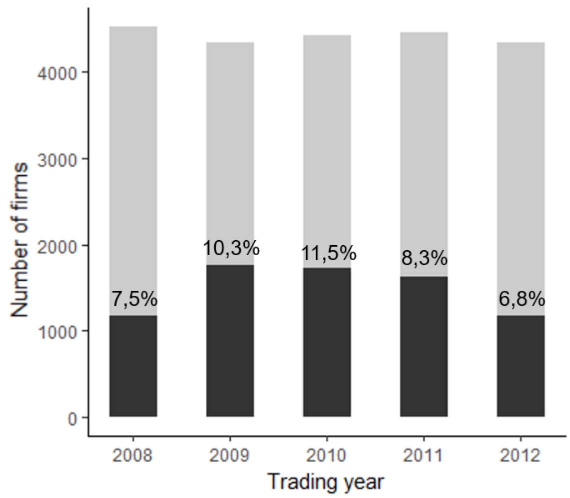
[30]Analyzing firms' trading performances, Liu et al. (2017) obtain a similar result: short firms are relatively more inclined to trade efficiently than long firms, among which that inclination is more heterogeneous.

[31]Unrestricted banking of issued EUAs is allowed since 2008 while borrowing is de facto permitted by the overlap between the compliance and allocation cycles, but limited. Specifically, free $(t+1)$-vintage allocation starts before year-t compliance is due, entailing that some $(t+1)$-vintage EUAs can be frontloaded to achieve compliance in year t (except across trading phases).

[32]The EUTL data does not specify whether transfers took place on an exchange or over the counter.

[33]Derivative trading represents the biggest share of all transactions but associated data requirements are prohibitive (private data for all operating exchanges would be needed and then linked to EUTL accounts). Because we aggregate EUTL data at the year level we capture the physical settlements of end-of-year derivative

Figure 1: Firms' annual market participation in Phase II



Note: Grey = participation, black = autarky. Slight year-on-year changes in the number of observations are due to plant closures and new entrants. Percentages indicate the proportions of autarkic firms in volume as a share of the total number of distributed permits each year.

II Member States only marginally exercised their right to auction up to 10% of all allowances with a hefty 96% realized share of free allocations (European Commission, 2015).[34]

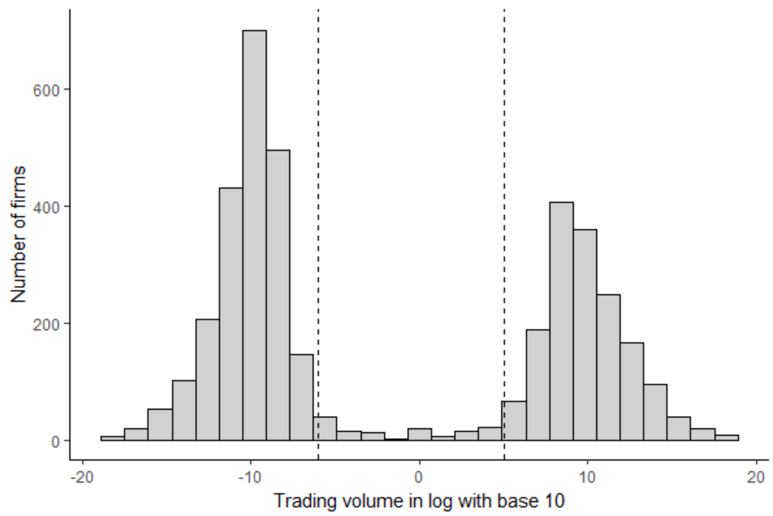
Compliance and transactions are recorded at the polluting site level (one account per installation). As the relevant unit of analysis is the firm (where trading, abatement and compliance decisions can be centralized and coordinated between subsidiary installations) we consolidate the EUTL database from the account to the firm level. Our consolidation methodology and results are described in Appendix B. Our consolidated dataset contains 5,145 firms, binned in six sectors, and transactions between them. The consolidation eliminates intra-firm transfers, which can used by firms as a primary tool to achieve compliance before having to trade on the market. Namely, a firm can pool the EUAs allocated to its installations in a central account and redistribute them back in accordance with installations' realized emissions. Despite that EUAs changed accounts, they have not explicitly been traded. Such intra-firm redistribution represents 27% of the total volume of transfers in Phase II.

Observed firms' behavior. We utilize the consolidated dataset to scrutinize firms' market behavior over Phase II. Figure 1 reports individual annual market participation, showing that around a third of firms did not register any trading activity, if not for annual allocation and

contracts but not those trades that clear in later calendar years.

[34]Accounting for auctioned EUAs would also be data intensive as it would require obtaining auctions data from all Member States and then link it to EUTL accounts.

Figure 2: Distribution of participating firms' trading $\log_{10}$ volumes in 2009



Note: The two vertical dotted lines demarcate the 5$^{\text{th}}$ percentiles, departing from zero, for the distributions of firm-level $\log_{10}$ volumes of cumulative annual purchases (on the right) and sales (on the left).

reconciliation.[35] Note that autarkic firms are relatively small in size (representing only 9% of overall yearly emissions caps on average) and average number of installations (see Table B.1) compared to active firms, whatever their sector. This was to be expected due to economies of scale and could point to the existence of *fixed* entry costs which can preclude some firms from participating to the market.

We find that about 80% of autarkic firms received more permits than their verified emissions. They held on to their surplus, de facto banking the entirety of their excess permits.[36] Their private bank at the end of Phase II amounted to 140% of their 2012 endowment on average. The remaining 20% of autarkic firms emitted in excess of their annual allocations and engaged in borrowing. On average, these firms frontloaded 30% of their future allocation on a year-on-year rolling basis.[37] Martin et al. (2015) find similar evidence of autarkic banking in Phase II and unveil a threshold effect: some firms start selling excess permits only when their surplus is large enough. As they argue, this behavior could be rationalized by a fixed cost of trading, controlling for other hedging and precautionary saving motives.

[35] Martin et al. (2015) find a similar share of autarkic firms in Phase II based on interviews for a subset of compliance entities. Martino & Trotignon (2013) also find evidence of autarkic behavior for 25% of regulated installations based on a similar analysis of transaction and compliance data in Phase I.

[36] Because we do not observe firms' abatement and cannot rule out the implementation of some abatement measures a priori, we cannot distinguish between firms' allocated surplus (i.e. possibly passive banking) and resulting surplus (possibly associated with a proactive abatement and banking strategy).

[37] Ellerman & Trotignon (2009) and Martino & Trotignon (2013) also evidence borrowing in Phase I.

Figure 2 depicts the distributions of the volumes of EUA purchases and sales at the firm level in log base 10 for active firms in 2009 (notice that a firm can both buy and sell at different points during the year). We find that active firms rarely engage in trades below some volume cut-off and that they trade infrequently (we record only 4 to 16 transactions per firm per year on average across sectors, see Table B.1).[38] This suggests that a wedge between sellers' and buyers' marginal abatement costs may persist in equilibrium and could point to the existence of variable costs that are *proportional* to the extent (and frequency) of trading.

To summarize, the fact that gains from trade go unrealized at the extensive margin (autarkic firms, Figure 1) as well as at the intensive margin (stifled trading, Figure 2) is suggestive of the existence of fixed and proportional trading costs.[39] As we will see in the next section in more detail, fixed costs only impact firms' participation in trading (extensive margin) while proportional costs also affect firms' extent of trading (intensive margin).

3 Model

We consider a unitary-mass continuum $\mathcal{I}$ of cost-minimizing firms indexed by $i \in \mathcal{I}$ regulated under a market for emission permits. The model is static and assumes away firms' production decisions, i.e. we rule out any incidence or indirect effect of the permit market on the goods markets that the firms serve. In the absence of the permit market, firm i releases u_i units of emission, its unregulated emission level, which can be abated at a cost of $C_i(u_i - e_i)$, where e_i are firm i's final emissions after end-of-pipe abatement. As is standard, we assume that $C_i', C_i'' > 0$ and let $C_i : a_i \mapsto \alpha_i a_i^2/2$ with $a_i = u_i - e_i \geq 0$, where we omit the linear term for analytical convenience and without loss of generality up to an innocuous translation of the results. The characteristics α_i and u_i are thus heterogeneous across firms.

Initial permit allocation is also firm-specific and denoted q_i for firm i. We assume that the overall cap on emissions $\mathcal{Q}$ is binding relative to overall 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 overallocation at the firm level, i.e. there exist some firms such that $u_i < q_i$, and we let $\beta_i = u_i - q_i \gtrless 0$ denote firm i's initial permit deficit. In this setting, the two firms'

[38]These observations are in line with the surveys' results of Sandoff & Schaad (2009) and Heindl (2012b). For instance, Heindl finds that about half of regulated German firms traded in 2009 and 2010, and two thirds of those that traded did so only once per year, usually as the compliance deadline drew near.

[39]Other behavioral, perceptual and temporal aspects may also hinder trade, as discussed in Section 1.1.

characteristics of interest are thus the α_i's and β_i's which we assume to be distributed over the bounded supports $[\underline{\alpha};\bar{\alpha}]$ and $[\underline{\beta};\bar{\beta}]$ where $0 < \underline{\alpha} < \bar{\alpha} < \infty$ and $\underline{\beta} < 0 < \bar{\beta} < \infty$. When firms cannot trade permits with one another, i.e. under autarky, firm i abates $a_i^0 = \max\{0;\beta_i\}$ with $p_i^0 = \alpha_i a_i^0$ the associated autarkic compliance shadow price. In other words, autarkic compliance implies that short firms abate just as much as to cover their permit deficits $\beta_i > 0$ while long firms do not use their surplus permits $-\beta_i > 0$. We next set forth the frictionless benchmark case before introducing fixed and proportional trading costs.

3.1 Benchmark: Frictionless equilibrium

Under frictionless conditions (i.e. unrestricted inter-firm permit trading, no trading costs), all firms equate their marginal abatement costs to the prevailing market price p, i.e. $\alpha_i(u_i - e_i) = p$ for any firm 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. Proposition 1 computes firms' net market positions and efficiency gains from permit trading relative to autarky for any given feasible permit price on the market.

Proposition 1. *Given a feasible market price for permits $p \in (0;\bar{\alpha}\bar{\beta})$, the sets of buying and selling firms are $\mathcal{D}(p) = \{i \,|\, \alpha_i\beta_i > p\}$ and $\mathcal{S}(p) = \{i \,|\, \alpha_i\beta_i < p\}$, 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 $p_i^0 = \alpha_i \max\{0;\beta_i\}$ is firm i's shadow price of autarkic compliance.

Proof. See Appendix A.1. □

Individual efficiency gains from permit trading in (1) consist of two non-negative components. The first is common to all firms and proportional to the squared distance in autarky-market prices. Specifically, selling (resp. buying) firms with $p_i^0 < p$ (resp. $p_i^0 > p$) find it profitable to abate more (resp. less) than under 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 only accrues to those firms that are initially overallocated as they sell the entirety of their initial surplus of permits at the market price at no cost.[40]

[40] This component complements the characterization of the effort-sharing gains in Doda et al. (2019).

Imposing market closure, i.e. $\int_{\mathcal{I}}(u_i - e_i)\mathrm{d}i = \mathcal{U} - \mathcal{Q}$, on top of firm-level optimality conditions defines the frictionless 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 so-called Coasean independence property, i.e. frictionless equilibrium outcomes does not hinge on the initial permit allocation and individual abatement efforts are efficiently reallocated by the market. Note, however, that $p^\star$ depends on the distribution of the α_i's, $\{\alpha_i\}_i$. Specifically, $p^\star$ is proportional to the stringency of the overall constraint on emissions set by the cap, i.e. $\mathcal{U} - \mathcal{Q}$, and the harmonic mean of $\{\alpha_i\}_i$. Therefore, the more skewed $\{\alpha_i\}_i$ towards lower values, the lower $p^\star$ and vice versa.

The individual efficiency gains defined in (1) with $p = p^\star$ stem from the cost-effective distribution of the total abatement effort $\mathcal{U} - \mathcal{Q}$ among all firms. Specifically, firm i abates in inverse proportion to α_i, i.e. $a_i^\star = u_i - e_i^\star = p^\star/\alpha_i > 0$, and all firms abate in equilibrium, even initially overallocated ones. We say that the frictionless equilibrium is cost-effective in the sense that (1) all firms are weakly better off participating to the market and (2) marginal abatement costs are equalized between them. As described below, this is no longer the case in the presence of pecuniary costs associated with permit trading.

3.2 Equilibrium with trading costs: Characterization

We consider that both permit buyers and sellers incur a market participation cost F and a proportional trading cost T. Following Singh & Weninger (2017), trading costs are assumed to be common to all firms and exogenously given.[41] That is, firms have to pay a fixed fee F to enter the market and trade permits and T is a mark-up on the permit price p, i.e. buying firms pay $p + T$ per permit purchased, selling firms receive $p - T$ per permit sold.[42]

In the presence of positive trading costs, i.e. $F > 0$ and/or $T > 0$, some firms can be better off under autarky: buying (resp. selling) firms in the frictionless equilibrium can remain buyers

[41] This assumption enables both analytical tractability and model calibration. In practice, trading costs can be firm specific and have non-zero curvature, e.g. convexity is generally considered in finance (e.g. Gârleanu & Pedersen, 2013; Dávila & Parlatore, 2020). To simplify, we assume that variable costs are linear in traded volume and we do not attempt to model how trading costs may arise endogenously (see e.g. Liski, 2001).

[42] In practice, the equilibrium price may also depend on how trading costs are distributed among buyers and sellers. Exogenously fixing how trading costs are shared between firms simplifies analytical computations in our multilateral trading framework, see Quemin & de Perthuis (2019) for a formal treatment of endogenously determined transaction prices in an analogous setting with regulatory restrictions on bilateral permit trading.

(resp. sellers) or prefer not to enter the market altogether (autarkic compliance). We consider that firms make and adjust decisions pertaining to their participation in and extent of trading to minimize individual compliance costs. That is, the only barrier to cost-effectiveness occurs in the form of trading costs (see Section 1.1 for a discussion of other barriers). Additionally, we assume that all firms fully acquit their compliance obligations (Stranlund, 2017).

Specifically, when $F > 0$ and $T = 0$, the market outcome is not cost-effective at the extensive margin (some firms do not participate in the market so that some trades that would otherwise be mutually beneficial go unrealized) but it remains cost-effective 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-effectiveness at the intensive margin further drops as participating firms abate in proportion to the actual permit price that they face, i.e. inclusive of the proportional trading cost, which drives a wedge between buyers' and sellers' marginal abatement costs in equilibrium. Specifically, the marginal abatement cost experienced by a buyer exceeds that experienced by a seller by $2T$.

Hence, given a market permit price p and a proportional 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 below defines market participation price thresholds for prospective buying and selling firms.

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. See Appendix A.2. □

Intuitively, firm i will purchase permits only if the market price is below its autarkic shadow price $p_i^0 = \alpha_i\beta_i$ adjusted for the fixed and proportional trading costs (note $\bar{p}_i$ 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 proportional trading cost T and a market permit price $p > T$, the sets of buying and selling firms are defined by*

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

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

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

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. $\mathcal{S}$).

Proof. See Appendix A.3. □

Proposition 2 extends Proposition 1 in the presence of fixed and proportional trading costs. Specifically, for p large (resp. low) enough one has $\mathcal{D} = \varnothing$ (resp. $\mathcal{S} = \varnothing$); for $F = T = 0$ one has $\mathcal{A} = \varnothing$, $\mathcal{D} = \{i|\alpha_i\beta_i > p\}$ and $\mathcal{S} = \{i|\alpha_i\beta_i < p\}$; 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 and a feasible price p, Figure 3 maps the zones where buying (red), selling (green) and autarkic (grey) firms are located in the (α, β)-space. The notions of admissible costs and feasible price are formalized in Lemma 2 below. Moreover, the blue (resp. yellow) arrows indicate how the participation frontiers move in response to an increase in F and T (resp. p).[43]

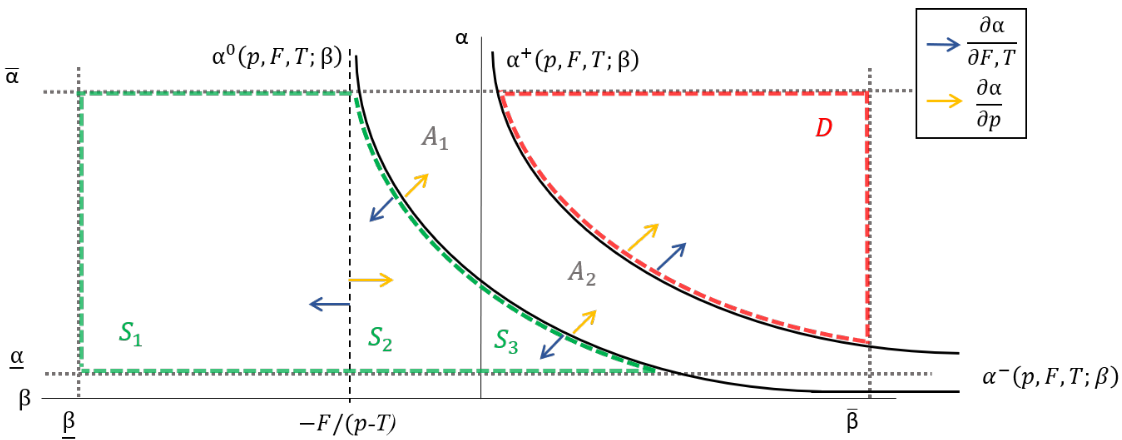
It is worthwhile describing each zone demarcated in Figure 3. We proceed from left to right and bottom to top:[44]

- $\mathcal{S}_1$ When $\beta_i \leq -F/(p-T)$, firm i more than recovers the fixed cost by just selling its initial surplus ($-\beta_i(p-T) \geq F$). Because this comes at no other cost for firm i, this

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

[44] In an intertemporal setting with banking and borrowing our categories would need be amended to reflect arbitrage opportunities, which would have a bearing on price formation and compliance costs. For instance, firms in $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{A}_1$ and $\mathcal{S}_3$ could bank some surplus permits for future compliance or sales (if they expect the discounted price to be higher or to sell them in larger batches to reduce trading costs) or firms in $\mathcal{A}_2$ and $\mathcal{D}$ could borrow future permits to cover a share of today's shortage and delay the full cost of compliance.

Figure 3: Market participation frontiers given price p and trading costs F and T



holds whatever its marginal abatement cost α_i. Moreover, firm i finds it profitable to also abate $(p-T)/\alpha_i > 0$ and sell the corresponding amount of freed-up permits.[45]

$\mathcal{S}_2$ When $-F/(p-T) < \beta_i \leq 0$, selling the initial surplus is not enough to cover the fixed cost. Because firm i can abate at a sufficiently low cost at the margin (i.e. $\alpha_i < \alpha^0$), it make profits by selling both surplus and freed-up permits $(p-T)/\alpha_i - \beta_i$.

$\mathcal{A}_1$ When $-F/(p-T) < \beta_i \leq 0$, selling the initial surplus is not enough to cover the fixed cost. Because firm i cannot abate at a sufficiently low cost at the margin (i.e. $\alpha_i > \alpha^0$), it is better off under autarky, i.e. not using its surplus permits and not abating.

$\mathcal{S}_3$ When $\beta_i > 0$ but small and abatement is sufficiently cheap at the margin (i.e. $\alpha_i < \alpha^-$) firm i abates to both meet compliance and sell some freed-up permits $(p-T)/\alpha_i - \beta_i$.

$\mathcal{A}_2$ When $\beta_i > 0$ is relatively larger and/or abatement is relatively less cheap at the margin (i.e. $\alpha^- \leq \alpha_i \leq \alpha^+$) firm i is better off abating its deficit only so as to comply without entering the market and incurring the associated trading costs.

$\mathcal{D}$ When $\beta_i > 0$ becomes larger and/or abatement becomes more expensive at the margin (i.e. $\alpha_i > \alpha^+$) firm i is better off incurring the trading costs so as to purchase permits to cover some portion of its deficit, the remainder being abated internally.

Importantly, note that with $F = 0$ and $T > 0$, the demand and supply sets in Proposition 2 rewrite $\mathcal{D} = \{i|\alpha_i > (p+T)/\beta_i \wedge \beta_i > 0\}$ and $\mathcal{S} = \{\{i|\alpha_i < (p-T)/\beta_i \wedge \beta_i > 0\} \cup \{i|\beta_i \leq 0\}\}$.

[45] With p given, one might think that firm i abates less than when $T = 0$ by T/α_i. However, because p in equilibrium hinges on the trading costs, one cannot conclude prima facie. This applies to all zones.

That is, overallocated firms would always find it profitable to sell their extra permits at a price $p > T$ so in our model a fixed entry cost is necessary to have that some overallocated firms prefer autarky over market participation. Additionally, 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 rise, 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.

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

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

When one of 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. See Appendix A.4. □

Trading costs are said admissible when positive supply and demand can emerge on the market – roughly speaking, when they are not too large. 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 set of feasible prices in (5) is not empty provided that trading costs are admissible, i.e. if they satisfy (4).

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 $a_i^*(x) = x/\alpha_i$ is participating firm i's optimal abatement decision given the net permit price $x = p \pm T$. If we denote by h the density function of the β_i's and by $g(\cdot|\beta)$ that of the

α_i's conditional on the β_i's, supply and demand rewrite

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

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

Lemma 3 characterizes how S and D vary with F, T and p alternatively.

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. See Appendix A.5. □

Equipped with Lemmas 2 and 3 and letting $V = S - D$ denote the net permit supply function, 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. See Appendix A.6. □

Proposition 3 ensures the existence and uniqueness of a market equilibrium in the presence of admissible fixed and proportional trading costs. Except with zero costs 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 in general. Below we thus seek to derive some properties of the equilibrium in the presence of trading costs using comparative statics.

3.3 Equilibrium with trading costs: Some properties

In this section, we leverage the equilibrium framework developed in Section 3.2 to extend the comparative static results in Stavins (1995) and Montero (1998). Specifically, we analyze the sensitivity of market equilibrium outcomes to incremental changes in the trading costs and firms' permit allocations. We complement our formal analysis with analytical and numerical examples to go beyond the impacts of incremental changes and gain further insight.

3.3.1 Impacts of a change in trading costs

Consider an arbitrarily small increase $\mathrm{d}K > 0$ in the trading cost $K = F$ or T. By virtue of the implicit function theorem, the resulting price response $\mathrm{d}\hat{p}$ in the vicinity of the equilibrium reads

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

which cannot unambiguously be signed in general. Indeed, Lemma 3 shows that $\partial V/\partial p > 0$ but the sign of $\partial V/\partial K$ is indefinite and hinges on the relative magnitudes of the demand and supply responses to the trading cost increase.[46] For instance, when demand is more responsive than supply, i.e. $|\partial D/\partial K| > |\partial S/\partial K|$, then $\mathrm{d}\hat{p}/\mathrm{d}K < 0$. In words, the equilibrium price is lowered as demand is relatively more constricted than supply, and vice versa. As a corollary, it follows that

$$\hat{p} \gtrless p^{\star}, \tag{9}$$

depending on the distributions of the firms' characteristics α_i and β_i and the levels of the trading costs. This showcases the break-down of the Coasean independence property in the presence of trading costs – in particular, equilibrium outcomes hinge on the initial allocation of permits across firms. Further taking the total differentials of supply and demand yields

$$\frac{\mathrm{d}S}{\mathrm{d}K} = \underbrace{\frac{\partial S}{\partial K}}_{<0} + \underbrace{\frac{\partial S}{\partial p}}_{>0} \underbrace{\frac{\mathrm{d}\hat{p}}{\mathrm{d}K}}_{\gtrless 0} \quad \text{and} \quad \frac{\mathrm{d}D}{\mathrm{d}K} = \underbrace{\frac{\partial D}{\partial K}}_{<0} + \underbrace{\frac{\partial D}{\partial p}}_{<0} \underbrace{\frac{\mathrm{d}\hat{p}}{\mathrm{d}K}}_{\gtrless 0}, \tag{10}$$

which, on the face of it, cannot unambiguously be signed either. Yet, $\mathrm{d}S/\mathrm{d}K = \mathrm{d}D/\mathrm{d}K$ must hold in equilibrium,[47] which in conjunction with (10) implies that $\mathrm{d}S/\mathrm{d}K = \mathrm{d}D/\mathrm{d}K < 0$. In words, an increase in the trading costs always lowers the equilibrium volume of trade.

As a result, the total regulatory control costs, i.e. the costs of abatement and trading summed over all firms, always increase as K rises. Because firms choose whether to trade or not, and if so the extent thereof, by minimizing their compliance costs, there exists a decreasing mapping between the total volume of trade (which measures the degree of cost-effective reallocation of abatement among firms) and the total control costs. In words, therefore, an increase in the trading costs negatively impacts welfare by consuming more monetary resources and stifling more trades which would have otherwise been mutually beneficial, at both the extensive and intensive margins. Proposition 4 summarizes the above results.

[46] Note that even taking (8) in the small as F and/or $T \to 0$ also yields an indefinite sign.
[47] One can formally see this by unpacking (8) with $V = S - D$ and term identification with (10).

Proposition 4. *In response to an increase in the fixed or proportional trading cost, the equilibrium volume of trade (resp. total regulatory control costs) always decreases (resp. increases) and vice versa. However, the market equilibrium price may increase or decrease.*

Proof. See above. □

To illuminate our general results, we exclusively focus on the equilibrium price impacts of a shift in the trading costs for two reasons. First, because Stavins (1995) and Montero (1998) shut down this channel in their comparative static analyses of trading costs. Second, because price impacts can go both ways, it is worthwhile investigating their determinants per se.

To go beyond the equilibrium sensitivity to incremental changes in the trading costs characterized in Proposition 4, below we provide both analytical and numerical illustrations. First, we consider two simple analytical examples for the distributions of the firms' characteristics whereby we are able to derive implicit closed-form solutions for the equilibrium price which lend themselves to economic interpretation. Second, we develop three numerical examples in more general cases to gain further insight into the price impacts of trading costs.

Analytical examples. We primarily focus on the price impacts of the fixed trading cost, i.e. we let $F \geq 0$ and $T = 0$. We also let firms be homogeneous in terms of initial net deficit, i.e. $\beta_i = \beta = \mathcal{U} - \mathcal{Q} > 0$ for all i in $\mathcal{I}$, and proceed with two alternative distributions 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 $g_{1,2}(x) = 0$ elsewhere. These density functions are normalized to a unitary mass and cherry-picked to ensure both analytical tractability and clear-cut results which are otherwise hard to come by.[48] They represent two opposite distributions of the α_i's: with g_1 (resp. g_2), $\{\alpha_i\}_i$ is skewed towards high (resp. low) values. The sketches of the derivations leading to the following results are gathered in Appendix A.7.

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

$$\hat{p}_1 + \frac{2F\sqrt{F(F + 2\beta\hat{p}_1)}}{\beta^3(\bar{\alpha} - \underline{\alpha})} = p_1^\star, \tag{11}$$

where $p_1^\star = \beta(\bar{\alpha} + \underline{\alpha})/2$. In this case, $\hat{p}_1 \leq p_1^\star$ with equality in $F = 0$ and $\mathrm{d}\hat{p}_1/\mathrm{d}F < 0$.

[48] For instance, when $\{\alpha_i\}_i$ is uniformly distributed we arrive at an analytically intractable transcendental equation in $\hat{p}$. The ordering between $\hat{p}$ and $p^\star$ thus depends on F in a non-straightforward way.

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

$$\hat{p}_2 - \frac{4F\bar{\alpha}^2\underline{\alpha}^2}{(\bar{\alpha}^2 - \underline{\alpha}^2)\hat{p}_2^3}\sqrt{F(F + 2\beta\hat{p}_2)} = p_2^\star, \tag{12}$$

where $p_2^\star = 2\beta\bar{\alpha}\underline{\alpha}/(\bar{\alpha} + \underline{\alpha})$. In this case, $\hat{p}_2 \geq p_2^\star$ with equality in $F = 0$ and $\mathrm{d}\hat{p}_2/\mathrm{d}F > 0$.

Observe that $p_1^\star > p_2^\star$. Indeed, with a homogeneous deficit β across firms the frictionless price is higher the more skewed the distribution of the α_i's towards high values. Crucially, a fixed trading cost tends to mitigate this. When the α_i's are tilted towards high values, introducing or increasing the fixed cost tends to evict more firms with a high α_i (i.e. demanders) than with a low α_i (i.e. suppliers), ceteris paribus. This entails that demand is more constricted than supply, hence a downward pressure on the price. The converse holds when the α_i's are skewed towards low values as a higher fixed cost shrinks supply more than demand, ceteris paribus. Hence, a fixed trading cost tends to have a tempering effect on the price when the frictionless price is 'high' – and conversely, it tends to hike $\hat{p}$ when $p^\star$ is 'low'.

Numerical examples. To expand the parameter space and enrich the analysis, we 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 fixed and proportional trading costs simultaneously. We set $\underline{\alpha} = 1$, $\bar{\alpha} = 10$, $\underline{\beta} = -5$ and $\bar{\beta} = 10$ and consider three cases: both $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ uniform (i.e. $\alpha_\alpha = \alpha_\beta = \beta_\alpha = \beta_\beta = 1$); $\{\alpha_i\}_i$ skewed towards high values (i.e. $\alpha_\alpha = 3$, $\beta_\alpha = 1$) with $\{\beta_i\}_i$ uniform; and $\{\beta_i\}_i$ skewed towards high values (i.e. $\alpha_\beta = 3$, $\beta_\beta = 1$) with $\{\alpha_i\}_i$ uniform.[49]

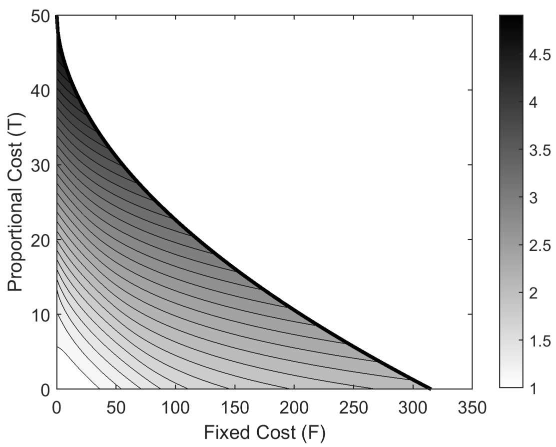
Given the trading costs F and T we can solve numerically for $\hat{p}_0$ such that $V(\hat{p}_0, F, T) = 0$, jointly verifying the cost admissibility conditions. Specifically, we seek $\hat{p}_0 = \min p$ such that $D - S > 0$ and $\hat{p}_0$ is feasible. The left column in Figure 4 depicts our results and shows the ratios $\hat{p}_0/p_0^\star$ in the (F, T)-space in the three cases, with $V(p_0^\star, 0, 0) = 0$. The thick black line delineates the admissible cost range defined by (4), i.e. the market breaks down above it.

With $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ both uniform (Figure 4a) the market price in the presence of trading costs is always larger than absent trading costs ($\hat{p}_0/p_0^\star \geq 1$) and it gets larger the larger F or T (contour lines are downward-sloping) but increases relatively more with T than F (contour lines are convex). For instance, $\hat{p}_0$ can be five times as large as $p_0^\star$ when $F = 0$ and T is large while it is at most twice as large when $T = 0$ and F is large. This is no longer so in the other two cases: $\hat{p}_0$ can be lower than $p_0^\star$ for some trading costs (e.g. when F is large and T small

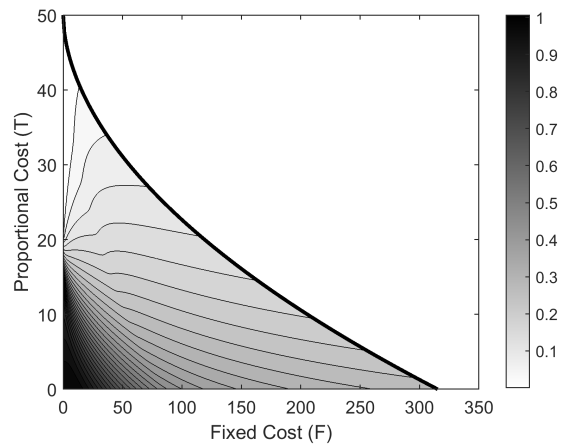
[49]The cases where $\{\alpha_i\}_i$ or $\{\beta_i\}_i$ are skewed towards low values lead to magnified but qualitatively similar results as for the case of $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ both uniform, hence omitted here.

Figure 4: Ratios $\hat{p}_0/p_0^\star$ (left) and $\frac{\hat{p}_t-\hat{p}_0}{\hat{p}_0}\Big/\frac{p_t^\star-p_0^\star}{p_0^\star}$ (right)

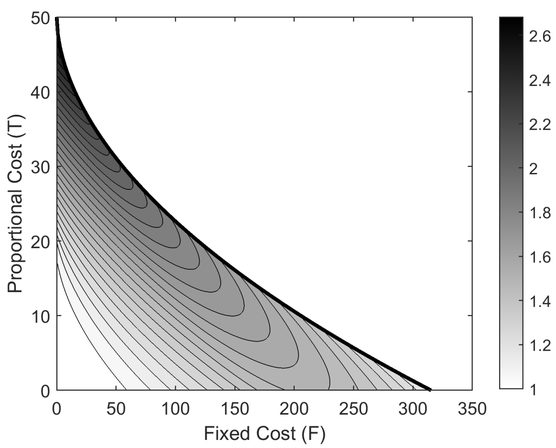
(a) $\alpha_\alpha = \beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p_0^\star = 9.8$)



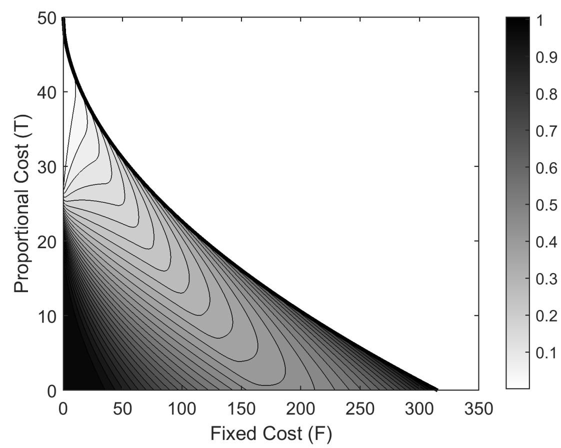
(b) $\alpha_\alpha = \beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p_t^\star = 13.7$)



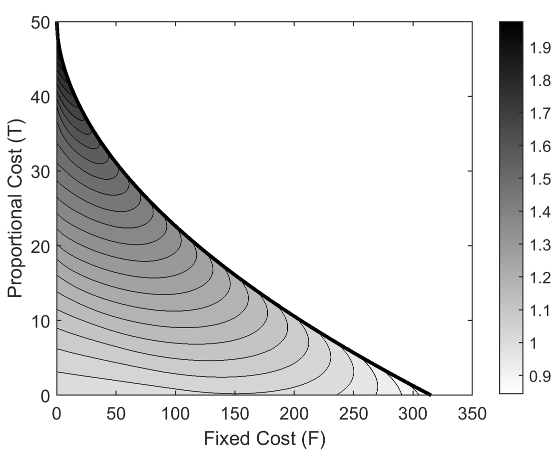
(c) $\alpha_\alpha = 3$, $\beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p_0^\star = 18.0$)



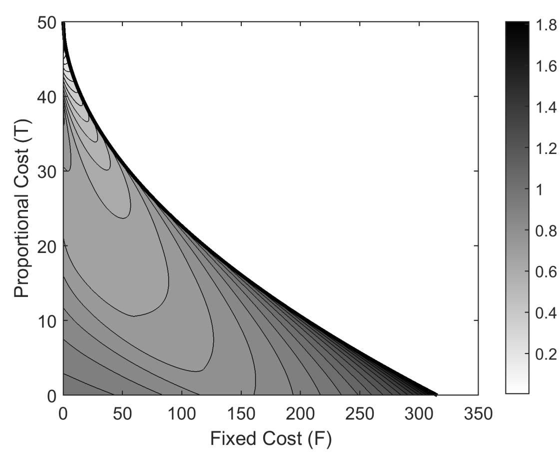
(d) $\alpha_\alpha = 3$, $\beta_\alpha = \alpha_\beta = \beta_\beta = 1$ ($p_t^\star = 25.2$)



(e) $\alpha_\beta = 3$, $\alpha_\alpha = \beta_\alpha = \beta_\beta = 1$ ($p_0^\star = 24.4$)



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



Note: The subscript 0 (resp. t) indicates the pre (resp. post) supply tightening situation.

in Figure 4e) and is non-monotonic in the trading costs (i.e. contour lines have distorted U shapes in Figures 4c and 4e). In particular, $\hat{p}_0$ increases with the trading costs when they are 'low enough' but the converse holds when they are 'large enough'.

Generally speaking, we find that the higher the frictionless price to start with, the relatively lower the market price in the presence of trading costs. Moreover, only when the frictionless price is sufficiently high can trading costs lead to a lower price level for some cost pairs. This relates to the aforementioned tempering price effect of trading costs.

3.3.2 Impacts of a change in total supply and individual allocations

We now study the equilibrium price impacts of a shift in total supply and firm-level allocations in the presence of constant trading costs. We refer the reader to Stavins (1995) and Montero (1998) for illustrations of how the total volume of trade and compliance costs vary with firms' initial allocations. Compared to these comparative static analyses taking the perspective of a given firm (i.e. assuming the market price is invariant), we use our equilibrium framework to analyze the case where all firms' allocations can vary simultaneously, which affects both the market price and firms' trading decisions. This in turn leads to different results. Contrary to Stavins (1995) for instance, the way permits are allocated among firms does affect equilibrium outcomes even with linear marginal trading costs.

We consider an arbitrarily small variation in $\{\beta_i\}_i$, i.e. we let β_i change to $\beta_i + \mathrm{d}\beta_i$ for all i in $\mathcal{I}$, possibly with $\mathrm{d}\beta_i \neq \mathrm{d}\beta_j$, and then let the price adapt to $\hat{p}+\mathrm{d}\hat{p}$ so that $V(\hat{p}+\mathrm{d}\hat{p}, F, T) = 0$ remains satisfied. This is tantamount to taking the total differential of $V(\hat{p}, F, T) = 0$ w.r.t. $\hat{p}$ and all β_i's. With a slight abuse of notation for the partial derivatives w.r.t. the β_i's, one has

$$\frac{\partial V(\hat{p}, F, T)}{\partial p}\mathrm{d}\hat{p} + \int_{\mathcal{I}} \frac{\partial V(\hat{p}, F, T)}{\partial \beta_i}\mathrm{d}\beta_i \mathrm{d}i = 0. \tag{13}$$

For the sake of the argument, we here only consider induced changes at the intensive margin (i.e. within $\mathcal{S}$ and $\mathcal{D}$ as given prior to the change in $\{\beta_i\}_i$) and ignore those at the extensive margin (i.e. changes due to the shifts in the locations of the participation frontiers in Figure 3). Extensive margin impacts, which render the exposition more complex but do not change the nature of the results, are formally treated and discussed in Appendix A.8.

Noting that $\dfrac{\partial V}{\partial \beta_i} = -\displaystyle\int_{\mathcal{S}(\hat{p},F,T)\cup\mathcal{D}(\hat{p},F,T)} \delta(j = i)\mathrm{d}j$ where $\delta(\cdot)$ denotes the Dirac distribution, (13) simplifies to

$$\frac{\partial V(\hat{p}, F, T)}{\partial p}\mathrm{d}\hat{p} = \int_{\mathcal{S}(\hat{p},F,T)\cup\mathcal{D}(\hat{p},F,T)} \mathrm{d}\beta_i \mathrm{d}i, \tag{14}$$

meaning that the sign and magnitude of $\mathrm{d}\hat{p}$ hinge upon the overall net change in the β_i's over $\mathcal{S}(\hat{p}, F, T) \cup \mathcal{D}(\hat{p}, F, T)$. This is because allocation changes for autarkic firms have no market impacts given our exclusive focus on the intensive margin impacts. Interestingly, note that the mere reshuffling of individual allocations, while keeping total supply invariant, may also influence the price outcome as the $\mathrm{d}\beta_i$'s may not cancel out over $\mathcal{S}(\hat{p}, F, T) \cup \mathcal{D}(\hat{p}, F, T)$.

Now consider that total supply is tightened by $\mathrm{d}\mathcal{Q} > 0$ and that the tightening is uniformly distributed among all firms.[50] Because $\mathcal{I}$ is of mass one, we have that $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}$ for all i. If we let $|\cdot|$ denote the mass (or measure) of a set, (14) then rewrites

$$\frac{\mathrm{d}\hat{p}}{\mathrm{d}\mathcal{Q}} = \frac{|\mathcal{S}(\hat{p}, F, T)| + |\mathcal{D}(\hat{p}, F, T)|}{\partial V(\hat{p}, F, T)/\partial p} > 0, \tag{15}$$

that is, the price response to the tightening is always positive in the presence of trading costs. But how does its magnitude compare to that in the frictionless case? Without trading costs, (14) reads

$$\frac{\partial V(p^\star, 0, 0)}{\partial p} \mathrm{d}p^\star = \int_{\mathcal{I}} \mathrm{d}\beta_i \mathrm{d}i = \mathrm{d}\mathcal{Q}, \tag{16}$$

so that

$$\frac{\mathrm{d}\hat{p}/\mathrm{d}\mathcal{Q}}{\mathrm{d}p^\star/\mathrm{d}\mathcal{Q}} = \underbrace{\Big(|\mathcal{S}(\hat{p}, F, T)| + |\mathcal{D}(\hat{p}, F, T)|\Big)}_{\text{distribution effect } \leq 1} \underbrace{\frac{\partial V(p^\star, 0, 0)/\partial p}{\partial V(\hat{p}, F, T)/\partial p}}_{\text{price effect } \geq 1} \gtrless 1. \tag{17}$$

The ordering of $\mathrm{d}\hat{p}$ and $\mathrm{d}p^\star$ is ambiguous in general and hinges on two countervailing forces. First, the overall impact on the net demand for permits (holding the price constant) relative to that without trading costs which ultimately depends on how the tightening is distributed among firms. In the intensive margin only case, for a given incidence of the tightening, this *distribution effect* always works to mitigate the price increase, all the more so that trading costs are large and the mass of autarkic firms is sizable. Second, the ratios of the sensitivities of the net supply functions to a price change with and without trading costs. In the intensive margin only case, this *price effect* always works to magnify the price increase, all the more so that trading costs are large.[51] This yields the overall ambiguous effect in (17).

We now state the following result in the general case of any change in supply inclusive of the induced impacts at both the intensive and extensive margins.

[50] Two alternative firm-level distributions of the tightening are considered in Appendix A.8.

[51] Since $|\mathcal{S}|+|\mathcal{D}| \leq |\mathcal{I}| = 1$, *distribution effect* $\leq$1 and decreases as $|\mathcal{A}|$ increases. $\partial V(p^\star, 0, 0)/\partial p = \int_{\mathcal{I}} \mathrm{d}i/\alpha_i$ while $\partial V(\hat{p}, F, T)/\partial p = \int_{\mathcal{S}\cup\mathcal{D}} \mathrm{d}i/\alpha_i$ so *price effect* $\geq$1. See Appendix A.8 for more details.

Proposition 5. *The market price response to a supply change can be amplified or dampened in the presence of trading costs relative to frictionless conditions. This hinges on a distribution effect (the relative impact on net permit demand holding the price constant) and a price effect (the relative price elasticity of net permit demand) which are generally countervailing.*

Proof. See Appendix A.8. □

In the general case the distribution and price effects can be greater or lower than one. Specifically, the distribution (resp. price) effect is likely to be predominantly lower (resp. greater) than one, except possibly when trading costs are small. In particular, when $F = 0$, the extensive margin effects are nil so that this case collapses to the above analysis. Appendix A.8 also discusses how the magnitude of the distribution effect depends on the type of firm-level incidence of the supply change. We now turn to analytical and numerical examples.

Analytical examples (cont'd). We consider that pursuant to some regulatory amendment overall supply $\mathcal{Q}$ is reduced by an arbitrarily small amount $\mathrm{d}\mathcal{Q} > 0$, which translates into a small increase $\mathrm{d}\beta = \mathrm{d}\mathcal{Q}$ in the firms' uniform deficit. We have the following results.

Case 1. The market price response to a small uniform supply tightening of $\mathrm{d}\beta$ is positive, i.e. $\mathrm{d}\hat{p}_1/\mathrm{d}\beta > 0$. However, only when $\hat{p}_1$ is not too small does it hold that $\mathrm{d}\hat{p}_1/\mathrm{d}\beta \geq \mathrm{d}p_1^\star/\mathrm{d}\beta > 0$, specifically when $(p_1^\star - 3F/\beta)/5 \leq \hat{p}_1 \leq p_1^\star$.

Case 2. The market price response to a small uniform supply tightening of $\mathrm{d}\beta$ is positive, i.e. $\mathrm{d}\hat{p}_2/\mathrm{d}\beta > 0$. However, only when $\hat{p}_2$ is not too large does it hold that $\mathrm{d}\hat{p}_2/\mathrm{d}\beta \geq \mathrm{d}p_2^\star/\mathrm{d}\beta > 0$, specifically when $p_2^\star \leq \hat{p}_2 \leq p_2^\star + \sqrt{p_2^\star(p_2^\star + 3F/(2\beta))}$.

Intuitively, tightening supply always implies higher price levels in the presence of a fixed cost, but the price rise can be magnified or dampened relative to frictionless conditions. The two cases suggest that the price response is more likely to be magnified in the presence of a fixed cost when $\hat{p}$ is not too distant from $p^\star$ prior to the tightening, ceteris paribus. We investigate this further with numerical examples.

Numerical examples (cont'd). We illustrate a uniform supply tightening with a shift in the support of the distribution $\{\beta_i\}_i$ from $[-5; 10]$ (indexed by 0) to $[-4; 11]$ (indexed by t). The right column in Figure 4 depicts our results and shows the ratios of the relative induced price increase $\frac{\hat{p}_t - \hat{p}_0}{\hat{p}_0}$ with trading costs to the relative induced price increase under frictionless conditions $\frac{p_t^\star - p_0^\star}{p_0^\star}$ in the (F, T)-space in the three cases considered.

We find that the relative induced price increase in the presence of trading costs is less pronounced the larger $\hat{p}_0$ relative to $p_0^\star$ to start with. Consequently, it tends to be larger under frictionless conditions than present trading costs most of the time in our simulations. Only in one case do we find that the relative price increase is larger with trading costs than without, namely in Figure 4f when F is large and $\hat{p}_0$ is close to or lower than $p_0^\star$.

4 Illustration

In this section, we illustrate our theoretical results based on actual market data. Specifically, we consider the universe of allocations, emissions, transactions and prices in Phase II of the EU ETS (2008-2012) to discipline both the calibration of model parameters and selection of practically relevant values for the fixed and proportional trading costs. We next leverage our calibrated model to compare the relative implications of various supply-tightening policies in terms of market price responses and compliance costs, with and without trading costs.

4.1 Calibration to EU ETS Phase II

We utilize our transaction and compliance databases consolidated at the firm level to calibrate the model parameters for each year in Phase II. We proceed in two steps. First, we infer yearly firms' characteristics $(\alpha_{i,t}, \beta_{i,t})$ conditional on given pairs of trading costs (F, T). Second, we select the trading cost pair that best rationalizes firms' observed participation in trading and, where applicable, the sign of their net market positions.

4.1.1 Inferring firms' characteristics with given trading costs

Yearly initial allocations and verified emissions are readily available at the polluting unit (or account) level from the EUTL which we consolidate at the firm level (see Appendix B for the methodology). We respectively denote them $q_{i,t}^r$ and $e_{i,t}^r$ for firm i in year t. We also compute yearly-averaged EUA spot prices p_t^r using ICE data. Except for the consolidation procedure, this is quite straightforward. We now need to make assumptions to set firms' yearly baseline emissions and marginal abatement cost slopes (which are both unobservable quantities) and control for banking and borrowing in our static setting. In a context where relevant quantities are either scarce or hard to reconstruct ex post, we opt for workable assumptions allowing for a first-pass yet reasonable model calibration.

We set year-t baseline emissions $u_{i,t}$ as the moving averages of i's verified emissions over the three preceding years $t-3$ to $t-1$. This captures the persistence in emissions demand over time and a steadily declining aggregate trend (e.g. Quemin & Trotignon, 2019). This proxy is further adjusted with yearly-fixed effects η_t introduced below. We exclude firms with implied negative abatement, i.e. with $u_{i,t} - e^r_{i,t} < 0$, which is the case for 30% of the firms on average across years. This leads to some changes in the size of the annual samples of firms.

Next, we set firms' marginal abatement cost slopes based on the equimarginal value principle applied to the total permit cost (i.e. permit price adjusted for the proportional trading cost) using observed prices p^r_t and implied abatement levels $u_{i,t} - e^r_{i,t}$. The imputed $\alpha_{i,t}$'s thus need to be conditioned on observed market participation decisions and net positions, that is

$$\alpha_{i,t} = \begin{cases} (p^r_t + T)/(u_{i,t} - e^r_{i,t}) & \text{if} \quad i \in \mathcal{D}^r_t \\ (p^r_t - T)/(u_{i,t} - e^r_{i,t}) & \text{if} \quad i \in \mathcal{S}^r_t \\ p^r_t/(u_{i,t} - e^r_{i,t}) & \text{if} \quad i \in \mathcal{A}^r_t \end{cases} \tag{18}$$

where $\mathcal{D}^r_t$, $\mathcal{S}^r_t$ and $\mathcal{A}^r_t$ are the sets of observed net buying, net selling and autarkic firms in year t, respectively. In (18) we assume that autarkic firms treat the permit price as a relevant signal to guide their abatement decisions though they do not effectively trade.

We then adjust firms' allocations for the temporal dimension that our static model does not capture, i.e. we compute effective allocation levels net of temporal intra-firm redistribution.[52] To that end, we begin by imputing firms' permit bank dynamics as

$$b^r_{i,t} = b^r_{i,t-1} + q^r_{i,t} + x^r_{i,t} - e^r_{i,t}, \tag{19}$$

where $b^r_{i,t}$ is firm i's observed bank carried over from year t to year $t+1$ with $b^r_{i,2007} = 0$, and $x^r_{i,t}$ is firm i's observed net permit purchase in year t. Then the banking-adjusted allocation for firm i in year t, denoted by $q^a_{i,t}$, is set as

$$q^a_{i,t} = \begin{cases} q^r_{i,t} - (b^r_{i,t} - b^r_{i,t-1}) = e^r_{i,t} - x^r_{i,t} & \text{if} \quad i \in \mathcal{D}^r_t \cup \mathcal{S}^r_t \\ q^r_{i,t} - \frac{1}{2}(b^r_{i,t} - b^r_{i,t-1}) = \frac{1}{2}(q^r_{i,t} + e^r_{i,t}) & \text{if} \quad i \in \mathcal{A}^r_t \end{cases} \tag{20}$$

For observed trading firms, effective allocations are simply raw allocations net of yearly bank increments $b^r_{i,t} - b^r_{i,t-1}$. That is, observed buying (resp. selling) firms are ex-ante short (resp. long) by their ex-post net traded volumes. For observed autarkic firms, we add a $\frac{1}{2}$ factor in front of the bank increment. Were it not for this factor, these firms would by construction

[52] This implies that (1) trading costs only affect the extent of annual inter-firm trading in isolation of other years and (2) each year temporal intra-firm trading is carried out before spatial inter-firm trading.

have no need to trade ex ante as $q^a_{i,t}$ would coincide with $e^r_{i,t}$ since $x^r_{i,t} = 0$. But because one of our aims is to recover autarkic compliance via the introduction of trading costs, this arbitrary factor leaves some trading opportunities open for these firms – here selling (or buying) up to half of their yearly permit surpluses (or deficits). Indeed, as discussed in Section 1.1, autarkic firms could in principle increase profits by trading on the market to some extent rather than exclusively exploiting the temporal flexibility margin.[53] From (20) we finally compute annual permit deficits as $\beta_{i,t} = u_{i,t} - q^a_{i,t}$ and Table C.1 contains some descriptive statistics on the inferred firms' characteristics $\{\alpha_{i,t}, \beta_{i,t}\}_{i,t}$ binned by sectors.

So equipped, we can populate the sets $\mathcal{D}_t$, $\mathcal{S}_t$ and $\mathcal{A}_t$ defined in Proposition 2 for any feasible price p and admissible trading costs (F, T). We consider a mesh where F and T respectively range from 0 to 200 k€ and 0 to 1.5 €/tCO$_2$ with steps of 1 k€ and 0.05 €/tCO$_2$, and p can vary freely within the feasibility region as per (5). This defines supply S_t and demand D_t as per (6-7) for any discretized pair (F, T). For any given pair, we can then solve for the year-t equilibrium price $\hat{p}_t$, namely $\hat{p}_t = \min p$ subject to $D_t - S_t > 0$.

4.1.2 Selecting relevant trading costs

As Carlson et al. (2000, p. 1319) observe, the failure of firms to realize cost savings through trading cannot be inferred simply by comparing price levels obtained under different modeling scenarios with those that actually prevailed. Specifically in our case, a multiplicity of trading cost pairs can replicate the observed price levels p^r_t. Additionally, price formation is influenced by a variety of other factors our model does not explicitly account for. As such, the ability to replicate observed prices is not robust enough a criterion to discriminate between cost pairs. Accordingly, we eliminate the initial difference between p^r_t and the $\hat{p}_t$'s for all pairs.[54]

To this end we introduce additive yearly fixed effects η_t adjusting firms' marginal abatement cost schedules to $\alpha_{i,t}(u_{i,t} - e_{i,t}) + \eta_t$, de facto shifting firms' initial permit deficits by $\eta_t/\alpha_{i,t}$. For instance, the η_t's can be thought of as partly controlling for common shocks to or trends in firms' baseline emissions, or for firms' intertemporal trading decisions, thereby improving on our first-pass proxies for baselines and banking-adjusted allocations. Specifically, $\eta_t > 0$ corrects for higher baselines or market-wide incentives to bank or both. We then pick the η_t

[53] The $\frac{1}{2}$ factor is arbitrary and reflects a lack of relevant empirical guidance. It affects our selected values for the trading costs thus: given the market-wide bank build-up in Phase II, a higher factor would imply less surplus permits available, hence lower trading costs to rationalize autarkic compliance, and vice versa.

[54] This implies that we ignore the direct impacts that trading costs may have on price formation when we select a cost pair but we do capture their indirect impacts i.e. as firms adjust their participation in, and their extent of, trading based on the cost levels.

Table 1: Typology of sorting errors

		Observations		
		Autarkic	Buyer	Seller
Model	Autarkic	–	$\mathcal{E}_1$	$\mathcal{E}_2$
	Buyer	$\mathcal{E}_3$	–	$\mathcal{E}_5$
	Seller	$\mathcal{E}_4$	$\mathcal{E}_6$	–

that eliminates the initial price wedge: for every cost pair (F, T) and year t there corresponds a unique η_t ensuring that $\hat{p}_t = p_t^r$ (if initially $\hat{p}_t < p_t^r$ then $\eta_t > 0$ and vice versa).

To make an educated guess about practically relevant trading costs values and discriminate between cost pairs, we propose to discipline the selection of trading costs by jointly minimizing the total number of modeling sorting errors and their dispersion across error types. That is, this selection criterion minimizes discrepancies between firms' participation in trading and their net market positions as predicted by the model vs. as observed in the data.

Among the six error types listed in Table 1, types 1-4 relate to the firms' market participation decisions while types 5-6 relate to their net market positions conditional on participation. For example, the set $\mathcal{E}_1$ (resp. $\mathcal{E}_5$) contains observed buyers (resp. sellers) mistakenly sorted as autarkic (resp. buyers) by the model given a triplet (F, T, η_t). When $F = T = 0$, no firm chooses autarky in the model so that $\mathcal{E}_1 = \mathcal{E}_2 = \varnothing$. As F and/or T rises and the autarkic zone in Figure 3 widens the cardinalities of these sets, $|\mathcal{E}_1|$ and $|\mathcal{E}_2|$, increase while both $|\mathcal{E}_3|$ and $|\mathcal{E}_4|$ decrease, indicating a trade-off in the cost levels. Recall that $F > 0$ is necessary to explain that some overallocated firms may remain autarkic, which causes $|\mathcal{E}_4|$ to shrink. We note that $|\mathcal{E}_5|$ and $|\mathcal{E}_6|$ are negligible relative to the numbers of participation-related errors.[55] This was to be expected because we set trading firms' effective allocations in line with their observed net market positions by construction in (20).

Formally, our twin objective is to (1) minimize the total number of sorting errors and (2) favor balanced distributions of these errors among error types. Goal (2) is congruent with maximizing Shannon's entropy applied to the distribution of error types $\{|\mathcal{E}_1|, \ldots, |\mathcal{E}_6|\}$. Specifically, letting $\mathcal{P}_i = |\mathcal{E}_i| \Big/ \sum_{j=1}^{6} |\mathcal{E}_j|$ denote the proportion of type-i errors, Shannon's entropy $\mathcal{H}$ is defined by

$$\mathcal{H} = -\sum_{i=1}^{6} \mathcal{P}_i \log(\mathcal{P}_i) \in [0; \log(6)],$$

and is maximal when the errors are evenly distributed, i.e. $\mathcal{P}_i = \mathcal{P}_j$ for all $i \neq j$. With N the

[55] Specifically, it suffices that F or T be positive but small for $|\mathcal{E}_5|$ and $|\mathcal{E}_6|$ to become nil.

Table 2: Annual calibration results (2008-2012)

	p_t^r	η_t	F	T	T/p_t^r	$\mathcal{H}/\log(6)$	$1-\sum_{i=1}^{6}\lvert\mathcal{E}_i\rvert/N$
2008	19.6	4.1	10	0.55	2.8%	0.74	0.90
2009	13.3	-0.3	18	1.40	10.5%	0.66	0.85
2010	14.3	8.1	5	0.55	3.8%	0.76	0.89
2011	13.1	0.3	16	1.30	9.9%	0.66	0.87
2012	7.4	0.3	8	0.60	8.1%	0.67	0.90

Note: p_t^r, η_t and T given in €/tCO_2. F given in k€.

total number of firms in the sample, we thus select (F, T) to maximize the normalized index

$$\Big(\mathcal{H}/\log(6)\Big) \times \Big(1 - \sum_{i=1}^{6} |\mathcal{E}_i|/N\Big) \in [0;1].$$

Given the aforementioned trade-off in trading cost levels and as detailed further in Appendix C, the normalized index is mostly determined by its entropy component which is inverted U shaped, hence globally concave. Our calibration results are reported in Table 2. In 2009, 2011 and 2012, η is close to zero, suggesting that baselines and effective allocations are on average satisfactorily parametrized for these years. In 2010, however, η is significantly larger, implying that an upward shift in the firms' initial deficits is necessary to reproduce observed prices. This can be explained by the economic downturn which dramatically decreased emissions in the three preceding years and thus our proxy for 2010 baselines.

The selected annual values for F and T vary between 5 and 18 k€, and 0.55 and 1.40 €/tCO_2 (or 2.8 and 10.5% of the EUA price) across years.[56] To substantiate the improvement relative to zero trading costs on average across years (see Appendix C for graphical illustrations and more details in a given year), the selected cost pairs decrease the number of sorting errors by 40%, rationalize 70% of individual autarkic compliance decisions and reduce the dispersion across sorting error types as measured by a 160% increase in Shannon's entropy.

Although our approach to selecting trading costs differs from the various methods used in the related empirical literature, our results are in the same range. For instance, Naegele (2018) estimates median and mean fixed permit market entry costs of 7 and 21 k€ across firms in Phase II, respectively.[57] Similarly, estimates of proportional trading costs are in the order of

[56] As a robustness check, we increase (resp. decrease) all firms' baseline emissions by 5%. As expected, the calibrated η_t's are smaller (resp. larger). However, the resulting variation in the selected values for F and T is negligible relative to the reference case.

[57] See also Table 1 in Naegele (2018) for a literature overview of transaction cost estimates in the EU ETS.

0.1 € per permit traded but can go up to 2 € for small firms (e.g. Jaraitė et al., 2010; Heindl, 2012b; Joas & Flachsland, 2016). Additionally, Frino et al. (2010) and Medina et al. (2014) find a bid-ask spread for Phase II futures contracts ranging from 1 to 10% of the EUA price, which can give a rough sense of the magnitude of proportional trading costs.

4.2 Supply control with vs. without trading costs

Trading costs affect equilibrium outcomes, which has important implications for policy design, evaluation and implementation. We thus utilize our calibrated model to appraise the market price responses to supply-curbing policies in the presence vs. absence of trading costs and how they depend on their incidence across firms.[58] We also evaluate how total compliance costs vary as supply is tightened depending on how firms change their trading behavior.

4.2.1 Impacts on market prices

We evaluate the price impacts of an arbitrary one-sixth tightening in (annual) supply in 2009 and 2012 for each select sample of firms.[59] We consider these two years because they feature negligible adjustment terms η_t and differing values for the trading cost pairs and market price levels p_t^r. We assume that permits are withdrawn directly from firms' allocations according to four alternative scenarios: (1) proportionally to their initial allocations, (2) uniformly across all firms, or uniformly across overallocated (3) or underallocated (4) firms exclusively. We take two alternative perspectives in this appraisal: one which is oblivious to the trading costs in the model calibration, the other in which trading costs are accounted for and selected as in Section 4.1.2. Our simulation results are reported in Table 3.

As expected, the incidence of the cutback across firms is neutral vis-à-vis the magnitude of the market price increase under the assumption of no trading costs. This is no longer the case when one accounts for trading costs: as Proposition 5 indicates, the magnitude of the market price increase, relative to the frictionless case, depends on a price effect and a distribution effect, i.e. how the tightening is distributed among firms.

Two findings emerge from our calibrated examples. First, the price increase is always larger when one accounts for trading costs than in the frictionless case, irrespective of the incidence

[58] This is a timely issue in a context where the Market Stability Reserve (MSR) is bound to reduce annual supply schedules in the short to mid term (e.g. Perino, 2018; Quemin & Trotignon, 2019).

[59] This is roughly the magnitude of the yearly reductions in auctions due to the MSR in 2019 and 2020, i.e. about 380 $MtCO_2$ for a total cap of about 1,850 $MtCO_2$ (European Commission, 2019). The magnitude of the tightening does not affect the qualitative nature of our results, see also Section 4.2.2.

Table 3: Price responses to a $\frac{1}{6}$ supply tightening with different incidence scenarios

		Incidence scenario			
		(1)	(2)	(3)	(4)
2009 (p^r=13.3)	$(p^\star - p^r)/p^r$	0.70	0.70	0.70	0.70
	$(\hat{p} - p^r)/p^r$	1.11	1.29	1.19	1.59
	$(\hat{p} - p^r)/(p^\star - p^r)$	1.59	1.84	1.70	2.27
2012 (p^r=7.4)	$(p^\star - p^r)/p^r$	1.04	1.04	1.04	1.04
	$(\hat{p} - p^r)/p^r$	2.14	2.47	2.27	2.79
	$(\hat{p} - p^r)/(p^\star - p^r)$	2.06	2.38	2.18	2.68

Note: p^r is the pre-tightening reference price in €/tCO_2, $p^\star$ (resp. $\hat{p}$) is the post-tightening price without (resp. with calibrated) trading costs. Incidence scenario: permits are withdrawn (1) proportionally to firms' allocations, uniformly across all (2), overallocated (3) or underallocated (4) firms in the annual samples.

scenario. This is because in our samples of firms some potential suppliers are autarkic due to the trading costs so that the market is initially tighter than in the frictionless case, which in turn tends to amplify the tightening-induced price increase. Additionally, the lower the price level to start with, the larger the relative tightening-induced price increase and the larger the absolute price increase in the absence vs. presence of trading costs – as previously hinted at in the analytical and numerical examples in Section 3.3. Relatedly and crucially, note that larger trading costs (as in 2009 w.r.t. 2012) should not be thought of as a sufficient condition to sustain larger price responses to a supply tightening.

Second, the incidence of the tightening has significant impacts on the resulting price increase, which can vary in size by 30-40% across incidence scenarios. Intuitively, we see that uniformly targeting the supply tightening on underallocated (resp. overallocated) firms leads to a larger (resp. smaller) price increase than when it is evenly distributed among all firms.[60] The lowest price increase obtains when the tightening is spread across firms in proportion to their initial permit endowments, a proxy for their size under grandfathering-based allocation. As Section 4.2.2 will confirm with a welfare analysis, this incidence type leaves less (costly) reallocations to occcur through the market than the others (relative to the least-cost optimum). As such, it mitigates induced additional market strain and thus the resulting price increase.

In summary, modeling assumptions (here considering trading costs or not) matter for supply policy evaluation (e.g. size of the price response) and implementation (e.g. role of the incidence

[60] Given a fixed market price, when targeting underallocated firms (*i*) demand from buyers in $\mathcal{D}$ rises; (*ii*) some autarkic firms in $\mathcal{A}_2$ become buyers, increasing demand further; (*iii*) supply by sellers in $\mathcal{S}_3$ declines and (*iv*) some of them become autarkic, reducing supply further. By contrast, only (*iii*) and (*iv*) applied to $\mathcal{S}_1 \cup \mathcal{S}_2$ occur when targeting overallocated firms.

on firms). Specifically, our simulation results suggest that a modeler/regulator who does not account for trading costs though they prevail in reality may underestimate the price impacts of supply-curbing policies, here by a factor of about two. This underestimation bias is slightly more pronounced the lower the pre-tightening price and varies with the incidence type.

4.2.2 Impacts on compliance costs

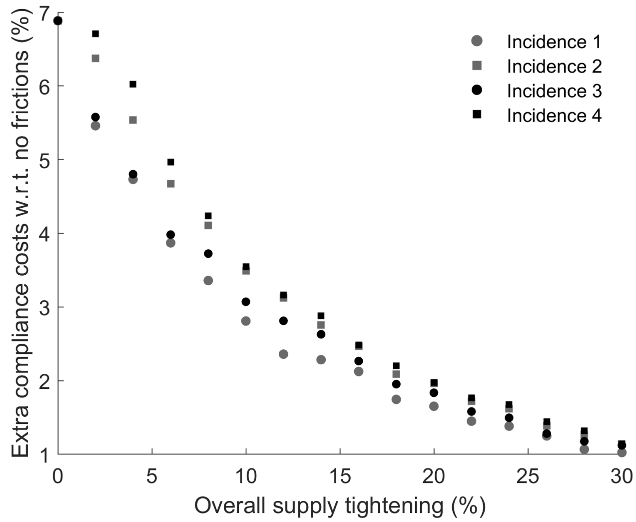
We evaluate compliance costs for a 0–30% range of tightening in supply for the 2009 select sample of firms and compare them depending on (1) whether trading costs are accounted for or not and (2) the type of incidence on firms. Overall compliance costs comprise abatement costs and incurred trading costs (if any) which we sum over all firms. By construction, they are increasing and convex in the stringency of the tightening and always higher in the presence of trading costs (Proposition 4). Relative to frictionless conditions, extra compliance costs result from incurred trading costs and foregone efficiency gains at the extensive and intensive margins. Figure 5 depicts these extra costs in relative terms as a function of the stringency of the tightening with the same four types of incidence as in Section 4.2.1.

Two findings emerge from our simulations. First, the extra compliance costs attributable to trading costs are in the order of 7% with the reference supply. This figure should be taken as illustrative only because it results from a comparison of modeling outputs under different modeling assumptions, not from a proper counterfactual analysis which we by construction cannot perform in our framework. As supply is tightened, the relative extra compliance costs decrease as the associated increase in compliance costs in absolute terms gradually dwarfs the difference in compliance costs with and without trading costs.

Second, the incidence of the tightening has notable impacts on the size of the extra compliance costs, which are higher by 15 to 35% between the most and least welfare-deteriorating type of incidence. Although there are many moving parts, most of this wedge can be explained by the relative changes in the number of autarkic firms as supply is tightened across incidence types. Indeed, autarkic firms' compliance costs are invariant when supply varies, namely nil when they are overallocated, positive but constant when underallocated. Simulations reveal that the number of autarkic firms (mostly overallocated) is relatively stable under incidence type (1) while it immediately collapses under types (2) and (4) as they induce the largest changes in market structure (here, initial individual deficits/surpluses).[61] An incidence proportional to firms' allocations thus appears as the least distortive incidence type.

[61] The share of autarkic firms under a uniform targeting of overallocated firms is also stable at first before starting to drop at a 7% cutback. This separation between incidence types (1) and (3) is visible in Figure 5.

Figure 5: Compliance costs for a 0–30% supply tightening with different incidence scenarios



Note: Based on the 2009 select sample of firms. Incidence scenario: permits are withdrawn (1) proportionally to firms' allocations, (2) uniformly across all, (3) overallocated or (4) underallocated firms.

Finally, we note that for a given emissions cap stringency, the simulated market price level is commensurate with the market strain which is reflected in the size of unrealized gains from trade and incurred trading costs. That is, the ranking of incidence types in terms of welfare loss visible in Figure 5 is identical to that in terms of price level given in Table 3, specifically incidence type (1) $\succ$ (3) $\succ$ (2) $\succ$ (4) where $\succ$ denotes welfare dominance.

5 Conclusion

This paper advances the frontier of research on permit markets with transaction costs and makes three contributions to the literature. First, we develop a consolidation procedure for annual transaction and compliance data allowing us to scrutinize firms' market behavior over EU ETS Phase II. This reveals two important empirical facts, which we interpret as pointing to the existence of fixed and variable trading costs: autarkic behavior is pervasive, especially among small or long firms, and those firms that engage in trade do so quite sparsely and only for sufficiently large volumes. Second, we incorporate fixed and proportional trading costs in a standard permit market model. In our equilibrium framework, the permit price and firms' participation in and extent of trading are determined endogenously. This allows us to analyze the sensitivity of the equilibrium to shifts in the trading costs and firms' allocations, and we characterize the properties of the market price impacts, as they are generally ambiguous and can go both ways. Third, we calibrate our model to EU ETS Phase II transaction data and show how trading costs in the order of 10 k€ per annum plus 1 € per permit traded noticeably reduce the discrepancies between observations and theoretical predictions for firms' behavior. Our simulations also suggest that ignoring trading costs may lead modelers to underestimate the price impacts of supply-curbing policies, with the size of this underestimation bias notably hinging on the specific incidence of such policies on firms.

It is important to acknowledge the caveats one must apply when interpreting our results, two of which we wish to highlight as alleys for future research – see also Section 1.1 for a broader discussion. First, while our measure of transaction costs captures all sorts of frictions (or the resultant thereof) one should seek to formally disentangle 'hard' financial trading costs from 'soft' behavioral factors such as the endowment effect or rational inattention. Second, one should aim to refine the modeling of the market structure to formally account for the temporal trading dimension and the presence of non-compliance entities such as intermediaries in order to understand their interaction with transaction costs and market efficiency. The theoretical and quantitative caveats of our modeling framework notwithstanding, we believe that it is a valuable contribution in the direction of bringing models closer to practical realities, and as such, making them better equipped for policy design and evaluation.

Acknowledgments

The authors would like to thank Baran Doda, Claire Gavard, Jean-Philippe Nicolaï, Michael Pahle, Grischa Perino, Lionel Ragot, Luca Taschini, Raphaël Trotignon, Frank Venmans and an anonymous referee from the FAERE for valuable conversations and helpful comments at various stages of this project, as well as participants at FSR Climate 2018, EAERE 2019, FAERE 2019 and seminars at EconomiX, LSE and PIK. The usual disclaimers apply.

All authors acknowledge funding from the Climate Economics Chair. Simon Quemin also acknowledges funding from LABEX Louis Bachelier Finance and Sustainable Growth (ANR 11-LABX-0019 Project) under the 'Investments for the Future' programme, in accordance with Article 8 of the Assigment Agreements for Communication Assistance, and from the Grantham Foundation for the Protection of the Environment, the UK Economic and Social Research Council (ESRC) through the Centre for Climate Change Economics and Policy.

References

Anderlini, L. & Felli, L. (2006). Transaction Costs and the Robustness of the Coase Theorem. *Economic Journal*, **116**(508), 223–45.

Balietti, A. C. (2016). Trader Types and Volatility of Emission Allowance Prices. Evidence from EU ETS Phase I. *Energy Policy*, **98**, 607–20.

Baumol, W. J. & Quandt, R. E. (1964). Rules of Thumb and Optimally Imperfect Decisions. *American Economic Review*, **54**(2), 23–46.

Bernard, A. B., Jensen, J. B., Redding, S. J. & Schott, P. K. (2012). The Empirics of Firm Heterogeneity and International Trade. *Annual Review of Economics*, **4**(1), 283–313.

Betz, R. & Schmidt, T. (2016). Transfer Patterns in Phase I of the EU Emissions Trading System: A First Reality Check Based on Cluster Analysis. *Climate Policy*, **16**(4), 474–95.

Borghesi, S. & Flori, A. (2018). EU ETS Facets in the Net: Structure and Evolution of the EU ETS Network. *Energy Economics*, **75**, 602–35.

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

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.

Chan, H. R., Chupp, B. A., Cropper, M. L. & Muller, N. Z. (2018). The Impact of Trading on the Costs and Benefits of the Acid Rain Program. *Journal of Environmental Economics & Management*, **88**, 180–209.

Charles, A., Darné, O. & Fouilloux, J. (2013). Market Efficiency in the European Carbon Markets. *Energy Policy*, **60**, 785–92.

Chen, C. (2018). *Slow Focus: Belief Evolution in the U.S. Acid Rain Program*. Working Paper, Harvard University.

Cludius, J. & Betz, R. (2020). The Role of Banks in EU Emissions Trading. *Energy Journal*, **41**(2), 275–99.

Coase, R. H. (1960). The Problem of Social Cost. *Journal of Law & Economics*, **3**(1), 1–44.

Coase, R. H. (1992). The Institutional Structure of Production. *American Economic Review*, **82**(4), 713–19.

Coggan, A., Whitten, S. M. & Bennett, J. (2010). Influences of Transaction Costs in Environmental Policy. *Ecological Economics*, **69**(9), 1777–84.

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.

Crocker, T. D. (1966). The Structuring of Atmospheric Pollution Control Systems. In H. Wolozin (ed.) *The Economics of Air Pollution*, (pp. 76–9). New York: Norton.

Crocker, T. D. (1972). On Air Pollution Control Instruments. *Loyola of Los Angeles Law Review*, **5**, 280–97.

Crossland, J., Li, B. & Roca, E. (2013). Is the European Union Emissions Trading Scheme (EU ETS) Informationally Efficient? Evidence from Momentum-Based Trading Strategies. *Applied Energy*, **109**, 10–23.

Dales, J. H. (1968). *Pollution, Property and Prices*. Toronto: University of Toronto Press.

Dávila, E. & Parlatore, C. (2020). Trading Costs and Informational Efficiency. *Journal of Finance*, forthcoming.

Deryugina, T., Moore, F. & Tol, R. S. J. (2020). *Applications of the Coase Theorem*. Working Paper 08-2020, University of Sussex.

Dixit, A. & Olson, M. (2000). Does Voluntary Participation Undermine the Coase Theorem? *Journal of Public Economics*, **76**(3), 309–35.

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

Dugast, J., Üslü, S. & Weill, P.-O. (2019). *A Theory of Participation in OTC and Centralized Markets*. Working Paper 25887, National Bureau of Economic Research.

Ellerman, A. D., Convery, F. J. & de Perthuis, C. (2010). *Pricing Carbon: The European Union Emissions Trading Scheme*. Cambridge, UK: Cambridge University Press.

Ellerman, A. D., Marcantonini, C. & Zaklan, A. (2016). The European Union Emissions Trading System: Ten Years and Counting. *Review of Environmental Economics & Policy*, **10**(1), 89–107.

Ellerman, A. D. & Trotignon, R. (2009). Cross Border Trading and Borrowing in the EU ETS. *Energy Journal*, **30**(2), 53–77.

European Commission (2015). *EU ETS Handbook*. Brussels, Climate Action.

European Commission (2019). *Report COM(2019) 557 final*. Brussels, 31.10.2019.

Foster, V. & Hahn, R. W. (1995). Designing More Efficient Markets: Lessons from Los Angeles Smog Control. *Journal of Law & Economics*, **38**(1), 19–48.

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.

Friedrich, M., Mauer, E.-M., Pahle, M. & Tietjen, O. (2019). *From Fundamentals to Financial Assets: The Evolution of Understanding Price Formation in the EU ETS*. Working Paper, Leibniz Information Center for Economics.

Frino, A., Kruk, J. & Lepone, A. (2010). Liquidity and Transaction Costs in the European Carbon Futures Market. *Journal of Derivatives & Hedge Funds*, **16**, 100–15.

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.

Gârleanu, N. & Pedersen, L. H. (2013). Dynamic Trading with Predictable Returns and Transaction Costs. *Journal of Finance*, **68**(6), 2309–40.

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

Grossman, S. J. & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review*, **70**(3), 393–408.

Hahn, R. W. & Hester, G. L. (1989). Marketable Permits: Lessons for Theory and Practice. *Ecology Law Quarterly*, **16**(2), 361–406.

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*. Discussion Paper 12-064, ZEW, Mannheim.

Heindl, P. (2012b). *Transaction Costs and Tradable Permits: Empirical Evidence from the EU Emissions Trading Scheme*. Discussion Paper 12-021, ZEW, Mannheim.

Heindl, P. (2017). The Impact of Administrative Transaction Costs in the EU Emissions Trading System. *Climate Policy*, **17**(3), 314–29.

Hintermann, B. & Ludwig, M. (2018). *Home Country Bias in International Emissions Trading: Indirect Evidence on Transaction Costs in the EU ETS*. Working Paper, CESifo.

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

Hortaçsu, A. & Puller, S. L. (2008). Understanding Strategic Bidding in Multi-Unit Auctions: A Case Study of the Texas Electricity Spot Market. *RAND Journal of Economics*, **39**(1), 86–114.

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.

Jo, A. (2019). *Trust and Compliance: Evidence from the EU Emissions Trading Scheme*. Working Paper 298, Grantham Research Institute, London School of Economics, *R&R at Journal of Law & Economics*.

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.

Joskow, P. L., Schmalensee, R. & Bailey, E. M. (1998). The Market for Sulfur Dioxide Emissions. *American Economic Review*, **88**(4), 669–85.

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1990). Experimental Tests of the Endowment

Effect and the Coase Theorem. *Journal of Political Economy*, **98**(6), 1325–48.

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

Kerr, S. & Maré, D. C. (1998). *Transaction Costs and Tradable Permit Markets: The United States Lead Phasedown*. Working Paper, MOTU Economic and Public Policy Research.

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

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

Krutilla, K. & Krause, R. (2011). Transaction Costs and Environmental Policy: An Assessment Framework and Literature Review. *International Review of Environmental & Resource Economics*, **4**(3-4), 261–354.

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

Liu, Y.-P., Guo, J.-F. & Fan, Y. (2017). A Big Data Study on Emitting Companies' Performance in the First Two Phases of the European Union Emission Trading Scheme. *Journal of Cleaner Production*, **142**, 1028–43.

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

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

McCann, L. (2013). Transaction Costs and Environmental Policy Design. *Ecological Economics*, **88**, 253–62.

Medema, S. G. (2020). The Coase Theorem at Sixty. *Journal of Economic Literature*, forthcoming.

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

Melitz, M. J. & Redding, S. J. (2014). Heterogeneous Firms and Trade. In G. Gopinath, E. Helpman & K. Rogoff (eds.) *Handbook of International Economics*, volume 4, chapter 1. North Holland.

Montagnoli, A. & de Vries, F. P. (2010). Carbon Trading Thickness and Market Efficiency. *Energy Economics*, **32**(6), 133–6.

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.

Murphy, J. J. & Stranlund, J. K. (2007). A Laboratory Investigation of Compliance Behavior under Tradable Emissions Rights: Implications for Targeted Enforcement. *Journal of*

Environmental Economics & Management, **53**(2), 196–212.

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

Palao, F. & Pardo, A. (2012). Assessing Price Clustering in European Carbon Markets. *Applied Energy*, **92**, 51–6.

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

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

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

Radner, R. (1996). Bounded Rationality, Indeterminacy, and the Theory of the Firm. *Economic Journal*, **106**, 1360–73.

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

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.

Schleich, J., Lehmann, S., Cludius, J., Abrell, J., Betz, R. & Pinkse, J. (2020). *Active or Passive? Companies' Use of the EU ETS*. Working Paper Sustainability and Innovation S07/2020, Fraunhofer ISI, Karlsruhe.

Schmalensee, R. & Stavins, R. N. (2013). The SO_2 Allowance Trading System: The Ironic History of a Grand Policy Experiment. *Journal of Economic Perspectives*, **27**(1), 103–22.

Schultz, E. & Swieringa, J. (2014). Catalysts for Price Discovery in the European Union Emissions Trading System. *Journal of Banking & Finance*, **42**(1), 112–22.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *American Economic Review*, **69**(4), 493–513.

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.

Stranlund, J. K. (2017). The Economics of Enforcing Emissions Markets: A Review of the Literature. *Review of Environmental Economics & Policy*, **11**(2), 227–246.

Toyama, Y. (2019). *Dynamic Incentives and Permit Market Equilibrium in Cap-and-Trade Regulation*. Working Paper, Waseda University, *R&R at AEJ: Microeconomics*.

Tversky, A. & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics*, **106**(4), 1039–61.

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

Appendices

A Analytical derivations and collected proofs

A.1 Proof of Proposition 1

Given a market price $p > 0$, firm i's optimal abatement is $a_i^*(p) = p/\alpha_i$ and its individual efficiency gains from permit trading are defined by

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

where $a_i^0 = \max\{0; \beta_i\}$. Recalling that $p_i^0 = \alpha_i a_i^0$ and $C_i(a_i) = \alpha_i a_i^2/2$, the above rewrites

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

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. $p = 0$). Hence, 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. $X^2 - 2\alpha_i\beta_i X - 2\alpha_i F > 0$ 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 nil for $F = 0$ and positive for $F > 0$.

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 positive. 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 positive provided that F is not too large, i.e. $F < \alpha_i\beta_i^2/2$. This must at least be true for the last potential buyer so $F < \bar{\alpha}\bar{\beta}^2/2$, see Lemma 2.

A.3 Proof of Proposition 2

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

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

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

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

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 $\mathcal{S}(p,F,T)$.

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

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, we compute the partial derivatives of the two terms of interest in (6) using Leibniz's rule in conjunction with Fubini's theorem. This yields

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

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

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

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

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 Proofs of analytical examples

After tedious but standard calculus (11) and (12) obtain by solving $D(p, F, 0) = S(p, F, 0)$ for p and $p^\star$ solves $D(p, 0, 0) = S(p, 0, 0)$. Below we sketch out the key steps of the computations for *Case 1* and omit those for *Case 2* as they follow the same lines. Define function J by

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

which is constant (in specie, nil) according to (11). The implicit function theorem yields

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

One then has $\mathrm{d}\hat{p}_1/\mathrm{d}F < 0$ and $\mathrm{d}\hat{p}_1/\mathrm{d}\beta > 0$ as it is easy to check that $\partial J/\partial F > 0$, $\partial J/\partial \hat{p}_1 > 0$ and $\partial J/\partial \beta < 0$. In particular, it is convenient to rewrite the second equality above as follows

$$\frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta}(1 + \beta^2 X) = X(3F + 5\beta\hat{p}_1) + \frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta} \text{ with } X = 2F^2 \Big/ \Big(\beta^4(\bar{\alpha} - \underline{\alpha})\sqrt{F(F + 2\beta\hat{p}_1)}\Big) > 0.$$

By linearity of $p_1^\star$ in β we have $\beta\frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta} = p_1^\star$ so that it finally comes

$$\Big(\frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta} - \frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta}\Big)\Big(1 + \beta X^2\Big) = X(3F + 5\beta\hat{p}_1 - \beta p_1^\star) \Rightarrow \frac{\mathrm{d}\hat{p}_1}{\mathrm{d}\beta} \geq \frac{\mathrm{d}p_1^\star}{\mathrm{d}\beta} \text{ iff } \hat{p}_1 \geq (p_1^\star - 3F/\beta)/5.$$

A.8 Proof of Proposition 5

We compute and determine the magnitudes of both the price and distribution effects in the face of a small variation in individual allocation levels accounting for induced changes at both the intensive and extensive margins. We study the two effects in turn.

Price effect. We aim to rank $\partial V(\hat{p}, F, T)/\partial p$ and $\partial V(p^\star, 0, 0)/\partial p$. We 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 the arguments in α^+ to reduce clutter. The intensive margin term captures the decrease in demand on the part of firms in $\mathcal{D}(\hat{p}, F, T)$. The extensive margin term captures the coexistent decrease in demand as the $\mathcal{A}_2$-$\mathcal{D}$ frontier moves to the northeast (see Figure 3). On that frontier the net demand $y - (\hat{p} + T)/\alpha^+$ is zero when $F = 0$ for any $T \geq 0$ since

$\alpha^+ = (\hat{p}+T)/y$; and positive whenever $F > 0$ (specifically, firms enter or exit $\mathcal{D}$ with positive individual demands). This means that the extensive margin drops to zero when $F = 0$.

We proceed similarly for S, see Appendix A.5 for some computation details. In total, we get

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

When $F = 0$ the extensive margin effects are nil so $\partial V(\hat{p},0,T)/\partial p < \partial V(p^\star,0,0)/\partial p$, i.e. the price effect is above one. It can however be below one for some pairs $(F > 0, T)$ for which the extensive margin effects surpass those at the intensive margin. This is more likely to occur for small costs as the intensive margin terms decrease with the 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_t evaluated at $(\hat{p},F,T)$ reads

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

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

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

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, $\lim_{\gamma\to 0}(D_t - D_0)/\gamma$ writes

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

as $\lim_{\gamma\to 0}\mathcal{O}(\gamma) = 0$. The intensive margin term captures the increase in demand on the part of firms in $\mathcal{D}$ prior to the tightening. The extensive margin term captures what happens at the $\mathcal{A}_2$-$\mathcal{D}$ 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 $T \geq 0$ when $F = 0$.

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

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

where $|\mathcal{I}| = \lim_{\gamma\to 0}(V_t(p^\star,0,0) - V_0(p^\star,0,0))/\gamma$. 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. $\lim_{\gamma\to 0}(V_t(\hat{p},F,T) - V_0(\hat{p},F,T))/\gamma < |\mathcal{I}|$ holds. With $F = 0$, this holds for all $T > 0$.

Finally, we consider alternative distributions of supply tightening in the intensive margin only case treated in the body of the paper. When uniformly distributed among all firms, $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}$ holds for all $i \in \mathcal{I}$. When uniformly targeted on all firms with positive (resp. negative) deficits, $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}/(|\bar{\mathcal{S}}_3| + |\bar{\mathcal{D}}| + |\bar{\mathcal{A}}_2|) > \mathrm{d}\mathcal{Q}$ (resp. $\mathrm{d}\beta_i = \mathrm{d}\mathcal{Q}/(|\bar{\mathcal{S}}_1| + |\bar{\mathcal{S}}_2| + \bar{\mathcal{A}}_1|) > \mathrm{d}\mathcal{Q}$) 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 is a shorthand meaning 'evaluated at $(\hat{p},F,T)$'. In these three cases the distribution effect is

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

The magnitude of the price increase in the face of a given supply tightening thus depends on the way it is allocated among firms. The ranking between the three incidence types presented above is unclear prima facie: it hinges on the levels of the trading costs F and T and on the distributions of the firms' characteristics $\{\alpha_i\}_i$ and $\{\beta_i\}_i$.

B Consolidation methodology

Data recorded in the European Union Transaction Log (EUTL) contains both compliance and trading related information (at the account level) in two separate databases: the compliance database keeps track of the initial allocation and reconciliation of allowances; the transaction database records every physical exchange completed across accounts (including the account holder names of trading parties, date and volume traded). There are three main categories of accounts: Operator Holding Accounts (OHAs, one per regulated installation), Person Holding Accounts (PHAs) and Trading Accounts. The latter two can be opened and managed by non-regulated entities with no compliance obligations (e.g. intermediaries, financiers).

These two databases need to be consolidated at the company level – the relevant granularity level for abatement, compliance, trading and wider economic decisions. However, two issues arise when trying to match accounts to parent companies. First, only limited or incomplete information is available on the firms and sectors associated to each account. For instance, no dedicated field in the account characteristics indicates the name of a parent company, when it exists (e.g. account holders must fill an 'Account Holder Name' field but it is uneven across accounts as to the precision of company-specific details). Second, there is no key to match the two databases so we need to create our own beforehand.

To get at the ownership structure within the EUTL, we first construct a list of parent company names from the compliance database which we then use as a key to consolidate the trading information from the transaction database at the company level. To that end, we first clean the 'Account Holder Name' fields in the compliance database, totaling about 17,000 accounts over all years. Specifically, we remove punctuation marks, prefixes, suffixes, etc and separate words. We then run a first round of matching for duplicates on the first word the character strings contain, and obtain a first-pass list which associates the so-extracted parent companies to their accounts. We gradually refine the list by repeating this procedure with the second, third and fourth words for the remaining unassigned accounts.

In practice, a company name – when explicitly specified – often appears in the first or second word of the search field so that our simple method allows us to get a reasonably goof idea of which company owns which accounts. After the fourth iteration of the matching procedure, around 10% of accounts are single. We manually assign them to a parent company (e.g. with dedicated web searches) and those for which manual matching is unsuccessful remain single. Our final parent company list contains 7,215 entries in total over Phase II (2008-2012).

In parallel, a total of 7,210 accounts recorded some trading activity (at least one exchange) in

Table B.1: Descriptive statistics for consolidated regulated firms (2009 sample)

Trading firms

Sector	Number of firms	⟨Number⟩ of accounts	% of total emissions	Median deficit	⟨Number⟩ of transactions	⟨Volume⟩ of transactions
Combustion	872	2.6	71.2	70,654	15.8	39,517
Refining	23	4.0	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	≈ 0	-5,600	4.0	1,494

Total number of observations: 1451.

Autarkic firms

Sector	Number of firms	⟨Number⟩ of accounts	% of total emissions	Median deficit
Combustion	848	2.3	62.6	3,665
Refining	14	2.1	11.0	267,178
Metallurgy	41	2.0	3.7	-52,835
Cement & Lime	165	2.3	13.6	-7,749
Chemicals	4	2.0	1.2	-13,266
Paper & Glass	217	2.4	8.9	-3,114
Other	1	1.0	≈ 0	-11,139

Total number of observations: 1290.

Note: Median deficits and average volumes of transactions given in tCO_2. $\langle \cdot \rangle$ denotes the average.

the transaction database over Phase II, some of which with no compliance obligations. Only considering compliance entities reduces the transaction database to 5,145 active accounts. This essentially amounts to keeping OHAs, or PHAs opened and run by a regulated company (typically to first pool allocations and later dispatch EUAs for compliance). Figure 1 is based on this select database, where (1) a compliance entity is deemed autarkic if it records no trade with compliance or non-compliance entities alike; (2) year-on-year changes in the number of observations occur due to installation/account closures and new entries as they occur.

We cross-check our consolidation outputs with those in Jaraitė-Kažukauskė & Kažukauskas (2015), Naegele (2018) or Hintermann & Ludwig (2018) who link EUTL accounts to the Orbis database (Bureau van Dijk) to match installations to parent companies. Their methodologies are similar to that underpinning the Ownership Links and Enhanced EUTL Dataset, hosted by the European University Institute. Although we were not aware of this publicly available database linking accounts to parent companies when we started our project, it allows us to perform an ex-post sanity check for our consolidation methodology. Our respective results are found to be similar, e.g. Naegele finds a close 4,578 firms with her method.

The Illustration requires us to merge the compliance and transaction databases as we want the allocation, verified emissions and trading activity at the regulated firm level. Because the 'Account Holder Name' field is present in both databases, this is in principle straightforward. Due to matching discrepancies, however, the merged database only contains 2,500 entries. It is used to plot Figure 2 but needs further cleaning to be used in the Illustration. Specifically, we exclude firms whose reported information is anomalous (e.g. emissions are nil) or missing (e.g. no allocation nor market position provided). This leads to slight changes in the number of yearly entries. Table B.1 below draws on these datasets. Finally, we remove firms with implied negative abatement and our yearly datasets are ready for use. This leads, again, to slight changes in the number of yearly entries, see Table C.1 for descriptive statistics.

C Calibration results

This Appendix provides additional details on the model parametrization (Section 4.1.1) and the selection of trading costs (Section 4.1.2). Specifically, Table C.1 provides basic descriptive statistics to help visualize the annual $\{\alpha_{i,t}, \beta_{i,t}\}_{i,t}$ inference outputs and Figure C.1 graphically depicts how our cost selection criteria evolve with F and T for the year 2009.

Table C.1: Annually inferred firms' characteristics (rounded)

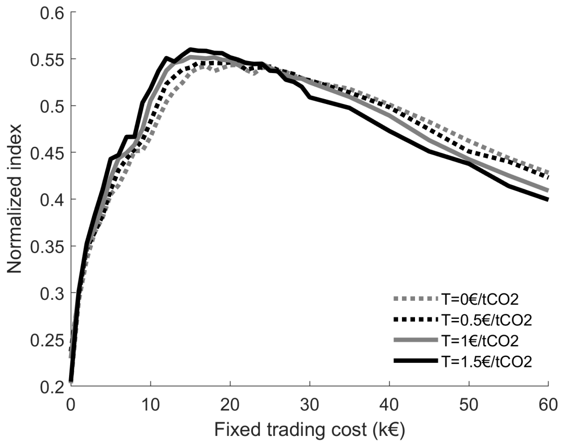
	Year	#Firms	Min	Max	Mean	Median	% Positive
α	2008	1,868	$2.1{\cdot}10^{-6}$	59	$9.4{\cdot}10^{-2}$	$7.1{\cdot}10^{-3}$	100
	2009	1,954	$1.3{\cdot}10^{-6}$	13	$2.9{\cdot}10^{-2}$	$3.6{\cdot}10^{-3}$	100
	2010	1,378	$1.6{\cdot}10^{-6}$	43	$7.1{\cdot}10^{-2}$	$5.1{\cdot}10^{-3}$	100
	2011	1,592	$3.8{\cdot}10^{-6}$	11	$5.6{\cdot}10^{-2}$	$5.5{\cdot}10^{-3}$	100
	2012	1,496	$3.4{\cdot}10^{-6}$	22	$4.5{\cdot}10^{-2}$	$3.1{\cdot}10^{-3}$	100
β	2008	1,868	$-4.0{\cdot}10^{6}$	$1.9{\cdot}10^{7}$	$23{\cdot}10^{3}$	-29	49
	2009	1,954	$-2.8{\cdot}10^{6}$	$2.2{\cdot}10^{7}$	$50{\cdot}10^{3}$	39	51
	2010	1,378	$-7.7{\cdot}10^{6}$	$1.8{\cdot}10^{7}$	$17{\cdot}10^{3}$	-1,100	38
	2011	1,592	$-2.2{\cdot}10^{6}$	$1.4{\cdot}10^{7}$	$26{\cdot}10^{3}$	-320	46
	2012	1,496	$-4.9{\cdot}10^{6}$	$7.5{\cdot}10^{6}$	$30{\cdot}10^{3}$	810	59

Note: α given in €/$(tCO_2)^2$ for $T = 0$. β given in tCO_2, not adjusted for year-on-year bank variations.

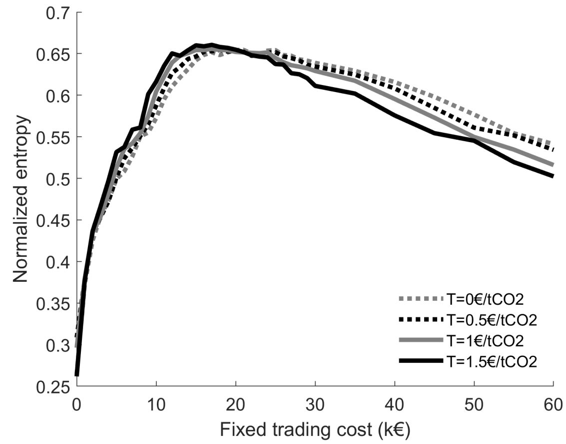
Let us elaborate on Figure C.1. As F and/or T increase the proportion of observed autarkic

Figure C.1: Selection criteria as functions of F and T (2009 sample)

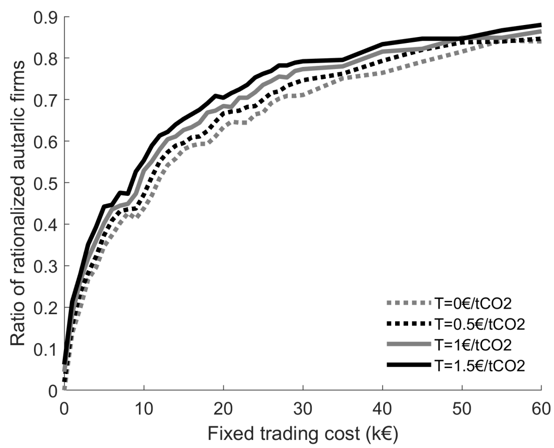
(a) Normalized index



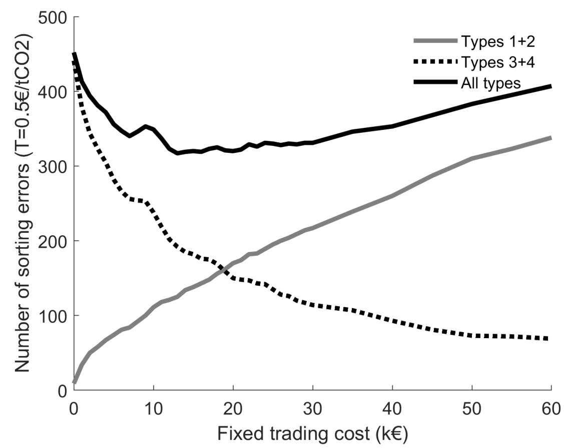
(b) Normalized entropy



(c) Ratio of rationalized autarkic firms



(d) Number of sorting errors (T=0.5€/tCO_2)



firms sorted as autarkic by the model increases (type 3-4 errors decrease). However, higher trading costs also imply that the model sorts more observed trading firms as autarkic (type 1-2 errors increase). As a result, for every value of T, Shannon's entropy hence has an inverted U shape and is maximal at some intermediate value of F. In turn, since the total number of errors is relatively stable, index variations are primarily driven by the entropy component – note that it can be locally non-concave due to the discrete nature of our problem.

We see that curves in Figure C.1c are ranked by increasing T values, i.e. the higher F and/or T the more autarkic compliance decisions are replicated by the model. In Figure C.1b this ranking only holds when F is not too large – specifically before the crossing in Figure C.1d. After this point, the ranking is reversed as an increase in F and/or T increases the imbalance between type 1-2 and type 3-4 sorting errors.